Statistica Sinica **18**(2008), 1131-1152

NONPARAMETRIC PREWHITENING ESTIMATORS FOR CONDITIONAL QUANTILES

Liangjun Su[1,2] and Aman Ullah[3]

[1]*Peking University,* [2]*Singapore Management University and* [3]*University of California, Riverside*

Abstract: We define a nonparametric prewhitening method for estimating conditional quantiles based on local linear quantile regression. We characterize the bias, variance and asymptotic normality of the proposed estimator. Under weak conditions our estimator can achieve bias reduction and have the same variance as the local linear quantile estimators. A small set of Monte Carlo simulations is carried out to illustrate the performance of our estimators. An application to US gross domestic product data demonstrates the usefulness of our methodology.

Key words and phrases: Local linear quantile regression, nonparametric quantile regression, prediction interval, prewhitening estimator, weighted Nadaraya-Watson estimator.

1. Introduction

Since the seminal work of Koenker and Bassett (1978), there has developed a large literature on (conditional) quantile estimation. There are a variety of approaches to estimating conditional quantiles. These can be divided into three categories: fully parametric, semiparametric, and purely nonparametric. For a recent account of the parametric approach, see Kim and White (2003) and Komunjer (2005). The second approach includes Koenker and Zhao (1996), He and Liang (2000), Lee (2003), and Engle and Manganelli (2004), whereas the third approach includes Chaudhuri (1991) Fan, Hu and Truong (1994), Portnoy (1997), Yu and Jones (1998), Cai (2002), and Hansen (2004a), among many others.

Prior to Kim and White (2003) and Komunjer (2005), most papers in parametric quantile regressions assume explicitly or implicitly that the conditional quantile regression model is correctly specified. When the parametric model is misspecified, Kim and White (2003) show that confidence intervals and hypothesis tests based on the conventional covariance matrix are invalid, and yet we can still estimate the pseudo-true parameters consistently, under certain regularity conditions, based on the principle of quasi-maximum likelihood estimation (QMLE). Komunjer (2005) establishes necessary conditions for QMLE to work. It is worth mentioning that, under misspecification, the resultant estimator for the conditional quantile is usually inconsistent with the true quantile function, even though the parameter estimator in the misspecified model is still consistent with the pseudo-true parameter. Nevertheless, one can estimate consistently the conditional quantile function nonparametrically.

In this paper, we propose a nonparametric prewhitening estimator for conditional quantiles based on local linear quantile regression. The principle of prewhitening has a long history in the time series literature. See Press and Tukey (1956) and Andrews and Monahan (1992), among many others. Recently the principle has been applied in the context of kernel density estimation, conditional mean regression, and conditional variance estimation by Hjort and Glad (1995) and Hjort and Jones (1996), Glad (1998), and Mishra, Su and Ullah (2006), respectively. All these authors start with a *parametric* specification and propose an *nonparametric* correction term for the parametric estimator. Consequently, these are two-stage estimators that can be viewed as semiparametric and thus termed as semiparametric prewhitening estimators. They outperform the purely nonparametric estimators in terms of bias reduction and mean squared error (MSE) under weak conditions (for example, the parametric start can capture some roughness feature of the underlying density, conditional mean or variance function). In case of misspecification, they are consistent whereas the parametric ones are not.

Despite the advantage of the semiparametric prewhitening estimators over their parametric or nonparametric analogues, their bias terms usually do not disappear when the optimal rates of bandwidth are applied. One exceptional case is when the first stage parametric specification is correct, so that the object of interest can be estimated consistently (and more efficiently) by the parametric model. In principle, one can extend the semiparametric prewhitening idea to the conditional quantile regression, and we conjecture that the result will be similar. To get rid of bias terms, we propose instead to adopt a nonparametric estimator in the first stage. Under weak conditions, the first stage nonparametric estimator is consistent. After another smoothing, the new estimator will have negligible bias and the same variance as that of the traditional local linear quantile estimator, even if we adopt the optimal rates of bandwidth. Because of the nonparametric nature, we call the resulting quantile estimator a nonparametric prewhitening estimator.

In comparison with Hjort and Glad (1995), Hjort and Jones (1996) and Glad (1998), our paper is different in three aspects. First, the estimators obtained in these early papers are all two-step *semiparametric* estimators where the first step starts with *parametric* model estimation, and the second step involves the nonparametric estimation of the correction term. In our case we are doing a two-step *nonparametric* estimator where, in both steps, *nonparametric* modeling and estimation are done.

Second, the two-step semiparametric estimator can keep the same variance as the ordinary nonparametric estimator and *potentially* has a smaller bias than the latter. The bias reduction can be achieved as long as the parametric pilot model is in the "neighborhood" of the true model. The bias *can also increase* if the parametric start is far away from the true model. In contrast, with our two-step nonparametric estimation we find that the asymptotic variance is the same as the one-step local quantile estimation, but our estimator *always* gives reduction in the bias.

Third, more generally, bias reduction in the two-step semiparametric estimation depends crucially on the degree of misspecification at the first stage. If the first stage parametric model is highly misspecified, then the gain will be very little and bias can actually increase in some cases. Such an issue of misspecification at the first stage does not arise in our two-step nonparametric estimation. In fact, we demonstrate through simulations that if the first step parametric model is highly misspecified, then the two-step nonparametric estimator will be more efficient than the two-step semiparametric estimator in terms of MSE.

We provide asymptotic theory for the normality of the nonparametric prewhitening quantile estimator. In terms of bias reduction and MSE, it dominates the local linear quantile estimator of Yu and Jones (1998) in the i.i.d. framework, and that of Honda (2000) and Lu, Hui and Zhao (2001) for dependent data. We provide both simulation and empirical data analysis to assess the strength and weakness of our nonparametric prewhitening quantile estimator. In simulations we consider a variety of data generating processes (DGPs) and fit parametric, nonparametric, semiparametric, and our nonparametric prewhitening quantile regression models to them. We compare them by the median length of 90% out-of-sample prediction intervals, coverage frequencies, and mean squared errors (MSEs), and we find significant gains can be achieved by using our modelling strategy. In the empirical data analysis, we illustrate the relative performance of various conditional quantile estimators with an application to the U.S. GDP growth rate.

The paper is structured as follows. In Section 2 we introduce our nonparametric prewhitening quantile estimators and their asymptotic properties. We conduct a Monte Carlo study to check the relative performance of the proposed estimator in Section 3. Section 4 provides an empirical data analysis. All technical details are relegated to Section 5.

2. The Nonparametric Prewhitening Estimation

2.1. The nonparametric prewhitening quantile (NPPQR) estimators

In this section, we propose a nonparametric prewhitening estimator of conditional quantiles that aims at reducing bias of the local linear quantile estimator. The data are given by $\{(Y_t, X_t),\ t = 1, \ldots, n\}$.

Let $\rho_\tau(z) = z(\tau - 1(z \le 0))$ be the "check" function. It is well known that the τ-th conditional quantile $q_\tau(x)$ of Y_t given $X_t = x$ satisfies

$$q_\tau(x) = \arg\min_f E\Big[\rho_\tau(Y_t - f(X_t)|X_t = x)\Big], \tag{2.1}$$

where we assume that the solution to the above minimization problem is unique and that f belongs to a space of measurable functions defined on $\mathbb{R}^d$. Denote by $\dot{q}_\tau(x) = (\partial q_\tau(x)/\partial x_1, \ldots, \partial q_\tau(x)/\partial x_d)'$ the first order derivative of $q_\tau(x)$ at $x = (x_1, \ldots, x_d)' \in \mathbb{R}^d$, where x' denotes the transpose of a vector x. The idea of the local linear fit is to approximate the unknown τ-th quantile $q_\tau(\cdot)$ by a linear function $q_\tau(z) \simeq q_\tau(x) + \dot{q}_\tau(x)'(z - x) \equiv \beta_0 + \beta_1'(z - x)$ for z in a neighborhood of x. Locally, estimating $q_\tau(x)$ is equivalent to estimating β_0, and estimating $\dot{q}_\tau(x)$ is equivalent to estimating β_1. This motivated Yu and Jones (1998) to define a local linear quantile regression ($LLQR$) estimator of $q_\tau(x)$ by $q_\tau^+(x) \equiv \beta_0^+$, where

$$\{\beta_0^+, \beta_1^+\} \equiv \arg\min_{\{\beta_0,\beta_1\}} n^{-1} \sum_{t=1}^n \rho_\tau\Big(Y_t - \beta_0 - \beta_1'(X_t - x)\Big) K_{h_0}(X_t - x), \tag{2.2}$$

$K_{h_0}(x) = K(x/h_0)$, K is a kernel function on $\mathbb{R}^d$, and $h_0 \equiv h_0(n)$ is the bandwidth.

Under suitable conditions, Lu, Hui and Zhao (2001) showed that $q_\tau^+(x)$ has the Bahadur representation

$$\sqrt{nh_0^d}\Big(q_\tau^+(x) - q_\tau(x)\Big) = \phi_\tau(x) \frac{1}{\sqrt{nh_0^d}} \sum_{t=1}^n \psi_\tau(Y_t^*(x,\tau) K_{h_0}(X_t - x) + o_P(1), \tag{2.3}$$

where $\psi_\tau(y) = \tau - 1(y \le 0)$, $Y_t^*(x,\tau) = Y_t - q_\tau(x) - \dot{q}_\tau(x)'(X_t - x)$, $\phi_\tau(x) = 1/[f_{Y|X}(q_\tau(x)|x) f_X(x)]$, $f_{Y|X}(\cdot|x)$ is the conditional density of Y given $X = x$, and $f_X(\cdot)$ is the marginal density of X. If one further assumes that the second order derivatives of $q_\tau(x)$ exist, then for an interior point x,

$$MSE(q_\tau^+(x)) \simeq \left\{\frac{1}{2} h_0^2 \mathrm{tr}\Big[\ddot{q}_\tau(x) \int uu' K(u) du\Big]\right\}^2 + \frac{\tau(1-\tau)\int K(u)^2 du}{nh_0^d [f_{Y|X}(q_\tau(x)|x)]^2 f_X(x)}, \tag{2.4}$$

where $\ddot{q}_\tau(x) = [\partial^2 q_\tau(x)/\partial x_i \partial x_j]$. Consequently, the optimal rate of bandwidth in terms of minimizing the MSE is proportional to $n^{-1/(d+4)}$. When x lies on the boundary of the support, the MSE formula looks similar. This reflects the two major advantages of local linear fitting and shows that these advantages apply to the local quantile regression too: (a) no dependence of the asymptotic bias on the density $f_X(x)$, and (b) automatic good behavior at boundaries. From the formula in (2.4), we can see that the asymptotic bias depends on the simple quantile curvature function $\ddot{q}_\tau$. When $\ddot{q}_\tau(x) = 0$, the MSE of $q^+_\tau(x)$ is dominated by the asymptotic variance. This motivates us to consider the following local linear quantile regression

$$\{\beta^*_0, \beta^*_1\} \equiv \arg \min_{\{\beta_0,\beta_1\}} n^{-1} \sum_{t=1}^{n} \rho_\tau\Big(\frac{Y_t q_\tau(x)}{q_\tau(X_t)} - \beta_0 - \beta'_1(X_t - x)\Big) K_h(X_t - x), \quad (2.5)$$

where $h \equiv h(n)$ is the bandwidth.

For simplicity, we assume that the observed Y_t are positive. (If not, let C_n be a positive number so that $\overline{Y}_t \equiv Y_t + C_n$ are all positive. One can first obtain the conditional quantile of $\overline{Y}_t$ given X_t, and then subtract C_n to get the conditional quantile of Y_t given X_t.) Denote by $nc_\tau(x)$ the τ-th conditional quantile of $Y_t/q_\tau(X_t)$ given $X_t = x$, and by $qc_\tau(x)$ the τ-th conditional quantile of $Y_t q_\tau(x)/q_\tau(X_t)$ given $X_t = x$. Then $nc_\tau(x) = 1$, $qc_\tau(x) = q_\tau(x)$, so that the pseudo-true parameter $\beta^0 = (\beta^0_0, \beta^{0'}_1)'$ for the regression in (2.5) is $\beta^0_0 = nc_\tau(x)q_\tau(x) = q_\tau(x)$ and $\beta^0_1 = \dot{nc}_\tau(x)q_\tau(x) = 0$. In comparison with β^+_0, the estimator β^*_0 can have the same variance as β^+_0 but negligible asymptotic bias because $nc_\tau(x)$ is a constant function whose derivatives are all zero.

Since $q_\tau(\cdot)$ is not observable, we propose to replace it in (2.5) by $q^+_\tau(\cdot)$. To summarize, we propose nonparametric prewhitening quantile estimators that are obtained in two steps:

1. *Obtain the usual local linear quantile estimator of* $q_\tau(\widetilde{x})$ *as in* (2.2) *and denote it by* $q^+_\tau(\widetilde{x})$ *for* $\widetilde{x} = X_1, \ldots, X_n$.

2. *Obtain the nonparametric prewhitening quantile regression* (*NPPQR*) *estimator of* $q_\tau(x)$ *by replacing* $q_\tau(\cdot)$ *in* (2.5) *with* $q^+_\tau(\cdot)$ *to obtain* $\widehat{q}_\tau(x) \equiv \widehat{\beta}_0$, *where*

$$\{\widehat{\beta}_0, \widehat{\beta}_1\} \equiv \arg \min_{\{\beta_0,\beta_1\}} n^{-1} \sum_{t=1}^{n} \rho_\tau\left(\frac{Y_t q^+_\tau(x)}{q^+_\tau(X_t)} - \beta_0 - \beta'_1(X_t - x)\right) K_h(X_t - x). \quad (2.6)$$

Remark 1. As Kauermann, Müller and Carroll (1998) remark, there are two general ways to reduce bias in nonparametric regression: (a) estimate the second derivative function ($\ddot{q}_\tau(x)$ here) and subtract a multiple of it from the usual nonparametric estimator; (b) reduce the bias indirectly either by undersmoothing or by the twicing technique. Method (a) is sensitive to the choice of a second bandwidth, whereas method (b) will increase the variance (say by a factor 1.44 for the Gaussian kernel and 1.42 for the Epanechnikov kernel, independent of the problem). In contrast, we show that our two-stage nonparametric estimator offers an effective way to reduce the bias and keep the variance of the one-step nonparametric estimator unchanged.

Remark 2. We show that $\widehat{q}_\tau(x)$ behaves similarly to β^*_0 and it dominates $q^+_\tau(x)$ in terms of MSE. To see why our two-step procedure helps, from (2.8) below one deduces that, uniformly in x,

$$\begin{aligned}
&q_\tau^+(x) - q_\tau(x) \\
&\quad = \frac{\phi_\tau(x)}{nh_0^d}\sum_{t=1}^n E[\psi_\tau(Y_t^*)K_{h_0}(X_t - x) + \frac{\phi_\tau(x)}{nh_0^d}\sum_{t=1}^n \Big\{\psi_\tau(Y_t^*)K_{h_0}(X_t - x) \\
&\qquad - E[\psi_\tau(Y_t^*)K_{h_0}(X_t - x)]\Big\} + o_P\Big(n^{-\frac{1}{2}}h_0^{-\frac{d}{2}}\Big) \\
&\quad \equiv b_n(x) + v_n(x) + o_P\Big(n^{-\frac{1}{2}}h_0^{-\frac{d}{2}}\Big). \qquad (2.7)
\end{aligned}$$

That is, $b_n(x)$ and $v_n(x)$ contribute to the asymptotic bias and variance of $q_\tau^+(x)$, respectively. Under Assumptions A.1-A.3 below, one can show that $\sup_x|b_n(x)| = O(h_0^2)$, $\sup_x|v_n(x)| = O_P(n^{-1/2}h_0^{-d/2}\sqrt{\log n}) = o_P(n^{-1/2}h^{-d/2})$, and hence

$$\frac{q_\tau^+(x)}{q_\tau(x)} = 1 + \frac{b_n(x)}{q_\tau(x)} + o_P\Big(n^{-\frac{1}{2}}h^{-\frac{d}{2}}\Big) \text{ uniformly in } x.$$

Consequently

$$\begin{aligned}
&\frac{q_\tau^+(X_t)}{q_\tau(X_t)} \times \frac{q_\tau(x)}{q_\tau^+(x)} \\
&\quad = \left[1 + \frac{b_n(x_t)}{q_\tau(X_t)} + o_P\Big(n^{-\frac{1}{2}}h^{-\frac{d}{2}}\Big)\right] \times \left[1 - \frac{b_n(x)}{q_\tau(x)} + O_p(h_0^4) + o_P\Big(n^{-\frac{1}{2}}h^{-\frac{d}{2}}\Big)\right] \\
&\quad = 1 + \left[\frac{b_n(x_t)}{q_\tau(X_t)} - \frac{b_n(x)}{q_\tau(x)}\right] + o_P\Big(n^{-\frac{1}{2}}h^{-\frac{d}{2}}\Big) \\
&\quad = 1 + o_P\Big(n^{-\frac{1}{2}}h^{-\frac{d}{2}}\Big) \text{ on the set } \{K_h(X_t - x) > 0\},
\end{aligned}$$

where the second equality follows from the fact that $h_0^4 = o(n^{-1/2}h^{-d/2})$ by Assumption A.3, and the last equality follows from Assumptions A.1(vi) and A.2. See Corollary 5.3 for details. The last result implies that $q_\tau^+(x)/q_\tau^+(X_t)$ is equal to $q_\tau(x)/q_\tau(X_t)$ to order $o_P(n^{-1/2}h^{-d/2})$, and replacing $q_\tau(x)/q_\tau(X_t)$ in (2.5) by $q_\tau^+(x)/q_\tau^+(X_t)$ in (2.6) has negligible effect on the estimation of β_0. In other words, $\widehat{q}_\tau(x)$ is asymptotically equivalent to β_0^*.

Remark 3. Our result relies on the crucial assumption that $(h/h_0)^d \log n = o(1)$, which implies that $h_0 \gg h$ and seems contradictary to bias reduction. For example, undersmoothing is frequently used in practice to correct for bias. Typical twicing techniques (e.g., Jones, Linton and Nielsen (1995), Kauermann, Müller and Carroll (1998)) choose the same bandwidth in both stages, which can make the leading bias terms from the first stage and second stage cancel out but inflate the variance at the same time. In our case, by choosing a larger bandwidth in the first-stage nonparametric regression, we can effectively control the variance of the first-stage estimator. The bias can be reduced because we find that only the ratio $q_\tau^+(x)/q_\tau^+(X_t)$ matters, and only those observations X_t in the neighborhood of x play key roles in the second-stage smoothing.

2.2. Asymptotic Theory for the Nonparametric-prewhitening Estimators

To introduce the theory, we make the following assumptions.

(A1) (i) The process $\{Y_t, X_t\}$ is strictly stationary α-mixing, with mixing coefficients $\alpha(j)$ satisfying $\sum_{j=1}^{\infty} j^a \alpha(j)^{\delta/(\delta+2)} < \infty$ for some $\delta > 0$ and $a > \delta/(\delta+2)$.

(ii) The marginal density function $f_X(\cdot)$ of X_t is continuous and $f_X(\widetilde{x}) > 0$ for all $\widetilde{x}$ on its bounded support $\mathcal{X}$.

(iii) The conditional density function $f_{Y|X}(y|\widetilde{x})$ of Y_t given $X_t = \widetilde{x}$ is continuous as a function of y in a neighborhood of $q_\tau(x)$, uniformly for all $\widetilde{x} \in \mathcal{X}$, and continuous as a function of $\widetilde{x}$ for all y in a neighborhood of $q_\tau(x)$. Also, $f_{Y|X}(q_\tau(\widetilde{x})|\widetilde{x}) > 0$ for all $\widetilde{x} \in \mathcal{X}$.

(iv) The joint density function $f_j(\widetilde{x}, x^*)$ of (X_1, X_{1+j}) is bounded uniformly in $j > 0$.

(v) The quantile function $q_\tau(\cdot)$ has continuous second order derivatives on its support.

(vi) The functions $f_X(\cdot)$, $f_{Y|X}(\cdot|\cdot)$, $q_\tau(\cdot)$, $\ddot{q}_\tau(\cdot)$ are Lipschitz continuous of degree 1.

(A2) The kernel K is a bounded density function with compact support on $\mathbb{R}^d$ such that $\int uK(u)du = 0$ and $\mu_2 = \int uu'K(u)du$ is positive definite.

(A3) As $n \to \infty$, (i) $(h/h_0)^d \log n \to 0$, $nh^d h_0^6 \to 0$, (ii) $nh^d \to \infty$.

Assumptions A1$-$A3 parallel those of Lu, Hui and Zhao (2001) but are stronger than theirs. Assumption A1(i) is standard in nonparametric regression whereas Assumptions A1(ii)$-$(v) are frequently assumed in nonparametric quantile regression. Assumption A1(vi) can be weakened to allow $f_X(\cdot)$, $f_{Y|X}(\cdot|\cdot)$, $q_\tau(\cdot)$ and $\ddot{q}_\tau(\cdot)$ to satisfy Lipschitz condition of degree κ $(0 < \kappa \le 1)$. In this case, the conditions for the bandwidth sequences will be modified correspondingly. As Hall, Wolf and Yao (1999) remark, the requirement in Assumption A2 that K is compactly supported can be removed at the cost of lengthier arguments used in the proofs and, in particular, Gaussian kernel is then allowed. Assumption A3 implies that $h \to 0$, $h_0 \to 0$ and $nh_0^d/\log n \to \infty$ as $n \to \infty$. It also implies that $n^{-1/2}h_0^{-d/2}\sqrt{\log n}$ and h_0^3 are both of smaller order than $n^{-1/2}h^{-d/2}$.

Next, we state a theorem that is used in the proof of our main theorem.

Theorem 2.1. *Under Assumptions* A1$-$A3,

$$\sqrt{nh_0^d}(q_\tau^+(x) - q_\tau(x)) = \phi_\tau(x)\frac{1}{\sqrt{nh_0^d}}\sum_{t=1}^{n}\psi_\tau(Y_t^*(x,\tau))K_{h_0}(X_t - x) + o_P(1), \quad (2.8)$$

where $o_P(1)$ holds uniformly in $x \in \mathcal{X}$.

Remark 4. Honda (2000) obtains uniform strong Bahadur representation for local polynomial estimators of conditional quantiles under a different set of conditions. In particular, he assumes that $q_\tau(x)$ is p-times differentiable and applies $h_0 \propto n^{-1/(d+2p)}$. One has to modify his condition on the bandwidth in order to adapt his result to our framework. Lu, Hui and Zhao (2001) obtain weak Bahadur representation for local linear quantile regression estimators. Their result is a pointwise result and can be extended to obtain the uniform result in the above theorem. Some other results for Bahadur representations are available in the literature. For example, He and Shao (1996) obtain strong Bahadur representation for a general class of M-estimators which include our quantile regression estimators. Portnoy (1997) establishes Bahadur representations for local parameters of the quantile smoothing splines.

We now state the main theorem.

Theorem 2.2. *Under Assumptions* A1$-$A3,

$$\sqrt{nh^d}(\widehat{q}_\tau(x) - q_\tau(x)) \xrightarrow{d} N\left(0, \frac{\tau(1-\tau)\int K(u)^2 du}{[f_{Y|X}(q_\tau(x)|x)]^2 f_X(x)}\right). \tag{2.9}$$

Remark 5. The above theorem implies that the asymptotic bias of $\sqrt{nh^d}\widehat{q}_\tau(x)$ is negligible, whereas its asymptotic variance remains the same as the asymptotic variance of $\sqrt{nh^d}q_\tau^+(x)$ when the same bandwidth and kernel are used in both cases. From Assumption A3, we see that we can take h to be proportional to $n^{-1/(d+6-\epsilon)}$ for arbitrarily small positive ϵ. This implies that $\widehat{q}_\tau(x)$ is infinitely more efficient than the conventional local linear quantile estimator, because $\widehat{q}_\tau(x)$ can converge to $q_\tau(x)$ at a rate approximating $n^{-1/2+d/[2(d+6)]}$ whereas the traditional one can converge to $q_\tau(x)$ at the best rate $n^{-1/2+d/[2(d+4)]}$.

3. Monte Carlo Simulations

We illustrate the performance of the proposed nonparametric prewhitening quantile estimator with a small set of simulations. In addition to the proposed estimator, we also study several other parametric, nonparametric and semiparametric quantile estimators.

3.1. Other Estimators for Conditional Quantiles

Parametric Quantile Regression (PQR) Estimator

A linear parametric quantile regression (PQR) estimator for $q_\tau(x)$ is given by $\widehat{q}_\tau^{pqr}(x) = \overline{x}'\overrightarrow{\beta}$, where $\overline{x} = (1, x')'$ and

$$\overrightarrow{\beta} \equiv \arg\min_{\beta\in\mathbb{R}^{d+1}} n^{-1}\sum_{t=1}^{n} \rho_\tau(Y_t - \overline{X}_t'\beta). \tag{3.1}$$

Weighted Nadaraya-Watson (WNW) Estimator

Denote by $F_{Y|X}(y|x)$ the conditional distribution function Y given $X = x$. Motivated by the good boundary properties of local polynomial estimators, Hall, Wolf and Yao (1999) suggest estimating $F(y|x)$ by a weighted version of the well known Nadaraya-Watson (NW) estimator:

$$\widehat{F}_{wnw}(y|x) = \frac{\sum_{t=1}^{n} p_t(x)K_{h_1}(X_t - x)1(Y_t \le y)}{\sum_{t=1}^{n} p_t(x)K_{h_1}(X_t - x)}, \tag{3.2}$$

where $h_1 \equiv h_1(n)$ is the bandwidth, and one chooses the nonnegative weight functions $p_t(x)$, $1 \le t \le n$, such that

$$\sum_{t=1}^{n} p_t(x) = 1, \quad \text{and} \quad \sum_{t=1}^{n} p_t(x)(X_t - x)K_{h_1}(X_t - x) = 0. \tag{3.3}$$

More recently, Cai (2002) proposed choosing $\{p_t(x)\}$ based on the idea of empirical likelihood, i.e., to maximize $\sum_{t=1}^{n} \log\{p_t(x)\}$ subject to the constraints specified in (3.3). He proposed inverting $\widehat{F}_{wnw}$ to get the conditional quantile estimator:

$$\widehat{q}_\tau^{wnw}(x) \equiv \inf\Big\{y \in \mathbb{R} : \widehat{F}_{wnw}(y|x) \ge \tau\Big\}. \tag{3.4}$$

Smoothed Local Linear (SLL) Estimator

Let l be a symmetric density function on $\mathbb{R}$ and L the corresponding distribution function. Yu and Jones (1998) propose a smoothed local linear estimator for conditional quantiles that is based on the observation that $E[L((y-Y_t)/h_2)|X_t = x] \to F(y|x)$as the bandwidth $h_2 \to 0$. To obtain the smoothed local linear (SLL) estimator for the conditional quantile function, one first obtains

$$(\overleftarrow{\beta}_0, \overleftarrow{\beta}_1) \equiv \arg\min_{\beta} \sum_{t=1}^{n} \Big\{L\Big(\frac{y - Y_t}{h_2}\Big) - \beta_0 - \beta_1'(X_t - x)\Big\}^2 K_{h_2}(X_t - x), \tag{3.5}$$

where $\beta = (\beta_0, \beta_1')' \in \mathbb{R} \times \mathbb{R}^d$, $h_2 \equiv h_2(n)$ is the bandwidth. Set $\widehat{F}_{sll}(y|x) = \widehat{\beta}_0$. Yu and Jones (1998) proposed inverting $\widehat{F}_{sll}$ to get the conditional quantile estimator

$$\widehat{q}_\tau^{sll}(x) \equiv \inf\{y \in \mathbb{R} : \widehat{F}_{sll}(y|x) \ge \tau\}. \tag{3.6}$$

Note that $\widehat{F}_{sll}(y|x)$ can range outside [0, 1]. In the special case where $d = 1$, it can be expressed as $\widehat{F}_{sll}(y|x) = \sum_{t=1}^{n} w_t(x)L((y - Y_t)/h_2) \sum_{s=1}^{n} w_s(x)$, where $w_t(x) = K_{h_1}(X_t - x)(1 - \widehat{\beta}_x(x - X_t))$, in which $\widehat{\beta}_x \equiv (\sum_{t=1}^{n} K_{h_1}(X_t - x)(x - X_t)^2)^{-1} \sum_{t=1}^{n} K_{h_1}(X_t - x)(x - X_t)$. To obtain an monotone estimator for $F(y|x)$

that lies between 0 and 1, Hansen (2004a) proposed replacing $w_t(x)$ by $w_t^*(x) \equiv K_{h_1}(X_t - x)(1 - \widehat{\beta}_x(x - X_t))1\{\widehat{\beta}_x(x - X_t) \leq 1\}$.

Semi-parametric Prewhitening Quantile Regression (SPPQR) Estimator

Motivated by the important work of Hjort and Glad (1995), Hjort and Jones (1996) and Glad (1998), we can start with parametric quantile regression (PQR) estimation in the first stage, and then proceed to the nonparametric estimation in the second stage. The resulting estimator is obtained as $\widehat{q}_\tau^{sppqr}(x) \equiv \widetilde{\beta}_0$, where

$$\{\widetilde{\beta}_0, \widetilde{\beta}_1\} \equiv \arg \min_{\{\beta_0, \beta_1\}} n^{-1} \sum_{t=1}^{n} \rho_\tau \left(\frac{Y_t \widehat{q}_\tau^{pqr}(x)}{\widehat{q}_\tau^{pqr}(X_t)} - \beta_0 - \beta_1'(X_t - x) \right) K_h(X_t - x). \quad (3.7)$$

3.2. Simulation

We first consider an ARCH(1) process:

$$Y_t = 0.6Y_{t-1} + \sqrt{0.3 + 0.9Y_{t-1}^2}\varepsilon_t,$$

where $\{\varepsilon_t\}$ are i.i.d. $N(0, 1)$. Here, $X_t = Y_{t-1}$. Despite the linear feature of the conditional mean function, the conditional quantile function is highly nonlinear so we expect that the linear quantile regression works poorly in this case. Like Cai (2002), we generated $n + 5$ data points $\{Y_t, X_t\}$, where the first n observations were used for estimation of the conditional quantiles, whereas the last 5 data points were used for out-of-sample evaluation. We chose $n = 100$, 200 and 500. The 90% prediction intervals (PI) $[q_{0.05}(Y_{n+j-1}), q_{0.95}(Y_{n+j-1})]$ $(1 \leq j \leq 5)$ were computed for each of the quantile estimators under study.

Hansen (2004a,b) proposed plug-in bandwidth selection for estimating $\widehat{F}_{wnw}(y|x)$ and $\widehat{F}_{sll}(y|x)$ in the case where X_t is a scalar random variable. Since $d = 1$ here, our preliminary choice of (h, h_1, h_2) is obtained according to Hansen (2004a,b). Since our interest is to estimate the conditional quantiles, we then adjusted the preliminary choice of bandwidth according to the rule of thumb recommended by Yu and Jones (1998): for example, $h = h^*\{\tau(1-\tau)[\phi(\Phi^{-1}(\tau))]^{-2}\}^{1/5}$, where h^* is the preliminary bandwidth obtained from Hansen (2004a), ϕ and Φ are the standard normal pdf and cdf, respectively. To obtain the nonparametric prewhitening estimator, we set $h_0 = 0.75hn^{-1/6}/n^{-1/5} = 0.75hn^{1/30}$ to guarantee $h \ll h_0$.

It is worth mentioning that according to Hansen (2004a,b), his choices of bandwidth are designed to minimize the (in-sample) integrated mean squared error of $\widehat{F}_{sll}(y|x)$. We used them to obtain other estimators for conditional quantiles with no attention to optimizing the performance of either the conventional nonparametric quantile estimators or our nonparametric prewhitening quantile

Table 1. The postsample prediction for the ARCH model: ratio of PI length vs. range of data

n	Estimators	Y_{n+1}	Y_{n+2}	Y_{n+3}	Y_{n+4}	Y_{n+5}
100	PQR	1	1	1	1	1
	WNW	0.674	0.673	0.673	0.676	0.686
	SLL	0.857	0.858	0.856	0.852	0.861
	LLQR	0.702	0.701	0.702	0.703	0.707
	NPPQR	0.606	0.614	0.604	0.612	0.615
	SPPQR	0.701	0.704	0.698	0.703	0.706
200	PQR	1	1	1	1	1
	WNW	0.643	0.638	0.642	0.647	0.644
	SLL	0.840	0.837	0.832	0.848	0.837
	LLQR	0.678	0.665	0.670	0.675	0.674
	NPPQR	0.574	0.578	0.567	0.580	0.571
	SPPQR	0.678	0.667	0.672	0.675	0.674
500	PQR	1	1	1	1	1
	WNW	0.613	0.616	0.617	0.619	0.619
	SLL	0.857	0.864	0.864	0.858	0.856
	LLQR	0.657	0.661	0.654	0.658	0.654
	NPPQR	0.561	0.568	0.551	0.557	0.556
	SPPQR	0.659	0.661	0.654	0.658	0.655

Note: The table reports the normalized ratio of the length of PI vs the IQR of the data for different quantile estimators: the entry elements are the PI length of different estimators over the PI length of the PQR estimator. Small numbers mean tighter PI. The number of repetitions is 1,000 for each n.

estimator. In all cases, we chose K and l (in obtaining the SLL quantile estimator) to be the standard normal density function. The number of repetitions was 1,000 for each n unless otherwise stated.

In Table 1 we present the median of the 1,000 values for the ratio of the length of PI versus the interquantile range (IQR) of the data $\{Y_t\}$, where the ratio of PQR PI versus IQR is normalized to be 1. We find the following interesting points: (1) in terms of PI length, the PQR is worst; (2) the NPPQR outperforms the LLQR and all other estimators, indicating potential gain of nonparametric prewhitening in reducing the length of prediction interval.

In Table 2 we present the coverage frequencies of 90% PI based on quantile estimation. We find: (1) the PQR and SLL estimation of PI tends to be overcovered for small sample sizes and, as sample sizes increase, the coverage frequency of the PQR estimate tends to the nominal 90%; (2) the WNW estimate of PI tends to be under-covered; (3) in terms of correct coverage ratio, the LLQR, NPPQR and SPPQR estimators work reasonably well.

Next, we considered the following DGPs in the i.i.d. framework:

Table 2. The postsample prediction for the ARCH model: coverage frequencies.

n	Estimators	Y_{n+1}	Y_{n+2}	Y_{n+3}	Y_{n+4}	Y_{n+5}
100	PQR	0.93	0.94	0.93	0.92	0.91
	WNW	0.85	0.85	0.85	0.84	0.82
	SLL	0.95	0.96	0.96	0.94	0.94
	LLQR	0.90	0.91	0.92	0.91	0.89
	NPPQR	0.88	0.89	0.88	0.88	0.87
	SPPQR	0.90	0.91	0.92	0.91	0.89
200	PQR	0.92	0.92	0.92	0.93	0.92
	WNW	0.84	0.84	0.82	0.84	0.83
	SLL	0.95	0.96	0.95	0.95	0.94
	LLQR	0.90	0.91	0.91	0.92	0.91
	NPPQR	0.88	0.90	0.89	0.89	0.89
	SPPQR	0.90	0.91	0.91	0.92	0.91
500	PQR	0.91	0.91	0.92	0.91	0.90
	WNW	0.80	0.82	0.80	0.80	0.82
	SLL	0.95	0.95	0.96	0.96	0.95
	LLQR	0.90	0.90	0.89	0.91	0.89
	NPPQR	0.89	0.90	0.88	0.89	0.88
	SPPQR	0.90	0.90	0.89	0.90	0.89

Note: The table reports the coverage frequencies for different quantile estimators to be compared with 90%. The numbers are calculated based on 1,000 repetitions.

DGPs 1−3: $Y_t = 1 + X_t + \varepsilon_t$, $X_t \sim$ uniform$[-2, 2]$, and $\varepsilon_t \sim N(0, 1)$, Gamma (1, 2), and Cauchy (0, 1) in DGPs 1−3, respectively;

DGPs 4−6: $Y_t = 1 + X_t + 2\cos(X_t^2)\sin(X_t) + \varepsilon_t$, $X_t \sim$ uniform$(-2, 2)$, and $\varepsilon_t \sim N(0, 1)$, Gamma (1, 2), and Cauchy (0, 1) in DGPs 4−6, respectively;

DGP 7: $Y_t = \sin(0.75X_t) + 1 + 0.3\sqrt{\sin(0.75X_t) + 1}\varepsilon_t$, $X_t \sim N(0, 0.0625)$, $\varepsilon_t \sim N(0, 1)$;

DGP 8: $Y_t = 2.5 + \sin(2X_t) + 2\exp(-16X_t^2) + 0.5\varepsilon_t$, $X_t \sim N(0, 1)$, $\varepsilon_t \sim N(0, 1)$;

DGP 9: $Y_t = 2 + 2\cos(2X_t) + \exp(-4X_t^2) + \varepsilon_t$, $X_t \sim N(0, 1)$, ε_t is exponential with mean 1;

DGP 10: $Y_t = 2 + X_t + \exp(-X_t)(\varepsilon_t - \log 2.6)$, $X_t \sim$ uniform$(0, 5)$, ε_t is exponential with mean 1.

DGPs 1−6 were studied in Min and Kim (2004) whereas DGPs 7−10 were studied in Yu and Jones (1998). Clearly, the true quantile function is linear for DGPs 1−3 and almost linear in DGPs 7 and 10 given the range of realizations of $\{X_t\}$. DGPs 2, 4, and 9−10 are used to examine the effect of skewness on quantile estimation, whereas DGPs 3 and 6 are for checking the effect of fat tails on quantile estimation. We have heteroskedastic errors in DGPs 7 and 10.

Table 3. Mean square errors for conditional median estimators

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8	DGP9	DGP10
PQR	**0.028**	**0.071**	**0.048**	1.016	1.081	1.070	**0.002**	0.663	0.967	**0.002**
WNW	0.087	0.157	0.134	0.294	0.481	0.575	0.100	0.388	0.712	0.375
SLL	0.037	0.103	0.086	0.152	0.348	0.386	0.003	0.202	0.141	0.009
LLQR	0.042	0.102	0.077	0.168	0.343	0.376	0.003	0.289	0.123	0.005
NPPQR	0.065	0.154	0.113	**0.108**	**0.275**	**0.193**	0.004	**0.143**	**0.066**	0.006
SPPQR	0.043	0.104	0.077	0.169	0.345	0.376	0.003	0.294	0.124	0.005

Note: The table reports the MSEs of conditional median estimators based on different quantile estimation techniques. The MSEs are averaged across 200 simulation repetitions and across all observations of X_t that lie within 1.5 standard deviations of the sample averages. Boldfaced elements indicate the corresponding estimator is best for the DGP in the column. $n = 100$.

We report in Table 3 the mean square errors (MSEs) of the conditional median estimators averaged over 200 simulation repetitions and over observations on X_t that lie within 1.5 standard deviations of the sample averages. To save on computation time, we limited ourselves to the case where $n = 100$. The results are interesting. (1) As expected, the PQR estimator is the best when the true quantile function is linear (DGPs 1−3), or almost linear (DGPs 7 and 10), and it is the worst when the true conditional quantile function is highly nonlinear. (2) NPPQR outperforms LLQR in cases where the true conditional median function is highly nonlinear (DGPs 4−6, 8 and 9), as well as all other estimators under study. This suggests that a test of linear conditional quantile function may be extremely helpful toward obtaining better conditional quantile estimators. If one rejects linearity with strong evidence, one should use the NPPQR estimator.

To check whether nonparametric prewhitening can help reduce the length of a prediction interval, we also estimated the 5th and 95th conditional percentiles at a fixed point $x = 0.25$, and then evaluated the 90% PI based on these two quantile estimators for all DGPs and all estimators under study. In this case, the true realization of Y was obtained by drawing the error term randomly from the corresponding distribution and taking the regressor to be 0.25. Table 4 reports the PI vs IQR and the coverage frequencies. First, we focus on the PI vs IQR. For ease of comparison, we normalize the PI vs IQR for the PQR estimator to be 1. A number smaller than 1 means reduction in PI length. As we can see from the top part of Table 4, for all DGPs, the NPPQR outperforms other estimators in most cases. It is worth mentioning that even if the true quantile function is linear (DGPs 1−3), one can still shorten the prediction interval by applying NPPQR. When the true quantile function is nonlinear, more gains can be obtained by applying NPPQR than LLQR in most cases.

For the 90% coverage frequency, we can see from the second part of Table 4: that (1) NPPQR is relatively more stable than other estimators, (2) for DGP7,

Table 4. The PI vs IQR and coverage frequencies

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8	DGP9	DGP10
PI/IQR										
PQR	1	1	1	1	1	1	1	1	1	1
WNW	1.31	1.17	0.97	1.13	1.09	0.96	1	0.97	0.91	1.56
SLL	1.31	1.17	0.98	1.14	1.10	0.97	1.07	0.90	0.94	1.98
LLQR	1	1	0.98	0.88	0.97	0.97	0.99	0.77	0.84	1.63
NPPQR	1	1	0.98	0.83	0.95	0.95	0.98	0.71	0.81	1.61
SPPQR	1	1	0.98	0.89	0.98	0.97	0.99	0.78	0.84	1.62
Coverage										
PQR	0.91	0.90	0.91	0.96	0.94	0.90	1	0.96	0.89	0.62
WNW	0.97	0.94	0.91	0.98	0.95	0.90	1	0.99	0.93	0.73
SLL	0.97	0.95	0.91	0.98	0.95	0.90	1	0.99	0.93	0.95
LLQR	0.91	0.90	0.91	0.94	0.95	0.90	1	0.98	0.92	0.85
NPPQR	0.91	0.90	0.91	0.91	0.94	0.89	1	0.98	0.93	0.85
SPPQR	0.91	0.90	0.91	0.93	0.95	0.89	1	0.98	0.92	0.85

Note: The table reports the PI vs IQR of the data, and coverage frequency for different quantile estimators. The PI vs IQR is normalized to be 1 for the PQR estimator. The theoretical coverage frequency is 90%. The numbers were calculated based on 1000 repetitions. $n = 200$.

the coverage frequency is 100% across all estimators because the signal/noise ratio is very large in this case (a similar story holds for DGP 8), and (3) for DGP 10, where both skewness and heteroskedasticity are present, even though the true quantile function is almost linear the coverage frequency for PQR and WNW breaks down.

4. Empirical Analysis

Here we check the relative performance of quantile estimators with application to the U.S. real gross domestic product (GDP) growth rate. Let GDP_t denote the level of quarterly real GDP. Let $Y_t \equiv 100(\ln(GDP_t) - \ln(GDP_{t-4}))$ denote the annual quarter-to-quarter growth rate. Set $X_t = Y_{t-1}$. We have the data on real GDP from the first quarter of 1946 to the first quarter of 2004, yielding 229 observations, the number of observations on $\{Y_t, X_t\}$ is 224.

To see how different quantile estimators can be used to predict future GDP growth rates, we leave the last 40 observations for forecasting evaluation and use the rolling window forecasting scheme. To be specific, the sample is divided into an in-sample part of size n_1, and an out-of-sample part of size n_2, where $n_1 = 184$ and $n_2 = 40$. At time $t \geq n_1$, we use all observations up to time t to form the quantile estimation, and then predict the conditional quantiles for time $t+1$, with $\tau = 0.05$, 0.50, and 0.95. We take the 50th conditional percentile as the point forecast, and the 5th and 95th conditional percentile as the lower and

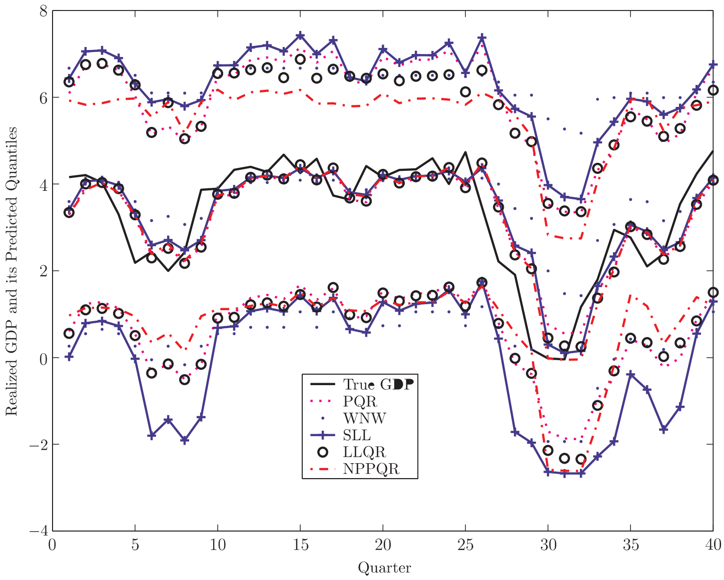

Figure 1. 5th, 50th and 95th 1-quarter-ahead predicted quantile estimators for various estimation techniques (1994Q2-2004Q1).

Table 5. The postsample prediction for the U.S. quarterly GDP growth rate: MAE, MSE and ratio of the length of PI vs the range of the data.

	5th		50th		95th		Ratio of PI vs
	MAE	MSE	MAE	MSE	MAE	MSE	range of data
PQR	2.684	7.705	0.556	0.477	2.700	7.767	0.343
WNW	2.912	9.233	0.710	0.882	3.069	10.51	0.381
SLL	3.326	11.673	**0.517**	0.463	3.020	9.689	0.404
LLQR	2.742	8.019	0.549	0.465	2.591	7.291	0.340
NPPQR	**2.568**	**7.300**	0.531	0.462	**2.294**	**6.061**	**0.310**
SPPQR	2.749	8.012	0.552	**0.461**	2.628	7.470	0.342

Note: The table reports the MAE and MSE of the 5th, 50th and 95th conditional percentile estimators calculated from the 40 out-of-sample predictions. The last column reports the average ratio of length of the PI versus the range of the data. Numbers in boldface denote the lowest value in each column.

upper bound. Figure 1 displays the true GDP growth rates for all time $t > n_1$, and the 5th, 50th and 95th conditional percentile estimators based on various techniques. (For clarity, we did not include the SPPQR estimator in the figure.) From the figure, we can see that (1) the 50th percentile predictors can trace the true GDP growth rates fairly well, and (2) the NPPQR estimator outperforms all other quantile estimators in that they tend to have shorter PIs than others.

Table 5 displays the MAE, MSE of the 5th, 50th and 95th conditional quantile predictors based on the 40 out-of-sample predictions. The last column of Table 5 reports the average ratio of the length of the 90% PI versus the range of the data. From the table, we see that in most cases, the nonparametric prewhitening quantile estimator helps reduce MAE, MSE, and the length of PI.

5. Proof of Theorems in Section 2

We use $\|\cdot\|$ to denote the Euclidean norm, C to signify a generic constant whose exact value may vary from case to case, and a' to denote the transpose of a. To save space, we frequently write the distribution or density functions $F_{Y|X}(\cdot|\cdot)$, $f_{Y|X}(\cdot|\cdot)$ as $F(\cdot|\cdot)$ and $f(\cdot|\cdot)$.

First we state a lemma that is used in the proof of our main theorem.

Lemma 5.1. *Let $V_n(\Delta)$ be a vector function that satisfies*
(i)$-\Delta' V_n(\lambda\Delta) \geq -\Delta' V_n(\Delta)$, $\lambda \geq 1$,
(ii)$\sup_{\|\Delta\|\leq M} \|V_n(\Delta) + f_{Y|X}(q_\tau(x)|x)D\Delta - A_n\| = o_p(1)$,
where $\|A_n\| = O_p(1)$, $0 < M < \infty$, $f_{Y|X}(q_\tau(x)|x) > 0$, *and D is a positive definite matrix. Suppose that Δ_n satisfies $\|V_n(\Delta_n)\| = o_p(1)$. Then, $\|\Delta_n\| = O_p(1)$ and*

$$\Delta_n = [f_{Y|X}(q_\tau(x)|x)]^{-1} D^{-1} A_n + o_p(1). \tag{5.1}$$

The above lemma is proved in Koenker and Zhao (1996, p.809). To apply it to Theorem 2.1, we need a stronger result. For this purpose, we strengthen Condition (ii) in Lemma 5.1 to

(ii*) $\sup_{x\in\mathcal{X}} \sup_{\|\Delta\|\leq M} \|V_n(\Delta) + f_{Y|X}(q_\tau(x)|x)D\Delta - A_n\| = o_p(1)$.
One can follow the proofs of Koenker and Zhao (1994, 1996) and prove the following lemma.

Lemma 5.2. *Let $V_n(\Delta)$ be a vector function that satisfies the conditions in Lemma* 5.1 *with Condition (ii) replaced by (ii*). Then, $\|\Delta_n\| = O_p(1)$ and*

$$\Delta_n = [f_{Y|X}(q_\tau(x)|x)]^{-1} D^{-1} A_n + o_p(1) \text{ uniformly in } x. \tag{5.2}$$

Proof of Theorem 2.1. The proof is analogous to the proof of Theorem 3.1 in Lu, Hui and Zhao (2001). The main difference is to verify Condition (ii*) in Lemma 5.2 instead of Condition (ii) in Lemma 5.1. This can be done by repeatedly using Bickel's (1975) standard chaining argument.

Corollary 5.3. *Suppose Assumptions* A1–A3 *hold. Then*

$$\max_{\{K_{hi}>0\}} \frac{q_\tau^+(X_i)q_\tau(x)}{q_\tau(X_i)q_\tau^+(x)} = 1 + o_P\Big(n^{-\frac{1}{2}}h^{-\frac{d}{2}}\Big), \tag{5.3}$$

where $K_{hi} = K(X_{hi})$, and $X_{hi} = (X_i - x)/h$.

Proof. Under Assumptions A1 and A3, Theorem 2.1 implies that, uniformly in $x \in \mathcal{X}$,

$$\frac{q_\tau^+(x)}{q_\tau(x)} = 1 + \Psi_n(x) + o_P\Big(n^{-\frac{1}{2}} h_0^{-\frac{d}{2}}\Big),$$
$$\frac{q_\tau(x)}{q_\tau^+(x)} = 1 - \Psi_n(x) + o_P\Big(n^{-\frac{1}{2}} h_0^{-\frac{d}{2}}\Big),$$

where $\Psi_n(x) = \phi_\tau(x) q_\tau(x)^{-1} (nh_0^d)^{-1} \sum_{t=1}^n \psi_\tau(Y_t^*(x,\tau)) K_{h_0}(X_t - x)$. Let $\overline{\Psi}_n(x) = E(\Psi_n(x))$. Following the standard argument of Masry (1996b), and using Assumptions A1$-$A3, one can show

$$\sup_{x\in\mathcal{X}} |\Psi_n(x) - \overline{\Psi}_n(x)| = O_P\Big(n^{-\frac{1}{2}} h_0^{-\frac{d}{2}} \sqrt{\log n}\Big) = o_P\Big(n^{-\frac{1}{2}} h^{-\frac{d}{2}}\Big),$$

and $\sup_{x\in\mathcal{X}} |\overline{\Psi}_n(x)| = O(h_0^2)$. Consequently, $q_\tau^+(x)/q_\tau(x) = 1 + \overline{\Psi}_n(x) + o_P(n^{-1/2} h^{-d/2})$, and $q_\tau(x)/q_\tau^+(x) = 1 - \overline{\Psi}_n(x) + o_P(n^{-1/2}h^{-d/2})$, uniformly in x. Then

$$\frac{q_\tau^+(X_i) q_\tau(x)}{q_\tau(X_i) q_\tau^+(x)} = 1 + \overline{\Psi}_n(X_i) - \overline{\Psi}_n(x) + o_P\Big(n^{-\frac{1}{2}} h^{-\frac{d}{2}}).$$

Since $K(\cdot)$ is a bounded density function with compact support, $\|X_{hi}\| \le C$ for some C when $K_{hi} > 0$. Then

$$\begin{aligned}
&|\overline{\Psi}_n(\widetilde{x}) - \overline{\Psi}_n(x)| \\
&= (nh_0^d)^{-1} \Bigg| \sum_{j=1}^n \Big\{ E\Big[\psi_\tau(Y_j^*(\widetilde{x},\tau))\alpha(\widetilde{x}) K_{h_0}(X_j - \widetilde{x})\Big] \\
&\quad - E\Big[\psi_\tau(Y_j^*(x,\tau))\alpha(x) K_{h_0}(X_j - x)\Big]\Big\}\Bigg| \\
&= \frac{1}{2} h_0^{-d} \Big| E\Big[f(q_\tau(X_1)|X_1)(\widetilde{x}-X_1)' \ddot{q}_\tau(X_1 + c_1(\widetilde{x}-X_1))(\widetilde{x}-X_1)\alpha(\widetilde{x}) K_{h_0}(X_1 - \widetilde{x})\Big] \\
&\quad - E\Big[f(q_\tau(X_1)|X_1)(x - X_1)' \ddot{q}_\tau(X_1 + c_2(x - X_1))(x - X_1)\alpha(x) K_{h_0}(X_1 - x)\Big]\Big| \\
&\quad + O(h_0^3) \\
&= \frac{h_0^2}{2} \Big| \int \Big[f(q_\tau(\widetilde{x} + h_0 u)|\widetilde{x} + h_0 u) u' \ddot{q}_\tau(\widetilde{x} + (1 - c_1) h_0 u) u \alpha(\widetilde{x}) K(u) du\Big] \\
&\quad - \int \Big[f(q_\tau(x + h_0 u)|x + h_0 u) u' \ddot{q}_\tau(x + (1 - c_2) h_0 u) u \alpha(x) K(u) du\Big]\Big| + O(h_0^3) \\
&= \frac{h_0^2}{2} \Big| \mathrm{tr}\Big\{ \int \Big[\alpha(\widetilde{x}) f(q_\tau(\widetilde{x}) \Big| \widetilde{x} + h_0 u) \ddot{q}_\tau(\widetilde{x} + (1 - c_1) h_0 u)
\end{aligned}$$

$$-\alpha(x)f(q_\tau(x)\Big|x+h_0u)\ddot{q}_\tau(x+(1-c_2)h_0u)]uu'K(u)du\Big\}\Bigg| + O(h_0^3),$$

where $\alpha(x)=\phi_\tau(x)q_\tau(x)^{-1}$, and c_1 and c_2 lie between 0 and 1. It follows from the Lipschitz continuity of $f_{Y|X}(\cdot|\cdot)$, $f_X(\cdot)$, $q_\tau(\cdot)$ and $\ddot{q}_\tau(\cdot)$, that $\sup_{\|\widetilde{x}-x\|\le Ch}|\overline{\Psi}_n(\widetilde{x})-\overline{\Psi}_n(x)| = O(h_0^2h+h_0^3)$, and by Assumptions A2−A3, $\max_{\{K_{hi}>0\}} q_\tau^+(X_i)q_\tau(x)/[q_\tau(X_i)q_\tau^+(x)] = 1+o_P(n^{-1/2}h^{-d/2})$.

Let $H_n=\sqrt{nh^d}$, $\overline{\theta}_n = H_n(\widehat{\beta}_0-q_\tau(x),h(\widehat{\beta}_1-0)')'$, and $\theta=H_n(\beta_0-q_\tau(x),h(\beta_1-0)')'$, where $(\beta_0,\beta_1')'$, $(\widehat{\beta}_0,\widehat{\beta}_1')'\in\mathbb{R}\times\mathbb{R}^d$. Define $Y_i^{**}=Y_iq_\tau^+(x)/q_\tau^+(X_i)-q_\tau(x)$, $Y_{ni}^*(\theta)=Y_i^{**}-U_{ni}(\theta)$, where $U_{ni}(\theta)=\theta'\mathcal{X}_{hi}/H_n$, $\mathcal{X}_{hi}=(1,X_{hi}')'$. Then

$$Y_{ni}^*(\theta)=\frac{Y_iq_\tau^+(x)}{q_\tau^+(X_i)}-q_\tau(x)-U_{ni}(\theta)=\frac{Y_iq_\tau^+(x)}{q_\tau^+(X_i)}-\beta_0-\beta_1'(X_i-x). \qquad (5.4)$$

Also, when $\|\theta\|\le M$ and $K_{hi}>0$, $|U_{ni}(\theta)|\le CH_n^{-1}\to 0$ as $n\to\infty$. It follows from (2.6) that

$$\overline{\theta}_n=\arg\min_{\theta\in\mathbb{R}^{1+d}}\sum_{i=1}^n\rho_\tau(Y_{ni}^*(\theta))K_{hi}. \qquad (5.5)$$

Set

$$V_n(\theta)=H_n^{-1}\sum_{i=1}^n\psi_\tau(Y_{ni}^*(\theta))\mathcal{X}_{hi}K_{hi}. \qquad (5.6)$$

The proof of Theorem 2.2 is based on the following lemma.

Lemma 5.4. *Suppose Assumptions* A1−A3 *hold. Then*
(i) $\sup_{\|\theta\|\le M}\|V_n(\theta)-V_n(0)-E[V_n(\theta)-V_n(0)]\|=o_P(1)$;
(ii) $\sup_{\|\theta\|\le M}\|E[V_n(\theta)-V_n(0)]+f_{Y|X}(q_\tau(x)|x)D\theta\|=o(1)$;
(iii) $\|V_n(\overline{\theta}_n)\|\le(d+1)H_n^{-1}\max_{i\le n}\|\mathcal{X}_{hi}'K_{hi}\|$;
(iv) $E[c'(V_n(0)-EV_n(0))]^2=\tau(1-\tau)f_X(x)\int(c_0+c_1'u)^2K^2(u)du+o(1)$,
where $D=f_X(x)diag(1,\int uu'K(u)du)$, *and* $c=(c_0,c_1')'\in\mathbb{R}\times\mathbb{R}^d$.

Proof. The proof of the lemma is similar to that of Lemmas B2−B5 in Lu, Hui and Zhao (2001), and is thus omitted. The main difference is that we need to use (5.3) and standard dominance convergence arguments repeatedly.

Proof of Theorem 2.2. To apply Lemma 5.1, take $A_n=V_n(0)$, $\Delta=\theta$ and $\Delta_n=\overline{\theta}_n$. Then Lemma 5.4(iv) implies that $A_n=O_p(1)$. By Lemmas 5.4(i)−(ii) $\sup_{\|\theta\|\le M}\|V_n(\theta)+f_{Y|X}(q_\tau(x)|x)D\theta-V_n(0)\|=o_P(1)$, whereas by Lemma 5.4(iii),

$\|V_n(\overline{\theta}_n)\| = o_P(1)$. So it suffices to verify that Condition (i) of Lemma 5.1 holds. This is true since ψ_τ increasing implies that

$$-\theta' V_n(\lambda\theta) = H_n^{-1} \sum_{i=1}^{n} \psi_\tau\Big(Y_i^{**} - \frac{\lambda\theta'\mathcal{X}_{hi}}{H_n}\Big)(-\theta'\mathcal{X}_{hi})K_{hi}$$

is increasing in λ. Consequently, $\overline{\theta}_n = [f_{Y|X}(q_\tau(x)|x)]^{-1}D^{-1}V_n(0) + o_P(1)$, implying that $\sqrt{nh^d}(\widehat{q}_\tau(x) - q_\tau(x)) = \varphi_\tau(x)H_n^{-1}\sum_{i=1}^n \psi_\tau(Y_i^{**})K_{hi} + o_P(1)$.

Let $\widetilde{Y}_i^{**} = Y_i - q_\tau(X_i)$. Then

$$\begin{aligned}
&\sqrt{nh^d}(\widehat{q}_\tau(x) - q_\tau(x)) \\
&= \varphi_\tau(x)H_n^{-1}\sum_{i=1}^{n}\psi_\tau(\widetilde{Y}_i^{**})K_{hi} + \varphi_\tau(x)H_n^{-1}\sum_{i=1}^{n}\Big(\psi_\tau(Y_i^{**}) - \psi_\tau(\widetilde{Y}_i^{**})\Big)K_{hi} + o_P(1) \\
&\equiv A_{n1} + A_{n2} + o_P(1). \qquad (5.7)
\end{aligned}$$

Let $v_i = h^{-d/2}\varphi_\tau(x)\psi_\tau(\widetilde{Y}_i^{**})K_{hi}$. Then

$$E(A_{n1} - EA_{n1})^2 = \mathrm{Var}(v_1) + 2\sum_{j=2}^{n}\Big(1 - \frac{j-1}{n}\Big)\mathrm{Cov}(v_1, v_j). \qquad (5.8)$$

Since $E(v_1) = h^{-d/2}\varphi_\tau(x)E[\{\tau - 1[Y_1 \le q_\tau(X_1)]\}K_{h1}] = 0$,

$$\begin{aligned}
\mathrm{Var}(v_1) = E(v_1^2) &= h^{-d}\varphi_\tau^2(x)E\Big\{\Big[\tau^2 + (1-2\tau)1(Y_i \le q_\tau(X_i))\Big]K_{hi}^2\Big\} \\
&= \tau(1-\tau)\varphi_\tau^2(x)f_X(x)\int K^2(u)du + o(1). \qquad (5.9)
\end{aligned}$$

To bound the second term on the right hand side of (5.8), we split it into two terms as follows:

$$\sum_{j=2}^{n}|\mathrm{Cov}(v_1, v_j)| = \sum_{j=2}^{d_n}|\mathrm{Cov}(v_1, v_j)| + \sum_{j=d_n+1}^{n}|\mathrm{Cov}(v_1, v_j)| \equiv J_1 + J_2, \qquad (5.10)$$

where d_n is a sequence of positive integers such that $d_nh^d \to 0$ as $n \to \infty$. Since for any $j > 1$, $E(v_1v_j) = O(h^d)$,

$$J_1 = O(d_nh^d) = o(1). \qquad (5.11)$$

By the Davydov inequality (e.g., Bosq (1998, p.19)) and Assumption A1, we have

$$J_2 \le \sum_{j=d_n+1}^{n} C[\alpha(j-1)]^{\frac{\delta}{2+\delta}}\Big\{E\Big[|v_1|^{2+\delta}\Big]\Big\}^{\frac{2}{2+\delta}}$$

$$\le Ch^{-\frac{\delta d}{2+\delta}} \sum_{j=d_n+1}^{n} [\alpha(j-1)]^{\frac{\delta}{2+\delta}}$$
$$\le Ch^{-\frac{\delta d}{2+\delta}} d_n^{-a} \sum_{j=d_n}^{\infty} j^a [\alpha(j)]^{\frac{\delta}{2+\delta}} = o(1), \tag{5.12}$$

by choosing d_n such that $d_n^a h^{\delta d/(2+\delta)} \to \infty$. The last condition can be simultaneously met with $d_n h^d \to 0$ for a well-chosen sequence $\{d_n\}$ because $a > \delta/(2+\delta)$ by Assumptions A.1 and A.3. Consequently, (5.8)−(5.12)imply that

$$E(A_{n1} - EA_{n1})^2 = \tau(1-\tau)\varphi_\tau^2(x) f_X(x) \int K^2(u)du + o(1).$$

Using the standard Doob small-block and large-block technique, (e.g., Cai, Fan and Yao (2000), Cai and Ould-Saïd (2003) and Masry (1996a)), we can show that

$$A_{n1} \xrightarrow{d} N\Big(0, \tau(1-\tau)\varphi_\tau^2(x) f_X(x) \int K^2(u)du\Big). \tag{5.13}$$

Noting that $(q_\tau^+(X_i) q_\tau(x)/q_\tau^+(x) - q_\tau(X_i))K_{hi} = o_P(n^{-1/2}h^{-d/2})$ by Corollary 5.3, we have

$$\begin{aligned}
&E\Big[|\psi_\tau(Y_i^{**} - \psi_\tau(\widetilde{Y}_i^{**})|K_{hi}\Big] \\
&= E\bigg[\bigg|1(Y_i \le q_\tau(X_i)) - 1\Big(Y_i \le \frac{q_\tau^+(X_i) q_\tau(x)}{q_\tau^+(x)}\Big)\bigg|K_{hi}\bigg] \\
&\le E\bigg[1\{|Y_i - q_\tau(X_i)| \le \bigg|\frac{q_\tau^+(X_i) q_\tau(x)}{q_\tau^+(x)} - q_\tau(X_i)\bigg|\bigg\}K_{hi}\bigg] \\
&= o\Big(n^{-\frac{1}{2}} h^{\frac{d}{2}}\Big).
\end{aligned}$$

Consequently, $E|A_{n2}| \le \varphi_\tau(x) H_n^{-1} \sum_{i=1}^n E[|\psi_\tau(Y_i^{**}) - \psi_\tau(\widetilde{Y}_i^{**})|K_{hi}] = o(1)$. It follows from the Chebyshev's inequality that

$$A_{n2} = o_P(1). \tag{5.14}$$

The conclusion of Theorem (2.2) follows from (5.7), (5.13) and (5.14).

Acknowledgement

The authors thank the editors and two anonymous referees for their truly helpful suggestions that led to a much improved presentation of the paper. L. Su gratefully acknowledges the financial support from the NSF of China under the grant numbers 70501001 and 70601001. A. Ullah gratefully acknowledges the financial support from the Academic Senate, UCR.

References

Andrews, D. W. K. and Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60**, 953-966.

Bickel, P. J. (1975). One-step Huber estimates in linear models. *J. Amer. Statist. Assoc.* **70**, 428-433.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Springer, New York.

Cai, Z. (2002). Regression quantiles for time series. *Econom. Theory* **18**, 169-192.

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.* **95**, 941–956.

Cai, Z. and Ould-Saïd, E. (2003). Local M-estimator for nonparametric time series. *Statist. Probab. Lett.* **65**, 433-449.

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

Engle, R. F. and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *J. Bus. Econom. Statist.* **22**, 367-381.

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

Glad, I. K. (1998). Parametrically guided non-parametric regression. *Scand. J. Statist.* **25**, 649-668.

Hall, P., Wolf, R. C. L. and Yao, Q. (1999). Methods of estimating a conditional distribution function. *J. Amer. Statist. Assoc.* **94**, 154-163.

Hansen, B. E. (2004a). Nonparametric estimation of smooth conditional distributions. *Working paper*, Dept. of Economics, University of Wisconsin, Madison.

Hansen, B. E. (2004b). Bandwidth selection for nonparametric distribution estimation. *Working paper*, Dept. of Economics, University of Wisconsin, Madison.

He, X. and Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statist. Sinica* **10**, 129-140.

He, X. and Shao, Q-M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic design. *Ann. Statist.* **24**, 2608-2630.

Hjort, N. L. and Glad, I. K. (1995). Nonparametric density estimation with a parametric start. *Ann. Statist.* **23**, 882-904.

Hjort, N. L. and Jones, M. C. (1996). Locally parametric nonparametric density estimation. *Ann. Statist.* **24**, 1619-1647.

Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Ann. Inst. Statist. Math.* **52**, 459-470.

Jones, M. C., Linton, O. and Nielsen, J. P. (1995). A simple bias reduction method for density estimation. *Biometrika* **82**, 327-338.

Kauermann, G., Müller, M. and Carroll, R. J. (1998). The efficiency of bias-corrected estimators for nonparametric kernel estimation based on local estimating equations. *Statist. Probab. Lett.* **37**, 41-47.

Kim, T-H. and White, H. (2003). Estimation, inference, and specification testing for possibly misspecified quantile regression. *Adv. Econometrics* **17**, 107-132.

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

Koenker, R. and Zhao, Q. (1994). L-estimation for linear heteroscedastic models. *J. Nonparametr. Stat.* **3**, 223-235.

Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econom. Theory* **12**, 793-813.

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *J. Econometrics* **128**, 137-164.

Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression models. *Econom. Theory* **19**, 1-31.

Lu, Z., Hui, Y. V. and Zhao, Q. (2001). Local linear quantile regression under dependence: Bahadur representation and applications. Mimeo, City University of Hong Kong.

Masry, E. (1996a). Multivariate regression estimation: local polynomial fitting for time series. *Stochastic Process. Appl.* **65**, 81-101.

Masry, E. (1996b). Multivariate local polynomial regression for time series: uniform strong consistency rates. *J. Time Ser. Anal.*, **17**, 571-599.

Min, I. and Kim, I. (2004). A Monte Carlo comparison of parametric and nonparametric quantile regressions. *Applied Economics Letters* **11**, 71-74.

Mishra, S., Su, L. and Ullah, A. (2006). Combined estimator of time series conditional heteroskedasticity. *Working paper*, Dept. of Economics, University of California, Riverside.

Portnoy, S. (1997). Local asymptotics for quantile smoothing splines. *Ann. Statist.* **25**, 414-434.

Press, H. and Tukey, J. W. (1956). Power spectral methods of analysis and their application to problems in airline dynamics. *Flight Test Manual*, NATO, Advisory Group for Aeronautical Research and Development, Vol., IV-C, 1-41.

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

Guanghua School of Management, Peking University, Beijing 100871, China.

E-mail: lsu@gsm.pku.edu.cn

School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903.

E-mail: ljsu@smu.edu.sg

Department of Economics, University of California, Riverside, CA 92521-0427, U.S.A.

E-mail: aman.ullah@ucr.edu

(Received August 2006; accepted January 2007)