Published in final edited form as:

Stat Sin. 2012 October ; 22(4): 1379–1401. doi:10.5705/ss.2010.199.

SEMIPARAMETRIC QUANTILE REGRESSION WITH HIGH-DIMENSIONAL COVARIATES

Liping Zhu,

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, 200433, P. R. China

Mian Huang, and

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, 200433, P. R. China

Runze Li

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

Liping Zhu: luz15@psu.edu; Mian Huang: huang.mian@mail.shufe.edu.cn; Runze Li: rli@stat.psu.edu

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 a surprising and interesting result 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 real 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 the 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 vector $\mathbf{x}$ 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 $\mathbf{x} = (X_1, \dots, X_p)^T$ be a covariate vector. In this paper, we consider the following heteroscedastic single-index model:

$$Y = G(\mathbf{x}^T \beta_0) + \sigma(\mathbf{x}^T \beta_0) \varepsilon, \quad (1.1)$$

where the error term ε is assumed to be independent of $\mathbf{x}$, $E(\varepsilon) = 0$ and $\text{var}(\varepsilon) = 1$. The functions $G(\cdot)$ and $\sigma(\cdot)$ are unspecified, nonparametric smoothing functions. The parameter of interest is the direction of β_0 . To ensure the identifiability, it is assumed throughout this paper that $\|\beta_0\| = 1$ with first nonzero element being positive. It is easy to see that model (1.1) includes linear quantile regression (Koenker and Bassett, 1978) as a special case thereof. 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 the 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 β_0 using the average derivative approach, which takes partial derivatives of the conditional quantile with respect to the covariates $\mathbf{x}$. The average derivative approach requires multivariate kernel regression, which is hence less useful in the presence of high-dimensional covariates. Wu, Yu and Yu (2010) proposed a back-fitting algorithm, which is computationally expensive when p is large. In this paper, we will propose a new estimation procedure for model (1.1). The newly proposed estimation procedure consists of two steps. We first estimate β_0 using the linear quantile regression. We show that the resulting estimate is consistent. We then employ the local linear regression (Fan and Gijbels, 1996) to estimate the conditional quantile function of Y given $\mathbf{x}^T \beta_0$. This paper makes the following major contributions to the literature:

- a. Under mild conditions, we show that for any link function $G(\cdot)$ and $\tau \in (0, 1)$, the τ -th linear quantile regression coefficient for $Y|\mathbf{x}$ is proportional to β_0 in the single-index model (1.1).
- b. We prove that the local linear regression estimate for the conditional quantile function based upon the root- n consistent estimate of β_0 has the same asymptotic bias and variance as the local linear estimate with the true value of β_0 .

The theoretical result in (a) implies that the ordinary linear quantile regression actually results in root- n consistent estimates for β_0 . Thus, the 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 reduce substantially the computational cost of the back-fitting algorithm (Wu, Yu and Yu, 2010). Thus our proposed procedure 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 β_0 iteratively once a root- n consistent estimator of β_0 is available. It also ensures our proposed procedure to work 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 rigorously established in this section. We demonstrate the methodologies through comprehensive simulation studies and an application to a real data set in Section 3. We conclude this paper with a brief discussion in Section 4. All proofs are given 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 β_0 . We further demonstrate that the quantile regression for model (1.1) can be reduced to the univariate quantile regression through replacing β_0 with $\hat{\beta}_0$, the resulting estimate of the linear quantile regression.

2.1. Estimation of β_0

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

$$(u_\tau, \beta_\tau) \stackrel{\text{def}}{=} \underset{u, \beta}{\operatorname{argmin}} \{\mathcal{L}_\tau(u, \beta)\}. \quad (2.1)$$

Theorem 1 below implies that the linear quantile regression offers a consistent estimate for β_0 in (1.1) under mild conditions.

Theorem 1—If the covariate vector $\mathbf{x}$ in model (1.1) satisfies the following linearity condition

$$E(\mathbf{x}|\beta_0^T \mathbf{x}) = \operatorname{var}(\mathbf{x})\beta_0\{\beta_0^T \operatorname{var}(\mathbf{x})\beta_0\}^{-1}\beta_0^T \mathbf{x}, \quad (2.2)$$

then $\beta_\tau = \kappa\beta_0$ for some constant κ , where β_τ is defined in (2.1).

The proof of Theorem 1 will be given in the Appendix. The linearity condition (2.2) plays an important role here. Theorem 1 implies that under the linearity condition, we are able to obtain a root n consistent estimator of the direction of β by the linear quantile regression. This will be further confirmed in Theorem 2. The linearity condition (2.2) is widely assumed in the context of sufficient dimension reduction. Li (1991) pointed out that (2.2) is satisfied when $\mathbf{x}$ follows an elliptically contour distribution. Hall and Li (1993) proved that this linearity condition 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.

It is worth pointing out that 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 may 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) may be used to select significant variables for model (1.1). It is important to correctly set the range of the tuning parameter in 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. Data-driven method for tuning parameter selection is recommended.

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

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

Let

$$(\hat{u}_{\tau}, \hat{\beta}_{\tau}) \stackrel{\text{def}}{=} \underset{u, \beta}{\operatorname{argmin}} \{\mathcal{L}_{\tau n}(u, \beta)\}. \quad (2.3)$$

The asymptotic normality of $\hat{\beta}_{\tau}$ is established in the following theorem, whose proof will be given in the Appendix.

Theorem 2—Assume that $\Lambda = E\{f^*(\zeta_{\tau}|\mathbf{x})\mathbf{x}\mathbf{x}^T\}$ is invertible, where $f^*(z|\mathbf{x})$ denotes the conditional density of $(Y - \mathbf{x}^T \beta_{\tau})$ given $\mathbf{x}$, and ζ_{τ} denotes the τ -th quantile of $Y|\mathbf{x}$. Then $n^{1/2}(\hat{\beta}_{\tau} - \beta_{\tau})$ is asymptotically normal with mean zero and covariance matrix $\tau(1 - \tau) \Lambda^{-1} E(\mathbf{x}\mathbf{x}^T) \Lambda^{-1}$. If the “partial residual” term $Y - \mathbf{x}^T \beta_{\tau}$ is independent of $\mathbf{x}$, the asymptotic variance matrix of $\hat{\beta}_{\tau}$ reduces to $\tau(1 - \tau) E(\mathbf{x}\mathbf{x}^T)^{-1} / \{f^*(\zeta_{\tau})\}^2$.

2.2. Estimation of conditional quantile function

By Theorem 1, β_{τ} is proportional to β_0 , and hence the conditional distribution of $Y|(\mathbf{x}^T \beta_0)$ is exactly the same as that of $Y|(\mathbf{x}^T \beta_{\tau})$. Consequently, Theorem 1 allows to estimate the quantile regression of $Y|(\mathbf{x}^T \beta_0)$ through the conditional distribution of $Y|(\mathbf{x}^T \beta_{\tau})$. We denote the τ -quantile function of $Y|(\mathbf{x}^T \beta_{\tau})$ by $G_{\tau}(\cdot)$.

By using data points $\{(\mathbf{x}_i^T \hat{\beta}_{\tau}, Y_i), i = 1, \dots, n\}$, we estimate $G_{\tau}(\cdot)$ by using the local linear regression. For ease of presentation, let $z = \mathbf{x}_0^T \beta_{\tau}$, $\hat{z} = \mathbf{x}_0^T \hat{\beta}_{\tau}$, $Z_i = \mathbf{x}_i^T \beta_{\tau}$, $\hat{Z}_i = \mathbf{x}_i^T \hat{\beta}_{\tau}$. 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

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

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

In the sequel we derive the asymptotic property for the nonparametric estimate of $G_{\tau}(z)$. For any fixed $\mathbf{x}_0 \in \mathbb{R}^p$, the asymptotic normality of the local linear estimator $\hat{G}_{\tau}(\hat{z})$, is presented in the following theorem, whose proof will be given in the Appendix.

Theorem 3—Under the regularity conditions (i)–(iv) in Appendix A, if $h \rightarrow 0$ and $nh \rightarrow \infty$, then the asymptotic conditional bias and variance of the local linear quantile regression estimator $\hat{G}_{\tau}(\hat{z})$ are given by

$$\begin{aligned} \text{bias}\{\hat{G}_{\tau}(\hat{z})\} &= G''_{\tau}(z) \mu_2 h^2 / 2 + o_p(h^2), \\ \text{var}\{\hat{G}_{\tau}(\hat{z})\} &= (nh)^{-1} \Delta \{1 + o_p(1)\}, \end{aligned}$$

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

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

is asymptotically normal with mean zero and asymptotic variance Δ .

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 $\{(\mathbf{x}_i^T \hat{\beta}_{\tau}, Y_i), i = 1, \dots, n\}$ with that of the estimator built upon $\{(\mathbf{x}_i^T \beta_{\tau}, Y_i), i = 1, \dots, n\}$ in Fan, Hu and Truong (1994), and found they performed equally well asymptotically. This implies that the resulting estimate of the newly proposed procedure performs as well as an oracle estimate with replacing $\hat{\beta}_{\tau}$ with its true value.

Next we discuss how to discuss the issue of bandwidth selection. One usually chooses the bandwidth which minimizes the mean squared error (MSE) of nonparametric estimation. Theorem 3 indicates that the leading term of MSE of $\hat{G}_{\tau}(\hat{z})$ is

$$\text{MSE}\{\hat{G}_{\tau}(\hat{z})\} = \frac{1}{4} \{\mu_2 G''_{\tau}(z)\}^2 h^4 + \frac{\nu_0}{nhf(z)} \frac{\tau(1-\tau)}{f_{\varepsilon}^2(\xi_{\tau}^*)} \sigma^2(z).$$

Thus, the optimal variable bandwidth which minimizes the asymptotic MSE of $\hat{G}_{\tau}(\hat{z})$ is

$$h^{\text{opt}}(\hat{z}) = \left[\frac{4\nu_0\tau(1-\tau)}{\{\mu_2 f_\varepsilon(\xi_\tau^*)\}^2} \frac{\sigma^2(z)/f(z)}{\{G''_\tau(z)\}^2} \right]^{1/5} n^{-1/5}. \quad (2.4)$$

This implies that the local quantile regression achieves the optimal rate of convergence $n^{2/5}$. To implement $h^{\text{opt}}(\hat{z})$, one has to replace all unknowns in (2.4) with their consistent estimators. This is usually computationally inefficient although all involved 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 the MSE of the form

$$\text{MSE} \{ \hat{G}_{\text{LS}}(\hat{z}) \} = \frac{1}{4} \{ \mu_2 G''(z) \}^2 h^4 + \frac{\nu_0}{n h f(z)} \sigma^2(z),$$

which results in the optimal bandwidth

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

Comparing (2.5) with (2.4), we obtain that

$$h^{\text{opt}}(\hat{z}) = h_m^{\text{opt}}(\hat{z}) \left\{ \frac{\tau(1-\tau)}{\{f_\varepsilon(\xi_\tau^*)\}^2} \frac{\{G''_\tau(z)\}^2}{\{G''(z)\}^2} \right\}^{1/5}.$$

We next 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 a standard normal random variable, then we can choose the approximately optimal bandwidth through

$$h^{\text{opt}}(\hat{z}) = h_m^{\text{opt}}(\hat{z}) \left\{ \frac{\tau(1-\tau)}{\{\varphi(\xi_\tau^*)\}^2} \right\}^{1/5}, \quad (2.6)$$

where $\varphi(\cdot)$ denotes the probability density function of the standard normal distribution.

There are many algorithms for calculating $h_m^{\text{opt}}(\hat{z})$ (Fan and Gijbels, 1996), thus $h^{\text{opt}}(\hat{z})$ can be rapidly calculated. One can estimate ξ_τ^* through the sample τ -th quantile of the residuals $\hat{\varepsilon}_i = \{Y_i - \hat{G}(\mathbf{x}_i^T \hat{\beta}_\tau)\} / \hat{\sigma}(\mathbf{x}_i^T \hat{\beta}_\tau)$, for $i = 1, \dots, n$. An alternative way is to replace ξ_τ^* in (2.6) with the τ -th quantile of the standard normal population to further simplify the calculation.

The idea of this approximation is originated from Yu and Jones (1998) and Yu and Lu (2004). Although this approximation 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 conduct simulations to compare the performance of the proposed methods with existing competitors and to illustrate the proposed methodology with a real data example.

3.1. Simulation

We generated 1000 datasets, each consisting of $n = 500$ observations, from

$$Y = \sin\{2(\mathbf{x}^T \beta_0)\} + 2\exp\{-16(\mathbf{x}^T \beta_0)^2\} + \sigma(\mathbf{x}^T \beta_0)\varepsilon, \quad (3.1)$$

where the index parameter $\beta_0 = (2, -2, -1, 1, 0, \dots, 0)^T / \sqrt{10}$ is a $p \times 1$ vector, and the covariate vector $\mathbf{x} = (X_1, \dots, X_p)^T$ is generated from a multivariate normal population with mean zero and variance-covariance matrix $\text{var}(\mathbf{x}) = (\sigma_{ij})_{p \times p}$ with $\sigma_{ij} = 0.5^{|i-j|}$. The conditional mean function in model (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 distribution $N(0, 1)$; (ii) a mixture of two normal distributions $0.8N(0, 1) + 0.2N(0, 9)$; (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 $\mathbf{x}$ are mutually independent. We considered two cases for $\sigma(\mathbf{x}^T \beta_0)$: $\sigma(\mathbf{x}^T \beta_0) = 1$ in homoscedasticity case (1), and $\sigma(\mathbf{x}^T \beta_0) = \exp(\mathbf{x}^T \beta_0)$ in heteroscedasticity case (2). We considered case (2) because it is often of interest to investigate the effect of heteroscedastic errors in quantile regressions. The response values contain outliers with larger probabilities in case (2) than in case (1). We compared the performance of our proposed procedures with the back-fitting algorithm proposed by Wu, Yu and Yu (2010). From our limited simulation experience, the back-fitting algorithm demands much computing time. We compared the simulation results of the back-fitting algorithm based on 100 and 1000 replications for one case, and found that the simulation results based on 100 replications are almost the same as those based on 1000 replications. Thus, we report the results of the back-fitting algorithm based on 100 replications in order to save computing time.

Performance in estimating β_0 —The index parameter β_0 was estimated via a series of quantile regressions with $\tau = 0.025, 0.05, 0.5, 0.95$ and 0.975 , respectively. Table 3.1 depicts the averages of mean squared errors (MSE) of the estimate $\hat{\beta}_\tau$, which is defined by $\text{MSE} = \|\hat{\beta}_\tau - \beta_0\|^2$. We include three competitors into our comparison 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 β_0 ; thus, it serves naturally as a competitor here. It can be seen from Table 3.1 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 β_0 . By contrast, the linear quantile regression assumes a linear quantile function to reduce the computational complexity. Yet the linear quantile regression has also a satisfactory performance. In the particular homoscedastic case (1) with Cauchy errors, the linear quantile regression performs even better than the back-fitting algorithm in terms of the MSE values. In the homoscedasticity 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 heteroscedasticity case (2) than in homoscedasticity case (1). This is expected because 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 often of interest to evaluate the accuracy of quantile regression in offering the prediction interval of Y given $\mathbf{x}^T\beta_0$. To this end, we can estimate the τ -th and $(1 - \tau)$ -th quantile of the conditional distribution of Y given $\mathbf{x}^T\beta_0$, which can be used as the confidence interval at the level of $(1 - 2\tau)$ for $\tau < 0.5$. That is, we use $[G^{-1}(\tau|\mathbf{x}^T\beta_0), G^{-1}(1 - \tau|\mathbf{x}^T\beta_0)]$ to construct the prediction interval for $Y|\mathbf{x}^T\beta_0$ at the $(1 - 2\tau) \times 100\%$ level, where $G^{-1}(\tau|\mathbf{x}^T\beta_0)$ is the inverse function of $G_\tau(Y|\mathbf{x}^T\beta_0)$. That is, $G^{-1}(\tau|\mathbf{x}^T\beta_0)$ is the τ -th quantile of $Y|\mathbf{x}^T\beta_0$. We choose $\tau = 0.025$ and 0.05 to provide the confidence intervals at 95% and 90%. We expect the empirical coverage probability of such confidence interval to be close to the nominal level of $(1 - 2\tau)$.

We included three competitors into our comparison: (i) The back-fitting algorithm (Wu, Yu and Yu, 2010) which can simultaneously estimate the index parameter β_τ and the quantile function; (ii) the quantile regression method introduced in Section 2.2 which is based on the estimated index parameter $\hat{\beta}_\tau$; and (iii) the oracle estimator which provides the prediction interval by using the true index parameter β_0 . The oracle estimator serves as a benchmark for comparison purposes. The average and the standard deviation of the empirical coverage probabilities are summarized in Table 3.2.

Both the homoscedasticity and the heteroscedasticity cases reported in Table 3.2 show very similar messages. The coverage probabilities of three competitors are all close to the nominal levels for different τ values, although the back-fitting algorithm uses a more accurate estimate of β_0 . The standard deviations of the empirical probabilities are very small, indicating that the quantile regression offers a very reliable coverage probability, which is comparable with its oracle version. This phenomenon verifies the theoretical results in Theorem 3.

Computational efficacy—It is of interest to compare the computational efficacy of the newly proposed method with that of the back-fitting algorithm proposed by Wu, Yu and Yu (2010). Table 3.3 summarizes the average computing time in seconds used for estimating the index parameter and calculating the prediction intervals for one replication. It can be easily seen from Table 3.3 that the newly proposed 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 from 10 to 20 and 50. Tables 3.4 and 3.5 report respectively the MSE values and the coverage probabilities for model (3.1) with $p = 20$ and 50. Table 3.4 together with Table 3.1 indicates that the MSE values of $\hat{\beta}_\tau$ increase significantly as the covariate dimension p increases due to the curse of dimensionality. However, the covariate dimension p has little effect on estimating the coverage probabilities of the prediction intervals, as can be seen from Tables 3.2 and 3.5. This benefits essentially from the single-index structure of the quantile function. These simulation results also partially support our theoretical investigations in Theorems 2 and 3, and confirm the advantage of our proposed 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 attributes such as miles per gallon and horsepower. Thus the 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 logarithm transformation of the MSRP in U.S. dollars as the response (Y). In addition, there are seven features which possibly affect the resulting price of vehicles, such as the engine size (X_1), number of cylinders (X_2), horsepower (X_3), average city miles per gallon (MPG, X_4) and 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 $\mathbf{x} = (X_1, \dots, X_7)^T$ are standardized marginally to have zero mean and unit variance in this empirical analysis. 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 . They yield similar estimates for the index parameter β_τ at different quantiles, indicating that similar patterns between the MSRP and different features are revealed at different quantiles. Specifically, the estimated index parameter at $\tau = 0.5$ is $\hat{\beta}_\tau = (0.3254, -0.1846, -0.6332, 0.3688, -0.3326, -0.4156, 0.1994)^T$. Table 3.6 reports that the standard deviation of $\hat{\beta}_\tau$ obtained from a nonparametric bootstrap procedure is $(0.052, 0.046, 0.051, 0.069, 0.057, 0.043, 0.032)^T$. 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 $\{(\mathbf{x}_i^T \hat{\beta}_\tau, Y_i), i = 1, \dots, n\}$, we first estimate the regression function $G(\mathbf{x}^T \beta_0)$ via the local least squares estimator. The estimated function and its 95% confidence interval are reported in Figure 3.1(C). We further conduct some exploratory data analysis on this dataset. The boxplot of the log-transformed MSRP, the response variable, is depicted in Figure 3.1(A), which shows that there are four vehicles whose prices are significantly higher than other vehicles. In addition, the histogram in Figure 3.1(B) reveals that the distribution of the log-transformed prices are highly skewed. This motivates us to further conduct empirical analysis of this dataset using the proposed procedure.

Next we apply the quantile regression to this dataset. Figure 3.1(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 3.1(D) presents the quantile regression at the level of 50%. It behaves similarly to the local least squares estimator in Figure 3.1(C) within the range of $\mathbf{x}^T \hat{\beta}_\tau$. When $\mathbf{x}^T \hat{\beta}_\tau$ 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 the proposed procedure in terms of description of the conditional distribution of Y given $\mathbf{x}$.

We further use a bootstrap procedure to demonstrate the performance of our method in the presence of outliers. Given the covariates we bootstrap new response variables by using the residuals. To calculate the residuals, we fix $\tau = 0.5$ and estimate β_τ and $G_\tau(\mathbf{x}^T \beta_\tau)$ to obtain $\hat{\varepsilon}_i = Y_i - \hat{G}_\tau(\mathbf{x}_i^T \hat{\beta}_\tau)$ for $i = 1, \dots, n$. For ease of illustration, we denote by F_n the empirical

distribution of the residuals ε_i . We bootstrap 500 samples each of form $\{(\mathbf{x}_i, Y_i^*), i = 1, \dots, n\}$ where $Y_i^* = \hat{G}_\tau(\mathbf{x}_i^T \hat{\beta}_\tau) + \hat{\varepsilon}_i^*$, and ε_i^* is generated from three different distributions: (1) F_n ; (2) $0.5F_n + 0.5N(0, 0.4332^2)$; we consider this mixture because the standard deviation of the original residuals is exactly 0.4332; and (3) $0.5F_n + 0.5\text{Cauchy}$. In case (3) the distribution of the response variable has a heavy tail.

We estimate the index parameter β_τ at different quantiles, and summarize the average and standard deviation of $\hat{\beta}_\tau$ 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 3.6. The averages of the estimated index parameter $\hat{\beta}_\tau$ are similar. This once again verifies that our method is robust to the presence of outliers. By contrast, the standard deviations of these three cases are slightly different. To be precise, the standard deviations of $\hat{\beta}_\tau$ in case (1) and (2) are very 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 retains 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 (i.e., $\sigma(\cdot)$ remains constant), and the error ε is symmetric about zero, then a byproduct of this paper is that our estimation procedure offers a robust estimation for the conditional mean function. The simulation studies in this article verify this point implicitly. In addition, it is not necessary that the error term ε in model (1.1) is independent of $\mathbf{x}$, in which case we can assume instead that $E(\mathbf{x}|\mathbf{x}^T\beta_0, \varepsilon)$ is a linear function of $\mathbf{x}^T\beta_0$. If the parameter of interest is the index parameter β_0 only, the composite quantile regression (Zou and Yuan, 2008) can be adapted to improve the efficiency in estimating β_0 .

In this paper 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 β_0 is known. Further research is needed to quantify the effect when β_0 is replaced with its root- n consistent estimate.

Acknowledgments

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 Pro-jet 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 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.

References

- Belloni A, Chernozhukov V. ℓ_1 -penalized quantile regression in high-dimensional sparse models. *Annals of Statistics*. 2011; 39:82–130.

- Bhattacharya PK, Gangopadhyay AK. Kernel and nearest-neighbor estimation of a conditional quantile regression. *Annals of Statistics*. 1990; 18:1400–1415.
- Chaudhuri P. Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics*. 1991; 19:760–777.
- Chaudhuri P, Doksum K, Samarov A. On average derivative quantile regression. *Annals of Statistics*. 1997; 25:715–744.
- De Gooijer JG, Zerom D. On additive conditional quantiles with high dimensional covariates. *Journal of the American Statistical Association*. 2003; 98:135–146.
- Fan, J.; Gijbels, I. *Local Polynomial Modelling and its Applications*. London: Chapman and Hall; 1996.
- Fan J, Hu TC, Truong YK. Robust non-parametric function estimation. *Scandinavian Journal of Statistics*. 1994; 21:433–446.
- Gannoun A, Girard S, Guinot C, Saracco J. Sliced inverse regression in reference curves estimation. *Computational Statistics and Data Analysis*. 2004; 46:103–122.
- Hall P, Li KC. On almost linearity of low dimensional projection from high dimensional data. *Annals of Statistics*. 1993; 21:867–889.
- Härdle WK, Song S. Confidence bands in quantile regression. *Econometric theory*. 2010; 26:1180–1200.
- Hjort, NL.; Pollard, D. Asymptotics for minimisers of convex processes. 1993. Preprint available at www.stat.yale.edu/~pollard/Papers/convex.pdf
- Horowitz JL, Lee S. Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*. 2005; 100:1238–1249.
- Ichimura H. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*. 1993; 58:71–120.
- Johnson RW. Kiplinger's personal finance. *Journal of Statistics Education*. 2003; 57:104–123.
- Kai B, Li R, Zou H. Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society Series B*. 2010; 72:49–69.
- Knight K. Limiting distribution for L_1 regression estimators under general conditions. *Annals of Statistics*. 1998; 26:755–770.
- Koenker, R. *Quantile Regression*. Cambridge University Press; 2005.
- Koenker R, Bassett G. Regression quantiles. *Econometrica*. 1978; 46:33–50.
- Koenker R, Ng P, Portnoys S. Quantiles smoothing splines. *Biometrika*. 1994; 81:673–680.
- Kong, E.; Xia, Y. Quantile estimation of a general single-index model. Working paper. 2010. Available at <http://arxiv.org/abs/0803.2474>
- Li KC. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*. 1991; 86:316–327.
- Li KC, Duan N. Regression analysis under link violation. *The Annals of Statistics*. 1989; 17:1009–1052.
- Pollard D. Asymptotics for least absolute deviation regression estimators. *Econometric Theory*. 1991; 7:186–199.
- Stone CJ. Consistent nonparametric regression (with discussion). *Annals of Statistics*. 1977; 5:595–645.
- Wang H, Zhu Z, Zhou J. Quantile regression in partially linear varying coefficient models. *Annals of Statistics*. 2009; 37:3841–3866.
- Wu Y, Liu Y. Variable selection in quantile regression. *Statistica Sinica*. 2009; 19:801–817.
- Wu T, Yu K, Yu Y. Single-index quantile regression. *Journal of Multivariate Statistics*. 2010; 101:1607–1621.
- Yu K, Lu Z. Local linear additive quantile regression. *Scandinavian Journal of Statistics*. 2004; 31:333–346.
- Yu K, Jones MC. Local linear quantile regression. *Journal of the American Statistical Association*. 1998; 93:228–237.

Zou H, Yuan M. Composite quantile regression and the oracle model selection theory. *Annals of Statistics*. 2008; 36:1108–1126.

Appendix: Technical Conditions and Proofs

A.1. Technical conditions

The following technical conditions are imposed. They are not the weakest possible conditions, but they 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 $(\mathbf{x}^T\beta)$ is positive and uniformly continuous for β in a neighborhood of β_0 . Furthermore, the density function of $(\mathbf{x}^T\beta_0)$ is continuous and bounded away from zero and infinity on its support.
- iii. The conditional density of Y given $(\mathbf{x}^T\beta_0)$, denoted by $f_Y(y|\mathbf{x}^T\beta_0)$, is continuous in $(\mathbf{x}^T\beta_0)$ for each $y \in \mathbb{R}$. Moreover, there exist positive constants ε and δ and a positive function $F(y|\mathbf{x}^T\beta_0)$ such that

$$\sup_{\|\mathbf{x}^T\beta - \mathbf{x}^T\beta_0\| \leq \varepsilon} f_Y(y|\mathbf{x}^T\beta) \leq F(y|\mathbf{x}^T\beta_0).$$

For any fixed value of $(\mathbf{x}^T\beta_0)$, we assume that $\int F(y|\mathbf{x}^T\beta_0) dy < \infty$ and

$$\int \{\rho_{\tau}(y-t) - \rho_{\tau}(y) - \rho'_{\tau}(y)t\}^2 F(y|\mathbf{x}^T\beta_0) dy = o(t^2), \text{ as } t \rightarrow 0.$$

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

We briefly discuss these conditions. 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 $f(\mathbf{x}^T\beta_0)$ 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 define that $\rho'_{\tau}(u) = \text{sign}(u) + (2\tau - 1)$ for $u \neq 0$ and $\rho'_{\tau}(0) = 0$. With Taylor's expansion, we obtain

$$\int \{\rho_{\tau}(y-t) - \rho_{\tau}(y) - \rho'_{\tau}(y)t\}^2 F(y|\mathbf{x}^T\beta_0) dy = t^2 \int \{\text{sign}(u^*) - \text{sign}(y)\}^2 F(y|\mathbf{x}^T\beta_0) dy$$

for some u^* lies between $y - t$ and y . Without loss of generality we assume $t > 0$, then

$$\int_{-\infty}^{\infty} \{\text{sign}(u^*) - \text{sign}(y)\}^2 F(y|\mathbf{x}^T\beta_0) dy = \int_0^t \{\text{sign}(u^*) - \text{sign}(y)\}^2 F(y|\mathbf{x}^T\beta_0) dy \leq 4 \int_0^t F(y|\mathbf{x}^T\beta_0) dy.$$

Thus, $\int \{\rho_{\tau}(y-t) - \rho_{\tau}(y) - \rho'_{\tau}(y)t\}^2 F(y|\mathbf{x}^T\beta_0) dy = o(t^2)$ is implied by $\int_0^t F(y|\mathbf{x}^T\beta_0) dy$ as $t \rightarrow 0$. 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 $u \in \mathbb{R}$, there exists a constant κ such that $\mathcal{L}_{\tau}(u, \beta) \leq \mathcal{L}_{\tau}(u, \kappa\beta_0)$. Specifically,

$$\begin{aligned}
\mathcal{L}_\tau(u, \beta) &= E[E\{\rho_\tau(Y - u - \beta^T \mathbf{x}) | \beta_0^T \mathbf{x}, \varepsilon\}] \\
&\geq E[\rho_\tau\{E(Y | \beta_0^T \mathbf{x}, \varepsilon) - u - E(\beta^T \mathbf{x} | \beta_0^T \mathbf{x}, \varepsilon)\}] \\
&= E[\rho_\tau\{Y - u - E(\beta^T \mathbf{x} | \beta_0^T \mathbf{x})\}] \\
&= E\{\rho_\tau(Y - u - \kappa \beta_0^T \mathbf{x})\},
\end{aligned}$$

where $\kappa = \beta^T \text{var}(\mathbf{x}) \beta_0 / \{\beta_0^T \text{var}(\mathbf{x}) \beta_0\}$. 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). Let us present the quadratic approximation lemma here to facilitate the proof.

Quadratic Approximation Lemma

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

Now we use the quadratic approximation lemma to prove Theorem 2. Let $\alpha_n = n^{-1/2}$ and $\mathbf{a} = n^{1/2}(\hat{\beta}_\tau - \beta_\tau)$. Further, let $\mathcal{L}_{\tau n}(u, \beta) = n^{-1} \sum_{i=1}^n \rho_\tau(Y_i - u - \mathbf{x}_i^T \beta)$. By applying the identity (Knight, 1998) that

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

it follows that

$$\begin{aligned}
&\mathcal{L}_{\tau n}(\xi_\tau + \alpha_n \mathbf{b}, \beta_\tau + \alpha_n \mathbf{a}) - \mathcal{L}_{\tau n}(\xi_\tau, \beta_\tau) = n^{-1} \sum_{i=1}^n \rho_\tau(Y_i - \xi_\tau - \alpha_n \mathbf{b} - \mathbf{x}_i^T \beta_\tau - \alpha_n \mathbf{x}_i^T \mathbf{a}) - n^{-1} \sum_{i=1}^n \rho_\tau(Y_i - \xi_\tau - \mathbf{x}_i^T \beta_\tau) \\
&= n^{-1} \sum_{i=1}^n \alpha_n (\mathbf{x}_i^T \mathbf{a} + \mathbf{b}) \{\mathbf{1}(Y_i - \mathbf{x}_i^T \beta_\tau \leq \xi_\tau) - \tau\} + n^{-1} \sum_{i=1}^n \int_0^{\alpha_n (\mathbf{x}_i^T \mathbf{a} + \mathbf{b})} \{\mathbf{1}(Y_i - \mathbf{x}_i^T \beta_\tau \leq \xi_\tau + t) - \mathbf{1}(Y_i - \mathbf{x}_i^T \beta_\tau \leq \xi_\tau)\} dt.
\end{aligned}$$

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

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

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

$$R_2 = \mathcal{L}_{\tau n}(\xi_\tau + \alpha_n \mathbf{b}, \beta_\tau + \alpha_n \mathbf{a}) - \mathcal{L}_{\tau n}(\xi_\tau, \beta_\tau) - E\{\mathcal{L}_{\tau n}(\xi_\tau + \alpha_n \mathbf{b}, \beta_\tau + \alpha_n \mathbf{a}) - \mathcal{L}_{\tau n}(\xi_\tau, \beta_\tau) | \mathbf{x}_1, \dots, \mathbf{x}_n\} = o_p(n^{-1}).$$

This, together with the quadratic approximation lemma of Hjort and Pollard (1993) and (A.2), leads to the asymptotic normality of $\hat{\beta}$.

A.4. Proof of Theorem 3

Recall the notations $z = \mathbf{x}_0^T \beta_\tau$, $\hat{z} = \mathbf{x}_0^T \hat{\beta}_\tau$, $Z_i = \mathbf{x}_i^T \beta_\tau$, $\hat{Z}_i = \mathbf{x}_i^T \hat{\beta}_\tau$. Let

$$a_0 = G_\tau(z) = G_\tau(\mathbf{x}_0^T \beta_\tau), b_0 = G'_\tau(z), \hat{K}_i = K\left\{\frac{\hat{Z}_i - \hat{z}}{h}\right\}, K_i = K\left\{\frac{Z_i - z}{h}\right\}, K_i = K\left\{\frac{Z_i - z}{h}\right\} \text{ and } K'_i = K'\left\{\frac{Z_i - z}{h}\right\} \text{ for } i = 1, \dots, n.$$

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

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

Define

$$\begin{aligned} R_{11} &= n^{-1} \sum_{i=1}^n K_i \left[\rho_\tau\{Y_i - a - b(\hat{Z}_i - \hat{z})\} - \rho_\tau\{Y_i - a - b(Z_i - z)\} \right], \\ R_{12} &= n^{-1} \sum_{i=1}^n (\hat{K}_i - K_i) \rho_\tau\{Y_i - a - b(Z_i - z)\}, \\ R_{13} &= E \left[\rho_\tau\{Y_i - a - b(\hat{Z}_i - \hat{z})\} - \rho_\tau\{Y_i - a - b(Z_i - z)\} \right] (\hat{K}_i - K_i). \end{aligned}$$

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

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

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

Next we deal with R_{12} . R_{12} can be decomposed into

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

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

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

which, together with the root- n consistency of $\hat{\beta}_b$ proves that $R_{12} = O_p(n^{-1/2})$

It remains to investigate the order of R_{13} . Following similar arguments to deal with R_{11} and R_{12} , we have $R_{13} = O_p(n^{-1/2})$.

Thus, (A.3) follows by combining the results of R_{1k} 's, which implies that

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

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

The rest of the proof follows literally from Theorem 3 of Fan, Hu, and Truong (1994). Based on the auxiliary data points $\{(\mathbf{x}_i^T \boldsymbol{\beta}_0, Y_i), i = 1, \dots, n\}$, 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 notational clarity, we write $\xi_i = (a - a_0) + (b - b_0)(Z_i - z)$. By using (A.1) we obtain immediately that

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

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

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

$$\begin{aligned} & 2 \sum_{i=1}^n \zeta_i \{ \mathbf{1}(Y_i - a_0 - b_0(Z_i - z) \leq 0) - \tau \} K_i \\ &= 2 \sum_{i=1}^n \zeta_i \{ F_{Y_i}(a_0 + b_0(Z_i - z) | Z_i) - \tau \} K_i \\ &+ 2 \sum_{i=1}^n \zeta_i \{ \mathbf{1}(Y_i - a_0 - b_0(Z_i - z) \leq 0) - F_{Y_i}(a_0 + b_0(Z_i - z) | Z_i) \} K_i. \end{aligned} \quad (\text{A.5})$$

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

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

This together with another Taylor's expansion entails that

$$\begin{aligned} F_{Y_i}(a_0 + b_0(Z_i - z)|Z_i) &= F_{Y_i}(G_\tau(Z_i) - G_\tau''(z)(Z_i - z)^2\{1/2 + o_p(1)\}|Z_i) \\ &= \tau - f_{Y_i}(G_\tau(Z_i)|Z_i)G_\tau''(z)(Z_i - z)^2\{1/2 + o_p(1)\}. \end{aligned}$$

With this result, this bias term can be simplified as follows.

$$2 \sum_{i=1}^n \zeta_i \{F_{Y_i}(a_0 + b_0(Z_i - z)|Z_i) - \tau\} K_i = - \sum_{i=1}^n \zeta_i f_{Y_i}(G_\tau(Z_i)|Z_i) G_\tau''(z)(Z_i - z)^2 \{1 + o_p(1)\} K_i,$$

which is of the form

$$-\frac{1}{\sqrt{nh}} \sum_{i=1}^n \beta^T \begin{pmatrix} 1 \\ (Z_i - z)/h \end{pmatrix} f_{Y_i}(G_\tau(Z_i)|Z_i) G_\tau''(z)(Z_i - z)^2 \{1 + o_p(1)\} K_i. \quad (\text{A.6})$$

The second term in the right hand of (A.5) can be approximated as follows.

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

which is again of the form

$$\frac{2}{\sqrt{nh}} \sum_{i=1}^n \beta^T \begin{pmatrix} 1 \\ (Z_i - z)/h \end{pmatrix} \{\mathbf{1}(Y_i \leq G_\tau(Z_i)) - \tau\} K_i \{1 + o_p(1)\}. \quad (\text{A.7})$$

Let $f_Y(y|Z)$ be the conditional density function of Y given Z . Then the second term in the right hand side of (A.4) can be approximated as follows.

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

which is now of the form

$$\frac{1}{nh} \sum_{i=1}^n f_{Y_i}(G_\tau(Z_i)|Z_i) \beta^T \begin{pmatrix} 1 & \{(Z_i - z)/h\} \\ \{(Z_i - z)/h\} & \{(Z_i - z)/h\}^2 \end{pmatrix} \beta K_i \{1 + o_p(1)\}. \quad (\text{A.8})$$

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

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

and

$$u_n = \sqrt{nh} \left[\frac{1}{nh} \sum_{i=1}^n \begin{pmatrix} 1 \\ \{(Z_i - z)/h\} \end{pmatrix} \{ \mathbf{1}(Y_i \leq G_\tau(Z_i)) - \tau \} K_i - \frac{1}{nh} \sum_{i=1}^n \begin{pmatrix} 1 \\ \{(Z_i - z)/h\} \end{pmatrix} f_{Y_i}(G_\tau(Z_i)|Z_i) G_\tau''(z) (Z_i - z)^2 K_i / 2 \right].$$

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

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

Because our target is the quantile function, it suffices to deal with the first element of $\mathbf{u}_n$ 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 $h^2 G_\tau''(z) \mu_2 \{ f_Y(G_\tau(z)|z) f_Z(z) \}$.

Let f_ε denote the density function of the error term ε , and ξ_τ^* denote its τ -th quantile. We can easily verify that $f_Y(G_\tau(z)|z) \sigma(z) = f_\varepsilon(\xi_\tau^*)$ 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.

Figure 3.1.

(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 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

Table 3.1

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

case (1): $\sigma(\mathbf{x}^T \hat{\beta}_0) = 1$					case (2): $\sigma(\mathbf{x}^T \hat{\beta}_0) = \exp(\mathbf{x}^T \hat{\beta}_0)$						
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
$\varepsilon \sim \text{Laplace distribution}$											
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
$\varepsilon \sim \text{Cauchy distribution}$											
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
$\varepsilon \sim \text{Laplace distribution}$											
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
$\varepsilon \sim \text{Cauchy distribution}$											
0.598	0.612	0.164	-	0.645	0.655	0.336	0.257	0.158	-	0.305	0.394

Table 3.2

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

case (1): $\sigma(x^T\beta_0) = 1$										case (2): $\sigma(x^T\beta_0) = \exp(x^T\beta_0)$									
quantile regression					Back-fitting					quantile regression					Back-fitting				
New		Oracle		algorithm		New		Oracle		New		Oracle		algorithm		New		Oracle	
Level	90%	95%	90%	95%	90%	95%	90%	95%	90%	Level	90%	95%	90%	95%	90%	Level	90%	95%	90%
$\varepsilon \sim N(0, 1)$																			
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							
$\varepsilon \sim \text{Laplace distribution}$																			
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							
$\varepsilon \sim \text{Cauchy distribution}$																			
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 3.3

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

Case	(1): $\sigma(\mathbf{x}^T \boldsymbol{\beta}_0) = 1$		(2): $\sigma(\mathbf{x}^T \boldsymbol{\beta}_0) = \exp(\mathbf{x}^T \boldsymbol{\beta}_0)$	
	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

Table 3.4

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

case (1): $\sigma(\mathbf{x}^T \hat{\beta}_0) = 1$						case (2): $\sigma(\mathbf{x}^T \hat{\beta}_0) = \exp(\mathbf{x}^T \hat{\beta}_0)$					
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
$\varepsilon \sim \text{Laplace distribution}$											
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
$\varepsilon \sim \text{Cauchy distribution}$											
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
$\varepsilon \sim \text{Laplace distribution}$											
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
$\varepsilon \sim \text{Cauchy distribution}$											
1.175	1.145	0.889	1.777	1.253	1.283	1.313	1.275	0.865	1.749	1.165	1.189

Table 3.5

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

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

Case	(1): $\sigma(x^T \beta_0) = 1$				(2): $\sigma(x^T \beta_0) = \exp(x^T \beta_0)$			
	New		Oracle		New		Oracle	
Error	Level	90%	95%	90%	95%	90%	95%	90%
Cauchy	aver.	90.6	95.5	90.4	95.4	90.6	95.5	91.1
	sidev	0.5	0.4	0.8	0.7	0.5	0.4	0.7
								0.6

Table 3.6

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

$\mathcal{E}_i^*$ 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
$.5F_n + .5N(0, .4332^2)$	aver.	0.305	-0.173	-0.630	0.351	-0.341	-0.428	0.191
	stdev.	0.058	0.051	0.051	0.073	0.060	0.047	0.036
$.5F_n + .5\text{Cauchy}$	aver.	0.303	-0.169	-0.628	0.314	-0.318	-0.421	0.184
	stdev.	0.088	0.081	0.129	0.136	0.110	0.098	0.062