Statistica Sinica **22** (2012), 1379-1401
doi:http://dx.doi.org/10.5705/ss.2010.199

SEMIPARAMETRIC QUANTILE REGRESSION WITH HIGH-DIMENSIONAL COVARIATES

Liping Zhu[1], Mian Huang[1] and Runze Li[2]

[1] *Shanghai University of Finance and Economics and*
[2] *Pennsylvania State University*

Abstract: This paper is concerned with quantile regression for a semiparametric regression model, in which both the conditional mean and conditional variance function of the response given the covariates admit a single-index structure. This semiparametric regression model enables us to reduce the dimension of the covariates and simultaneously retains the flexibility of nonparametric regression. Under mild conditions, we show that the simple linear quantile regression offers a consistent estimate of the index parameter vector. This is interesting because the single-index model is possibly misspecified under the linear quantile regression. With a root-n consistent estimate of the index vector, one may employ a local polynomial regression technique to estimate the conditional quantile function. This procedure is computationally efficient, which is very appealing in high-dimensional data analysis. We show that the resulting estimator of the quantile function performs asymptotically as efficiently as if the true value of the index vector were known. The methodologies are demonstrated through comprehensive simulation studies and an application to a dataset.

Key words and phrases: Dimension reduction, heteroscedasticity, linearity condition, local polynomial regression, quantile regression, single-index model.

1. Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression has become an important statistical analytic tool. Koenker (2005) provided a comprehensive review of the topic of quantile regression. Nonparametric quantile regression has been extensively studied in situations where the covariate vector is univariate. For instance, Bhattacharya and Gangopadhyay (1990) studied kernel estimation and nearest neighborhood estimation, and Fan, Hu, and Truong (1994) and Yu and Jones (1998) proposed local linear polynomial quantile regression. Koenker, Ng, and Portnoys (1994) proposed regression spline approaches for estimating the conditional quantile regression. When the covariate $\mathbf{x}$ is multivariate, Stone (1977) and Chaudhuri (1991) proposed fully nonparametric quantile regression. Due to the "curse of dimensionality", however, the fully nonparametric quantile regression may not be very useful in practice. Thus, researchers considered alternatives by assuming structures on the regression function. De Gooijer and Zerom (2003), Yu and Lu (2004), and Horowitz and Lee (2005) proposed quantile regression for additive models; Wang, Zhu, and Zhou (2009) considered quantile regression for partially linear varying-coefficient models; Wu, Yu and Yu (2010) suggested a back-fitting algorithm for single-index models. Based on their estimation equation, Kong and Xia (2010) investigated the Bahadur representation of single-index parameter estimators. Gannoun et al. (2004) considered quantile regression for multi-index models.

Let Y be a response variable, and $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}}$ be a covariate vector. In this paper, we consider the heteroscedastic single-index model

$$Y = G(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0) + \sigma(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)\varepsilon, \tag{1.1}$$

where the error term ε is assumed to be independent of $\mathbf{x}$, $E(\varepsilon) = 0$, and $\mathrm{Var}(\varepsilon) = 1$. The functions $G(\cdot)$ and $\sigma(\cdot)$ are unspecified, nonparametric smoothing functions. The parameter of interest is the direction of $\boldsymbol{\beta}_0$. To ensure identifiability, it is assumed throughout that $\|\boldsymbol{\beta}_0\| = 1$ with first nonzero element being positive. It is easy to see that (1.1) includes linear quantile regression (Koenker and Bassett (1978)) as a special case. Model (1.1) becomes a nonparametric regression model when $p = 1$, and the traditional single-index model (Ichimura (1993)) when $\sigma(\cdot)$ is a positive constant. Here we allow for heteroscedasticity. The single-index structure in (1.1) retains the flexibility of nonparametric regression and allows the presence of high-dimensional covariates.

Chaudhuri, Doksum, and Samarov (1997) proposed an estimation procedure for $\boldsymbol{\beta}_0$ using the average derivative approach that takes partial derivatives of the conditional quantile with respect to the covariates $\mathbf{x}$. The average derivative approach requires multivariate kernel regression and is less useful in the presence of high-dimensional covariates. Wu, Yu and Yu (2010) proposed a back-fitting algorithm that is computationally expensive when p is large. In this paper, we propose a new estimation procedure for model (1.1) that consists of two steps. We first estimate $\boldsymbol{\beta}_0$ using linear quantile regression and show that the resulting estimate is consistent. We then employ local linear regression (Fan and Gijbels (1996)) to estimate the conditional quantile function of Y given $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0$. We make the following contributions to the literature

(a) Under mild conditions, for any link function $G(\cdot)$ and $\tau \in (0,1)$, the τ-th linear quantile regression coefficient for $Y \mid \mathbf{x}$ is proportional to $\boldsymbol{\beta}_0$ in the single-index model (1.1).

(b) The local linear regression estimate for the conditional quantile function based upon the root-n consistent estimate of $\boldsymbol{\beta}_0$ has the same asymptotic bias and variance as the local linear estimate with the true value of $\boldsymbol{\beta}_0$.

The theoretical result in (a) implies that the ordinary linear quantile regression actually results in root-n consistent estimates for $\boldsymbol{\beta}_0$. Thus, quantile regression provides an effective tool to reduce the dimension of the covariate in the presence of high-dimensional covariates. This property enables us to substantially reduce the computational cost of the back-fitting algorithm (Wu, Yu and Yu (2010)); this is appealing in high-dimensional data analysis. We demonstrate this issue in detail in our numerical analysis. In addition, variable selection procedures for quantile regression (Zou and Yuan (2008); Wu and Liu (2009); Belloni and Chernozhukov (2011)) may be applicable to model (1.1) with high-dimensional covariates when many covariates are not significant. In general, the regularization parameter in the penalized quantile regression should depend on the quantile index τ (Belloni and Chernozhukov (2011)). The result in (b) implies that the dimensionality of covariate does not affect the performance of the local linear estimate asymptotically. Thus, it may not be necessary to update the estimate of $\boldsymbol{\beta}_0$ iteratively once a root-n consistent estimator of $\boldsymbol{\beta}_0$ is available. It also ensures that our proposed procedure works properly in high-dimensional settings.

The rest of this article is organized as follows. In Section 2 we propose a two-step procedure to estimate the conditional quantile function. Asymptotic properties of the resulting estimates are also established in this section. We demonstrate the methodologies through comprehensive simulation studies and an application to a dataset in Section 3. We conclude this paper with a brief discussion in Section 4. All proofs are in the Appendix.

2. A New Estimation Procedure

In this section we study the estimation of quantile regression for model (1.1). We prove that the linear quantile regression yields consistent estimate for $\boldsymbol{\beta}_0$. We further demonstrate that the quantile regression for model (1.1) can be reduced to the univariate quantile regression through replacing $\boldsymbol{\beta}_0$ with $\hat{\boldsymbol{\beta}}_0$, the resulting estimate of the linear quantile regression.

2.1. Estimation of $\boldsymbol{\beta}_0$

Let $\rho_\tau(r) = \tau r - r\mathbf{1}(r < 0)$ be the check loss function, and $\mathcal{L}_\tau(\mathrm{u}, \boldsymbol{\beta}) = E\left\{\rho_\tau\left(Y - \mathrm{u} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}$. Take

$$(\mathrm{u}_\tau, \boldsymbol{\beta}_\tau) \stackrel{\text{def}}{=} \underset{\mathrm{u},\boldsymbol{\beta}}{\operatorname{argmin}} \left\{\mathcal{L}_\tau(\mathrm{u}, \boldsymbol{\beta})\right\}. \tag{2.1}$$

Theorem 1. *If the covariate vector* $\mathbf{x}$ *at* (1.1) *satisfies*

$$E\left(\mathbf{x} \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}\right) = \mathrm{Var}\,(\mathbf{x})\boldsymbol{\beta}_0 \left\{\boldsymbol{\beta}_0^{\mathrm{T}}\mathrm{Var}\,(\mathbf{x})\boldsymbol{\beta}_0\right\}^{-1} \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}, \tag{2.2}$$

then $\boldsymbol{\beta}_\tau = \kappa\boldsymbol{\beta}_0$ *for some constant* κ*, where* $\boldsymbol{\beta}_\tau$ *is defined in* (2.1)*.*

The linearity condition (2.2) plays an important role here. With it, we are able to obtain a root n consistent estimator of the direction of $\boldsymbol{\beta}$ by the linear quantile regression, see Theorem 2. The linearity condition (2.2) is widely assumed in the context of sufficient dimension reduction. Li (1991) pointed out that it is satisfied when $\mathbf{x}$ follows an elliptically contour distribution, and Hall and Li (1993) proved that it always holds to a good approximation in single-index models of the form (1.1) when the dimension p of the covariates diverges. Thus, the linearity condition is typically regarded as mild, particularly when p is fairly large.

Similar results can be established for general convex loss functions under the linearity condition. When the quadratic loss function is used, however, the resulting estimate cannot recover any information beyond the conditional mean function. Similarly, the absolute loss function is designed to identify information in the conditional median function. More generally, the check loss function at any given τ is specifically designed to find the information relevant exclusively to the τ-th quantile function. Theorem 1 justifies partially the validity of the procedure suggested by Gannoun et al. (2004). However, their method can be excessive because the dimension-reduction step captures all information relevant to the whole conditional distribution function.

Theorem 1 also implies that the existing statistical procedures for the linear quantile regression can be directly applied to model (1.1). In particular, the existing variable selection procedure for the linear quantile regression with high-dimensional covariates (Wu and Liu (2009); Zou and Yuan (2008); Belloni and Chernozhukov (2011)) can be used to select significant variables. It is important to correctly set the range of the tuning parameter in the variable selection procedure based on a penalized linear quantile regression. The range of the tuning parameter may explicitly or implicitly depend on the link function G, the quantile index τ, and the density of error distribution. A data-driven method for tuning parameter selection is recommended.

Let $(\mathbf{x}_i, Y_i), i = 1, \ldots, n$, be a random sample from $(\mathbf{x}, Y)$, and take

$$\mathcal{L}_{\tau n}(\mathrm{u}, \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \left\{ \rho_\tau \left(Y_i - \mathrm{u} - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} \right) \right\}.$$

Let

$$\left(\widehat{\mathrm{u}}_\tau, \widehat{\boldsymbol{\beta}}_\tau \right) \stackrel{\mathrm{def}}{=} \operatorname*{argmin}_{\mathrm{u}, \boldsymbol{\beta}} \left\{ \mathcal{L}_{\tau n}(\mathrm{u}, \boldsymbol{\beta}) \right\}. \tag{2.3}$$

Theorem 2. *Assume that* $\boldsymbol{\Lambda} = E\left\{ f^*(\xi_\tau \mid \mathbf{x}) \mathbf{x}\mathbf{x}^{\mathrm{T}} \right\}$ *is invertible, where* $f^*(z \mid \mathbf{x})$ *denotes the conditional density of* $\left(Y - \mathbf{x}^{\mathrm{T}} \boldsymbol{\beta}_\tau \right)$ *given* $\mathbf{x}$*, and* ξ_τ *denotes the* τ*-th quantile of* $Y \mid \mathbf{x}$*. Then* $n^{1/2}(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau)$ *is asymptotically normal with mean zero* *and covariance matrix* $\tau(1-\tau)\,\mathbf{\Lambda}^{-1}E\left(\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)\mathbf{\Lambda}^{-1}$. *If the "partial residual" term* $Y-\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_\tau$ *is independent of* $\mathbf{x}$, *the asymptotic variance matrix of* $\widehat{\boldsymbol{\beta}}_\tau$ *reduces to* $\tau(1-\tau)\,E\left(\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)^{-1}/\left\{f^*(\xi_\tau)\right\}^2$.

2.2. Estimation of the conditional quantile function

By Theorem 1, $\boldsymbol{\beta}_\tau$ is proportional to $\boldsymbol{\beta}_0$, and hence the conditional distribution of $Y \mid \left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)$ is that of $Y \mid \left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_\tau\right)$. Consequently, Theorem 1 allows one to estimate the quantile regression of $Y \mid \left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)$ through the conditional distribution of $Y \mid \left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_\tau\right)$. We denote the τ-quantile function of $Y \mid \left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_\tau\right)$ by $G_\tau(\cdot)$.

By using data points $\{(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau, Y_i), i=1,\ldots,n\}$, we estimate $G_\tau(\cdot)$ by local linear regression. For ease of presentation, let $z=\mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}_\tau$, $\widehat{z}=\mathbf{x}_0^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau$, $Z_i=\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau$, $\widehat{Z}_i=\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau$. Local linear regression is used to approximate $G_\tau(Z)$ by

$$G_\tau(Z) \approx G_\tau(z) + G_\tau'(z)(Z-z)$$

for Z in the neighborhood of z. Let

$$\left(\widehat{a},\widehat{b}\right) \stackrel{\text{def}}{=} \underset{a,b}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left\{Y_i - a - b\left(\widehat{Z}_i-\widehat{z}\right)\right\} K\left\{\frac{(\widehat{Z}_i-\widehat{z})}{h}\right\}.$$

Then, $\widehat{G}_\tau(\widehat{z})=\widehat{a}$ and $\widehat{G}_\tau'(\widehat{z})=\widehat{b}$.

Theorem 3. *Under the regularity conditions* (i)−(iv) *in Appendix A, if* $h\to 0$ *and* $nh\to\infty$, *then the asymptotic conditional bias and variance of the local linear quantile regression estimator* $\widehat{G}_\tau(\widehat{z})$ *are*

$$\begin{aligned}\operatorname{bias}\{\widehat{G}_\tau(\widehat{z})\} &= G_\tau''(z)\mu_2\frac{h^2}{2} + o_p(h^2),\\ \operatorname{Var}\left\{\widehat{G}_\tau(\widehat{z})\right\} &= (nh)^{-1}\mathbf{\Delta}\left\{1+o_p(1)\right\},\end{aligned}$$

where $\mathbf{\Delta}=\tau(1-\tau)\nu_0/\left\{f(z)f_Y^2(\xi_\tau\mid z)\right\}=\tau(1-\tau)\nu_0\sigma^2(z)/\left\{f(z)f_\varepsilon^2(\xi_\tau^\star)\right\}$, $\nu_0=\int_{-1}^{1}K^2(u)du$, f_ε *is the density function of* ε, *and* $\xi_\tau^\star$ *is its* τ*-th quantile. Furthermore, as* $n\to\infty$, $h\to 0$ *and* $nh\to\infty$,

$$(nh)^{1/2}\left\{\widehat{G}_\tau(\widehat{z}) - G_\tau(z) - G_\tau''(z)\mu_2\frac{h^2}{2}\right\}$$

is asymptotically normal with mean zero and asymptotic variance $\mathbf{\Delta}$.

For univariate $\mathbf{x}$, the asymptotic bias and variance of nonparametric quantile regression can be found in Fan, Hu, and Truong (1994) and Yu and Jones (1998). We compared the asymptotic bias and variance of the local linear quantile regression estimator built upon $\{(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau, Y_i), i=1,\ldots,n\}$ with that of the estimator built upon $\{(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau, Y_i), i=1,\ldots,n\}$ in Fan, Hu, and Truong (1994), and found they performed equally well asymptotically. This implies that the resulting estimate of our procedure performs as well as an oracle estimate that replaces $\hat{\boldsymbol{\beta}}_\tau$ with its true value.

We turn to the issue of bandwidth selection. One usually chooses it to minimize the mean squared error (MSE) of nonparametric estimation. Theorem 3 indicates that the leading term of the MSE of $\widehat{G}_\tau(\widehat{z})$ is

$$\mathrm{MSE}\left\{\widehat{G}_\tau(\widehat{z})\right\} = \frac{1}{4}\left\{\mu_2 G_\tau''(z)\right\}^2 h^4 + \frac{\nu_0}{nhf(z)}\frac{\tau(1-\tau)}{f_\varepsilon^2(\xi_\tau^\star)}\sigma^2(z).$$

Thus, the bandwidth that minimizes the asymptotic MSE of $\widehat{G}_\tau(\widehat{z})$ is

$$h^{\mathrm{opt}}(\widehat{z}) = \left[\frac{4\nu_0\tau(1-\tau)}{\{\mu_2 f_\varepsilon(\xi_\tau^\star)\}^2}\frac{\sigma^2(z)/f(z)}{\{G_\tau''(z)\}^2}\right]^{1/5} n^{-1/5}. \tag{2.4}$$

This implies that the local quantile regression achieves the optimal rate of convergence $n^{2/5}$. To implement $h^{\mathrm{opt}}(\widehat{z})$, one has to replace all unknowns in (2.4) with their consistent estimators. This is usually computationally inefficient although all the unknown quantities are univariate nonparametric functions.

In the sequel we introduce a computationally efficient way to calculate the bandwidth for the quantile regression. We notice from Theorem 3 that the leading term of the asymptotic bias for the local linear quantile regression is the same as that for the local linear least squares estimator, whereas their asymptotic variances are different. The local least squares estimator (Fan and Gijbels (1996)) has MSE of the form

$$\mathrm{MSE}\left\{\widehat{G}_{\mathrm{LS}}(\widehat{z})\right\} = \frac{1}{4}\left\{\mu_2 G''(z)\right\}^2 h^4 + \frac{\nu_0}{nhf(z)}\sigma^2(z),$$

which results in the optimal bandwidth

$$h_m^{\mathrm{opt}}(\widehat{z}) = \left[\frac{4\nu_0}{\mu_2^2}\frac{\sigma^2(z)/f(z)}{\{G''(z)\}^2}\right]^{1/5} n^{-1/5}. \tag{2.5}$$

Comparing (2.5) with (2.4),

$$h^{\mathrm{opt}}(\widehat{z}) = h_m^{\mathrm{opt}}(\widehat{z})\left\{\frac{\tau(1-\tau)}{\{f_\varepsilon(\xi_\tau^\star)\}^2}\frac{\{G''(z)\}^2}{\{G_\tau''(z)\}^2}\right\}^{1/5}.$$

We introduce a rule of thumb bandwidth selector. If the curvatures of the τ-th quantile function $G_\tau''(z)$ and the conditional mean function $G''(z)$ are similar, and the error ε is close to standard normal, then we take

$$h^{\mathrm{opt}}(\widehat{z}) = h_m^{\mathrm{opt}}(\widehat{z})\left\{\frac{\tau(1-\tau)}{\{\phi\left(\xi_\tau^\star\right)\}^2}\right\}^{1/5}, \tag{2.6}$$

where $\phi(\cdot)$ denotes the probability density function of the standard normal distribution. There are many algorithms for calculating $h_m^{\text{opt}}(\widehat{z})$ (Fan and Gijbels (1996)), hence $h^{\text{opt}}(\widehat{z})$. One can estimate $\xi_\tau^\star$ through the sample τ-th quantile of the residuals $\widehat{\varepsilon}_i = \{Y_i - \widehat{G}(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau)\}/\widehat{\sigma}(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau)$, for $i = 1, \ldots, n$. An alternative way is to replace $\xi_\tau^\star$ in (2.6) with the τ-th quantile of the standard normal to further simplify the calculation.

The idea of this approximation originated with Yu and Jones (1998) and Yu and Lu (2004). Although it is built upon several assumptions, it provides a computationally efficient way to calculate the bandwidth for the quantile regression. We find it performs quite well in usual practice.

3. Simulations and Application

In this section we report on simulations to compare the performance of the proposed methods with existing competitors and to illustrate the proposed methodology with a data example.

3.1. Simulation

We generated 1,000 datasets, each consisting of $n = 500$ observations, from

$$Y = \sin\left\{2\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)\right\} + 2\exp\left\{-16\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)^2\right\} + \sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)\varepsilon, \tag{3.1}$$

where the index parameter $\boldsymbol{\beta}_0 = (2, -2, -1, 1, 0, \ldots, 0)^{\mathrm{T}}/\sqrt{10}$ is a $p \times 1$ vector, and the covariate vector $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}}$ was generated as multivariate normal with mean zero and covariance matrix $\mathrm{Var}(\mathbf{x}) = (\sigma_{ij})_{p\times p}$ with $\sigma_{ij} = 0.5^{|i-j|}$. The conditional mean function at (3.1) was designed by Kai, Li, and Zou (2010). In our simulations, we chose $p = 10$, 20, and 50, and considered five error distributions for ε: (i) the standard normal $N(0, 1)$; (ii) the mixture $0.8N(0, 1) + 0.2N(0, 9)$; (iii) the Laplace distribution; (iv) the student-t distribution with 3 degrees of freedom $t(3)$; and (v) the Cauchy distribution. The error term ε and the covariates $\mathbf{x}$ were mutually independent. We took $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right) = 1$ in the homoscedasticity case (1), and $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right) = \exp\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)$ in the heteroscedasticity case (2). The response values contain outliers with larger probabilities in case (2) than in case (1). We compared the performance of our procedures with the back-fitting algorithm proposed by Wu, Yu and Yu (2010). Though, from our limited experience, the back-fitting algorithm demands much computing time. We compared the simulation results of the back-fitting algorithm based on 100 and 1,000 replications for one case, and found that the simulation results based on 100 replications were almost the same as those based on 1,000 replications. Thus, we report the results of the back-fitting algorithm based on 100 replications in order to save computing time.

Performance in estimating $\boldsymbol{\beta}_0$. The index parameter $\boldsymbol{\beta}_0$ was estimated via a series of quantile regressions with $\tau = 0.025$, 0.05, 0.5, 0.95 and 0.975, respectively. Table 1 depicts the averages of mean squared errors (MSE) of the estimate $\widehat{\boldsymbol{\beta}}_\tau$, defined by $\text{MSE} = \left\| \widehat{\boldsymbol{\beta}}_\tau / \|\widehat{\boldsymbol{\beta}}_\tau\| - \boldsymbol{\beta}_0 \right\|^2$. We included three competitors when $p = 10$: (i) the linear quantile regression formulated in (2.3); (ii) the back-fitting algorithm proposed by Wu, Yu and Yu (2010); and (iii) the ordinary least squares estimate (LSE). Li and Duan (1989) proved that this LSE is a consistent estimator of the direction of $\boldsymbol{\beta}_0$; thus, it serves naturally as a competitor here. It can be seen from Table 1 that the back-fitting algorithm performs the best in most scenarios. This is expected in that the back-fitting algorithm updates the nonparametric quantile function in a data-driven manner while estimating $\boldsymbol{\beta}_0$. By contrast, the linear quantile regression assumes a linear quantile function to reduce computational complexity. Yet the linear quantile regression also has a satisfactory performance. In case (1) with Cauchy errors, the linear quantile regression performs even better than the back-fitting algorithm in terms of the MSE values. In case (1) with standard normal errors, the LSE performs slightly better than the linear quantile regression. However, the linear quantile regression is superior to the LSE in all other cases. The improvement of linear quantile regression over the LSE is more significant in case (2) than in case (1). This is expected in that the performance of the LSE is sensitive to the presence of outliers. The quantile regression offers a more robust estimation in most scenarios.

Performance in coverage probability of prediction intervals. It is of interest to evaluate the accuracy of quantile regression in its prediction interval of Y given $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0$. Here the τ-th and $(1-\tau)$-th quantile of the conditional distribution of Y given $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0$ can be used as a confidence interval at the level of $(1-2\tau)$ for $\tau < 0.5$. We choose $\tau = 0.025$ and 0.05 to provide confidence intervals at 95% and 90%, and expect the empirical coverage probability to be close to the nominal level of $(1-2\tau)$.

We include three competitors in our comparison: (i) the back-fitting algorithm (Wu, Yu and Yu (2010)) which can simultaneously estimate the index parameter $\boldsymbol{\beta}_\tau$ and the quantile function; (ii) the quantile regression method introduced in Section 2.2 based on the estimated index parameter $\widehat{\boldsymbol{\beta}}_\tau$; and (iii) the oracle estimator that provides the prediction interval by using the true index parameter $\boldsymbol{\beta}_0$. The average and the standard deviation of the empirical coverage probabilities are summarized in Table 2.

Both the homoscedasticity and the heteroscedasticity cases reported in Table 2 show similar messages. The coverage probabilities of the three competitors are all close to the nominal levels for different τ values, although the back-fitting algorithm uses a more accurate estimate of $\boldsymbol{\beta}_0$. The standard deviations of the

Table 1. The averages of MSEs of $\widehat{\boldsymbol{\beta}}_\tau$ for model (3.1) with $p = 10$.

case (1): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right) = 1$						case (2): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right) = \exp\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)$					
the τ-th quantile in %						*the τ-th quantile in %*					
2.5	5	50	LSE	95	97.5	2.5	5	50	LSE	95	97.5
Linear quantile regression											
$\varepsilon \sim N(0,1)$											
0.200	0.191	0.140	0.113	0.292	0.305	0.173	0.164	0.107	0.265	0.230	0.245
$\varepsilon \sim 0.8N(0,1) + 0.2N(0,9)$											
0.259	0.246	0.173	0.197	0.347	0.362	0.236	0.223	0.145	0.484	0.289	0.301
$\varepsilon \sim$ Laplace distribution											
0.144	0.136	0.097	0.110	0.219	0.230	0.127	0.121	0.082	0.255	0.186	0.197
$\varepsilon \sim t(3)$											
0.249	0.239	0.171	0.217	0.344	0.359	0.222	0.210	0.139	0.498	0.290	0.303
$\varepsilon \sim$ Cauchy distribution											
0.376	0.359	0.237	1.441	0.452	0.474	0.371	0.354	0.215	1.430	0.420	0.442
Back-fitting algorithm											
$\varepsilon \sim N(0,1)$											
0.137	0.118	0.069	–	0.164	0.195	0.105	0.094	0.048	–	0.117	0.135
$\varepsilon \sim 0.8N(0,1) + 0.2N(0,9)$											
0.319	0.232	0.086	–	0.284	0.316	0.136	0.112	0.069	–	0.163	0.184
$\varepsilon \sim$ Laplace distribution											
0.119	0.104	0.045	–	0.128	0.167	0.098	0.087	0.035	–	0.115	0.125
$\varepsilon \sim t(3)$											
0.269	0.218	0.089	–	0.260	0.314	0.162	0.146	0.083	–	0.142	0.168
$\varepsilon \sim$ Cauchy distribution											
0.598	0.612	0.164	–	0.645	0.655	0.336	0.257	0.158	–	0.305	0.394

empirical probabilities are very small, indicating that the quantile regression offers a very reliable coverage probability that is comparable with its oracle version.

Computational efficacy. It is of interest to compare the computational efficacy of our method with that of the back-fitting algorithm proposed by Wu, Yu and Yu (2010). Table 3 summarizes the average computing time in seconds used for estimating the index parameter and calculating the prediction intervals for one replication. It can be seen from Table 3 that our method is much faster than the back-fitting algorithm. In addition, the computing time of the back-fitting algorithm varies over the error distributions.

Performance in high dimension. Next we investigate the performance of the proposed method when p is relative large. We still use model (3.1) for illustration purposes, but increase the dimension p of the covariates. Tables 4 and 5 report, respectively, the MSE values and the coverage probabilities for model (3.1) with $p = 20$ and 50. Table 4 together with Table 1 indicates that the MSE values of $\widehat{\boldsymbol{\beta}}_\tau$ increase significantly as the covariate dimension p increases. However, the covariate dimension p has little effect on estimating the coverage probabilities of

Table 2. The empirical coverage probabilities for model (3.1) with $p = 10$.

	case (1): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)=1$						case (2): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)=\exp\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)$					
	quantile regression				Back-fitting		quantile regression				Back-fitting	
	New		Oracle		algorithm		New		Oracle		algorithm	
Level	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\varepsilon \sim N(0,1)$												
aver.	89.4	94.5	89.8	94.8	90.5	95.5	90.1	95.1	90.9	95.5	91.1	95.9
stdev	0.5	0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.5	0.4
$\varepsilon \sim 0.8N(0,1)+0.2N(0,9)$												
aver.	89.7	94.7	89.9	94.7	90.6	95.5	90.3	95.0	90.7	95.9	91.1	95.8
stdev	0.4	0.3	0.5	0.3	0.5	0.4	0.4	0.4	0.5	0.4	0.4	0.5
$\varepsilon \sim$ Laplace distribution												
aver.	89.7	94.7	89.8	94.8	90.6	95.6	90.3	95.1	90.9	95.5	91.1	95.9
stdev	0.4	0.3	0.5	0.4	0.6	0.4	0.4	0.4	0.5	0.4	0.5	0.4
$\varepsilon \sim t(3)$												
aver.	89.7	94.7	89.8	94.8	90.6	95.5	90.3	95.1	90.8	95.4	91.2	95.9
stdev	0.4	0.3	0.5	0.4	0.6	0.4	0.4	0.4	0.5	0.4	0.5	0.4
$\varepsilon \sim$ Cauchy distribution												
aver.	90.4	95.4	90.4	95.7	90.9	95.7	90.8	95.6	91.0	95.7	91.2	95.9
stdev	0.5	0.4	0.5	0.5	0.7	0.4	0.5	0.4	0.5	0.4	0.5	0.5

Table 3. The averages of computing times (in seconds) for model (3.1) with $p = 10$.

Case	(1): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)=1$		(2): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)=\exp\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)$	
Error	New	back-fitting	New	back-fitting
$N(0,1)$	0.11	1018.57	0.10	1029.58
$0.8N(0,1)+0.2N(0,9)$	0.11	1472.64	0.11	1490.92
Laplace distribution	0.11	1531.51	0.09	1551.81
$t(3)$	0.12	1386.17	0.10	1401.70
Cauchy distribution	0.13	1429.88	0.11	1426.16

the prediction intervals, as can be seen from Tables 2 and 5. This benefits essentially from the single-index structure of the quantile function. These simulation results support our methodology in high-dimensional setting.

3.2. An application

We illustrate the proposed procedures by an empirical analysis of an automobile dataset (Johnson (2003)). It is of interest to know how the manufactures' suggested retail price (MSRP) of vehicles depends upon different levels of such attributes as miles per gallon and horsepower. Thus, MSRP serves naturally as the response variable. This is the manufacturer's assessment of a vehicle's worth, and includes adequate profit for the manufacturer and the dealer. We use the logarithmic transformation of the MSRP in U.S. dollars as the response (Y).

Table 4. The averages of MSEs of $\widehat{\boldsymbol{\beta}}_\tau$ for model (3.1) with $p = 20$ and 50.

case (1): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right) = 1$						case (2): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right) = \exp\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)$					
the τ-th quantile in %						*the τ-th quantile in %*					
2.5	5	50	LSE	95	97.5	2.5	5	50	LSE	95	97.5
$p = 20$											
$\varepsilon \sim N(0,1)$											
0.380	0.365	0.271	0.230	0.516	0.535	0.371	0.353	0.231	0.439	0.450	0.468
$\varepsilon \sim 0.8N(0,1) + 0.2N(0,9)$											
0.476	0.458	0.345	0.380	0.619	0.640	0.483	0.460	0.300	0.722	0.539	0.561
$\varepsilon \sim$ Laplace distribution											
0.294	0.280	0.213	0.232	0.438	0.458	0.290	0.275	0.185	0.446	0.400	0.419
$\varepsilon \sim t(3)$											
0.444	0.427	0.331	0.387	0.606	0.628	0.470	0.445	0.290	0.736	0.535	0.559
$\varepsilon \sim$Cauchy distribution											
0.651	0.624	0.447	1.633	0.754	0.778	0.701	0.671	0.431	1.597	0.708	0.733
$p = 50$											
$\varepsilon \sim N(0,1)$											
0.757	0.733	0.586	0.504	0.941	0.967	0.840	0.804	0.539	0.780	0.863	0.888
$\varepsilon \sim 0.8N(0,1) + 0.2N(0,9)$											
0.895	0.868	0.689	0.718	1.041	1.067	1.029	0.991	0.660	1.065	0.970	0.995
$\varepsilon \sim$ Laplace distribution											
0.670	0.643	0.496	0.506	0.878	0.906	0.704	0.671	0.465	0.756	0.817	0.844
$\varepsilon \sim t(3)$											
0.890	0.859	0.667	0.734	1.008	1.034	1.009	0.972	0.656	1.083	0.972	0.997
$\varepsilon \sim$Cauchy distribution											
1.175	1.145	0.889	1.777	1.253	1.283	1.313	1.275	0.865	1.749	1.165	1.189

In addition, there are seven features which possibly affect the resulting price of vehicles: engine size (X_1), number of cylinders (X_2), horsepower (X_3), average city miles per gallon (MPG, X_4), average highway MPG (X_5), weight in pounds (X_6), and wheel base in inches (X_7). This dataset consists of 428 observations, sixteen of which have missing values. We remove those observations with missing values in our subsequent analysis. The remaining dataset has a total of eight variables and 412 observations. Both the response variable Y and the covariate vector $\mathbf{x} = (X_1, \ldots, X_7)^{\mathrm{T}}$ are standardized marginally to have zero mean and unit variance. The fully nonparametric regression is not applicable to this particular case because the sample size is small compared with the dimensionality of the covariates.

We specify the heteroscedastic single-index model (1.1) for analyzing the new vehicle data. The index parameter is estimated at five different quantiles: $\tau = 0.025$, 0.05, 0.50, 0.95 and 0.975. This yields similar estimates for the index parameter $\boldsymbol{\beta}_\tau$ at different quantiles, indicating that similar patterns between the

Table 5. The empirical coverage probabilities for model (3.1) with $p = 20$ and 50.

	Case	(1): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right) = 1$				(2): $\sigma\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right) = \exp\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)$			
		New		Oracle		New		Oracle	
Error	Level	90%	95%	90%	95%	90%	95%	90%	95%
					$p = 20$				
N(0,1)	aver.	90.3	95.3	90.9	95.9	90.8	95.7	91.4	96.1
	stdev	0.4	0.4	0.5	0.4	0.5	0.3	0.5	0.4
$0.8N(0,1)$	aver.	90.4	95.3	90.9	95.7	90.8	95.6	91.3	95.9
$+0.2N(0,9)$	stdev	0.4	0.3	0.5	0.4	0.4	0.3	0.5	0.5
Laplace	aver.	90.4	95.4	91.0	95.8	90.8	95.7	91.4	96.0
	stdev	0.4	0.3	0.5	0.5	0.4	0.3	0.5	0.5
$t(3)$	aver.	90.3	95.3	90.8	95.7	90.8	95.6	91.2	96.0
	stdev	0.4	0.4	0.5	0.4	0.5	0.4	0.5	0.4
Cauchy	aver.	90.6	95.3	90.6	95.5	90.6	95.4	91.0	95.7
	stdev	0.7	0.4	0.9	0.7	0.5	0.4	0.4	0.4
					$p = 50$				
$N(0,1)$	aver.	90.3	95.3	90.9	95.8	90.7	95.6	91.3	96.2
	stdev	0.4	0.3	0.5	0.4	0.4	0.4	0.4	0.5
$0.8N(0,1)$	aver.	90.4	95.3	90.9	95.7	90.6	95.5	91.2	95.9
$+0.2N(0,9)$	stdev	0.4	0.3	0.5	0.4	0.5	0.3	0.5	0.4
Laplace	aver.	90.3	95.3	91.0	95.8	90.6	95.5	91.3	96.1
	stdev	0.4	0.3	0.6	0.5	0.4	0.3	0.5	0.4
$t(3)$	aver.	90.4	95.3	90.9	95.7	90.7	95.5	91.3	96.0
	stdev	0.4	0.3	0.5	0.4	0.5	0.4	0.5	0.4
Cauchy	aver.	90.6	95.5	90.4	95.4	90.6	95.5	91.1	95.8
	stdev	0.5	0.4	0.8	0.7	0.5	0.4	0.7	0.6

MSRP and different features are revealed at different quantiles. Specifically, the estimated index parameter at $\tau = 0.5$ is $\widehat{\boldsymbol{\beta}}_\tau = (0.3254, -0.1846, -0.6332, 0.3688, -0.3326, -0.4156, 0.1994)^{\mathrm{T}}$. Table 6 reports that the standard deviation of $\widehat{\boldsymbol{\beta}}_\tau$, obtained from a nonparametric bootstrap procedure, is $(0.052, 0.046, 0.051, 0.069, 0.057, 0.043, 0.032)^{\mathrm{T}}$. This implies that the horsepower (X_3) is perhaps the most important factor that affects the suggested retail price (Y), followed by the weight in pounds (X_6).

Using the data $\{(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau, Y_i), i = 1, \ldots, n\}$, we first estimated the regression function $G(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$ via the local least squares estimator. The estimated function and its 95% confidence interval are reported in Figure 1(C). We further conducted some exploratory data analysis on this dataset. The boxplot of the log-transformed MSRP, the response variable, is depicted in Figure 1(A), which shows that there are four vehicles whose prices are significantly higher than other vehicles. In addition, the histogram in Figure 1(B) reveals that the distribution

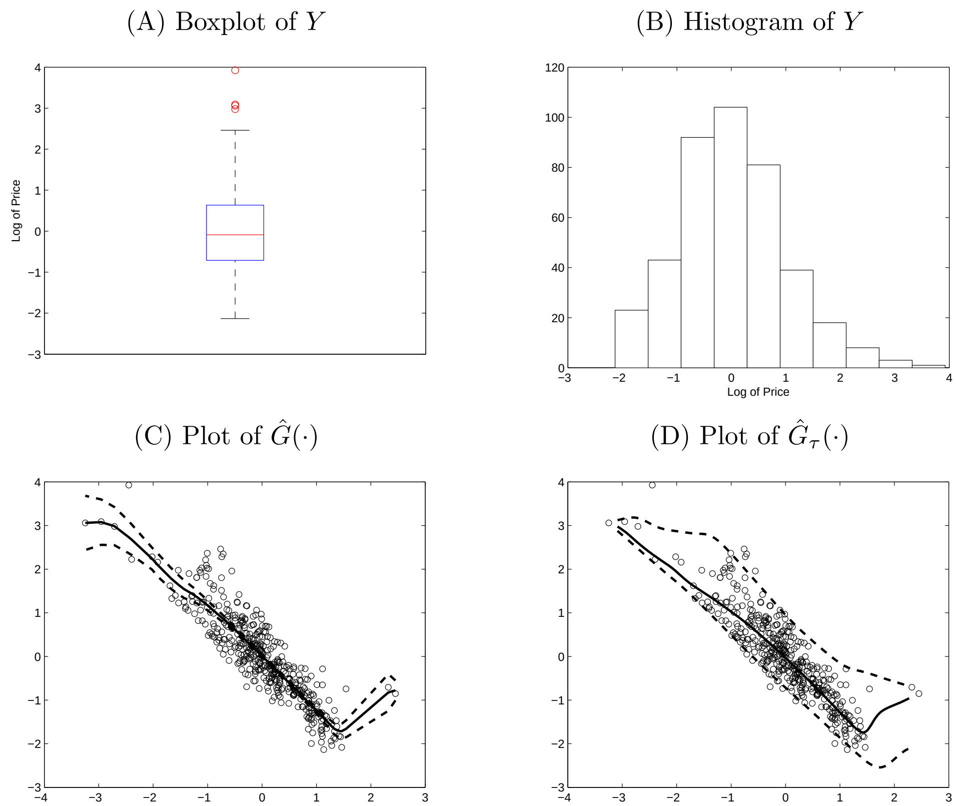

Figure 1. (A) Boxplot of log-transformed retail price. (B) Histogram of log-transformed retail price. (C) Plot of $\hat{G}(\cdot)$ and its 95% confidence interval using the local linear least squares method. (D) Plot of $\hat{G}_\tau(\cdot)$ with $\tau = 2.5\%$, 50% and 97.5%. The 95% prediction interval of Y given $\mathbf{x}$ covers 96% points

of the log-transformed prices are highly skewed. This motivated us to further conduct empirical analysis of this dataset using the proposed procedure.

Next we applied quantile regression to this dataset. Figure 1(D) presents the 2.5% and 97.5% quantiles to give an approximate 95% prediction interval of Y. This prediction interval covers 95.87% of the data points, which is close to the nominal level. The line in the middle of Figure 1(D) presents the quantile regression at the level of 50%. It behaves similarly to the local least squares estimator in Figure 1(C) within the range of $\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau$. When $\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau$ is near to the boundary, however, the local least squares estimate is clearly more sensitive to the outliers than the quantile estimate. This example demonstrates the effectiveness of our procedure in terms of description of the conditional distribution of Y given $\mathbf{x}$.

We further used a bootstrap procedure to demonstrate the performance of

Table 6. The average and standard deviation of the estimated index parameter based on 500 bootstrap samples of the new vehicle data.

ε_i^* sampled from		X_1	X_2	X_3	X_4	X_5	X_6	X_7
F_n	aver.	0.325	-0.185	-0.633	0.369	-0.333	-0.416	0.199
	stdev.	0.052	0.046	0.051	0.069	0.057	0.043	0.032
$0.5F_n+$	aver.	0.305	-0.173	-0.630	0.351	-0.341	-0.428	0.191
$0.5N(0, 0.4332^2)$	stdev.	0.058	0.051	0.051	0.073	0.060	0.047	0.036
$.5F_n$ +	aver.	0.303	-0.169	-0.628	0.314	-0.318	-0.421	0.184
.5Cauchy	stdev.	0.088	0.081	0.129	0.136	0.110	0.098	0.062

our method in the presence of outliers. Given the covariates we bootstrapped new response variables by using the residuals. To calculate the residuals, we fixed $\tau = 0.5$ and estimated $\boldsymbol{\beta}_\tau$ and $G_\tau(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_\tau)$ to obtain $\widehat{\varepsilon}_i = Y_i - \widehat{G}_\tau(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau)$ for $i = 1, \ldots, n$. For ease of illustration, we denote by F_n the empirical distribution of the residuals $\widehat{\varepsilon}_i$. We bootstrapped 500 samples each of form $\{(\mathbf{x}_i, Y_i^*), i = 1, \ldots, n\}$, where $Y_i^* = \widehat{G}_\tau(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau) + \widehat{\varepsilon}_i^*$, and ε_i^* was generated from three different distributions: (1) F_n; (2) $0.5F_n + 0.5N(0, 0.4332^2)$, since the standard deviation of the original residuals was exactly 0.4332; and (3) $0.5F_n + 0.5$Cauchy. In case (3) the distribution of the response variable has a heavy tail.

We estimated the index parameter $\boldsymbol{\beta}_\tau$ at different quantiles, and summarized the average and standard deviation of $\widehat{\boldsymbol{\beta}}_\tau$ based on the bootstrap samples. Because the results for different quantiles show similar messages, we only report the results for $\tau = 0.5$, which are depicted in Table 6. The averages of the estimated index parameter $\widehat{\boldsymbol{\beta}}_\tau$ are similar. This once again suggests that our method is robust to the presence of outliers. By contrast, the standard deviations of these three cases are slightly different, the standard deviations of $\widehat{\boldsymbol{\beta}}_\tau$ in case (1) and (2) are similar, both of which have smaller standard deviations than case (3). This can be interpreted to mean that the bootstrapped residuals in case (3) have substantially larger variance than in cases (1) and (2).

4. Discussion

To estimate the quantile regression with high-dimensional covariates, we impose a heteroscedastic single-index structure on the regression function. The heteroscedastic single-index structure allows us to reduce the dimension of covariates and simultaneously retain the flexibility of nonparametric regression. We propose a computationally efficient two-step estimation procedure to estimate the parameters involved in the quantile regression function. Asymptotic properties of the proposed procedures are studied.

It is remarkable here that, if (1.1) reduces to the homoscedastic single-index model, and the error ε is symmetric about zero, then a byproduct is that our estimation procedure offers a robust estimation for the conditional mean function. Simulation studies support this point implicitly. In addition, it is not necessary that the error term ε in model (1.1) be independent of $\mathbf{x}$; we can assume instead that $E\left(\mathbf{x} \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0, \varepsilon\right)$ is a linear function of $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0$. If the parameter of interest is the index parameter $\boldsymbol{\beta}_0$ only, the composite quantile regression (Zou and Yuan (2008)) can be adapted to improve the efficiency in estimating $\boldsymbol{\beta}_0$.

We suggest quantile regression to construct prediction intervals. This is in spirit a local pointwise prediction interval. How to construct global confidence band is an interesting question. Härdle and Song (2010) investigated the confidence band in univariate quantile regression. Their results are readily applicable to the single-index model (1.1) when $\boldsymbol{\beta}_0$ is known. Further research is needed to quantify the effect when $\boldsymbol{\beta}_0$ is replaced with its root-n consistent estimate.

Acknowledgements

Zhu was supported by a National Natural Science Foundation of China grant 11071077 and a National Institute on Drug Abuse (NIDA) grant R21-DA024260. Huang is the corresponding author and was supported by a funding through Projet 211 Phase 3 of SHUFE, and Shanghai Leading Academic Discipline Project, B803. Li was supported by NSF grant DMS 0348869, National Natural Science Foundation of China grants 11028103 and and 10911120395, and National Institute on Drug Abuse (NIDA) grant P50-DA10075. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NIH.

Appendix: Technical Conditions and Proofs

A.1. Technical conditions

The following technical conditions are imposed. They are not the weakest possible conditions, but are imposed to facilitate the proofs.

(i) The quantile function $G_\tau(\cdot)$ has a continuous and bounded second derivative.

(ii) The density function $f(\cdot)$ of $(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})$ is positive and uniformly continuous for $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$. Furthermore, the density function of $(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$ is continuous and bounded away from zero and infinity on its support.

(iii) The conditional density of Y given $(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$, denoted by $f_Y(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$, is continuous in $(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$ for each $y \in \mathbb{R}$. Moreover, there exist positive constants ε and δ and a positive function $F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$ such that

$$\sup_{\|\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}-\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\| \le \varepsilon} f_Y(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) \le F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0).$$

For any fixed value of $(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$, $\int F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)\,dy < \infty$; and as $t \to 0$, $\int \left\{\rho_\tau(y-t) - \rho_\tau(y) - \rho_\tau'(y)\,t\right\}^2 F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)\,dy = o(t^2)$.

(iv) The kernel function $K(\cdot)$ is symmetric and has compact support $[-1, 1]$. It satisfies the first-order Lipschitz condition.

Condition (i) is commonly assumed for a link function in the literature. Condition (ii) guarantees the existence of any ratio terms with the density function $f(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$ appearing in the denominator. Condition (iii) is weaker than the Lipschitz condition of the function $\rho'_\tau(\cdot)$. The check loss function $\rho_\tau(\cdot)$ is piecewise linear and non-differentiable at 0. We take $\rho'_\tau(u) = \mathrm{sign}(u) + (2\tau - 1)$ for $u \neq 0$ and $\rho'_\tau(0) = 0$. With Taylor's expansion, we obtain $\int \{\rho_\tau(y-t) - \rho_\tau(y) - \rho'_\tau(y)\, t\}^2 F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)\, dy = t^2 \int \{\mathrm{sign}(u^*) - \mathrm{sign}(y)\}^2 F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0) dy$ for some u^* between $y - t$ and y. Without loss of generality we assume $t > 0$. Then

$$\begin{aligned}
&\int_{-\infty}^{\infty} \{\mathrm{sign}(u^*) - \mathrm{sign}(y)\}^2 F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0) dy \\
&\quad = \int_0^t \{\mathrm{sign}(u^*) - \mathrm{sign}(y)\}^2 F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0) dy \leq 4 \int_0^t F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0) dy.
\end{aligned}$$

Thus, $\int \{\rho_\tau(y-t) - \rho_\tau(y) - \rho'_\tau(y)\, t\}^2 F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)\, dy = o(t^2)$ is implied by $\int_0^t F(y \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0) dy$ as $t \to 0$. Thus condition (iii) is mild. Condition (iv) requires that the kernel function be a proper density function with compact support.

A.2. Proof of Theorem 1

Theorem 1 follows from the convexity of the check loss function and the linearity condition (2.2). It suffices to prove that, for any constant $\mathrm{u} \in \mathbb{R}$, there exists a constant κ such that $\mathcal{L}_\tau(\mathrm{u}, \boldsymbol{\beta}) \geq \mathcal{L}_\tau(\mathrm{u}, \kappa\boldsymbol{\beta}_0)$. Specifically,

$$\begin{aligned}
\mathcal{L}_\tau(\mathrm{u}, \boldsymbol{\beta}) &= E\left[E\left\{\rho_\tau\left(Y - \mathrm{u} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}, \varepsilon\right\}\right] \\
&\geq E\left[\rho_\tau\left\{E\left(Y \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}, \varepsilon\right) - \mathrm{u} - E\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}, \varepsilon\right)\right\}\right] \\
&= E\left[\rho_\tau\left\{Y - \mathrm{u} - E\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} \mid \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}\right)\right\}\right] \\
&= E\left\{\rho_\tau\left(Y - \mathrm{u} - \kappa\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}\right)\right\},
\end{aligned}$$

where $\kappa = \boldsymbol{\beta}^{\mathrm{T}}\mathrm{Var}(\mathbf{x})\boldsymbol{\beta}_0 / \left\{\boldsymbol{\beta}_0^{\mathrm{T}}\mathrm{Var}(\mathbf{x})\boldsymbol{\beta}_0\right\}$. The first equality follows from the iterative law of conditional expectation; the first inequality follows from Jensen's inequality and the convexity of ρ_τ; the second equality is true in view of model (1.1), and the last equality holds true by invoking the linearity condition (2.2) and the error ε is independent of $\mathbf{x}$. This completes the proof.

A.3. Proof of Theorem 2

This proof follows from the quadratic approximation lemma (Hjort and Pollard (1993)).

Quadratic Approximation Lemma. Suppose $\mathcal{L}_{\tau n}(\boldsymbol{\beta})$ is convex and can be represented as $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{B}\boldsymbol{\beta}/2+\mathrm{u}_n^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{a}_n+R_n(\boldsymbol{\beta})$, where $\mathbf{B}$ is symmetric and positive definite, u_n is stochastically bounded, $\mathbf{a}_n$ is arbitrary, and $R_n(\boldsymbol{\beta})$ goes to zero in probability for each $\boldsymbol{\beta}$. Then $\widehat{\boldsymbol{\beta}}$, the minimizer of $\mathcal{L}_{\tau n}(\boldsymbol{\beta})$, is only $o_p(1)$ away from $-\mathbf{B}^{-1}\mathrm{u}_n$. If u_n converges in distribution to u, then $\widehat{\boldsymbol{\beta}}$ converges in distribution to $-\mathbf{B}^{-1}\mathrm{u}$ accordingly.

Let $\alpha_n=n^{-1/2}$ and $\mathbf{a}=n^{1/2}(\widehat{\boldsymbol{\beta}}_\tau-\boldsymbol{\beta}_\tau)$. Further, let $\mathcal{L}_{\tau n}(\mathrm{u},\boldsymbol{\beta})=n^{-1}\sum_{i=1}^n \rho_\tau(Y_i-\mathrm{u}-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$. By applying the identity (Knight (1998)) that

$$\rho_\tau(x-y)-\rho_\tau(x)=2\left[y\left\{\mathbf{1}(x\le 0)-\tau\right\}+\int_0^y\left\{\mathbf{1}(x\le t)-\mathbf{1}(x\le 0)\right\}dt\right], \quad \text{(A.1)}$$

it follows that

$$\begin{aligned}
&\mathcal{L}_{\tau n}(\xi_\tau+\alpha_n\mathbf{b},\boldsymbol{\beta}_\tau+\alpha_n\mathbf{a})-\mathcal{L}_{\tau n}(\xi_\tau,\boldsymbol{\beta}_\tau)\\
&\quad=n^{-1}\sum_{i=1}^n\rho_\tau(Y_i-\xi_\tau-\alpha_n\mathbf{b}-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau-\alpha_n\mathbf{x}_i^{\mathrm{T}}\mathbf{a})-n^{-1}\sum_{i=1}^n\rho_\tau(Y_i-\xi_\tau-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau)\\
&\quad=n^{-1}\sum_{i=1}^n\alpha_n\left(\mathbf{x}_i^{\mathrm{T}}\mathbf{a}+\mathbf{b}\right)\left\{\mathbf{1}(Y_i-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau\le\xi_\tau)-\tau\right\}\\
&\qquad+n^{-1}\sum_{i=1}^n\int_0^{\alpha_n\left(\mathbf{x}_i^{\mathrm{T}}\mathbf{a}+\mathbf{b}\right)}\left\{\mathbf{1}(Y_i-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau\le\xi_\tau+t)-\mathbf{1}(Y_i-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau\le\xi_\tau)\right\}dt.
\end{aligned}$$

The objective function $\mathcal{L}_{\tau n}(\xi_\tau+\alpha_n\mathbf{b},\boldsymbol{\beta}_\tau+\alpha_n\mathbf{a})-\mathcal{L}_{\tau n}(\xi_\tau,\boldsymbol{\beta}_\tau)$ is convex in $(\mathbf{a},\mathbf{b})$. With a slight change of notation, $F(t\mid\mathbf{x})=E\left\{\mathbf{1}(Y-\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_\tau\le t)\mid\mathbf{x}\right\}$. By Taylor's expansion, it follows that

$$\begin{aligned}
&E\left\{\mathcal{L}_{\tau n}(\xi_\tau+\alpha_n\mathbf{b},\boldsymbol{\beta}_\tau+\alpha_n\mathbf{a})-\mathcal{L}_{\tau n}(\xi_\tau,\boldsymbol{\beta}_\tau)\mid\mathbf{x}_1,\ldots,\mathbf{x}_n\right\}\\
&\quad=n^{-1}\sum_{i=1}^n\left[\alpha_n\left(\mathbf{x}_i^{\mathrm{T}}\mathbf{a}+\mathbf{b}\right)\left\{F(\xi_\tau\mid\mathbf{x}_i)-\tau\right\}\right.\\
&\qquad\left.+\int_{\xi_\tau}^{\xi_\tau+\alpha_n\left(\mathbf{x}_i^{\mathrm{T}}\mathbf{a}+\mathbf{b}\right)}F(t\mid\mathbf{x}_i)-F(0\mid\mathbf{x}_i)dt\right]\\
&\quad=\frac{\alpha_n}{n}\sum_{i=1}^n\left[\left(\mathbf{x}_i^{\mathrm{T}}\mathbf{a}+\mathbf{b}\right)\left\{F(\xi_\tau\mid\mathbf{x}_i)-\tau\right\}+\frac{\alpha_n^2}{2}f(\xi_\tau\mid\mathbf{x}_i)\left(\mathbf{b}+\mathbf{a}^{\mathrm{T}}\mathbf{x}_i\right)^2\right]+o_p(\alpha_n^2). \quad \text{(A.2)}
\end{aligned}$$

The first two terms are of order $O_p(1/n)$. The last term of (A.2) admits a quadratic function of $(\mathbf{a},\mathbf{b})$. Following standard arguments, we obtain that

$$\begin{aligned}
R_2&=\mathcal{L}_{\tau n}(\xi_\tau+\alpha_n\mathbf{b},\boldsymbol{\beta}_\tau+\alpha_n\mathbf{a})-\mathcal{L}_{\tau n}(\xi_\tau,\boldsymbol{\beta}_\tau)\\
&\quad-E\left\{\mathcal{L}_{\tau n}(\xi_\tau+\alpha_n\mathbf{b},\boldsymbol{\beta}_\tau+\alpha_n\mathbf{a})-\mathcal{L}_{\tau n}(\xi_\tau,\boldsymbol{\beta}_\tau)\mid\mathbf{x}_1,\ldots,\mathbf{x}_n\right\}=o_p\left(n^{-1}\right).
\end{aligned}$$

This, together with the quadratic approximation lemma and (A.2), leads to the asymptotic normality of $\widehat{\boldsymbol{\beta}}$.

A.4. Proof of Theorem 3

Recall the notations $z = \mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}_\tau$, $\widehat{z} = \mathbf{x}_0^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau$, $Z_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau$, $\widehat{Z}_i = \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau$. Let $a_0 = G_\tau(z) = G_\tau(\mathbf{x}_0^{\mathrm{T}}\boldsymbol{\beta}_\tau)$, $b_0 = G'_\tau(z)$, $\widehat{K}_i = K\{(\widehat{Z}_i - \widehat{z})/h\}$, $K_i = K\{(Z_i - z)/h\}$ and $K'_i = K'\{(Z_i - z)/h\}$ for $i = 1, \ldots, n$.

We first show that, for $\widehat{\boldsymbol{\beta}}_\tau$ to be a root-n consistent estimate of $\boldsymbol{\beta}_\tau$,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\rho_\tau\{Y_i - a - b(\widehat{Z}_i - \widehat{z})\}\widehat{K}_i - \rho_\tau\{Y_i - a - b(Z_i - z)\}K_i\right] = O_p(n^{-1/2}). \quad \text{(A.3)}$$

Define

$$R_{11} = n^{-1}\sum_{i=1}^{n} K_i\left[\rho_\tau\left\{Y_i - a - b\left(\widehat{Z}_i - \widehat{z}\right)\right\} - \rho_\tau\{Y_i - a - b(Z_i - z)\}\right],$$

$$R_{12} = n^{-1}\sum_{i=1}^{n}\left(\widehat{K}_i - K_i\right)\rho_\tau\{Y_i - a - b(Z_i - z)\},$$

$$R_{13} = E\left[\rho_\tau\left\{Y_i - a - b\left(\widehat{Z}_i - z\right)\right\} - \rho_\tau\{Y_i - a - b(Z_i - z)\}\right]\left(\widehat{K}_i - K_i\right).$$

Then, the left hand side of (A.3) can be written as $R_{11} + R_{12} + R_{13}$. Using the identity $\rho_\tau(x - y) - \rho_\tau(x) = 2\left[y\{\mathbf{1}(x \le 0) - \tau\} + \int_0^y \{\mathbf{1}(x \le z) - \mathbf{1}(x \le 0)\}\,dz\right]$, we have $|\rho_\tau(x - y) - \rho_\tau(x)| \le 4|y|$. Invoking the root-$n$ consistency of $\widehat{\boldsymbol{\beta}}_\tau$, we obtain

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\left[K_i\left|\rho_\tau\left\{Y_i - a - b\left(\widehat{Z}_i - \widehat{z}\right)\right\} - \rho_\tau\{Y_i - a - b(Z_i - z)\}\right|\right] \\ & \quad \le 4b\sup_u |K(u)| n^{-1}\sum_{i=1}^{n}\left|\widehat{Z}_i - Z_i\right| = 4bn^{-1}\sum_{i=1}^{n}\left|\mathbf{x}_i^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau\right)\right| = O_p(n^{-1/2}). \end{aligned}$$

This indicates that $|R_{11}| = O_p(n^{-1/2})$.

Next we deal with R_{12}. R_{12} can be written as

$$\left(\widehat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau\right)^{\mathrm{T}}(nh)^{-1}\sum_{i=1}^{n} K'\left(\frac{Z_i - z}{h}\right)(\mathbf{x}_i - \mathbf{x}_0)\,\rho_\tau\{Y_i - a - b(Z_i - z)\}\{1 + o_p(1)\}.$$

Let $\mathbf{z}_i = (\mathbf{x}_i - \mathbf{x}_0)/h$. Then

$$\begin{aligned} & (nh)^{-1}\sum_{i=1}^{n} K'\left(\frac{Z_i - z}{h}\right)(\mathbf{x}_i - \mathbf{x}_0)\,\rho_\tau\{Y_i - a - b(Z_i - z)\} \\ & \quad = n^{-1}\sum_{i=1}^{n} K'\left(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau\right)\mathbf{z}_i\rho_\tau\left(Y_i - a - bh\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}_\tau\right) = O_p(1), \end{aligned}$$

which, together with the root-n consistency of $\widehat{\boldsymbol{\beta}}_\tau$, proves that $R_{12} = O_p\left(n^{-1/2}\right)$.

It remains to investigate the order of R_{13}, but following similar arguments, we have $R_{13} = O_p(n^{-1/2})$.

Thus, (A.3) follows and implies that

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\rho_\tau\left\{Y_i - a - b\left(\widehat{Z}_i - \widehat{z}\right)\right\}\widehat{K}_i \\
&\quad = n^{-1}\sum_{i=1}^{n}\rho_\tau\left\{Y_i - a - b\left(Z_i - z\right)\right\}K_i + o_p\left\{(nh)^{-1/2}\right\}.
\end{aligned}
$$

That is, the quantile regression estimate of $\widehat{G}(\cdot)$ based on $\{(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\tau, Y_i),\ i = 1, \ldots, n\}$ is asymptotically as efficient as that based on $\{(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0, Y_i), i = 1, \ldots, n\}$.

The rest of the proof follows literally from Theorem 3 of Fan, Hu, and Truong (1994). Based on the auxiliary data points $\left\{(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0, Y_i), i = 1, \ldots, n\right\}$, the technique for establishing the asymptotic normality involves the convexity lemma (Pollard (1991)) and the quadratic approximation lemma (Hjort and Pollard (1993)). We only sketch the outline below.

For the sake of clarity, we write $\zeta_i = (a - a_0) + (b - b_0)(Z_i - z)$. By using (A.1) we obtain immediately that

$$
\begin{aligned}
&\sum_{i=1}^{n}\rho_\tau\left\{Y_i - a - b\left(Z_i - z\right)\right\}K_i - \sum_{i=1}^{n}\rho_\tau\left\{Y_i - a_0 - b_0\left(Z_i - z\right)\right\}K_i \\
&\quad = 2\sum_{i=1}^{n}\Big[\zeta_i\left\{\mathbf{1}(Y_i - a_0 - b_0(Z_i - z) \le 0) - \tau\right\} \qquad\qquad (\mathrm{A.4})\\
&\qquad + \int_0^{\zeta_i}\left\{\mathbf{1}(Y_i - a_0 - b_0(Z_i - z) \le t) - \mathbf{1}(Y_i - a_0 - b_0(Z_i - z) \le 0)\right\}dt\Big]K_i.
\end{aligned}
$$

Let $\boldsymbol{\beta} = \sqrt{nh}\left\{a - a_0, h\left(b - b_0\right)\right\}^{\mathrm{T}}$. In the sequel, we show the first quantity in (A.4) has approximately the form $2\mathrm{u}_n^{\mathrm{T}}\boldsymbol{\beta}$, and the second quantity has the form $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{B}\boldsymbol{\beta}$. Then we use the quadratic approximation lemma to complete the proof.

Denote by $F_Y(y \mid Z)$ the conditional distribution of Y given Z. Then

$$
\begin{aligned}
&2\sum_{i=1}^{n}\zeta_i\left\{\mathbf{1}(Y_i - a_0 - b_0(Z_i - z) \le 0) - \tau\right\}K_i \\
&\quad = 2\sum_{i=1}^{n}\zeta_i\left\{F_{Y_i}\left(a_0 + b_0(Z_i - z) \mid Z_i\right) - \tau\right\}K_i \qquad\qquad (\mathrm{A.5})\\
&\qquad + 2\sum_{i=1}^{n}\zeta_i\left\{\mathbf{1}(Y_i - a_0 - b_0(Z_i - z) \le 0) - F_{Y_i}\left(a_0 + b_0(Z_i - z) \mid Z_i\right)\right\}K_i.
\end{aligned}
$$

The first term in (A.5) is a bias term. To view it, we use the Taylor's expansion

$$G_\tau(Z_i) = a_0 + b_0(Z_i - z) + G''_\tau(z)(Z_i - z)^2 \left\{ \frac{1}{2} + o_p(1) \right\}, \text{ for } |Z_i - z| \le h.$$

This together with another Taylor's expansion entails that

$$\begin{aligned} F_{Y_i}\left(a_0 + b_0(Z_i - z) \mid Z_i\right) &= F_{Y_i}\left(G_\tau(Z_i) - G''_\tau(z)(Z_i - z)^2 \left\{ \frac{1}{2} + o_p(1) \right\} \mid Z_i \right) \\ &= \tau - f_{Y_i}(G_\tau(Z_i) \mid Z_i) G''_\tau(z)(Z_i - z)^2 \left\{ \frac{1}{2} + o_p(1) \right\}. \end{aligned}$$

With this result, the bias term can be simplified as

$$\begin{aligned} &2\sum_{i=1}^{n} \zeta_i \left\{ F_{Y_i}\left(a_0 + b_0(Z_i - z) \mid Z_i\right) - \tau \right\} K_i \\ &\quad = -\sum_{i=1}^{n} \zeta_i f_{Y_i}(G_\tau(Z_i) \mid Z_i) G''_\tau(z)(Z_i - z)^2 \left\{ 1 + o_p(1) \right\} K_i, \end{aligned}$$

which is of the form

$$-\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \boldsymbol{\beta}^{\mathrm{T}} \begin{pmatrix} 1 \\ \frac{(Z_i - z)}{h} \end{pmatrix} f_{Y_i}(G_\tau(Z_i) \mid Z_i) G''_\tau(z)(Z_i - z)^2 \left\{ 1 + o_p(1) \right\} K_i. \tag{A.6}$$

The second term on the right side of (A.5) can be approximated as

$$2\sum_{i=1}^{n} \zeta_i \left\{ \mathbf{1}(Y_i \le G_\tau(Z_i)) - \tau \right\} K_i \left\{ 1 + o_p(1) \right\},$$

which is again of the form

$$\frac{2}{\sqrt{nh}} \sum_{i=1}^{n} \boldsymbol{\beta}^{\mathrm{T}} \begin{pmatrix} 1 \\ \frac{(Z_i - z)}{h} \end{pmatrix} \left\{ \mathbf{1}(Y_i \le G_\tau(Z_i)) - \tau \right\} K_i \left\{ 1 + o_p(1) \right\}. \tag{A.7}$$

Let $f_Y(y \mid Z)$ be the conditional density function of Y given Z. Then the second term on the right side of (A.4) can be approximated as

$$\begin{aligned} &2\sum_{i=1}^{n} \Big[\int_0^{\zeta_i} \left\{ \mathbf{1}(Y_i - a_0 - b_0(Z_i - z) \le t) - \mathbf{1}(Y_i - a_0 - b_0(Z_i - z) \le 0) \right\} dt \Big] K_i \\ &\quad = \sum_{i=1}^{n} f_{Y_i}\left(a_0 + b_0(Z_i - z) \mid Z_i\right) \zeta_i^2 K_i \left\{ 1 + o_p(1) \right\} \\ &\quad = \sum_{i=1}^{n} f_{Y_i}\left(G_\tau(Z_i) \mid Z_i\right) \zeta_i^2 K_i \left\{ 1 + o_p(1) \right\}, \end{aligned}$$

which is now of the form

$$\frac{1}{nh}\sum_{i=1}^{n} f_{Y_i}\left(G_\tau(Z_i)\mid Z_i\right)\boldsymbol{\beta}^{\mathrm{T}}\begin{pmatrix} 1 & \left\{\frac{(Z_i-z)}{h}\right\} \\ \left\{\frac{(Z_i-z)}{h}\right\} & \left\{\frac{(Z_i-z)}{h}\right\}^2 \end{pmatrix}\boldsymbol{\beta}K_i\left\{1+o_p(1)\right\}. \tag{A.8}$$

Combining the results of (A.5)−(A.8), we observe that (A.4) is now approximated by a quadratic function of $\boldsymbol{\beta}$. By the quadratic approximation lemma, $\boldsymbol{\beta}$ is only $o_p(1)$ away from $-\mathbf{B}^{-1}\mathrm{u}_n$, where

$$\mathbf{B} = \frac{1}{nh}\sum_{i=1}^{n} f_{Y_i}\left(G_\tau(Z_i)\right)\begin{pmatrix} 1 & \left\{\frac{(Z_i-z)}{h}\right\} \\ \left\{\frac{(Z_i-z)}{h}\right\} & \left\{\frac{(Z_i-z)}{h}\right\}^2 \end{pmatrix}K_i,$$

$$\mathrm{u}_n = \sqrt{nh}\left[\frac{1}{nh}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ \left\{\frac{(Z_i-z)}{h}\right\} \end{pmatrix}\left\{\mathbf{1}(Y_i\le G_\tau(Z_i))-\tau\right\}K_i \right.$$
$$\left. -\frac{1}{nh}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ \left\{\frac{(Z_i-z)}{h}\right\} \end{pmatrix} f_{Y_i}(G_\tau(Z_i)\mid Z_i)G''_\tau(z)(Z_i-z)^2\frac{K_i}{2}\right].$$

It can be easily shown that, as $n\to\infty$, $\mathbf{B}$ converges in probability to

$$f_Y\left(G_\tau(z)\mid z\right)f_Z(z)\begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}.$$

Because our target is the quantile function, it suffices to deal with the first element of u_n because the second element is concerned instead with its derivative. Without much difficulty, we can prove that the first quantity in the curly parenthesis converges in distribution to a normal population with zero mean and variance $\tau(1-\tau)f_Z(z)\int_{-1}^{1}K^2(u)du$, and the second quantity in the curly parenthesis converges in probability to $h^2G''_\tau(z)\mu_2\{f_Y(G_\tau(z)\mid z)f_Z(z)\}$.

Let f_ε denote the density function of the error term ε, and $\xi^\star_\tau$ denote its τ-th quantile. We can easily verify that $f_Y(G_\tau(z)\mid z)\sigma(z) = f_\varepsilon(\xi^\star_\tau)$ when model (1.1) is true. Then the asymptotic variance has a different appearance.

The proof of Theorem 3 is completed by combining the above results.

References

Belloni, A. and Chernozhukov, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39**, 82-130.

Bhattacharya, P. K. and Gangopadhyay, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile regression. *Ann. Statist.* **18**, 1400-1415.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahahur representation. *Ann. Statist.* **19**, 760-777.

Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *Ann. Statist.* **25**, 715-744.

De Gooijer, J. G. and Zerom, D. (2003). On additive conditional quantiles with high dimensional covariates. *J. Amer. Statist. Assoc.* **98**, 135-146.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* Chapman and Hall, London.

Fan, J., Hu, T.-C. and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Statist.* **21**, 433-446

Gannoun, A., Girard, S., Guinot, C. and Saracoo, J. (2004). Sliced inverse regression in reference curves estimation. *Comput. Statist. Data Anal.* **46**, 103-122.

Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projection from high dimensional data. *Ann. Statist.* **21**, 867-889.

Härdle, W. K, and Song, S. (2010). Confidence bands in quantile regression *Econom. Theory* **26**, 1180-1200

Hjort, N. L. and Pollard, D. (1993). Asymptotics for minimisers of convex processes. Preprint available at `www.stat.yale.edu/~pollard/Papers/convex.pdf`

Horowitz, J. L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.* **100**, 1238-1249.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58**, 71-120.

Johnson, R. W. (2003). Kiplinger's personal finance. *J. Statist. Education* **57**, 104-123.

Kai, B., Li, R. and Zou, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *J. Roy. Statist. Soc. Ser. B* **72**, 49-69.

Knight, K. (1998). Limiting distribution for L_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755-770.

Koenker, R. (2005). *Quantile Regression.* Cambridge University Press.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrika* **46**, 33-50.

Koenker, R., Ng, P. and Portnoys, S. (1994). Quantiles smoothing splines. *Biometrika* **81**, 673-680.

Kong, E., and Xia, Y. (2010). Quantile estimation of a general single-index model. Working paper available at `http://arxiv.org/abs/0803.2474.`

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86**, 316-327.

Li, K.-C. and Duan, N. (1989). Regression analysis under link violation. *Ann. Statist.* **17**, 1009-1052.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econom. Theory* **7**, 186-199.

Stone, C. J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.* **5**, 595-645.

Wang, H., Zhu, Z. and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *Ann. Statist.* **37**, 3841-3866.

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statist. Sinica* **19**, 801-817.

Wu, T., Yu, K. and Yu, Y. (2010). Single-index quantile regression. *J. Multiariate Statist.* **101**, 1607-1621.

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

Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. *Scand. J. Statist.* **31**, 333-346.

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

School of Statistics and Management and the key Laboratory of Mathematical Economics, Ministry of Education, Shanghai University of Finance and Economics, Shanghai, 200433, P. R. China.

E-mail: zhu.liping@mail.shufe.edu.cn

School of Statistics and Management and the key Laboratory of Mathematical Economics, Ministry of Education, Shanghai University of Finance and Economics, Shanghai, 200433, P. R. China.

E-mail: huang.mian@mail.shufe.edu.cn

Department of Statistics and The Methodology Center, The Pennsylvania State University, University Park, Pennsylvania, 16802-2111, USA.

E-mail: rli@stat.psu.edu

(Received September 2010; accepted October 2011)