VARIABLE SELECTION IN HETEROSCEDASTIC SINGLE INDEX QUANTILE REGRESSION MODEL

E. Christou    M. G. Akritas

Department of Statistics
The Pennsylvania State University

ISNPS Conference, Avignon, 12 June 2016
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Single Index Quantile Regression</td>
</tr>
<tr>
<td></td>
<td>- The Proposed Estimator</td>
</tr>
<tr>
<td></td>
<td>- Main Results</td>
</tr>
<tr>
<td>3</td>
<td>Numerical Studies</td>
</tr>
<tr>
<td>4</td>
<td>Conclusions</td>
</tr>
</tbody>
</table>
Least Squares Regression: Models the relationship between a $d$-dimensional vector of covariates $\mathbf{X}$ and the conditional mean of the response $Y$ given $\mathbf{X} = \mathbf{x}$.

Least Squares Estimator:

1. provides only a single summary measure for the conditional distribution of the response given the covariates.
2. sensitive to outliers and can provide a very poor estimator in many non-Gaussian and especially long-tailed distributions.
**Quantile Regression (QR):** Models the relationship between a \( d \)-dimensional vector of covariates \( \mathbf{X} \) and the \( \tau \)th conditional quantile of the response \( Y \) given \( \mathbf{X} = \mathbf{x} \), \( Q_\tau(Y | \mathbf{x}) \).

**Quantile Regression:**

1. provides a more complete picture for the conditional distribution of the response given the covariates.
2. useful for modeling data with heterogeneous conditional distributions, especially data where extremes are important.
Introduction
Possible Models

- **Linear QR**: \( Q_\tau(Y|x) = \beta'x \), for a \( d \)-dimensional vector of unknown parameters \( \beta \).

  **Disadvantage**: linear assumption not always valid.

- **Nonparametric QR**: \( Q_\tau(Y|x) = g(x) \), for an unspecified function \( g: \mathbb{R}^d \to \mathbb{R} \).

  **Disadvantage**: meets the 'data sparseness' problem for high dimensional data.

- **Single-Index QR**: \( Q_\tau(Y|x) = g(\beta'x) \), depends on \( x \) only through the single linear combination \( \beta'x \).

  **Advantage**: reduces the dimension and maintains some nonparametric flexibility.
Introduction
Possible Models

- **Linear QR:** \( Q_\tau(Y|x) = \beta'x \), for a \( d \)-dimensional vector of unknown parameters \( \beta \).
  **Disadvantage:** linear assumption not always valid.

- **Nonparametric QR:** \( Q_\tau(Y|x) = g(x) \), for an unspecified function \( g : \mathbb{R}^d \rightarrow \mathbb{R} \).
  **Disadvantage:** meets the ’data sparseness’ problem for high dimensional data.
Introduction
Possible Models

- **Linear QR:** \( Q_\tau(Y|x) = \beta'x \), for a \( d \)-dimensional vector of unknown parameters \( \beta \).
  
  **Disadvantage:** linear assumption not always valid.

- **Nonparametric QR:** \( Q_\tau(Y|x) = g(x) \), for an unspecified function \( g : \mathbb{R}^d \rightarrow \mathbb{R} \).
  
  **Disadvantage:** meets the 'data sparseness' problem for high dimensional data.

- **Single-Index QR:** \( Q_\tau(Y|x) = g(\beta'x) \), depends on \( x \) only through the single linear combination \( \beta'x \).
  
  **Advantage:** reduces the dimension and maintain some nonparametric flexibility.
However,

- Ubiquity of high dimensional data.
However,

- Ubiquity of high dimensional data.
- Including unnecessary variables Vs Variable selection.
However,

- Ubiquity of high dimensional data.
- Including unnecessary variables Vs Variable selection.
- Variable selection is also important in Quantile regression.
However,

- Ubiquity of high dimensional data.
- Including unnecessary variables Vs Variable selection.
- Variable selection is also important in Quantile regression.
Outline

1 Introduction

2 Single Index Quantile Regression
   - The Proposed Estimator
   - Main Results

3 Numerical Studies

4 Conclusions
Let \( \{ Y_i, X_i \}_{i=1}^n \) be independent and identically distributed (iid) observations that satisfy
\[
Y_i = Q_{\tau}(Y|X_i) + \epsilon_i,
\]
where \( Q_{\tau}(Y|x) = Q_{\tau,\beta}(Y|\beta'x) \) and the error term satisfies \( Q_{\tau}(\epsilon_i|x) = 0 \).
Let \( \{Y_i, X_i\}_{i=1}^n \) be independent and identically distributed (iid) observations that satisfy

\[
Y_i = Q_\tau(Y|X_i) + \epsilon_i,
\]

where \( Q_\tau(Y|x) = Q_{\tau,\beta_1}(Y|\beta_1'x) \) and the error term satisfies \( Q_\tau(\epsilon_i|x) = 0 \). For identifiability, we impose certain conditions on \( \beta_1 \):

- the most common one: \( \|\beta_1\| = 1 \) with its first coordinate positive.
- the one adopted here: \( \beta_1 = (1, \beta')' \), for \( \beta \in \mathbb{R}^{d-1} \).
Sparsity Assumption:

- \( \mathbf{X}_i = (\mathbf{X}_{i1}', \mathbf{X}_{i2}')' \), \( \mathbf{X}_{i1} \in \mathbb{R}^{d*} \), \( \mathbf{X}_{i2} \in \mathbb{R}^{d-d*} \), \( d* \leq d \) is an integer that specifies the \( d* \) relevant variables.

- \( \mathbf{\beta}_1 = (1, \mathbf{\beta}')' = (1, \mathbf{\beta}'_{11}, \mathbf{\beta}'_{12})' = (1, \mathbf{\beta}'_{11}, \mathbf{0}')' \), where \( \mathbf{\beta}_{11} \in \mathbb{R}^{d* - 1} \) is the non-zero parametric vector and \( \mathbf{\beta}_{12} \in \mathbb{R}^{d-d*} \) is the zero vector corresponding to the irrelevant variables.
The true parametric vector $\beta$ satisfies

$$\beta = \arg \min_b E(\rho_\tau(Y - Q_\tau(Y|b_1'X))),$$

where $b_1 = (1, b')'$ and $\rho_\tau(u) = (\tau - I(u < 0))u$. 
The Proposed Estimator

Iterative algorithms

As in the single index mean regression (SIMR) problem, the unknown $Q_\tau(Y|b'_1X)$ must be replaced with an estimator. Unlike the SIMR problem, however, there is no closed form expression for the estimator of $Q_\tau(Y|b'_1X)$, and this has led to iterative algorithms for estimating $\beta$ (Wu, Yu, and Yu, 2010; Kong and Xia, 2012).
For any given $b \in \mathbb{R}^{d-1}$, define the function $g(u|b) : \mathbb{R} \to \mathbb{R}$ as

$$g(u|b) = E(Q_{\tau}(Y|X)|b_1'X = u),$$

where $b_1 = (1, b')'$. 
For any given $b \in \mathbb{R}^{d-1}$, define the function $g(u|b) : \mathbb{R} \to \mathbb{R}$ as

$$g(u|b) = E(Q_\tau(Y|X)|b'_1X = u),$$

where $b_1 = (1, b')'$. Noting that, under the SIQR model,

$$g(\beta'_1X|\beta) = Q_{\tau,\beta_1}(Y|\beta'_1X) = Q_\tau(Y|X),$$

$\beta$ also satisfies

$$\beta = \arg \min_b E(\rho_\tau(Y - g(b'_1X|b))),$$

where $b_1 = (1, b')'$. 

E. Christou, M. G. Akritas (PSU)
The Proposed Estimator
The New Approach

The sample level version of (1) consists of minimizing

\[ S_n(\tau, b) = \sum_{i=1}^{n} \rho_\tau(Y_i - g(b_1'X_i | b)). \]
The sample level version of (1) consists of minimizing

$$S_n(\tau, b) = \sum_{i=1}^{n} \rho_{\tau}(Y_i - g(b'_1 X_i|b)).$$

Again, $g(\cdot|b)$ is unknown but it can be estimated, in a non-iterative fashion, by first obtaining estimators $\hat{Q}_{\tau}(Y|X_i)$, for $i = 1, \ldots, n$, and forming the Nadaraya-Watson-type estimator

$$\hat{g}^{NW}(t|b) = \sum_{i=1}^{n} \frac{\hat{Q}_{\tau}(Y|X_i)K\left(\frac{t-b'_1 X_i}{h}\right)}{\sum_{k=1}^{n} K\left(\frac{t-b'_1 X_k}{h}\right)},$$

where $K(\cdot)$ is a univariate kernel function and $h$ is a bandwidth.
The Proposed Estimator

The New Approach

For example, \( \hat{Q}_\tau(Y|X_i) \), can be the local linear conditional quantile estimator (Guerre and Sabbah, 2012). Specifically, for a multivariate kernel function \( K^*(x) = K^*(x_1, ..., x_d) \) and a univariate bandwidth \( h^* \), let

\[
L_n((\alpha_0, \alpha_1); \tau, x) = \frac{1}{nh^*d} \sum_{i=1}^{n} \rho_\tau(Y_i - \alpha_0 - \alpha_1'(X_i - x)) K^* \left( \frac{X_i - x}{h^*} \right)
\]

and define \( \hat{Q}_\tau(Y|x) \) as \( \hat{\alpha}_0(\tau; x) \), where \( \hat{\alpha}_0(\tau; x) \) is defined through

\[
(\hat{\alpha}_0(\tau; x), \hat{\alpha}_1(\tau; x)) = \arg \min_{(\alpha_0, \alpha_1)} L_n((\alpha_0, \alpha_1); \tau, x).
\]
The proposed estimator is obtained by

$$\hat{\beta} = \arg \min_{b \in \Theta} \sum_{i=1}^{n} \rho_{\tau} \left( Y_i - \hat{g}^{NW} (b_1' X_i | b) \right),$$

where $\Theta \subset \mathbb{R}^{d-1}$ is a compact set, $\beta$ is in the interior of $\Theta$. 


The Proposed Estimator

New Approach

Goals:

- For high dimensional data, it is possible to improve the estimation of the local polynomial conditional quantile estimators $\hat{Q}_\tau(Y|X_i)$. Consider the use of penalty term for variable selection to estimate $Q_\tau(Y|x)$. 
The Proposed Estimator
New Approach

Goals:

- For high dimensional data, it is possible to improve the estimation of the local polynomial conditional quantile estimators \( \hat{Q}_\tau(Y|X_i) \). Consider the use of penalty term for variable selection to estimate \( Q_\tau(Y|x) \).

- Consider the penalty term for producing the estimator of the parametric component \( \beta \) of the SIQR model.
The Proposed Estimator

New Approach

Steps:

- Step 1 can be thought of as an initial variable selection method used to achieve better estimation of the conditional quantiles $Q_\tau(Y|X_i)$. 
Steps:

- Step 1 can be thought of as an initial variable selection method used to achieve better estimation of the conditional quantiles $Q_{\tau}(Y|X_i)$.
- Step 2 uses these improved estimators to perform simultaneous variable selection and estimation of the parametric component of the SIQR model.
Specifically, for Step 1 we minimize with respect to $b$ the objective function

$$
\hat{L}_n(\tau, b) = \sum_{i=1}^{n} \rho_{\tau}(Y_i - \hat{g}^{NW}(b_1' X_i | b)) + n \sum_{j=2}^{d} p_{\lambda_1}(|b_j|),
$$

where $b_1 = (1, b')' = (1, b_2, ..., b_d)$ and $\lambda_1$ is the tuning parameter for the SCAD penalty (Fan and Li, 2001) $p_{\lambda_1}(\cdot)$.  


The Proposed Estimator
New Approach - Step 1

Denote with $\hat{\beta}^\text{VS}$ the minimizer resulting from minimizing the objective function (4).

This estimator can be used for variable selection.

- Let $\hat{A}_n = \{j \in (2, ..., d) : \hat{\beta}_j^\text{VS} \neq 0\}$ of cardinality $\hat{d}^* - 1$.

- Construct improved estimators, $\hat{Q}_\tau^\text{VS}(Y|x)$, of the conditional quantiles.

In particular, define $\hat{Q}_\tau^\text{VS}(Y|x)$ as the local polynomial conditional quantile estimator on the relevant variables.

$\hat{Q}_\tau^\text{VS}(Y|x)$ has indeed a faster rate of convergence than $\hat{Q}_\tau(Y|x)$. 
The Proposed Estimator - Step 2

For Step 2, we set

\[ \hat{g}_{VS}^{NW}(t|\bf{b}) = \sum_{i=1}^{n} \frac{\hat{Q}^{VS}(Y|\bf{X}_i)K\left(\frac{t-b'_1X_i}{h}\right)}{\sum_{k=1}^{n} K\left(\frac{t-b'_1X_k}{h}\right)} \]

and define the minimization problem

\[ \hat{S}_n(\tau, \bf{b}) = \sum_{i=1}^{n} \rho_{\tau}\left(Y_i - \hat{g}_{VS}^{NW}(b'_1X_i|\bf{b})\right) + n \sum_{j=2}^{d} p_{\lambda_2}(|b_j|), \]

where \( \bf{b}_1 = (1, b')' = (1, b_2, ..., b_d) \) and \( \lambda_2 > 0 \) is the tuning parameter for the SCAD penalty \( p_{\lambda_2}(\cdot) \).
The Proposed Estimator - Step 2

For Step 2, we set

\[
\hat{g}^{NW}_{VS}(t|b) = \sum_{i=1}^{n} \frac{\hat{Q}^{VS}_{\tau}(Y|X_i)K\left(\frac{t-b'_1X_i}{h}\right)}{\sum_{k=1}^{n} K\left(\frac{t-b'_1X_k}{h}\right)}
\]

and define the minimization problem

\[
\hat{S}_n(\tau, b) = \sum_{i=1}^{n} \rho_{\tau} \left( Y_i - \hat{g}^{NW}_{VS}(b'_1X_i|b) \right) + n \sum_{j=2}^{d} p_{\lambda_2}(|b_j|),
\]

where \( b_1 = (1, b'_1) = (1, b_2, ..., b_d) \) and \( \lambda_2 > 0 \) is the tuning parameter for the SCAD penalty \( p_{\lambda_2}(\cdot) \). The proposed estimator is obtained by

\[
\hat{\beta}^{SCAD} = \arg \min_{b \in \Theta} \hat{S}_n(\tau, b),
\]

where \( \Theta \in \mathbb{R}^{d-1} \) is a compact set assumed to contain the true value of \( \beta \).
Remark:

- An alternative version of Step 1 is to first perform dimension reduction and use the estimated directions as the new covariates on which to estimate the conditional quantiles $Q_\tau(Y|X_i)$. This version requires the assumption that for some $d \times \bar{d}$ matrix $B$, $Y \perp \perp X | B^T X$, i.e., that $Y$ is independent from $X$ given $B^T X$. Simulations suggest that this alternative version of Step 1 also works well, but the corresponding theory has not been established.
1 Introduction

2 Single Index Quantile Regression
   - The Proposed Estimator
   - Main Results

3 Numerical Studies

4 Conclusions
Main Results

Theorem

**Model Selection Consistency:** Let $\hat{\beta}^{VS} = (\hat{\beta}_2^{VS}, ..., \hat{\beta}_d^{VS})$ be the minimizer of the objective function $\hat{L}_n(\tau, b)$ defined in (4). Define $\hat{A}_n$ the set $\{j \in \{2, ..., d\} : \hat{\beta}_j^{VS} \neq 0\}$ and $A$ the set $\{j \in \{2, ..., d\} : \beta_j \neq 0\}$.

Under some regularity conditions and the conditions $\lambda_1 \to 0$ and $\sqrt{n\lambda_1} \to \infty$ as $n \to \infty$, we have

$$\lim_{n \to \infty} P(\hat{A}_n = A) = 1.$$
Main Results

Proposition 1

Let \( \hat{Q}_\tau^{VS}(Y|x) \) denote the local polynomial conditional quantile estimator on the relevant variables. Then, under some regularity assumptions,

\[
\sup_{x \in \mathcal{X}_0} | \hat{Q}_\tau^{VS}(Y|x) - Q_\tau(Y|x) | = O_p \left( \frac{\log n}{n} \right)^{s/(2s+d^*)} = O_p(a^{**})
\]

if the bandwidth \( h^* \) is asymptotically proportional to \( (\log n/n)^{1/(2s+d^*)} \).
Main Results

Proposition 1

Let \( \hat{Q}^{VS}_\tau(Y|x) \) denote the local polynomial conditional quantile estimator on the relevant variables. Then, under some regularity assumptions,

\[
\sup_{x \in X_0} |\hat{Q}^{VS}_\tau(Y|x) - Q_\tau(Y|x)| = O_P \left( \frac{\log n}{n} \right)^{s/(2s+d^*)} = O_P(a_n^{**})
\]

if the bandwidth \( h^* \) is asymptotically proportional to \( (\log n/n)^{1/(2s+d^*)} \).

Proposition 2

Let \( \hat{\beta}^{SCAD} \) be as defined in (5). Then, under the assumptions of Proposition 1 and the assumption that \( \lambda_2 \to 0 \) as \( n \to \infty \), \( \hat{\beta}^{SCAD} \) is \( \sqrt{n} \)-consistent estimator of \( \beta \).
**Oracle Property:** Let $\hat{\beta}^{SCAD}$ be as defined in (5) and set $\hat{\beta}^{SCAD} = (\hat{\beta}_{11}^{SCAD}, \hat{\beta}_{12}^{SCAD})'$, where $\hat{\beta}_{11}^{SCAD}$ is of cardinality $d^* - 1$. If the assumptions of Proposition 1 hold and moreover $\lambda_2 \to 0$ and $\sqrt{n}\lambda_2 \to \infty$ as $n \to \infty$, then with probability tending to one,

1. **Sparsity:** $\hat{\beta}_{12}^{SCAD} = 0$

2. **Asymptotic Normality:**

$$\sqrt{n}(\hat{\beta}_{11}^{SCAD} - \beta_{11}) \xrightarrow{d} N(0, \tau(1 - \tau)\mathbb{V}_{11}^{-1}\Sigma_{11}\mathbb{V}_{11}^{-1}),$$

where

$$\mathbb{V}_{11} = E\left( (g'(\beta_1'X_1|\beta))^2(X_{1,-1} - E(X_{1,-1}|\beta_1'X))(X_{1,-1} - E(X_{1,-1}|\beta_1'X))^\prime f_\epsilon|X(0|X) \right)$$

and

$$\Sigma_{11} = E\left( (g'(\beta_1'X_1|\beta))^2(X_{1,-1} - E(X_{1,-1}|\beta_1'X))(X_{1,-1} - E(X_{1,-1}|\beta_1'X))^\prime \right),$$

for $g'(t|b) = (\partial/\partial t)g(t|b)$ and $X_{1,-1}$ the $(d^* - 1)$-dimensional vector.
Numerical Studies

- NWQR denotes the unpenalized estimator.
- SCAD-NWQR denotes the proposed estimator.
- LASSO-AY denotes the LASSO estimator of Alkenani and Yu (2013).
- ALASSO-AY denotes the adaptive-LASSO estimator of Alkenani and Yu (2013).
Numerical Studies

How to evaluate the performance?

- size of the estimated parametric component (number of nonzero coefficients),
- number of its correct and incorrect zeros,
- absolute estimation error of the estimated parametric component,
- and mean squared error (MSE) of $\hat{\beta}_{1}^{SCAD'}X$, defined as

$$MSE(\hat{\beta}_{1}^{SCAD'}X) = \frac{1}{n} \sum_{i=1}^{n} (\hat{\beta}_{1}^{SCAD'}X_i - \beta'_1X_i)^2.$$ 

For the final conditional quantile estimator $\hat{Q}_{\tau}^{LL}(Y|\hat{\beta}_{1}^{SCAD'}X_i)$, we consider the mean check based absolute residuals

$$R_{\tau}(\hat{Q}_{\tau}^{LL}) = \frac{1}{n} \sum_{i=1}^{n} \rho_{\tau}(Y_i - \hat{Q}_{\tau}^{LL}(Y|\hat{\beta}_{1}^{SCAD'}X_i)).$$
Consider the model

\[ Y = 5 \cos(\beta_1' X) + \exp(- (\beta_1' X)^2) + \epsilon, \quad (6) \]

where \( X = (X_1, \ldots, X_5)' \), \( X_i \sim U(0, 1) \) are iid, \( \beta_1 = (1, 2, 0, 0, 0)' \), the residual \( \epsilon \) follows an exponential distribution with mean 2, and \( X_i \)'s and \( \epsilon \) are mutually independent.
Consider the model

\[ Y = 5 \cos(\beta_1' X) + \exp(-(\beta_1' X)^2) + \epsilon, \]  

(6)

where \( X = (X_1, \ldots, X_5)' \), \( X_i \sim U(0, 1) \) are iid, \( \beta_1 = (1, 2, 0, 0, 0)' \), the residual \( \epsilon \) follows an exponential distribution with mean 2, and \( X_i \)'s and \( \epsilon \) are mutually independent.

We fit the SIQR model for five different quantile levels, \( \tau = 0.1, 0.25, 0.5, 0.75, 0.9 \).

We use a sample size of \( n = 400 \) and perform \( N = 100 \) replications.
**Numerical Studies**

**Example 1**

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>size</td>
<td>1.27</td>
<td>1.28</td>
<td>1.32</td>
<td>1.24</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>(0.4462)</td>
<td>(0.4513)</td>
<td>(0.4688)</td>
<td>(0.5707)</td>
<td>(0.6256)</td>
</tr>
<tr>
<td># corr. zeros</td>
<td>2.73</td>
<td>2.72</td>
<td>2.68</td>
<td>2.76</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td>(0.4462)</td>
<td>(0.4513)</td>
<td>(0.4688)</td>
<td>(0.5707)</td>
<td>(0.6256)</td>
</tr>
<tr>
<td># incor. zeros</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>$R(\hat{\beta}_{SCAD})$</td>
<td>0.3702</td>
<td>0.5190</td>
<td>0.6535</td>
<td>0.7779</td>
<td>1.0298</td>
</tr>
<tr>
<td>$R_\tau(\hat{Q}_\tau^{LL})$</td>
<td>0.1955</td>
<td>0.4414</td>
<td>0.6991</td>
<td>0.6865</td>
<td>0.4425</td>
</tr>
</tbody>
</table>

**Table 2:** mean values and standard deviations (in parenthesis) for the size and the number of correct and incorrect zeros of the estimated parametric component $\hat{\beta}_{SCAD}$. Also, mean values for $R(\hat{\beta}_{SCAD})$ and $R_\tau(\hat{Q}_\tau^{LL})$ for Model (6).
Comments: The proposed methodology

- selects correctly $X_2$ as the significant covariate 100% of the times for all quantile levels (0 average number of incorrect zeros),
- gives an average number of correct zeros close to the true value of 3.
- The average $R(\hat{\beta}^{SCAD})$ of the estimated parametric component seems to increase with increasing quantile level; the asymptotic variance formula of the sample quantiles involves the inverse density of the error term evaluated at the specific quantile of interest.
- Overall performance is good.
## Example 1

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCAD-NWQR</td>
<td>0.0015</td>
<td>0.0014</td>
<td>0.0023</td>
<td>0.0035</td>
<td>0.0174</td>
</tr>
<tr>
<td></td>
<td>(0.0041)</td>
<td>(0.0034)</td>
<td>(0.0047)</td>
<td>(0.0059)</td>
<td>(0.0997)</td>
</tr>
<tr>
<td>LASSO-AY</td>
<td>0.0006</td>
<td>0.0022</td>
<td>0.0046</td>
<td>0.0335</td>
<td>0.0581</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td>(0.0026)</td>
<td>(0.0064)</td>
<td>(0.0454)</td>
<td>(0.0734)</td>
</tr>
<tr>
<td>ALASSO-AY</td>
<td>0.0005</td>
<td>0.0020</td>
<td>0.0046</td>
<td>0.0311</td>
<td>0.0509</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0022)</td>
<td>(0.0065)</td>
<td>(0.0443)</td>
<td>(0.0702)</td>
</tr>
</tbody>
</table>

**Table 3:** mean values and standard deviations (in parenthesis) for the $\text{MSE}(\hat{\beta}_1'X)$ for the SCAD-NWQR, LASSO-AY, and ALASSO-AY estimated parametric components for Model (6).
Comments:

- the proposed methodology outperforms the LASSO and adaptive-LASSO estimators of Alkenani and Yu (2013) for all quantile levels, except for the lowest one.
Consider the model

\[ Y = \sin(2\pi \beta_1' X) + \frac{(1 + \beta_2' X)^2}{4} \epsilon, \]  

(7)

where \( X = (X_1, ..., X_5)' \), \( X_i \sim U(0, 1) \) are iid, \( \beta_1 = (1, 2, 0, 0, 0)' \), \( \beta_2 = (1, 1, 1, 0, 0)' \), the residual \( \epsilon \) follows an exponential distribution with mean 1.
Consider the model

\[ Y = \sin(2\pi \beta_1' X) + \frac{(1 + \beta_2' X)^2}{4}\epsilon, \]  

(7)

where \( X = (X_1, ..., X_5)' \), \( X_i \sim U(0, 1) \) are iid, \( \beta_1 = (1, 2, 0, 0, 0)' \), \( \beta_2 = (1, 1, 1, 0, 0)' \), the residual \( \epsilon \) follows an exponential distribution with mean 1.

We fit the SIQR model for five different quantile levels, \( \tau = 0.1, 0.25, 0.5, 0.75, 0.9 \). We use a sample size of \( n = 400 \) and perform \( N = 100 \) replications.
Consider the model

\[ Y = \sin(2\pi \beta_1' X) + \frac{(1 + \beta_2' X)^2}{4} \epsilon, \]  

(7)

where \( X = (X_1, ..., X_5)' \), \( X_i \sim U(0, 1) \) are iid, \( \beta_1 = (1, 2, 0, 0, 0)' \), \( \beta_2 = (1, 1, 1, 0, 0)' \), the residual \( \epsilon \) follows an exponential distribution with mean 1.

We fit the SIQR model for five different quantile levels, \( \tau = 0.1, 0.25, 0.5, 0.75, 0.9 \). We use a sample size of \( n = 400 \) and perform \( N = 100 \) replications.

We compare the penalized estimated parametric component (SCAD-NWQR) with the unpenalized one (NWQR).
Table 4: mean values and standard deviations (in parenthesis) for the size and the number of correct and incorrect zeros of the estimator parametric component $\hat{\beta}^{SCAD}$ for Model (7).
<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MSE(\hat{\beta}_1'X)$</td>
<td>NWQR</td>
<td>1.4696</td>
<td>0.4422</td>
<td>0.1259</td>
<td>0.1401</td>
</tr>
<tr>
<td></td>
<td>SCAD-NWQR</td>
<td>(0.9675)</td>
<td>(0.5094)</td>
<td>(0.2012)</td>
<td>(0.2992)</td>
</tr>
<tr>
<td></td>
<td>0.3283</td>
<td>0.1094</td>
<td>0.0536</td>
<td>0.0239</td>
<td>0.0683</td>
</tr>
<tr>
<td></td>
<td>(0.4606)</td>
<td>(0.2201)</td>
<td>(0.1486)</td>
<td>(0.0461)</td>
<td>(0.1468)</td>
</tr>
<tr>
<td>$R(\hat{\beta}_{SCAD})$</td>
<td>NWQR</td>
<td>2.3519</td>
<td>1.3141</td>
<td>0.8491</td>
<td>1.1313</td>
</tr>
<tr>
<td></td>
<td>SCAD-NWQR</td>
<td>0.8644</td>
<td>0.4743</td>
<td>0.2843</td>
<td>0.2711</td>
</tr>
<tr>
<td>$R_{\tau}(\hat{Q}_{LL})$</td>
<td>NWQR</td>
<td>0.1486</td>
<td>0.2563</td>
<td>0.3406</td>
<td>0.3239</td>
</tr>
<tr>
<td></td>
<td>SCAD-NWQR</td>
<td>0.1248</td>
<td>0.2264</td>
<td>0.3213</td>
<td>0.3130</td>
</tr>
</tbody>
</table>

**Table 5:** mean values and standard deviations (in parenthesis) for the $MSE(\hat{\beta}_1'X)$, average $R(\hat{\beta})$ and average $R_{\tau}(\hat{Q}_{LL})$ for NWQR and SCAD-NWQR estimated parametric components for Model (7).
Numerical Studies
Boston Housing Data

- 506 observations.
- 14 variables, response: \textit{medv}, the median value of owner-occupied homes in $1000$’s, predictors: measurements on the 506 census tracts in suburban Boston from the 1970 census.
- available in the \texttt{MASS} library in \texttt{R}.
- Collinearity in the data set; Breiman and Friedman (1985) applied ACE method and selected the four covariates \textit{RM}, \textit{TAX}, \textit{PTRATIO}, and \textit{LSTAT}.
exclude CHAS and RAD,

standardize the response and the remaining 11 predictor variables,

denote the predictor variables by $X_1, X_2, ..., X_{11}$, and

consider the SIQR model:

$$Q_{\tau}(\text{medv}|\mathbf{X}) = g(X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_{11} X_{11}|\beta), \quad (8)$$

for $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$, which assumes that RM is a significant predictor.
Table 6: Parameter estimates for the Boston housing data for fitting the SIQR model (8) to each quantile level.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>RM</th>
<th>CRIM</th>
<th>ZN</th>
<th>INDUS</th>
<th>NOX</th>
<th>AGE</th>
<th>DIS</th>
<th>TAX</th>
<th>PTRATIO</th>
<th>BLACK</th>
<th>LSTAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1</td>
<td>-0.1047</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.3385</td>
<td>-1.5e-5</td>
<td>-0.4494</td>
<td>-0.0897</td>
<td>0.2672</td>
<td>-0.0367</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>-0.2216</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.2964</td>
<td>-0.0737</td>
<td>-0.3736</td>
<td>-0.2238</td>
<td>0.2876</td>
<td>-0.0048</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>-0.2187</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-2.0e-5</td>
<td>-0.4243</td>
<td>-0.2986</td>
<td>0.4840</td>
<td>-0.5273</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.1260</td>
<td>-4.7e-5</td>
<td>0</td>
<td>-0.2783</td>
<td>0.3893</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>-0.1080</td>
<td>0</td>
<td>0</td>
<td>-0.2015</td>
<td>0</td>
<td>-0.2220</td>
<td>0</td>
<td>-0.1908</td>
<td>0.1814</td>
<td>0</td>
</tr>
</tbody>
</table>
### Tables 7 and 8: Number of zero coefficients and mean check based absolute residuals, $R_\tau(\hat{Q}_\tau^{LL})$, for SCAD-NWQR, LASSO-AY, and ALASSO-AY for fitting the SIQR model (8) to each quantile level.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of zeros</td>
<td>SCAD-NWQR</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>LASSO-AY</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ALASSO-AY</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCAD-NWQR</td>
<td>0.0641</td>
<td>0.1125</td>
<td>0.1455</td>
<td>0.1410</td>
<td>0.0882</td>
</tr>
<tr>
<td>LASSO-AY</td>
<td>0.0503</td>
<td>0.0990</td>
<td>0.1355</td>
<td>0.1299</td>
<td>0.0892</td>
</tr>
<tr>
<td>ALASSO-AY</td>
<td>0.0527</td>
<td>0.1092</td>
<td>0.1355</td>
<td>0.1288</td>
<td>0.0888</td>
</tr>
</tbody>
</table>
Numerical Studies
Boston Housing Data

Comments:

- different significant predictor variables,
- the proposed methodology gives more sparse models than the LASSO and adaptive-LASSO methodologies of Alkenani and Yu (2013),
- the mean check based absolute residuals resulting from the proposed SCAD-NWQR is slightly larger than that of the other methods for all except the 90th quantile. The difference is probably due to the different penalty functions used by the different methods, but could also be explained by the different levels of sparsity attained.
Conclusions

- conditional quantiles provide a more complete picture of the conditional distribution and share nice properties such as robustness.
Conclusions

- Conditional quantiles provide a more complete picture of the conditional distribution and share nice properties such as robustness.
- Consider a single index model to reduce the dimensionality and maintain some nonparametric flexibility.
Conclusions

- conditional quantiles provide a more complete picture of the conditional distribution and share nice properties such as robustness.
- consider a single index model to reduce the dimensionality and maintain some nonparametric flexibility.
- iterative algorithms for estimating the parametric vector; convergence issues.
Conclusions

- conditional quantiles provide a more complete picture of the conditional distribution and share nice properties such as robustness.
- consider a single index model to reduce the dimensionality and maintain some nonparametric flexibility.
- iterative algorithms for estimating the parametric vector; convergence issues.
- proposed a method that directly estimates the parametric component non-iteratively and have derived its asymptotic distribution under heteroscedastic errors.
Conclusions

- Conditional quantiles provide a more complete picture of the conditional distribution and share nice properties such as robustness.
- Consider a single index model to reduce the dimensionality and maintain some nonparametric flexibility.
- Iterative algorithms for estimating the parametric vector; convergence issues.
- Proposed a method that directly estimates the parametric component non-iteratively and have derived its asymptotic distribution under heteroscedastic errors.
- For high-dimensional data, introduce penalty terms both for variable selection to estimate $Q_\tau(Y|x)$ and for producing the estimator of the parametric component of the SIQR model.
References I

A. Alkenani and K. Yu
Penalized single-index quantile regression.

L. Breiman and J.H. Friedman
Estimating Optimal Transformations for multiple regression and correlation.

A. Bücher, A. El Ghouch and I. Van Keilegom
Single-index quantile regression models for censored data.
*Unpublished manuscript*,

E. Christou and M.G. Akritas
Single Index Quantile Regression for Censored Data.
W. Gonzalez-Manteiga and C. Cadarso-Suarez
Asymptotic properties of a generalized Kaplan-Meier estimator with some applications.

E. Guerre and C. Sabbah.
Uniform bias study and Bahadure representation for local polynomial estimators of the conditional quantile function.

R. Koenker.
Additive Models for Quantile Regression: Model Selection and Confidence Bandaids.
References III

E. Kong and Y. Xia.
A single-index quantile regression model and its estimation.

E. Kong, O. Linton and Y. Xia
Global Bahadur Representation for nonparametric Censored Regression Quantiles and its Applications.

J. D. Opsomer and D. Ruppert.
A fully automated bandwidth selection for additive regression model.

H. Wang, G. Li and G. Jian
Robust Regression Shrinkage and Consistent Variable Selection through the LAD-Lasso.
References IV

Y. Wu and Y. Liu
Variable Selection in Quantile Regression.

Single index quantile regression.

K. Yu and Z. Lu.
Local linear additive quantile regression.
Thank you!
Many thanks to my advisor, Prof. Michael Akritas.
Numerical Studies
Example 3

Consider the model

$$Y = \exp(\beta'_1 X) + \epsilon,$$

(9)

where $X = (X_1, \ldots, X_{10})'$, $X_i \sim N(0, 1)$ are iid, $\beta_1 = (1, 1, 1, 0, \ldots, 0)'$, the residual $\epsilon$ follows a standard normal distribution, and $X_i$’s and $\epsilon$ are mutually independent.
Consider the model

$$Y = \exp(\beta_1^T X) + \epsilon,$$  \hspace{1cm} (9)

where $X = (X_1, \ldots, X_{10})'$, $X_i \sim N(0, 1)$ are iid, $\beta_1 = (1, 1, 1, 0, \ldots, 0)'$, the residual $\epsilon$ follows a standard normal distribution, and $X_i$'s and $\epsilon$ are mutually independent. The objective of this example is to demonstrate the alternative procedure discussed in Remark.
Consider the model

\[ Y = \exp(\beta_1' X) + \epsilon, \]  

(9)

where \( X = (X_1, \ldots, X_{10})' \), \( X_i \sim N(0, 1) \) are iid, \( \beta_1 = (1, 1, 1, 0, \ldots, 0)' \), the residual \( \epsilon \) follows a standard normal distribution, and \( X_i \)'s and \( \epsilon \) are mutually independent. The objective of this example is to demonstrate the alternative procedure discussed in Remark. The sufficient dimension is been calculated using SIR and the conditional quantiles are been calculated as local linear conditional quantile estimators on the estimated directions. Simultaneous variable selection and parameter estimation is then performed using Step 2. We denote with SCAD-SIR the resulting estimator.
Numerical Studies

Example 3

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>size</td>
<td>2.51</td>
<td>2.49</td>
<td>2.53</td>
<td>2.54</td>
<td>2.59</td>
</tr>
<tr>
<td></td>
<td>(1.0298)</td>
<td>(0.8102)</td>
<td>(0.6883)</td>
<td>(0.9147)</td>
<td>(1.0550)</td>
</tr>
<tr>
<td># corr. zeros</td>
<td>6.49</td>
<td>6.51</td>
<td>6.47</td>
<td>6.46</td>
<td>6.41</td>
</tr>
<tr>
<td></td>
<td>(1.0298)</td>
<td>(0.8102)</td>
<td>(0.6883)</td>
<td>(0.9147)</td>
<td>(1.0550)</td>
</tr>
<tr>
<td># incor. zeros</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>( R(\hat{\beta}^{SCAD}) )</td>
<td>0.1004</td>
<td>0.0882</td>
<td>0.0813</td>
<td>0.0860</td>
<td>0.0910</td>
</tr>
<tr>
<td>( R_{\tau}(\hat{Q}_{LL}^{\tau}) )</td>
<td>0.1711</td>
<td>0.3140</td>
<td>0.3961</td>
<td>0.3148</td>
<td>0.1720</td>
</tr>
</tbody>
</table>

Mean values and standard deviations (in parenthesis) for the size and the number of correct and incorrect zeros of the estimated parametric component \( \hat{\beta}^{SCAD} \). Also, mean values for \( R(\hat{\beta}^{SCAD}) \) and \( R_{\tau}(\hat{Q}_{LL}^{\tau}) \) for Model (9).
### Numerical Studies

#### Example 3

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCAD-SIR</td>
<td>0.0016</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0013</td>
<td>0.0013</td>
</tr>
<tr>
<td>LASSO-AY</td>
<td>(0.0016)</td>
<td>(0.0013)</td>
<td>(0.0015)</td>
<td>(0.0013)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>ALASSO-AY</td>
<td>0.0565</td>
<td>0.0452</td>
<td>0.0336</td>
<td>0.0370</td>
<td>0.0453</td>
</tr>
<tr>
<td>ALASSO-AY</td>
<td>(0.0479)</td>
<td>(0.0278)</td>
<td>(0.0300)</td>
<td>(0.0298)</td>
<td>(0.0345)</td>
</tr>
</tbody>
</table>

Mean values and standard deviations (in parenthesis) for the $MSE(\hat{\beta}_1'X)$ for the SCAD-SIR, LASSO-AY, and ALASSO-AY estimated parametric components for Model (9).