SIQR: An R Package for Single-index Quantile Regression

by Tianhai Zu and Yan Yu

Abstract We develop an R package **SIQR** that implements the single-index quantile regression (SIQR) models via an efficient iterative local linear approach in Wu et al. (2010). Single-index quantile regression models are important tools in semiparametric regression to provide a comprehensive view of the conditional distributions of a response variable. It is especially useful when the data is heterogeneous or heavy-tailed. The package provides functions that allow users to fit SIQR models, predict, provide standard errors of the single-index coefficients via bootstrap, and visualize the estimated univariate function. We apply the R package **SIQR** to a well-known Boston Housing data.

Introduction

Single-index quantile regression (Wu et al., 2010) generalizes the seminal work of linear quantile regression of Koenker and Bassett (1978) by projecting the d-dimensional covariate **x** to a univariate index $\mathbf{x}\boldsymbol{\beta}$ and allowing a flexible univariate function $g(\mathbf{x}\boldsymbol{\beta})$. Quantile regression is often of great interest, especially when heterogeneity is present. Applications lie in a variety of fields, such as growth curves and reference charts in medicine; survival analysis when a given covariate may have a different effect on individuals with different levels of risks; value at risk calculation and wage and income studies in financial economics; high peak electricity demand in terms of weather characteristics in utility and energy; modeling rainfall, river flow, and air pollution in environmental modeling (see a survey in Yu et al. 2003).

Single-index quantile regression (SIQR) is a flexible semiparametric quantile regression model for analyzing heterogeneous data. The SIQR model has some appealing features: (i) It can provide a comprehensive view of the conditional distribution of a response variable given d-dimensional covariates by examining the full spectrum of conditional quantiles. This is especially important for complex heterogeneous data. (ii) The single-index structure is flexible to accommodate nonlinearity while avoiding the curse of dimensionality. It can also implicitly model some interactions among the covariates. Some interesting interpretations of the single-index parameter may be preserved. (iii) The quantile regression approach is robust to heavy-tailed distributions.

We present a package **SIQR** in R that implements the iterative local linear approach to the single-index quantile regression in Wu et al. (2010). The unknown univariate function is estimated by local linear estimation. The key algorithm can be decomposed into two efficient estimation steps on augmented data through local linear approximation and some equivalent formulation of the expected loss. Essentially, it iterates between two linear quantile regressions utilizing the state-of-the-art R package **quantreg**.

We apply our R package, **SIQR**, to the well-known Boston Housing data (1978) that is available in the R default library. The data has a total of 506 observations, and the response variable of interest is the median price of owner-occupied homes on the census tracts in suburban Boston from the 1970 census. The response variable and some covariates are left-skewed. Clearly, quantile regression is a natural tool to analyze the data (e.g., Chaudhuri et al. 1997; Yu and Lu 2004; Wu et al. 2010; Kong and Xia 2012). We organize the rest of the paper as follows. In the next section, we review the SIQR models. Next, we discuss the estimation algorithms implemented in this package. The section following describes the main features of the functions provided. Section "Real Data Analysis and Simulation" illustrates the use of **SIQR** in R for Boston housing data and a simulation study. The last section concludes the paper.

An overview for single-index quantile regression

Data structure and model settings

We develop an R package for the single-index quantile regression for semiparametric estimation with d-dimensional covariates. Let Y be the response variable and $\mathbf{X}$ be the covariate vector. Suppose there are n observations $\left\{(\mathbf{x}_i, y_i)\right\}_{i=1}^{n}$ of $(\mathbf{X} = \mathbf{x}, Y = y)$. Given $\tau \in (0, 1)$ and covariates $\mathbf{x}_i$, the single-index quantile model for the τ-th conditional quantile of the i-th observation is

$$q_\tau(Y = y_i | \mathbf{X} = \mathbf{x}_i) = g_\tau(\mathbf{x}_i \boldsymbol{\beta}_\tau), \tag{1}$$

where y_i is a real valued response, covariate $\mathbf{x}_i$ is a d-dimensional row vector, the single-index parameter $\boldsymbol{\beta}_\tau$ is a column vector in $\mathbf{R}^d$, and the univariate function $g_\tau : \mathbf{R} \to \mathbf{R}$ is subject to different τ. For identifiability, the single index parameter $\|\boldsymbol{\beta}_\tau\| = 1$ and the first non-zero element of $\boldsymbol{\beta}_\tau$ is positive (Yu and Ruppert, 2002). The projection $\mathbf{x}\boldsymbol{\beta}$ is often termed as the "single index". When g_τ is linear, single-index quantile regression model (1) reduces to the seminal work of linear quantile regression of Koenker and Bassett (1978).

Review of local linear estimation for single-index quantile regression

We implement the local linear estimation for single-index quantile regression (1) (Wu et al., 2010). For notational convenience, we omit the subscript τ in g_τ and $\boldsymbol{\beta}_\tau$. The true parameter vector $\boldsymbol{\beta}$ is the minimizer of

$$E\left[\rho_\tau\left(y - g(\mathbf{x}\boldsymbol{\beta})\right)\right], \tag{2}$$

where $\rho_\tau(u) = |u| + (2\tau - 1)u$ is the loss function, often termed as the "check" function in quantile regression. $g(\cdot)$ is the unknown univariate function. Constraint $\|\boldsymbol{\beta}\| = 1, \beta_1 > 0$ is imposed for identifiability. The above expected loss can be equivalently written as

$$E\left\{E\left[\rho_\tau\left(y - g(\mathbf{x}\boldsymbol{\beta})\right)|\mathbf{x}\boldsymbol{\beta}\right]\right\}, \tag{3}$$

where $E\left[\rho_\tau\left(y - g(\mathbf{x}\boldsymbol{\beta})\right)|\mathbf{x}\boldsymbol{\beta}\right]$ is the conditional expected loss and $g(\cdot)$ is the τth conditional quantile given the single-index parameter $\boldsymbol{\beta}$.

We adopt a local linear approximation. In particular, for $\mathbf{x}_i\boldsymbol{\beta}$ "close" to u, we can approximate the τth conditional quantile at $\mathbf{x}_i\boldsymbol{\beta}$ linearly via

$$g(\mathbf{x}_i\boldsymbol{\beta}) \approx g(u) + g'(u)(\mathbf{x}_i\boldsymbol{\beta} - u) = a + b(\mathbf{x}_i\boldsymbol{\beta} - u),$$

where we define $a \equiv g(u)$ and $b \equiv g'(u)$.

Now, we can minimize the sample analogue of (2) below as in Yu and Jones (1998) with respect to (a, b) with local linear estimation

$$\sum_{i=1}^n \rho_\tau\left(y_i - a - b(\mathbf{x}_i\boldsymbol{\beta} - u)\right) K\left(\frac{\mathbf{x}_i\boldsymbol{\beta} - u}{h}\right), \tag{4}$$

where $K(\cdot)$ is the kernel function and h is the bandwidth.

We further average (4) over u and obtain the sample analog of (3). The objective function below is used to estimate our single-index quantile regression model (1),

$$\sum_{j=1}^n \sum_{i=1}^n \rho_\tau\left(y_i - a_j - b_j(\mathbf{x}_i\boldsymbol{\beta} - \mathbf{x}_j\boldsymbol{\beta})\right) \omega_{ij}, \tag{5}$$

where

$$\omega_{ij} = \frac{K_h(\mathbf{x}_i\boldsymbol{\beta} - \mathbf{x}_j\boldsymbol{\beta})}{\sum_{k=1}^n K_h(\mathbf{x}_k\boldsymbol{\beta} - \mathbf{x}_j\boldsymbol{\beta})} \tag{6}$$

and $K_h(\cdot) = K(\cdot/h)/h$. We implement minimizing (5) iteratively with a detailed algorithm described next.

Bandwidth is a critical smoothing parameter that tunes the smoothness of the fitted function in local estimation. We implement the choice of the optimal bandwidth h_τ as advocated in Wu et al. (2010) through a computationally-expedient rule-of-thumb:

$$h_\tau = h_m \left\{\tau(1 - \tau)/\phi\left(\Phi^{-1}(\tau)\right)^2\right\}^{1/5}, \tag{7}$$

where $\phi(\cdot)$ is the probability density function and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Here, $h_m = \left\{\frac{[\int K^2(v)dv][var(y|\mathbf{x}\boldsymbol{\beta}=u)]}{n[\int v^2 K(v)dv]^2[\frac{d^2}{du^2}E(y|\mathbf{x}\boldsymbol{\beta}=u)]^2[f_{u_0}(u)]}\right\}^{1/5}$ is the optimal bandwidth in mean regression, which is easily obtainable from many existing packages (Ruppert et al., 1995).

Algorithm

We present the main algorithm for fitting the single-index quantile regression (SIQR) with local linear estimation in detail as following:

Input: Quantile level $\tau \in (0, 1)$, d-dimensional covariate vector $\mathbf{X} = \mathbf{x}$, and a response vector $Y = \mathbf{y}$.

Output: The estimated quantile single-index parameter $\widehat{\boldsymbol{\beta}}_\tau$ and fitted conditional quantile $\hat{q}_\tau(Y = \mathbf{y}|\mathbf{X} = \mathbf{x})$. The univariate function estimate $\hat{g}_\tau(\cdot)$.

1 Obtain an initial estimate $\hat{\boldsymbol{\beta}}^{(0)}$ of the quantile single-index parameter $\boldsymbol{\beta}$ from a linear quantile regression model (default) or a user-provided initial list. Standardize the initial estimate such that $||\hat{\boldsymbol{\beta}}^{(0)}|| = 1$ and $\hat{\beta}_1^{(0)} > 0$.

2 Given $\hat{\boldsymbol{\beta}}$, obtain $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ by solving a series of the following

$$\min_{(a_j, b_j)} \sum_{i=1}^n \rho_\tau \left(y_i - a_j - b_j(\mathbf{x}_i - \mathbf{x}_j)\hat{\boldsymbol{\beta}} \right) \omega_{ij}, \tag{8}$$

where the weights ω_{ij} is defined in (6). The bandwidth h is chosen optimally following a rule-of-the-thumb criterion in (7).

3 Given $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$, obtain $\hat{\boldsymbol{\beta}}$ by solving

$$\min_{\boldsymbol{\beta}} \sum_{j=1}^n \sum_{i=1}^n \rho_\tau \left(y_i - \hat{a}_j - \hat{b}_j(\mathbf{x}_i - \mathbf{x}_j)\boldsymbol{\beta} \right) \omega_{ij}, \tag{9}$$

with ω_{ij} evaluated at $\boldsymbol{\beta}$ and h from step 2.

4 Repeat Steps 2 and 3 until convergence.

5 Finally, we estimate $g(\cdot)$ at any u by $\hat{g}(\cdot; h, \hat{\boldsymbol{\beta}}) = \hat{a}$, where

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b)} \sum_{i=1}^n \rho_\tau \left(y_i - a - b(\mathbf{x}_i\hat{\boldsymbol{\beta}} - u) \right) K_h(\mathbf{x}_i\hat{\boldsymbol{\beta}} - u).$$

Obtain the final fitted conditional quantile $\hat{q}_\tau(Y = \mathbf{y}|\mathbf{X} = \mathbf{x})$ from model (1).

The above algorithm effectively decomposes (5) into two steps that can be achieved by two standard linear quantile regression procedures in Steps 2 and 3. In Step 3, we further note that (9) can be written as

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \arg \min_{\boldsymbol{\beta}} \sum_{j=1}^n \sum_{i=1}^n \rho_\tau \left(y_i - \hat{a}_j - \hat{b}_j(\mathbf{x}_i - \mathbf{x}_j)\boldsymbol{\beta} \right) \omega_{ij} \\
&= \arg \min_{\boldsymbol{\beta}} \sum_{j=1}^n \sum_{i=1}^n \rho_\tau \left(y_{ij}^* - \mathbf{x}_{ij}^* \boldsymbol{\beta} \right) \omega_{ij},
\end{aligned}
$$

where $y_{ij}^* = y_i - \hat{a}_j$, $\mathbf{x}_{ij}^* = \hat{b}_j(\mathbf{x}_i - \mathbf{x}_j)$, and ω_{ij} evaluated at the previous step, $i, j = 1, \cdots, n$. Given $\hat{a}_j$'s and $\hat{b}_j$'s, we can estimate $\boldsymbol{\beta}$ through usual linear quantile regression without intercept (*regression-through-origin*) 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 $\boldsymbol{\beta}$ from the previous iteration.

We can see that (9) is an alternative to (8). Adopting (9) yields some advantages: (i) It uses all the data and is more efficient in estimation; (ii) The double sum in (9) effectively increases the "augmented" sample size to n^2, similar to the minimum average variance estimation (MAVE) in the mean regression (Xia and Härdle, 2006).

The SIQR package

The R package **SIQR** consists of one core estimation function siqr and some supporting functions such as visualization tool plot.siqr and summary function summary.siqr. The R package SIQR depends on the R packages **stats**, **quantreg**, **KernSmooth**.

Main fitting function

The main estimation function siqr implements the iterative local linear approach to the single-index quantile regression in Wu et al. (2010).

The usage and input arguments of the main fitting function siqr are summarized as follows:

```
siqr(y, X, tau=0.5, h=NULL, beta.initial=NULL, se.method = NULL, maxiter=30, tol=1e-8)
```

This function takes two required arguments: the response variable y in vector format, the covariate matrix X. Please note that all the input covariates are required to be numeric variables.

This function also takes several optional arguments for finer controls. The optional argument tau is the quantile index, which specifies the left-tail probability. The default value of tau is 0.5, which refers to a single-index median regression. The optional argument h is the bandwidth in local linear quantile regression. Users can either provide a bandwidth or let the algorithm decide the optimal bandwidth as advocated in Wu et al. (2010) by setting this argument to NULL as default. The optional argument beta.initial is a numeric vector of the same length as the dimensionality of covariates. The users can use this argument to pass in any appropriate user-defined initial single-index coefficients based on prior information or domain knowledge. The default value is NULL, which instructs the function to estimate the initial single-index coefficients by linear quantile regression. The optional argument se.method is a character variable that specifies the method to obtain the standard error of estimated single-index coefficients. The default value is NULL to skip the calculation of standard error while the bootstrap-based method is available with "bootstrap". The optional argument maxiter and tol are control parameters that specify the criteria to terminate the iteration process. Although the algorithm normally converges quickly, the default maxiter and tol are set to 30 and 1e-8, respectively.

Other functions

We also provide several supporting functions:

```
summary.siqr(siqr.object)
print.summary.siqr(siqr.object)
```

The functions summary.siqr and print.summary.siqr provide detailed information related to the fitted model and summarize the results as illustrated in the next section. These two functions can be called directly by applying functions print and summary to the siqr.object.

```
plot.siqr(siqr.object, data.points = TRUE, bootstrap.interval=FALSE)
```

This function plots the fitted quantiles against the single-index term from an SIQR-fitted model object. By default, this function will also plot the observed data points in addition to the fitted quantiles to visualize the fitness of the model. One can remove the data points by setting the optional argument data.points to FALSE. Pointwise confidence interval will be added to the plot if the optional argument bootstrap.interval is set to TRUE.

```
simulation_data <- generate.data(n, true.theta=NULL, sigma=0.1,
setting="setting1", ncopy=1)
```

To help perform simulation studies, the function generate.data generates a size n data from two different settings: (i) a sine-bump model; and (ii) a location-scale model as in Wu et al. (2010). Users can define the single-index coefficients β via the argument true.beta and the noise level via sigma. If no true.beta was provided, the function will use $(1,1,1)^\mathsf{T}/\sqrt{3}$ for setting 1 and $(1,2)^\mathsf{T}/\sqrt{5}$ for setting 2 as the default. The last optional argument ncopy generates multiple copies of data for Monte Carlo simulations.

Real Data and Simulations

Boston Housing data

We consider the Boston housing data to demonstrate the real data application of the proposed R package **SIQR**. This dataset contains the median value of houses (in \$1000's), medv, in 506 tracts in Boston and 13 other socio-demographic related variables. This data has been investigated by many studies. Heterogeneity and some non-linear dependence of medv on predictor variables have been found by previous researchers. The dataset is maintained at the StatLib library of Carnegie Mellon University and can be found at the R built-in package **MASS**.

We focus on the following four covariates: RM, the average number of rooms per dwelling; TAX, the full-value property tax (in \$) per \$10,000; PTRATIO, the pupil-teacher ratio by town; and LSTAT, percentage of the lower status of the population as in Opsomer and Ruppert (1998), Yu and Lu (2004), and Wu et al. (2010). Following previous studies, we take logarithmic transformations on TAX and LSTAT and center the dependent variable medv around zero.

We use the following codes to load data from **MASS** and pre-process as discussed above. We fit a single-index quantile regression with $\tau = 0.25, 0.50, 0.75$ to the data and report fitted single-index coefficients for each variable.

```
library(SIQR)
#load data from MASS
library(MASS)
medv<- Boston$medv
RM <- Boston$rm
logTAX <- log(Boston$tax)
PTRATIO <- Boston$ptratio
logLSTAT <- log(Boston$lstat)

X <- cbind(RM,logTAX,PTRATIO,logLSTAT)
y0 <- medv - mean(medv)
beta0 <- NULL
tau.vec <- c(0.25,0.50,0.75)
est.coefficient <- matrix(NA, nrow = length(tau.vec), ncol = 5)
est.coefficient[,1] <- tau.vec
for (i in 1:length(tau.vec)){
est <- siqr(y0,X,beta.initial = beta0, tau=tau.vec[i],maxiter = 30,tol = 1e-8)
est.coefficient[i,2:5] <- est$beta
}
colnames(est.coefficient) <- c("quantile tau",colnames(X))
est.coefficient

#>      quantile tau        RM     logTAX     PTRATIO    logLSTAT
#> [1,]         0.25 0.3358285 -0.5243025 -0.06856117 -0.7795033
#> [2,]         0.50 0.3129182 -0.4294159 -0.06640472 -0.8445558
#> [3,]         0.75 0.2385613 -0.1933015 -0.07860687 -0.9484429
```
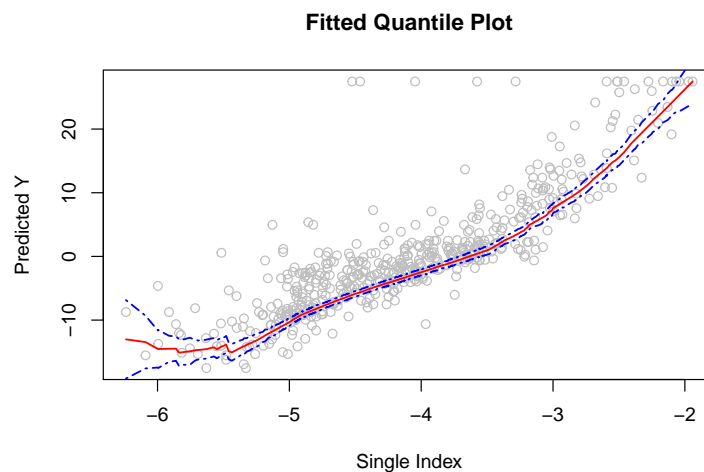
The estimated 0.25, 0.50, and 0.75 quantiles and their 95% pointwise confidence bounds are plotted with the following codes and outputs.

```
est.tau25 <- siqr(y0,X,beta.initial = NULL, tau=0.25)
plot.siqr(est.tau25,bootstrap.interval = TRUE)
```


Fitted Quantile Plot

Figure 1: The R output of plot.siqr with estimated 0.25 quantiles and the 95% pointwise confidence bounds.

```
est.tau50 <- siqr(y0,X,beta.initial = NULL, tau=0.50)
plot.siqr(est.tau05,bootstrap.interval = TRUE)
```

```
est.tau75 <- siqr(y0,X,beta.initial = NULL, tau=0.75)
plot.siqr(est.tau75,bootstrap.interval = TRUE)
```

As the estimated single-index function curves are almost monotonically increasing across different quantiles, variables that contribute positively to the single index affect the response variable (medv) positively. Based on the estimated coefficients and above plots, we found that the number of rooms

Fitted Quantile Plot

Figure 2: The R output of `plot.siqr` with estimated 0.50 quantiles and the 95% pointwise confidence bounds.

Fitted Quantile Plot
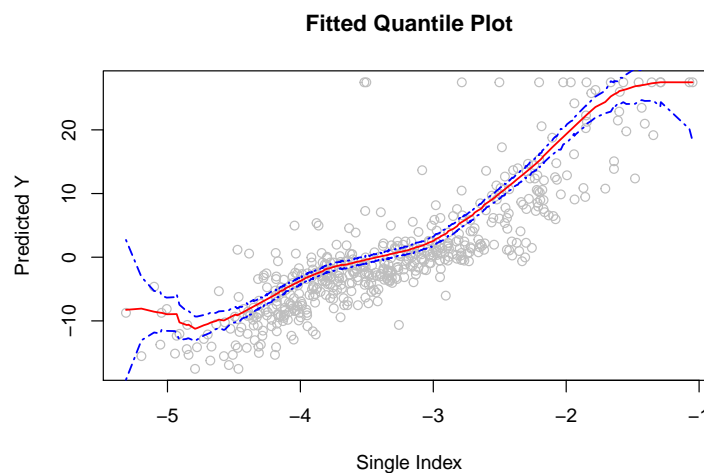
Figure 3: The R output of `plot.siqr` with estimated 0.75 quantiles and the 95% pointwise confidence bounds.

per house (rm) positively affects different quantiles. This matches the intuition that people value large spaces and multi-functional rooms. The property tax rate ln(tax) has a negative impact on housing prices across different quantiles. However, the influence of the tax rate is not significant at higher quantile $\tau = 0.75$. That suggests the tax rate may be less concerned for higher-income households, possibly due to tax deduction towards their income tax. Both the pupil-teacher ratio (ptratio) and the percentage of the lower (educational) status of the population ln(lstat) show negative influences on housing values, especially for the higher quantiles. It may suggest that potential buyers prefer areas featuring solid educational resources for their children and neighbors with higher education degrees and that preference grows more vital for more expensive houses.

Simulation

We consider two simulation settings. In the first simulation example, we use a sine-bump model with homoscedastic errors:

$$y = 5 \sin \left(\frac{\pi (\mathbf{x}\boldsymbol{\beta} - A)}{C - A} \right) + 0.1Z, \tag{10}$$

where $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}, C = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$, $\mathbf{x}$ is an $n \times 3$ design matrix that draws from an independent uniform distribution with min of 0 and max of 1, and the residual Z follows a standard normal

distribution. The true single-index parameter $\boldsymbol{\beta} = (1,1,1)^{\mathsf{T}}/\sqrt{3}$.

	Estimate	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
	mean	0.5782	0.5727	0.5725
$\tau = 0.25$	s.e.	0.0131	0.0281	0.0293
	bias	0.0009	-0.0046	-0.0048
	mean	0.5787	0.5755	0.5774
$\tau = 0.50$	s.e.	0.0115	0.0105	0.0111
	bias	0.0014	-0.0018	0.0003
	mean	0.5803	0.5756	0.5757
$\tau = 0.75$	s.e.	0.0119	0.0110	0.0118
	bias	0.0029	-0.0017	-0.0016

Table 1: Summary of parameter estimates for sine-bump simulation example 1 of sample size $n = 400$. True $\boldsymbol{\beta} = (1,1,1)^{\mathsf{T}}/\sqrt{3}$. The sample mean, standard error (s.e.), and bias of the parameter estimates of single-index coefficients from 200 replications.

The single-index coefficients are estimated via a series of quantile regressions with $\tau = 0.25, 0.50, 0.75$. Table 1 reports the mean, standard error (s.e.), and bias for each parameter estimate with sample size $n = 400$ over $M = 200$ replications on the simulation example 1. One can see that the algorithm for our R package **SIQR** is effective as the estimates are close to the true values.

For demonstration purposes, we show codes to generate data from (10) and fit the SIQR model using $\tau = 0.50$ with 200 replications as follows:

```
n <- 400
beta0 <- c(1, 1, 1)/sqrt(3)
n.sim <- 200
tau <- 0.50
data <- generate.data(n, true.theta=beta0, setting = "setting1",ncopy = n.sim)
sim.results.50 <- foreach(m = 1:n.sim,.combine = "rbind") %do% {
X <- data$X
Y <- data$Y[[m]]
est <- siqr(Y, X, beta.initial = c(2,1,0), tau=0.50,maxiter = 30,tol = 1e-8)
return(est$beta)
}
```

Note that this process has been repeated for the cases with $\tau = 0.25, 0.75$. We obtain a box plot of estimated single-index coefficients for $\tau = 0.25, 0.50, 0.75$, respectively, by applying the following code snippet.

```
boxplot(data.frame((sim.results.25)), outline=T,notch=T,range=1,
main = "Boxplots of Coefficient Estimates, tau = 0.25",horizontal = F)


boxplot(data.frame((sim.results.50)), outline=T,notch=T,range=1,
main = "Boxplots of Coefficient , tau = 0.50",horizontal = F)


boxplot(data.frame((sim.results.75)), outline=T,notch=T,range=1,
main = "Boxplots of Coefficient Estimates, tau = 0.75",horizontal = F)
```

Next, we consider a location-scale model as simulation example 2, where both the location and the scale depend on a common index $u = \mathbf{x}\boldsymbol{\beta}$. The quantiles are "almost-linear-in-index" as in Yu and Jones (1998) when the single index u is close to zero:

$$y = 5\cos(\mathbf{x}\boldsymbol{\beta}) + \exp(-(\mathbf{x}\boldsymbol{\beta})^2) + E, \tag{11}$$

where $\mathbf{x}$ is an $n \times 2$ design matrix that draws from an independent normal distribution with a standard deviation of 0.25, and the residual E follows an exponential distribution with a mean 2. The single-index parameter $\boldsymbol{\beta} = (1,2)^{\mathsf{T}}/\sqrt{5}$.

The simulated data are generated with the following codes. The sample size $n = 400$ with 100 replications. We only present the case when $\tau = 0.50$ for demonstration purposes.

```
n <- 400
beta0 <- c(1, 2)/sqrt(5)
```
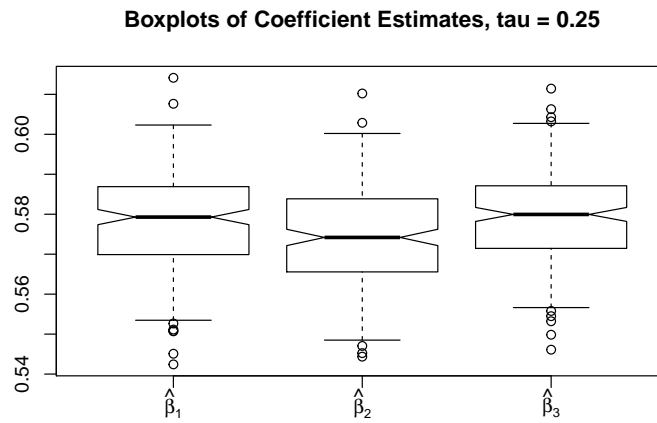
Boxplots of Coefficient Estimates, tau = 0.25

Figure 4: The box plot of estimated single-index coefficients for $\tau = 0.25$ from example 1.

Boxplots of Coefficient Estimates, tau = 0.25
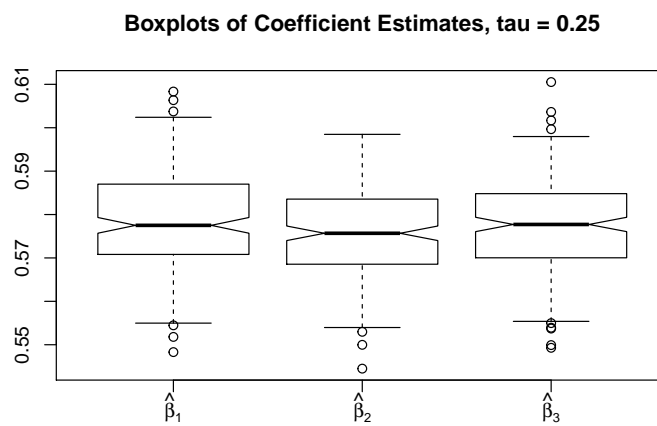
Figure 5: The box plot of estimated single-index coefficients for $\tau = 0.50$ from example 1.

Boxplots of Coefficient Estimates, tau = 0.25
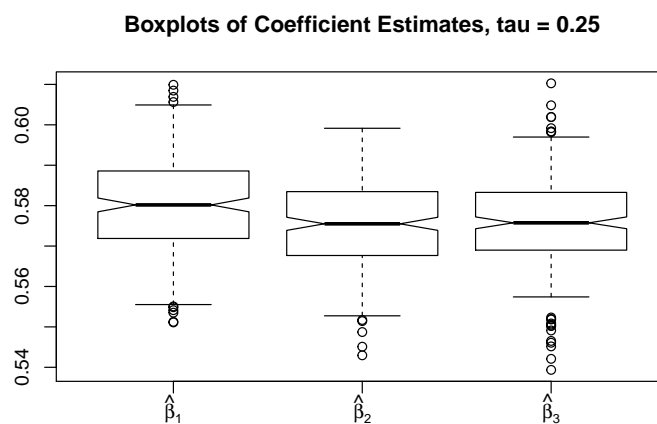
Figure 6: The box plot of estimated single-index coefficients for $\tau = 0.75$ from example 1.

```
n.sim <- 100
tau <- 0.5
data <- generate.data(n, true.theta=beta0, setting = "setting3",ncopy = n.sim)
sim.results <- foreach(m = 1:n.sim,.combine = "rbind") %do% {
X <- data$X
Y <- data$Y[[m]]
est <- siqr(Y, X, beta.initial = NULL, tau=tau,maxiter = 30,tol = 1e-8)
est$beta
}
est.mean <- c(tau,apply(sim.results,2,mean))
names(est.mean) <- c("tau","beta1.hat","beta2.hat")
est.mean

est.mean <- cbind(p_vec,apply(sim_results,c(1,2),sd))
colnames(est.mean) <- c("quantile tau","X1","X2","X3")
est.mean

#> tau beta1.hat beta2.hat
#> 0.5 0.4515909 0.8917233
```

The average estimated single-index coefficients shown above are close to the true single-index parameter $\beta = (1,2)^\top/\sqrt{5} \approx (0.4472, 0.8944)$. On top of that, the simulation standard error is also reported as below:

```
est.se <- c(tau,apply(sim.results,2,sd))
names(est.se) <- c("tau","beta1.se.hat","beta1.se.hat")
est.se

#>   tau beta1.se.hat beta1.se.hat
#>   0.5  0.02682211   0.01359602
```
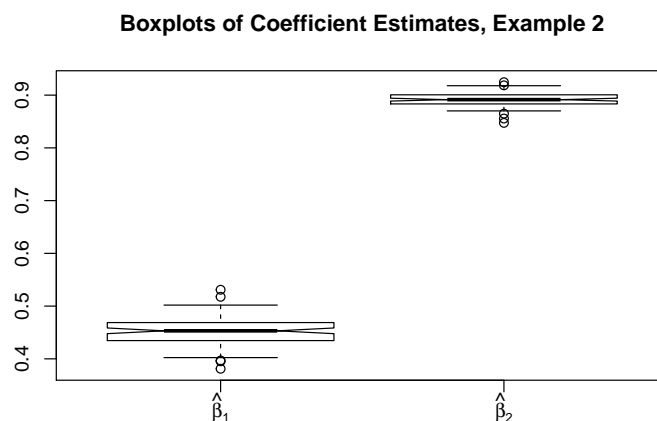
Meanwhile, the following box plots show that the estimated single-index coefficients are close to the true parameters with small deviations.

```
boxplot(data.frame((sim.results)), outline=T,notch=T,range=1,
main = "Boxplots of Coefficient Estimates (100 replications)",horizontal = F)
```
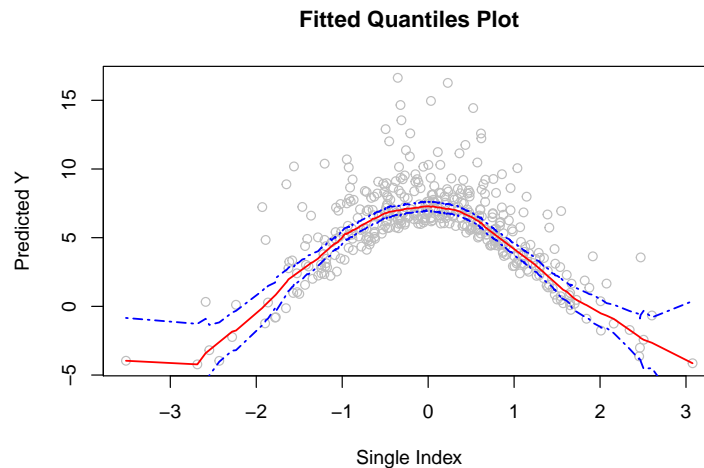

Figure 7: The box plot of estimated single-index coefficients for $\tau = 0.50$ from example 2.

Similarly, we plot the estimated quantiles and their 95% pointwise confidence bounds with the provided plot function `plot.siqr`. The observed data points are also plotted.

```
est.sim.50 <- siqr(data$Y[[1]],data$X,beta.initial = NULL, tau=0.5)
plot.siqr(est.sim.50,bootstrap.interval = TRUE)
```

Figure 8: The R output of `plot.siqr` with estimated 0.50 quantiles and the 95% pointwise confidence bounds from example 2.

Summary

In this paper, we present the R package **SIQR** for the local linear approach to single-index quantile regression models in Wu et al. (2010). We demonstrate the package applications to a popular Boston-housing data application and two simulation studies. It is our hope that the package will be useful to a variety of applications, especially for complex heterogeneous data where flexible quantile regression modeling is desirable.

Bibliography

P. Chaudhuri, K. Doksum, and A. Samarov. On average derivative quantile regression. *Annals of Statistics*, 25(2):715–744, Apr. 1997. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1031833670. URL https://projecteuclid.org/euclid.aos/1031833670. Publisher: Institute of Mathematical Statistics. [p460]

R. Koenker and G. Bassett. Regression Quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 0012-9682. doi: 10.2307/1913643. URL https://www.jstor.org/stable/1913643. Publisher: [Wiley, Econometric Society]. [p460, 461]

E. Kong and Y. Xia. A SINGLE-INDEX QUANTILE REGRESSION MODEL AND ITS ESTIMATION. *Econometric Theory*, 28(4):730–768, 2012. ISSN 0266-4666. URL http://www.jstor.org/stable/23257656. Publisher: Cambridge University Press. [p460]

J. D. Opsomer and D. Ruppert. A Fully Automated Bandwidth Selection Method for Fitting Additive Models. *Journal of the American Statistical Association*, 93(442):605–619, 1998. ISSN 0162-1459. doi: 10.2307/2670112. URL http://www.jstor.org/stable/2670112. Publisher: [American Statistical Association, Taylor & Francis, Ltd.]. [p463]

D. Ruppert, S. J. Sheather, and M. P. Wand. An Effective Bandwidth Selector for Local Least Squares Regression. *Journal of the American Statistical Association*, 90(432):1257–1270, 1995. ISSN 0162-1459. doi: 10.2307/2291516. URL http://www.jstor.org/stable/2291516. Publisher: [American Statistical Association, Taylor & Francis, Ltd.]. [p461]

T. Z. Wu, K. Yu, and Y. Yu. Single-index quantile regression. *Journal of Multivariate Analysis*, 101 (7):1607–1621, Aug. 2010. ISSN 0047-259X. doi: 10.1016/j.jmva.2010.02.003. URL http://www.sciencedirect.com/science/article/pii/S0047259X10000333. [p460, 461, 462, 463, 469]

Y. Xia and W. Härdle. Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, 97(5):1162–1184, May 2006. ISSN 0047-259X. doi: 10.1016/j.jmva.2005.11.005. URL http://www.sciencedirect.com/science/article/pii/S0047259X05001995. [p462]

K. Yu and M. C. Jones. Local Linear Quantile Regression. *Journal of the American Statistical Association*, 93(441):228–237, 1998. ISSN 0162-1459. doi: 10.2307/2669619. URL http://www.jstor.org/stable/2669619. Publisher: [American Statistical Association, Taylor & Francis, Ltd.]. [p461, 466]

K. Yu and Z. Lu. Local Linear Additive Quantile Regression. *Scandinavian Journal of Statistics*, 31(3): 333–346, 2004. ISSN 0303-6898. URL http://www.jstor.org/stable/4616834. Publisher: [Board of the Foundation of the Scandinavian Journal of Statistics, Wiley]. [p460, 463]

K. Yu, Z. Lu, and J. Stander. Quantile Regression: Applications and Current Research Areas. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52(3):331–350, 2003. ISSN 0039-0526. URL http://www.jstor.org/stable/4128208. Publisher: [Royal Statistical Society, Wiley]. [p460]

Y. Yu and D. Ruppert. Penalized Spline Estimation for Partially Linear Single-Index Models. *Journal of the American Statistical Association*, 97(460):1042–1054, 2002. ISSN 0162-1459. URL https://www.jstor.org/stable/3085829. Publisher: [American Statistical Association, Taylor & Francis, Ltd.]. [p461]

Tianhai Zu
University of Cincinnati
2906 Woodside Drive
Cincinnati, OH 45221
(ORCiD if desired)
zuti@mail.uc.edu

Yan Yu
University of Cincinnati
2906 Woodside Drive
Cincinnati, OH 45221
https://orcid.org/0000-0002-2859-3093
Yan.YU@uc.edu