Bootstrap Partial Linear Quantile Regression and Confidence Bands*

Wolfgang K. Härdle [†], Ya'acov Ritov [‡], Song Song [§]

June 7, 2009

Abstract

In this paper uniform confidence bands are constructed for nonparametric quantile estimates of regression functions. The method is based on the bootstrap, where resampling is done from a suitably estimated empirical density function (edf) for residuals. It is known that the approximation error for the uniform confidence band by the asymptotic Gumbel distribution is logarithmically slow. It is proved that the bootstrap approximation provides a substantial improvement. The case of multidimensional and discrete regressor variables is dealt with using a partial linear model. Comparison to the classic asymptotic uniform bands is presented through a simulation study. An economic application considers the labour market discrimination effect w.r.t. different education levels.

Keywords: Bootstrap, Quantile Regression, Confidence Bands, Nonparametric Fitting, Kernel Smoothing, Partial Linear Model

JEL classification: C14; C21; C31; J01; J31; J71

---

*The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin is gratefully acknowledged. Ya'acov Ritov's research is supported by an ISF grant.

[†] Professor at Humboldt-Universität zu Berlin and Director of CASE - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Berlin, Germany.

[‡] Professor at The Hebrew University of Jerusalem, Department of Statistics, Jerusalem, Israel.

[§] Corresponding author. Research associate at the Institute for Statistics and Econometrics of Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany. Email: songsong@cms.hu-berlin.de.

1 Introduction

Quantile regression, as first introduced by Koenker and Bassett (1978), is "gradually developing into a comprehensive strategy for completing the regression prediction" as claimed by Koenker and Hallock (2001). Quantile smoothing is an effective method for estimation of quantile curves in a flexible nonparametric way. Since this technique makes no structural assumptions on the underlying curve, it is very important to have a device for understanding when observed features are significant, e.g. a question often asked in this context is whether or not an observed peak or valley is actually a feature of the underlying regression function or is only an artifact of the observational noise. For such issues, confidence intervals should be used that are simultaneous (i.e., uniform over location) in nature.

To construct a uniform confidence band, Härdle and Song (2009) used strong approximations of the empirical process and extreme value theory. However, the very poor convergence rate of extremes of a sequence of n independent normal random variables is well documented. It was first noticed and investigated by Fisher and Tippett (1928), and was discussed in greater detail by Hall (1991). In the latter paper it was shown that the rate of the convergence to its limit (the suprema of a stationary Gaussian process) can be no faster than $(\log n)^{-1}$. For example, the supremum of a nonparametric quantile estimate can converge to its limit no faster than $(\log n)^{-1}$. These results may make extreme value approximation of the distributions of suprema somewhat doubtful, for example in the context of the uniform confidence band construction for a nonparametric quantile estimate.

This paper proposes and analyzes a method of obtaining any number of uniform confidence bands for quantile estimates. The method is simple to implement, does not rely on the evaluation of quantities which appear in asymptotic distributions and also takes the bias properly into account (at least asymptotically). More importantly, we show that the bootstrap approximation to the distribution of the supremum of a quantile estimate is accurate to within $n^{-2/5}$ which represents a significant improvement relative to $(\log n)^{-1}$. Previous research by Hahn (1995) showed consistency of a bootstrap approximation to the cumulative density function (cdf) without assuming independence of the error and regressor terms. Horowitz (1998) showed bootstrap methods for median regression models based on a smoothed least-absolute-deviations (SLAD) estimate.

Let (X_1, Y_1), (X_2, Y_2), …, (X_n, Y_n) be a sequence of independent identically distributed bivariate random variables with joint pdf $f(x, y)$, joint cdf $F(x, y)$, conditional pdf $f(y|x)$, $f(x|y)$, conditional cdf $F(y|x)$, $F(x|y)$ for Y given X and X given Y respectively, and marginal pdf $f_X(x)$ for X, $f_Y(y)$

for Y. At the first stage we assume that $x \in J^*$, and $J^* = (a, b)$ for some $0 < a < b < 1$. Let $l(x)$ denote the p-quantile curve, i.e. $l(x) = F^{-1}_{Y|x}(p)$.

In economics, discrete or categorial regressors are very common, e.g. one of the most prominent tasks in labour market analysis is to find out how revenues depend on the age of the employee for different education levels, sex and nationality etc, i.e. whether discrimination effects exist. This motivates the extension to multivariate covariables by partial linear modelling (PLM). This is convenient especially when we have categorial elements of the X vector. Partial linear models, which were first considered by Green and Yandell (1985), Denby (1986), Speckman (1988) and Robinson (1988), are gradually developing into a class of commonly used and studied semiparametric regression models, which can retain the flexibility of nonparametric models and ease of interpretation of linear regression models while avoiding the "curse of dimensionality". Recently Liang and Li (2009) used penalized quantile regression for variable selection of partially linear models with measurement errors.

In this paper, we propose an extension of the quantile regression model to $x = (u, v)^\top \in \mathbb{R}^d$ with $u \in \mathbb{R}^{d-1}$ and $v \in J^* \subset \mathbb{R}$. The quantile regression curve we consider is: $\tilde{l}(x) = F^{-1}_{Y|x}(p) = u^\top \beta + l(v)$. Now the multivariate confidence band can be constructed based on the univariate uniform confidence band plus the estimated linear part which we will prove is more accurately ($\sqrt{n}$ consistency) estimated. This makes various tasks in economics, e.g. labour market discrimination effect investigation, multivariate model specification test, and investigation to the distribution of income and wealth across regions, countries or to distribution across households possible. Additionally, since the natural link between quantile and expectile regression, as developed by Newey and Powell (1987), we could furthur extend our result into expectile regression for various tasks, e.g. demography risk research or expectile-based Value at Risk (EVAR) as in Kuan et al. (2009).

The rest of this article is organized as follows. In Section 2, the bootstrap approximation rate for the uniform confidence band in quantile regression is presented through a coupling argument. An extension to multivariate covariance X with partial linear modelling is shown in Section 3. In Section 4, in the Monte Carlo study we compare the bootstrap uniform confidence band with the one based on the asymptotic theory and investigate the behaviour of partial linear estimates with the corresponding confidence band. In Section 5, an application considers the labour market discrimination effect. All proofs are sketched in Section 6.

2 Bootstrap confidence bands

Suppose $Y_i = l(X_i) + \varepsilon_i$, $i = 1, \ldots, n$, where ε_i has distribution function $F(\cdot|X_i)$ with $F(0|X_i) = p$. $F(\xi|x)$ is smooth as a function of x and ξ for any x, and for any ξ in the neighborhood of 0. We assume:

(A1). $X_1, \ldots, X_n$ are an i.i.d. sample, and $f_X(x) \geq \lambda_0$. The quantile function satisfies: $|l'(\cdot)| \leq \lambda_1$, $|l''(\cdot)| \leq \lambda_2$.

(A2). $F(t|x)$ have a density $f(t|x) \geq \lambda_3 > 0$, continuous in x, and in t in a neighborhood of 0. More exactly, we have the following Taylor expansion, for some $A(\cdot)$ and $f_0(\cdot)$, and for every (x', t):

$$F(t|x') = p + f_0(x)t + A(x)(x' - x) + R(t, x'; x), \tag{1}$$

where

$$\sup_{t,x,x'} \frac{|R(t, x'; x)|}{t^2 + |x' - x|^2} < \infty.$$

Let K be a symmetric density function with compact support and $d_K = \int u^2 K(u)du < \infty$. Let $l_h(\cdot) = l_{n,h}(\cdot)$ be the nonparametric p-quantile estimate of $Y_1, \ldots, Y_n$ with weight function $K\{(X_i - \cdot)/h\}$ for some global bandwidth $h = h_n$ ($K_h(u) = h^{-1}K(u/h)$), that is, a solution of:

$$\frac{\sum_{i=1}^n K_h(x - X_i)\mathbf{1}\{Y_i < l_h(x)\}}{\sum_{i=1}^n K_h(x - X_i)} < q \leq \frac{\sum_{i=1}^n K_h(x - X_i)\mathbf{1}\{Y_i \leq l_h(x)\}}{\sum_{i=1}^n K_h(x - X_i)}. \tag{2}$$

Generally, the bandwidth may also depend on x. The local adaptive bandwidth selection issue deserves further research.

Note that by assumption ((A1)), $l_h(x)$ is the quantile of a discrete distribution, which is equivalent to a sample of size $\mathcal{O}_p(nh)$ from a distribution with p-quantile whose bias is $\mathcal{O}_p(h^2)$ relative to the true value. Let δ_n be the local rate of convergence of the function l_h, essentially $\delta_n = h^2 + (nh)^{-1/2} = \mathcal{O}(n^{-2/5})$ with optimal bandwidth choice $h = h_n = \mathcal{O}(n^{-1/5})$. We employ also an auxiliary estimate $l_g = l_{n,g}$, essentially one similar to $l_{n,h}$ but with a slightly larger bandwidth $g = g_n = h_n n^\zeta$ (a heuristic explanation of why it is essential to oversmooth g is given later), where ζ is some small number. The asymptotically optimal choice of ζ as shown later is $4/45$.

(A3). The estimate l_g satisfies:

$$\begin{aligned} \sup_{x \in J^*} |l_g''(x) - l''(x)| &= \scriptstyle\mathcal{O}_p(1), \\ \sup_{x \in J^*} |l_g'(x) - l'(x)| &= \scriptstyle\mathcal{O}_p(\delta_n/h). \end{aligned} \tag{3}$$

Assumption (A3) is only stated to overwrite the issue here. It actually follows from the assumptions on (g,h). We will use S_n to denote any slowly varying function, e.g. $S_n^2 = S_n$ is a valid expression, etc. λ_i and C_i are generic constants throughout this paper and the subscripts have no specific meaning. Note that there is no S_n term in (3) exactly because the bandwidth g_n used to calculate l_g is slightly larger than that used for l_h. As a result l_g, as an estimate of the quantile function, has a slightly worse rate of convergence, but its derivatives converge faster.

We also consider a family of estimates $\hat{F}_i$, $i = 1, \dots, n$, estimating respectively $F(\cdot|X_i)$ and satisfying $\hat{F}_i(0) = p$, where $\hat{F}_i(t) = \hat{F}(t|X_i)$. For example we can take the distribution with a point mass $c^{-1}K\{\alpha_n(X_j - X_i)\}$ on $Y_j - l_h(X_i)$, $j = 1, \dots, n$, where $c = \sum_{j=1}^n K\{\alpha_n(X_j - X_i)\}$, where $\alpha_n \approx h^{-1}$. We assume:

(A4). $f_X(x)$ is twice continuously differentiable and $f(t|x)$ is uniformly bounded in x and t by, say, λ_4.

LEMMA 2.1 *[Franke and Mwita (2003), p14] If assumptions (A1, A2, A4) hold, then for any small enough $\varepsilon > 0$,*

$$\sup_{|t|<\varepsilon, i=1,\dots,n, X_i \in J^*} |\hat{F}_i(t) - F(t|X_i)| = \mathcal{O}_p\{S_n \delta_n \varepsilon^{1/2} + \varepsilon^2\}. \tag{4}$$

Note that the result in Lemma 2.1 is natural, since by definition, there is no error at $t = 0$, since $\hat{F}_i(0) = p$. For $t \in (0, \varepsilon)$, $\hat{F}_i(t)$, like l_h, is based on a sample of size $\mathcal{O}_p(nh)$. Hence, the random error is $\mathcal{O}_p\{(nh)^{-1/2}t^{1/2}\}$, while the bias is $\mathcal{O}_p(\varepsilon h^2) = \mathcal{o}_p(\delta_n)$. The S_n term takes care of the maximization.

Let $F^{-1}(\cdot|\cdot)$ and $\hat{F}_{\cdot}^{-1}(\cdot)$ be the inverse function of the conditional cdf and its estimate. We consider the following bootstrap procedure: Let $U_1, \dots, U_n$ be i.i.d. uniform $[0,1]$ variables. Let

$$Y_i^* = l_g(X_i) + \hat{F}_i^{-1}(U_i), \quad i = 1, \dots, n \tag{5}$$

be the bootstrap sample. We couple this sample to a sample from the true conditional distribution:

$$Y_i^{\#} = l(X_i) + F^{-1}(U_i|X_i), \quad i = 1, \dots, n. \tag{6}$$

Note that given $X_1, \dots, X_n$, $Y_1, \dots, Y_n$ and $Y_1^{\#}, \dots, Y_n^{\#}$ are equally distributed. We are interested in the values of $Y_i^{\#}$ and Y_i^* near the appropriate quantile, that is, only if $|U_i - p| < S_n\delta_n$. But then, by equation (1), Lemma 2.1 and the inverse function theorem, we have:

$$\begin{aligned} & \max_{i:|F^{-1}(U_i|X_i) - F^{-1}(p)| < S_n\delta_n} |F^{-1}(U_i|X_i) - \widehat{F}_i^{-1}(U_i)| \\ = & \max_{i:|Y_i^{\#} - l(X_i)| < S_n\delta_n} |Y_i^{\#} - l(X_i) - Y_i^* + l_g(X_i)| = \mathcal{O}_p\{S_n \delta_n^{-3/2}\}. \end{aligned} \tag{7}$$

Let now $q_{hi}(Y_1, \ldots, Y_n)$ be the solution of the local quantile as given by (2) at X_i, with bandwidth h. Note that by (3), if $|X_i - X_j| = \mathcal{O}(h)$, then

$$\max_{|X_i - X_j| < ch} |l_g(X_i) - l_g(X_j) - l(X_i) + l(X_j)| = \mathcal{O}_p(\delta_n) \tag{8}$$

Let l_h^* and $l_h^\#$ be the local bootstrap quantile and its coupled sample analogue. Then

$$\begin{aligned} l_h^*(X_i) - l_g(X_i) &= q_{hi}[\{Y_j^* - l_g(X_i)\}_{j=1}^n] \\ &= q_{hi}[\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\})_{j=1}^n], \end{aligned} \tag{9}$$

while

$$l_h^\#(X_i) - l(X_i) = q_{hi}[\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n]. \tag{10}$$

From (7) – (10) we conclude that

$$\max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) - l(X_i)| = \mathcal{O}_p(\delta_n). \tag{11}$$

Based on (11), we obtain the following theorem (the proof is given in the appendix):

THEOREM 2.1 *If assumptions (A1 - A3) and Lemma 2.1 hold, then*

$$\sup_{x \in J^*} |l_h^*(x) - l_g(x) - l_h^\#(x) - l(x)| = \mathcal{O}_p(\delta_n) = \mathcal{O}_p(n^{-2/5}).$$

A number of replications of $l_h^*(x)$ can be used as the basis for simultaneous error bars because the distribution of $l_h^\#(x) - l(x)$ is approximated by the distribution of $l_h^*(x) - l_g(x)$, as Theorem 2.1 shows.

While bootstrap methods are well-known tools for assessing variability, more care must be taken to properly account for the type of bias encountered in nonparametric curve estimation. The choice of the bandwidth is crucial here. The next issue is how to fine tune the choice of the pilot bandwidth g. While it is true that the bootstrap works with a rather crude choice of g, it is intuitively clear that specification of g will play a role here. Since the main role of the pilot bandwidth is to provide a correct adjustment for the bias, we use the goal of bias estimation as a criterion. In particular, similar to Härdle and Marron (1991), recall that the bias in the estimation of $l(x)$ by $l_h^\#(x)$ is given by

$$b_h(x) = \mathsf{E}\, l_h^\#(x) - l(x).$$

The bootstrap bias of the estimate constructed from the resampled data is

$$\hat{b}_{h,g}(x) = \mathsf{E}\, l_h^*(x) - l_g(x).$$

The following theorem gives an asymptotic representation of the mean squared error for the problem of estimating $b_h(x)$ by $\hat{b}_{h,g}(x)$. It is then straightforward to find g to minimize this representation. Such a choice of g will make the quantiles of the original and coupled bootstrap distributions close to each other. In addition to the technical assumptions before, we also need:

(A5). l and f are four times continuously differentiable.

(A6). K is twice continuously differentiable.

THEOREM 2.2 *Under assumptions (A1 - A6), for any $x \in J^*$*

$$\mathsf{E}\left[\left\{\hat{b}_{h,g}(x) - b_h(x)\right\}^2 | X_1, \dots, X_n\right] \sim h^4(C_1 g^4 + C_2 n^{-1} g^{-5}) \qquad (12)$$

in the sense that the ratio between the RHS and the LHS tends in probability to 1 *for some constants C_1, C_2.*

An immediate consequence of Theorem 2.2 is that the rate of convergence of g should be $n^{-1/9}$. This makes precise the previous intuition which indicated that g should slightly oversmooth. Under our assumptions, reasonable choices of h will be of the order $n^{-1/5}$ as in Yu and Jones (1998). Hence, (12) shows once again that g should tend to zero more slowly than h. Note that Theorem 2.2 is not stated uniformly over h. The reason is that we are only trying to give some indication of how the pilot bandwidth g should be selected.

3 Bootstrap confidence bands in PLMs

The case of multivariate regressors may be handled via a semiparametric specification of the quantile regression curve. More specifically we assume that with $x = (u, v)^\top \in \mathbb{R}^d$, $v \in \mathbb{R}$:

$$\tilde{l}(x) = u^\top \beta + l(v)$$

In this section we show how to proceed in this multivariate setting and how - based on Theorem 2.1 - a multivariate confidence band may be constructed. We first describe the numerical procedure for obtaining estimates of β and l, where l denotes - as in the earlier sections - the one-dimensional conditional quantile curve. We then move on to the theoretical properties. First note that the PLM quantile estimation problem can be seen as estimating (β, l) in

$$\begin{aligned} y &= u^\top \beta + l(v) + \varepsilon \\ &= \tilde{l}(x) + \varepsilon \end{aligned}$$

where the p-quantile of ε conditional on both u and v is 0.

In order to estimate β, let a_n denote an increasing sequence of positive integers and set $b_n = a_n^{-1}$. For each $n = 1, 2, \ldots$, partition the unit interval $[0,1]$ for v in a_n intervals $I_{ni}, i = 1, \ldots, a_n$, of equal length b_n and let m_{ni} denote the midpoint of I_{ni}. In each of these small intervals $I_{ni}, i = 1, \ldots, a_n$, $l(v)$ can be regarded as a constant, which is actually an ANOVA idea. Therefore, the two stage estimation procedure is as follows:

1) The linear quantile regression inside each partition is used to estimate $\hat{\beta}_i, i = 1, \ldots, a_n$. Their weighted mean yields $\hat{\beta}$. More exactly, consider the parametric quantile regression of y on $u, \mathbf{1}\big(v \in [0, b_n)\big), \mathbf{1}\big(v \in [b_n, 2b_n)\big), \ldots, \mathbf{1}\big(v \in [1 - b_n, 1]\big)$. That is, let

$$\psi(t) \stackrel{\text{def}}{=} (1-p)\mathbf{1}(t < 0) + p\mathbf{1}(t > 0).$$

Then let

$$\hat{\beta} = \arg\min_{\beta} \min_{l_1, \ldots, l_{a_n}} \sum_{i=1}^{n} \psi\{Y_i - \beta^\mathsf{T} U_i - \sum_{j=1}^{a_n} l_j \mathbf{1}\big(V_i \in I_{ni}\big)\}$$

2) Define the smooth quantile estimate $\hat{l}_h(v)$ from $(V_i, Y_i - U_i^\top \hat{\beta})_{i=1}^n$.

The following theorem states the asymptotic distribution of $\hat{\beta}$.

THEOREM 3.1 *There exist positive definite matrices D_n, C_n, such that*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N\{0, p(1-p) D_n^{-1} C_n D_n^{-1}\} \text{ as } n \to \infty.$$

Similar to Härdle and Song (2009), $l(v)$, $l_h(v)$ (quantile smoother based on $(v, y - u^\top \beta)$) and $\hat{l}_h(v)$ can be treated as a zero (w.r.t. θ, $\theta \in I$ where I is a possibly infinite, or possibly degenerate, interval in $\mathbb{R}$) of the functions

$$\widetilde{H}(\theta, v) \stackrel{\text{def}}{=} \int_{\mathbb{R}} f(v, \tilde{y}) \psi(\tilde{y} - \theta) d\tilde{y}, \tag{13}$$

$$\widetilde{H}_n(\theta, v) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^{n} K_h(v - V_i) \psi(\widetilde{Y}_i - \theta), \tag{14}$$

$$\widetilde{\widetilde{H}}_n(\theta, v) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^{n} K_h(v - V_i) \psi(\widetilde{\widetilde{Y}}_i - \theta), \tag{15}$$

where

$$\widetilde{Y}_i \stackrel{\text{def}}{=} Y_i - U_i^\top \beta$$
$$\widetilde{\widetilde{Y}}_i \stackrel{\text{def}}{=} Y_i - U_i^\top \hat{\beta} = Y_i - U_i^\top \beta + U_i^\top(\beta - \hat{\beta}) \stackrel{\text{def}}{=} \widetilde{Y}_i + Z_i.$$

From Theorem 3.1 we know that $\hat{\beta} - \beta = \mathcal{O}_p(1/\sqrt{n})$ and $||Z_i||_\infty = \mathcal{O}_p(1/\sqrt{n})$. Under the following assumption:

(A7). The conditional densities $f(\cdot|y),\ y \in \mathbb{R}$, are uniformly local Lipschitz continuous of order $\tilde{\alpha}$ (ulL-$\tilde{\alpha}$) on J, uniformly in $y \in \mathbb{R}$, with $0 < \tilde{\alpha} \leqslant 1$, and $(nh)/\log n \to \infty$.

For some constant C_3 not depending on n, Lemma 2.1 in Härdle and Song (2009) shows a.s. as $n \to \infty$:

$$\sup_{\theta \in I} \sup_{v \in J^*} |\widetilde{H}_n(\theta, v) - \widetilde{H}(\theta, v)| \quad \leq \quad C_3 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\}.$$

Observing that $\sqrt{h/\log n} = \mathcal{o}(1)$, we then have:

$$\begin{aligned}
\sup_{\theta \in I} \sup_{v \in J^*} |\widetilde{\widetilde{H}}_n(\theta, v) - \widetilde{H}(\theta, v)| \quad &\leq \quad \sup_{\theta \in I} \sup_{v \in J^*} |\widetilde{H}_n(\theta, v) - \widetilde{H}(\theta, v)| \\
&\quad + \underbrace{\sup_{\theta \in I} \sup_{v \in J^*} |\widetilde{H}_n(\theta, v) - \widetilde{\widetilde{H}}_n(\theta, v)|}_{\leq \mathcal{O}_p(1/\sqrt{n}) \sup_{v \in J} |n^{-1} \sum K_h|} \\
&\leq \quad C_4 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \qquad (16)
\end{aligned}$$

for a constant C_4 which can be different from C_3. To show the uniform consistency of the quantile smoother, we shall reduce the problem of strong convergence of $\hat{l}_h(v) - l(v)$, uniformly in v, to an application of the strong convergence of $\widetilde{H}_n(\theta, v)$ to $\widetilde{H}(\theta, v)$, uniformly in x and θ. For our result on $l_h(\cdot)$, we shall also require

(A8). $\inf_{v \in J^*} \left| \int \psi\{y - l(v) + \varepsilon\} dF(y|v) \right| \geqslant \tilde{q}|\varepsilon|, \quad$ for $|\varepsilon| \leqslant \delta_1$,

where δ_1 and $\tilde{q}$ are some positive constants, see also Härdle and Luckhaus (1984). This assumption is satisfied if there exists a constant $\tilde{q}$ such that $f\{l(v)|v\} > \tilde{q}/p$, $x \in J$. Härdle and Song (2009) showed the following:

LEMMA 3.1 *Under assumptions* (A7) *and* (A8)*, we have a.s. as* $n \to \infty$

$$\sup_{v \in J^*} |\hat{l}_h(v) - l(v)| \leq C_5 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{17}$$

with another constant C_5 *not depending on* n*. If additionally* $\tilde{\alpha} \geqslant \{\log(\sqrt{\log n}) - \log(\sqrt{nh})\}/\log h$, (17) *can be further simplified to:*

$$\sup_{v \in J^*} |\hat{l}_h(v) - l(v)| \leq C_5\{(nh/\log n)^{-1/2}\}.$$

Since the convergence rate for the parametric part is $\mathcal{O}_p(n^{-1/2})$ which is smaller than the bootstrap approximation error for the nonparametric part $\mathcal{O}_p(n^{-2/5})$ as shown in Theorem 2.1, we can estimate the parametric part more accurately. This makes the construction of uniform confidence bands for multivariate $x \in \mathbb{R}^d$ with a partial linear model possible which is stated in the following corollary.

COROLLARY 3.1 *Under the assumptions* (A1) - (A8)*, an approximate* $(1-\alpha) \times 100\%$ *confidence band over* $\mathbb{R}^{d-1} \times [0,1]$ *is*

$$u^\top \hat{\beta} + l_h(v) \quad \pm \quad \left[\hat{f}\{l(x)|x\}\sqrt{\hat{f}_X(x)}\right]^{-1} d_\alpha^*,$$

where d_α^* *is based on the bootstrap sample which we will specify later and* $\hat{f}\{l(x)|x\}$, $\hat{f}_X(x)$ *are consistent estimators of* $f\{l(x)|x\}$, $f_X(x)$ *with use of* $f(y|x) = f(x,y)/f_X(x)$.

4 A Monte Carlo study

This section is divided into two parts. First we concentrate on the univariate regressor variable x, summarize the basic steps for the bootstrap procedure together with settings in the specific example, check the validity and compare with the asymptotic uniform bands in Härdle and Song (2009). Second we incorporate the partial linear model to handle the multivariate case of $x \in \mathbb{R}^d$.

Below is the summary of the simulation procedure:

1) Simulate $(X_i, Y_i), i = 1, \ldots, n$ according to their joint pdf $f(x,y)$.

   In order to compare with earlier results in the literature, we choose the joint pdf of bivariate data $\{(X_i, Y_i)\}_{i=1}^n$, $n = 1000$ as:

$$f(x,y) = f_{y|x}(y - \sin x)\mathbf{1}(x \in [0,1]), \tag{18}$$

   where $f_{y|x}(x)$ is the pdf of $\mathrm{N}(0,x)$ with an increasing heteroscedastic structure. Thus the theoretical quantile is $l(x) = \sin(x) + \sqrt{x}\Phi^{-1}(p)$. Based on this normality property, all the assumptions can be seen to be satisfied.

2) Compute the local quantile smoother $l_h(x)$ of $Y_1, \ldots, Y_n$ with bandwidth h and obtain residuals $\hat{\varepsilon}_i = Y_i - l_h(X_i),\ i = 1, \ldots, n$.

   If we choose $p = 0.9$, then $\Phi^{-1}(p) = 1.2816$, $l(x) = \sin(x) + 1.2816\sqrt{x}$ and the bandwidth is $h = 0.05$.

3) Compute the conditional edf:

$$F_n(t|x) = \frac{\sum_{i=1}^n K_h(x - X_i)\mathbf{1}\{\hat{\varepsilon}_i \leqslant t\}}{\sum_{i=1}^n K_h(x - X_i)}$$

   with the quartic kernel

$$K(u) = \frac{15}{16}(1 - u^2)^2, \quad (|u| \leqslant 1).$$

4) For each $i = 1, \ldots, n$, generate random variables $\varepsilon^{\star}_{i,b} \sim F_{n|x}$, $b = 1, \ldots, B$ and construct the bootstrap sample $Y^{\star}_{i,b}, i = 1, \ldots, n, b = 1, \ldots, B$ as follows:

$$Y^{\star}_{i,b} = l_g(X_i) + \varepsilon^{\star}_{i,b},$$

   where $l_g(X_i)$ is defined as in (2) with $g = 0.2$.

5) For each bootstrap sample $\{(X_i, Y^{*}_{i,b})\}_{i=1}^n$, compute l^*_h and the random variable

$$d_b \stackrel{\text{def}}{=} \sup_{X \in J^*} \left[\hat{f}\{l(x)|x\}\sqrt{\hat{f}_X(x)}|l^*_h(X) - l_g(X)|\right]. \tag{19}$$

   where $\hat{f}\{l(x)|x\},\ \hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\},\ f_X(x)$ with use of $f(y|x) = f(x, y)/f_X(x)$.

6) Calculate the $(1 - \alpha)$ quantile d^*_α of $d_1, \ldots, d_B$.

7) Construct the bootstrap uniform confidence band centered around $l_h(x)$, i.e. $l_h(x) \pm \left[\hat{f}\{l(x)|x\}\sqrt{\hat{f}_X(x)}\right]^{-1} d^*_\alpha$.

In Figure 1 the theoretical 0.9 quantile curve, 0.9 quantile estimate with corresponding 95% uniform confidence band from the asymptotic theory and confidence band from the bootstrap are displayed. The real 0.9 quantile curve is marked as the black dotted line. We then compute the classic local quantile estimate $l_h(x)$ (cyan solid) with its corresponding 95% uniform confidence band (magenta dashed) based on asymptotic theory according to Härdle and Song (2009). The 95% confidence band from the bootstrap is displayed as

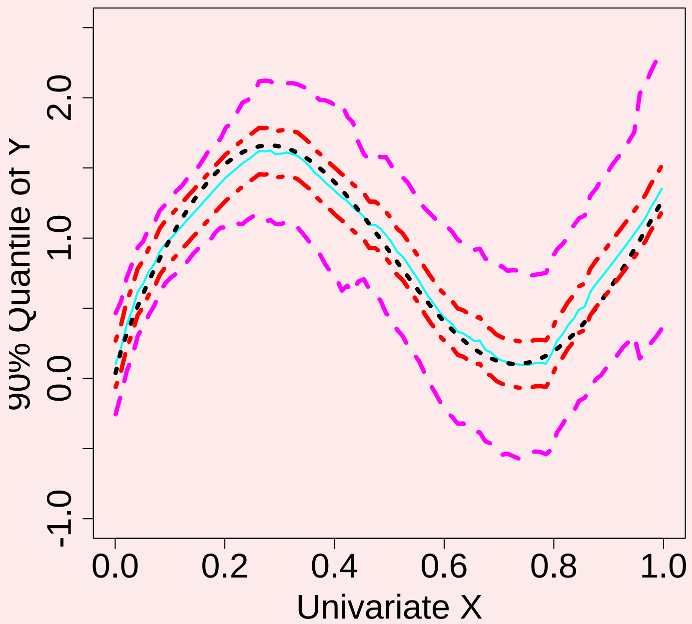

Figure 1: The real 0.9 quantile curve, 0.9 quantile estimate with corresponding 95% uniform confidence band from asymptotic theory and confidence band from bootstrapping.

red dashed-dot lines. At first sight, the quantile smoother together with corresponding bands captures the heteroscedastic structure, and the width of the bootstrap confidence band is smaller than the one based on asymptotic theory in Härdle and Song (2009).

We now extend x to the multivariate case and use a different quantile function to verify our method. Choose $x = (u, v)^\top \in \mathbb{R}^d$, $v \in \mathbb{R}$, and generate the data $\{(U_i, V_i, Y_i)\}_{i=1}^n, n = 1000$ with:

$$y = 2u + v^2 + \varepsilon - 1.2816, \tag{20}$$

where u and v are uniformly distributed random variables in $[0, 2]$ and $[0, 1]$ respectively. ε has a standard normal distribution. The theoretical 0.9-quantile curve is $\tilde{l}(x) = 2u + v^2$. Since the choice of a_n is uncertain here, we test different choices of a_n for different n by simulation. To this end, we modify the theoretical model as follows:

$$y = 2u + v^2 + \varepsilon - \Phi^{-1}(p)$$

such that the real β is always equal to 2 no matter if p is 0.01 or 0.99. The result is displayed in Figure 2 for $n = 1000$, $n = 8000$, $n = 261148$

(number of observations for the data set used in the following application part). Different lines correspond to different a_n, i.e. $n^{1/3}/8$, $n^{1/3}/4$, $n^{1/3}/2$, $n^{1/3}$, $n^{1/3}\cdot 2$, $n^{1/3}\cdot 4$ and $n^{1/3}\cdot 8$. At first, it seems that the choice of a_n doesn't matter too much. To further investigate this, we calculate the SSE $(\sum_1^{99}\{\hat{\beta}(i/100)-\beta\})$ where $\hat{\beta}(i/100)$ denotes the estimate corresponding to the $i/100$ quantile. Results are displayed in Table 1. Obviously a_n has much less effect than n on SSE. Considering computational cost, which increases with a_n, and estimation performance, empirically we suggest $a_n = n^{1/3}$.

a_n	$n = 1000$	$n = 8000$	$n = 261148$
$n^{1/3}/8$			$3.6 * 10^{-3}$
$n^{1/3}/4$	$5.4 * 10^{-1}$	$4.0 * 10^{-2}$	$3.3 * 10^{-3}$
$n^{1/3}/2$	$6.1 * 10^{-1}$	$3.5 * 10^{-2}$	$3.2 * 10^{-3}$
$n^{1/3}$	$6.2 * 10^{-1}$	$3.6 * 10^{-2}$	$3.1 * 10^{-3}$
$n^{1/3}\cdot 2$	$8.0 * 10^{-1}$	$3.9 * 10^{-2}$	$2.9 * 10^{-3}$
$n^{1/3}\cdot 4$	$4.9 * 10^{-1}$	$3.6 * 10^{-2}$	$2.8 * 10^{-3}$
$n^{1/3}\cdot 8$			$3.4 * 10^{-3}$

Table 1: SSE of $\hat{\beta}$ with respect to a_n for different numbers of observations.

Thus for the specific model (20), we have $a_n = 10$, $\hat{\beta} = 1.997, h = 0.2$ and $g = 0.7$. In Figure 3 the theoretical 0.9 quantile curve with respect to v, and the 0.9 quantile estimate with corresponding uniform confidence band are displayed. The real 0.9 quantile curve is marked as the black dotted line. We then compute the quantile smoother $l_h(x)$ (magenta solid). The 95% bootstrap uniform confidence band is displayed as red dashed lines and cover the true quantile curve quite well.

5 A labour market application

To study the labour market discrimination effect w.r.t. different education levels, we use data from the German national pension office (Deutsche Rentenversicherung Bund) for the following group: males aged $25-59$, born between 1939 and 1942 who began receiving a pension in 2004 or 2005, with at least 30 yearly observations. It has $n = 261148$ observations. The data for the entire German population is available since the German reunification for years 1991 to 2005. For the period from 1939 to 1990 data for the Federal states in former West Germany and those in former East Germany are available separately, which was combined later. We observe a complete earnings history, so this is a true panel, not a pseudo-panel. We have the following four education categories: -9 "no answer", 1 "low education", 2 "apprenticeship" and 3 "university" for variable u, while variable v is the age of the employee.

A log transformation to hourly real wages (unit: EUR, in year 2000 prices) is carried out first. In the data all ages ($25 \sim 59$) are reported as integers, they are categorised in one-year groups. We rescaled them into $[0, 1]$ by dividing 40, with corresponding bandwidth 0.059 for nonparametric quantile smoothers. This is equivalent to setting a bandwidth 2 for the original age data. This makes sense because to detect whether a discrimination effect for different education levels exists or not, we compare the corresponding uniform confidence bands, i.e. differences indicate that the discrimination effect may exist for different education levels in the German labour market for that specific labour group.

Following an application of the partial linear model in Section 3, Figure 4 displays the $\hat{\beta}$ with respect to different quantiles. At first sight, the $\hat{\beta}$ curve is quite surprising, since it is not, as in mean regression, a positive constant, but rather varies a lot, e.g. $\hat{\beta}(0.05) = 0.01059365$, $\hat{\beta}(0.50) = 3.418906 \cdot 10^{-5}$ and $\hat{\beta}(0.90) = -8.4 * 10^{-3}$. It seems that high end labour in the "university" group earns less than the high end ones in the "no answer" group. To judge whether these differences are significant, we use the uniform confidence band techniques discussed in Section 2 which are displayed in Figure 5 - 7 corresponding to 0.05, 0.50 and 0.90 quantiles respectively.

The quantile estimates for the "low education" group are marked as magenta solid lines, together with corresponding 95% uniform confidence bands from bootstrapping (red dashed lines). The uniform confidence bands for "apprenticeship", "university" and "no answer" groups are displayed as blue dotted, black dashed-dot and darkgreen dashed lines respectively. For 0.05-quantiles in Figure 5, "university", "low education" and "no answer" bands significantly differ from each other and become progressively lower, which indicates high education offered a safety line for low end labour. However, for 0.50-quantiles in Figure 6, the corresponding bands do not differ significantly resulting from the very small $\hat{\beta}(0.50) = 3.418906 \cdot 10^{-5}$, which indicates, for this specific group of people (born between 1939 and 1942) that on average education doesn't have a significant effect on income, while increasing age seems the main driving force. For 0.90-quantiles in Figure 7, "university", "low education" and "no answer" significantly differ from each other (there is some overlap between "university" and "low education" at the 5% significance level, but they differ at the 6% significance level) and increase, which reveals the fact that many high income people for this specific group of people do not have high education. However, this does not indicate causality.

Interesting subsequent problem will be the causality test that whether the following conclusions hold true: "high education is the causality effect that drives low end labour earn more than low education's" and "low education is not the causality effect that drives high end labour earn more than high education's". To this end, methods similar to Jeong et al. (2009) could be employed which deserve further research.

We need to notice that all the above conclusions are based on the specific group of people in the German labour market: born between 1939 and 1942. With the progress of the times, especially the flying technology level, more and more high income jobs require high educated people. We expect that if this method is applied, for example, to people born in 1980s, we may observe a quite different $\hat{\beta}$ curve as in Figure 4. Investigating time variation of the $\hat{\beta}$ curve deserves further research.

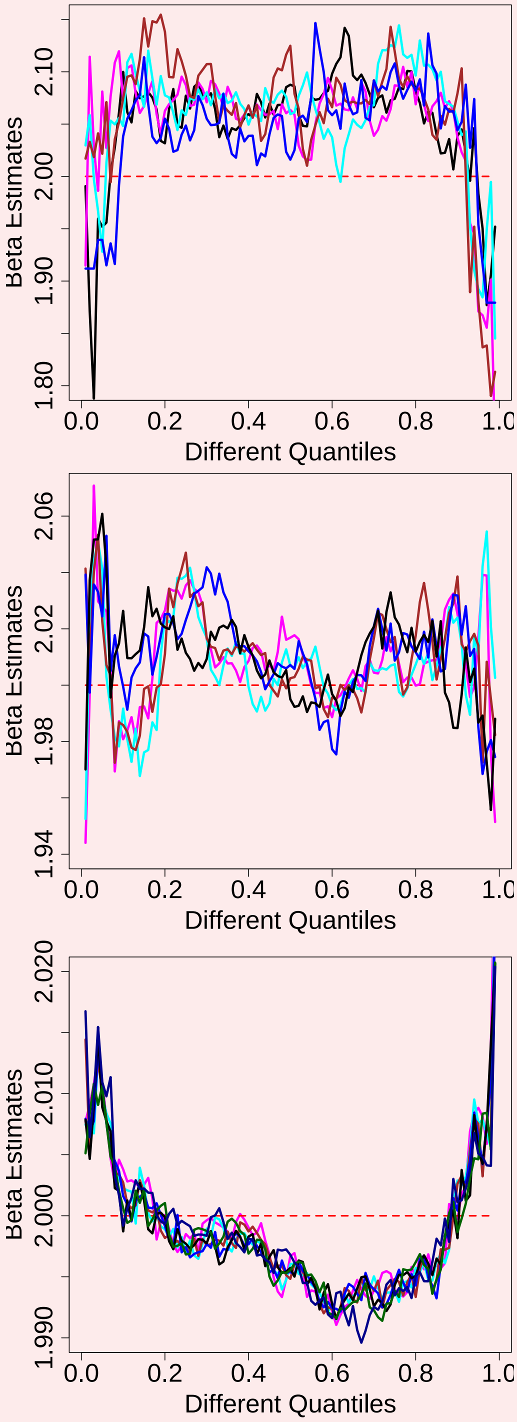

Figure 2: $\hat{\beta}$ with respect to different quantiles for different numbers of observations, i.e. $n = 1000$, $n = 8000$, $n = 261148$.

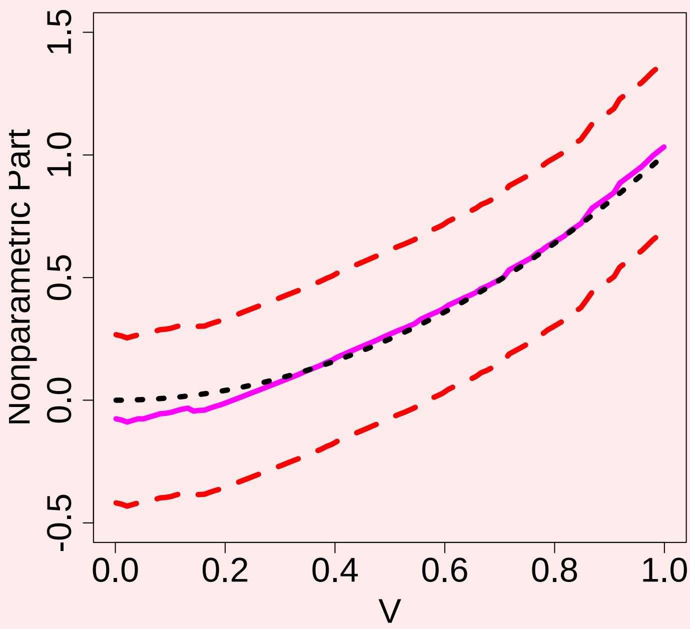

Figure 3: Nonparametric part smoothing, real 0.9 quantile curve with respect to v, 0.9 quantile smoother with corresponding 95% bootstrap uniform confidence band.

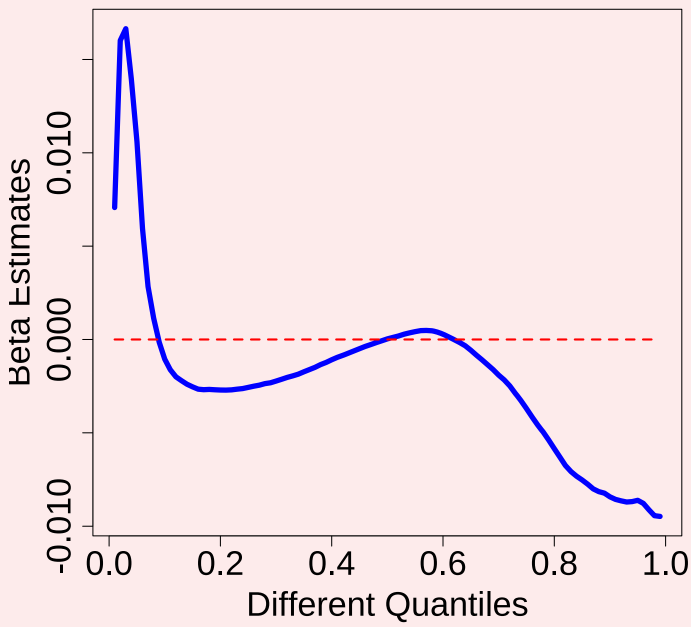

Figure 4: $\hat{\beta}$ corresponding to different quantiles.

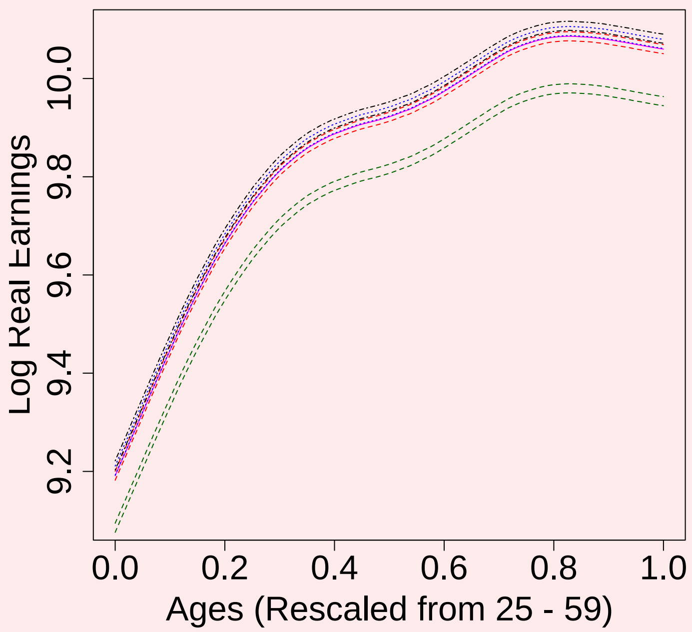

Figure 5: 95% uniform confidence bands for 0.05-quantile smoothers with 4 different education levels

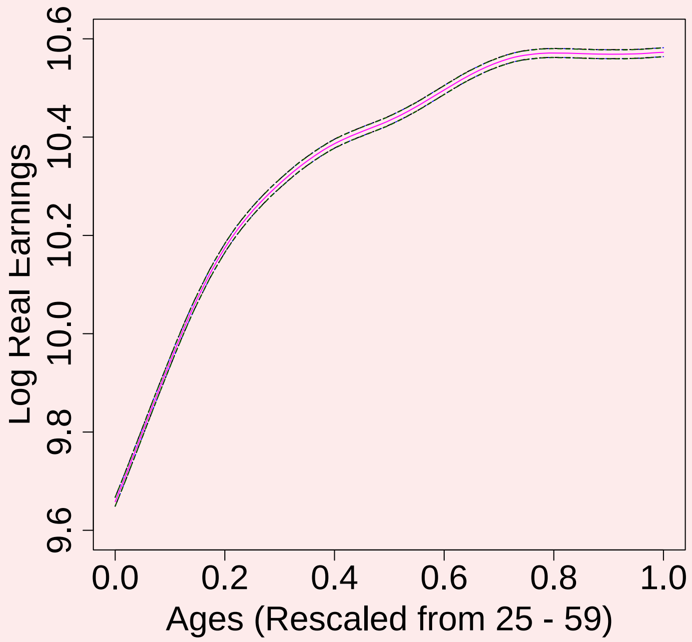

Figure 6: 95% uniform confidence bands for 0.50-quantile smoothers with 4 different education levels

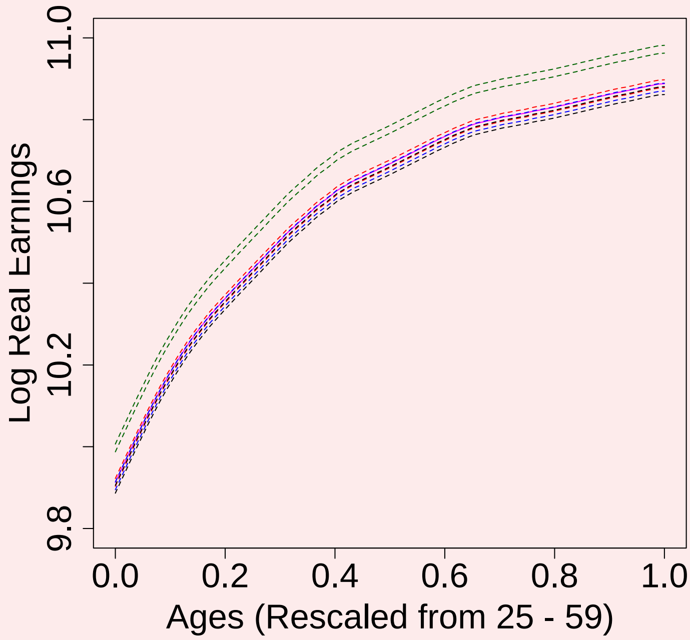

Figure 7: 95% uniform confidence bands for 0.90-quantile smoothers with 4 different education levels

6 Appendix

Proof of Theorem 2.1. Without loss of generality, based on assumption (A1), we reorder the original observations $\{X_i, Y_i\}_{i=1}^n$, such that $X_1 \leqslant X_2 \leqslant \ldots, \leqslant X_n$. First decompose:

$$\sup_{x\in J^*} |l_h^*(x) - l_g(x) - l_h^{\#}(x) - l(x)| = \max_i |l_h^*(X_i) - l_g(X_i) - l_h^{\#}(X_i) - l(X_i)| + \max_i \sup_{x\in[X_i,X_{i+1}]} |l_h^*(x) - l_g(x) - l_h^{\#}(x) - l(x)|. \quad (21)$$

From assumption (A1) we know $l'(\cdot) \le \lambda_1$ and $\max_i(X_{i+1} - X_i) = \mathcal{O}_p(S_n/n)$. By the mean value theorem, we conclude that the second term of (21) is of lower order than the first term. Together with equation (11) we have

$$\begin{aligned} &\sup_{x\in J^*} |l_h^*(x) - l_g(x) - l_h^{\#}(x) - l(x)| \\ &= \quad \mathcal{O}\{\max_i |l_h^*(X_i) - l_g(X_i) - l_h^{\#}(X_i) - l(X_i)|\} = \mathcal{O}_p(\delta_n), \end{aligned}$$

which means that the supremum of the approximation error over all x is of the same order of the maximum over the discrete observed X_i. $\square$

Proof of Theorem 2.2. The proof of (12) uses methods related to those in the proof of Theorem 3 of Härdle and Marron (1991), so only the main steps are explicitly given. The first step is a bias-variance decomposition,

$$\mathsf{E}\left[\left\{\hat{b}_{h,g}(x) - b_h(x)\right\}^2 |X_1, ...X_n\right] = \mathcal{V}_n + \mathcal{B}_n^2 \quad (22)$$

where

$$\begin{aligned} \mathcal{V}_n &= \mathsf{Var}\left[\hat{b}_{h,g}(x)|X_1, ...X_n\right], \\ \mathcal{B}_n^2 &= \mathsf{E}\left[\hat{b}_{h,g}(x) - b_h(x)|X_1, ...X_n\right]. \end{aligned}$$

Following the uniform Bahadur representation techniques for quantile regression as in Theorem 3.2 of Kong et al. (2008), we have the following linear approximation for the quantile smoother as a local polynomial smoother corresponding to a specific loss function:

$$l_h^{\#}(x) - l(x) = L_n + o_p(L_n),$$

where

$$L_n = \frac{n^{-1} \sum K_h(x - X_i)\psi\{Y_i - l(x)\}}{f\{l(x)|x\} f_X(x)}$$

for

$$\begin{aligned}
\psi(u) &= p\mathbf{1}\{u \in (0, \infty)\} - (1-p)\mathbf{1}\{u \in (-\infty, 0)\} \\
&= p - \mathbf{1}\{u \in (-\infty, 0)\}, \\
l(x-t) - l(x) &= l'(x)(-t) + l''(x)t^2 + \mathcal{O}(t^2), \\
\{l(x-t) - l(x)\}' &= l''(x)(-t) + l'''(x)t^2 + \mathcal{O}(t^2), \\
f(x-t) &= f(x) + f'(x)(-t) + f''(x)(t^2) + \mathcal{O}(t^2), \\
f'(x-t) &= f'(x) + f''(x)(-t) + f'''(x)t^2 + \mathcal{O}(t^2), \\
\int K_h(t)t\,dt &= 0, \\
\int K_h(t)t^2 dt &= h^2 d_K, \\
\int K_h(t)\mathcal{O}(t^2)dt &= \mathcal{O}(h^2).
\end{aligned}$$

Then we have

$$\mathcal{B}_n = \mathcal{B}_{n1} + \mathcal{O}(\mathcal{B}_{n1}),$$

where

$$\mathcal{B}_{n1} = \frac{\int K_g(x-t)\mathcal{U}_h(t)dt - \mathcal{U}_h(x)}{f_X(x) f\{l(x)|x\}}$$

for

$$\begin{aligned}
\mathcal{U}_h(x) &= \int K_h(x-s)\psi\{l(s) - l(x)\} f(s)ds \\
&= \int K_h(t)\psi\{l(x-t) - l(x)\} f(x-t)dt.
\end{aligned}$$

By differentiation, a Taylor expansion and properties of the kernel K (see assumption (A2)),

$$\begin{aligned}
\mathcal{U}_h'(x) &= \int K_h(t)[\psi'\{l(x-t) - l(x)\}' f(x-t) \\
&\quad + \psi\{l(x-t) - l(x)\} f'(x-t)]dt.
\end{aligned}$$

Collecting terms, we get

$$\begin{aligned}
\mathcal{U}_h'(x) &= \int K_h(t)\{\psi' l''(x) f_X'(x)t^2 + \psi' l''' f_X(x)t^2 \\
&\quad + a f'''(x)t^2 + \mathcal{O}(t^2)\}dt \\
&= \int K_h(t)\left\{C_0 t^2 + o(t^2)\right\} dt = h^2 d_K \cdot C_0 + \mathcal{O}(h^2),
\end{aligned}$$

where a is a constant with $|a| < 1$ and $C_0 = \psi' l''(x) f'_X(x) + \psi' l''' f_X(x) + a f'''(x)$.

Hence, by another substitution and Taylor expansion, for the first term in the numerator of $\mathcal{B}_{n1}$, we have

$$\mathcal{B}_{n2} = g^2 h^2 (d_K)^2 \cdot C_0 + o(g^2 h^2).$$

Thus, along almost all sample sequences,

$$\mathcal{B}_n^2 = C_1 g^4 h^4 + o(g^4 h^4) \tag{23}$$

for $C_1 = (d_K)^4 C_0^2 / [f_X^2(x) f^2 \{l(x)|x\}]$.

For the variance term, calculation in a similar spirit shows that

$$\mathcal{V}_n = \mathcal{V}_{n1} + o(\mathcal{V}_{n1}),$$

where

$$\mathcal{V}_{n1} = \frac{\int K_g^2(x-t)\mathcal{W}_h(t)dt - \{\int K_g(x-t)\mathcal{U}_h(t)dt\}^2 f_X(x) f\{l(x)|x\}}{f_X(x) f\{l(x)|x\}}$$

for

$$\begin{aligned}\mathcal{W}_h(x) &= \int K_h^2(x-s)\psi\{l(s)-l(x)\}^2 f(s)ds \\ &= \int K_h^2(t)\psi\{l(x-t)-l(x)\}^2 f(x-t)dt.\end{aligned}$$

Hence, by Taylor expansion, collecting items and similar calculation, we have

$$\mathcal{V}_n = n^{-1} h^4 g^{-5} C_2 + o(n^{-1} h^4 g^{-5}) \tag{24}$$

for a constant C_2. This, together with (22) and (23) completes the proof of Theorem 2.2. □

Proof of Theorem 3.1. In case the function l is known, the estimate $\hat{\beta}_I$ is:

$$\hat{\beta}_I = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \psi\{Y_i - l(V_i) - U_i^\top \beta\}.$$

Since l is unknown, in each of these small intervals I_{ni}, $l(V_i)$ could be regarded as a constant $\alpha = l(m_{ni})$ for some i whose corresponding interval I_{ni} covers V_i. From assumption (A1), we know that $|l(V_i) - \alpha_i| \le \lambda_1 b_n < \infty$. If we define our first step estimate $\hat{\beta}_i$ inside each small interval as

$$(\hat{\alpha}_i, \hat{\beta}_i) = \underset{\alpha,\,\beta}{\operatorname{argmin}} \sum \psi(Y_i - \alpha - U_i^\top \beta),$$

$|\{Y_i - l(V_i) - U_i^\top \beta\} - (Y_i - \alpha - U_i^\top \beta)| \leq \lambda_1 b_n < \infty$ indicates that we could treat $\hat{\beta}_i$ as $\hat{\beta}_I$ inside each partition. If we use d_i to denote the number of observations inside partition I_{ni} (based on the i.i.d. assumption as in assumption (A1), on average $d_i = n/a_n$). For each of the $\hat{\beta}_i$ inside interval I_{ni}, various parametric quantile regression literature, e.g. the convex function rule in Pollard (1991) and Knight (2001) yields

$$\sqrt{d_i}(\hat{\beta}_i - \beta) \xrightarrow{\mathcal{L}} \mathrm{N}\{0, p(1-p){D'}_i^{-1}(p) C'_i {D'}_i^{-1}(p)\} \tag{25}$$

with the matrices $C'_i = {d_i}^{-1} \sum_{i=1}^{d_i} U_i^\top U_i$ and $D'_i(p) = {d_i}^{-1} \sum_{i=1}^{d_i} f\{l(V_i)|v\} U_i^\top U_i$.

To get $\hat{\beta}$, our second step is to take the weighted mean of $\hat{\beta}_1, \ldots, \hat{\beta}_{a_n}$ as:

$$\begin{aligned} \hat{\beta} &= \arg\min_{\beta} \sum_{i=1}^{a_n} d_i (\hat{\beta}_i - \beta)^2 \\ &= \sum_{i=1}^{a_n} d_i \hat{\beta}_i / n \end{aligned}$$

Please note that under this construction, $\hat{\beta}_1, \ldots, \hat{\beta}_{a_n}$ are independent but not identical. Thus we intend to use the Lindeberg's condition for the central limit theorem. To this end, use s_n^2 to denote $\mathsf{Var}(\sum_{i=1}^{a_n} d_i \hat{\beta}_i / n)$, and we need furthur check whether the following "Lindeberg's condition" holds:

$$\lim_{a_n \to \infty} \frac{1}{s_n^2} \sum_{i=1}^{a_n} \int_{(|d_i \hat{\beta}_i / n - \beta| > \varepsilon s_n)} (\hat{\beta}_i - \beta)^2 \, dF = 0, \quad \text{for all } \varepsilon > 0. \tag{26}$$

Since

$$\begin{aligned} \mathsf{Var}(\sum_{i=1}^{a_n} d_i \hat{\beta}_i / n) &= \sum_{i}^{a_n} p(1-p) \Big\{ \Big[n/d_i \sum_{j=1}^{d_i} f\{l(V_j)|v\} U_j^\top U_j \Big]^{-1} \\ &\quad \times \sum_{i=1}^{d_i} U_i^\top U_i \Big[n/d_i \sum_{j=1}^{d_i} f\{l(V_j)|v\} U_j^\top U_j \Big]^{-1} \Big\} \\ &\approx p(1-p) \Big[\sum_{j=1}^{n} f\{l(V_j)|v\} U_j^\top U_j \Big]^{-1} \\ &\quad \times \sum_{i=1}^{n} U_i^\top U_i \Big[\sum_{j=1}^{n} f\{l(V_j)|v\} U_j^\top U_j \Big]^{-1} \\ &\stackrel{\text{def}}{=} \frac{1}{n} p(1-p) D_n^{-1} C_n D_n^{-1}, \end{aligned}$$

where $D_n = \frac{1}{n}\sum_{j=1}^{n} f\{l(V_j)|v\}U_j^\top U_j$ and $C_n = \frac{1}{n}\sum_{i=1}^{n} U_i^\top U_i$, together with the normality of $\hat{\beta}_i$ as in (25) and properties of the tail of the normal distribution, e.g. Exe. 14.3 – 14.4 of Borak et al. (2010), (26) follows.

Thus as n, $a_n \to \infty$ (although at a lower rate than n), we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} \mathrm{N}\{0, p(1-p)D_n^{-1}C_nD_n^{-1}\}. \tag{27}$$

References

Borak, S., Härdle, W., and Lopez, B. (2010). Statistics of Financial Markets Exercises and Solutions. Springer-Verlag, Heidelberg.

Denby, L. (1986). Smooth regression functions. Statistical report 26, AT&T Bell Laboratories.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Cambridge Philos. Soc., 24:180–190.

Franke, J. and Mwita, P. (2003). Nonparametric estimates for conditional quantiles of time series. Report in Wirtschaftsmathematik 87, University of Kaiserslautern.

Green, P. J. and Yandell, B. S. (1985). Semi-parametric generalized linear models. In Proceedings 2nd International GLIM Conference, volume 32 of Lecture Notes in Statistics 32, pages 44–55, New York. Springer.

Hahn, J. (1995). Bootstrapping quantile regression estimators. Econometric Theory, 11(1):105–121.

Hall, P. (1991). On convergence rates of suprema. Probab. Th. Rel. Fields, 89:447–455.

Härdle, W. and Luckhaus, S. (1984). Uniform consistency of a class of regression function estimators. Ann. Statist., 12:612–623.

Härdle, W. and Marron, J. (1991). Bootstrap simultaneous error bars for nonparametric regression. Ann. Statist., 19:778–796.

Härdle, W. and Song, S. (2009). Confidence bands in quantile regression. Econometric Theory, forthcoming.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. Econometrica, 66(6):1327–1351.

Jeong, K., Härdle, W., and Song, S. (2009). A consistent nonparametric test for causality in quantile. Econometric Theory, revise and resubmit.

Knight, K. (2001). Comparing conditional quantile estimators: first and second order considerations. Unpublished manuscript.

Koenker, R. and Bassett, G. W. (1978). Regression quantiles. Econometrica, 46:33–50.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. Journal of Econometric Perspectives, 15(4):143–156.

Kong, E., Linton, O., and Xia, Y. (2008). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. Econometric Theory, accepted.

Kuan, C., Yeh, J., and Hsu, Y. (2009). Assessing value at risk with care, the conditional autoregressive expectile models. Journal of Econometrics, 150:261–270.

Liang, H. and Li, R. (2009). Variable selection for partially linear models with measurement errors. J. Amer. Statist. Assoc., forthcoming.

Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. Econometrica, 55:816C847.

Pollard, D. (1991). Asymptotics for least absolute deviation estimators. Econometric Theory, 7:186–199.

Robinson, P. M. (1988). Semiparametric econometrics: A survey. Journal of Applied Econometrics, 3:35–51.

Speckman, P. E. (1988). Regression analysis for partially linear models. 50:413–436.

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