Partial Linear Quantile Regression and Bootstrap Confidence Bands

Wolfgang Karl Härdle*
Ya'acov Ritov**
Song Song*

BERLIN

ECONOMIC RISK

SFB 649

* Humboldt-Universität zu Berlin, Germany
** Hebrew University of Jerusalem, Israel

Partial Linear Quantile Regression and Bootstrap Confidence Bands*

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

November 28, 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 classic asymptotic uniform bands is presented through a simulation study. An economic application considers the labour market differential effect with respect to different education levels.

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

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

[†]Professor at Humboldt-Universität zu Berlin and Director of C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Berlin, Germany. & Department of Finance, National Central University, Taipei, Taiwan, R.O.C.

[‡]Professor at The Hebrew University of Jerusalem, Department of Statistics and the Center for the Study of Rationality, 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 to estimate 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 and deciding between functional forms, for example 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. Moreover, uniform confidence bands give an idea about the global variability of the estimate.

In the previous work the theoretical focus has mainly been on obtaining consistency and asymptotic normality of the quantile smoother, thereby providing the necessary ingredients to construct its pointwise confidence intervals. This, however, is not sufficient to get an idea about the global variability of the estimate, neither can it be used to correctly answer questions about the curve's shape, which contains the lack of fit test as an immediate application. This motivates us to construct the confidence bands. To this end, Härdle and Song (2010) 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 and was first noticed and investigated by Fisher and Tippett (1928), and 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

2

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. With some abuse of notation we use the letters f and F to denote different pdf's and cdf's respectively. The exact distribution will be clear from the context. 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. $l(x) = F_{Y|x}^{-1}(p)$.

In economics, discrete or categorial regressors are very common. An example is from labour market analyse where one tries to find out how revenues depend on the age of the employee (for different education levels, labour union status, genders and nationalities), i.e. in econometric analysis one targets for the differential effects. For example, Buchinsky (1995) examined the U.S. wage structure by quantile regression techniques. 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 the interpretation of linear regression models while avoiding the "curse of dimensionality". Recently Liang and Li (2009) used penalised 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: $\tilde{l}(x) = F_{Y|x}^{-1}(p) = u^\top \beta + l(v)$. The multivariate confidence band can now 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 differential effect investigation, multivariate model specification tests and the investigation of the distribution of income and wealth across regions, countries or the distribution across households possible. Additionally, since the natural link between quantile and expectile regression was developed by Newey and Powell (1987), we can further extend our result into expectile regression for various tasks, e.g. demog-

3

raphy risk research or expectile-based Value at Risk (EVAR) as in Kuan et al. (2009). For high-dimensional modelling, Belloni and Chernozhukov (2009) recently investigated high-dimensional sparse models with L_1 penalty (LASSO). Additionally, by simple calculations, our result can be further extended to intersection bounds (one side confidence bands), which is similar to Chernozhukov et al. (2009).

The rest of this article is organised as follows. To keep the main idea transparent, we start with Section 2, as an introduction to the more complicated situation, the bootstrap approximation rate for the uniform confidence band (univariate case) in quantile regression is presented through a coupling argument. An extension to multivariate covariance X with partial linear modelling is shown in Section 3 with the actual type of confidence bands and their properties. 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 differential effect. The discussion is restricted to the semiparametric extension. We do not discuss the general nonparametric regression. We conjecture that this extension is possible under appropriate conditions. All proofs are sketched in Section 6.

2 Bootstrap confidence bands in the univariate case

Suppose $Y_i = l(X_i) + \varepsilon_i$, $i = 1, \ldots, n$, where ε_i has distribution function $F(\cdot|X_i)$. For simplicity, but without any loss of generality, we assume that $F(0|X_i) = p$. $F(\xi|x)$ is smooth as a function of x and ξ for any x, and for any ξ in the neighbourhood of 0. We assume:

(A1). $X_1, \ldots, X_n$ are an i.i.d. sample, and $\inf_x f_X(x) = \lambda_0 > 0$. The quantile function satisfies: $\sup_x |l^{(j)}(x)| \leq \lambda_j < \infty$, $j = 1, 2$.

(A2). The distribution of Y given X has a density and $\inf_{x,t} f(t|x) \geq \lambda_3 > 0$, continuous in x, and in t in a neighbourhood of 0. More exactly, we have the following Taylor expansion, for some $A(\cdot)$ and $f_0(\cdot)$, and for every x, x', t:

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

where

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

4

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:

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

Generally, the bandwidth may also depend on x. A local (adaptive) bandwidth selection though deserves future 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 $\mathcal{O}(h^2)$ 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 $l_g \overset{\text{def}}{=} l_{n,g}$, 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:

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

Assumption (A3) is only stated to overwrite the issue here. It actually follows from the assumptions on (g, h). A sequence $\{a_n\}$ is slowly varying if $n^{-\alpha}a_n \to 0$ for any $\alpha > 0$. With some abuse of notation we will use S_n to denote *any* slowly varying function which may change from place to place e.g. $S_n^2 = S_n$ is a valid expression (since if S_n is a slowly varying function, then S_n^2 is slowly varying as well). 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 $\hat{F}(\cdot|X_i)$, $i = 1, \ldots, n$, estimating respectively $F(\cdot|X_i)$ and satisfying $\hat{F}(0|X_i) = p$. For example we can take the distribution with a point mass $c^{-1}K\{\alpha_n(X_j - X_i)\}$ on $Y_j - l_h(X_i)$, $j = 1, \ldots, n$, where $c = \sum_{j=1}^n K\{\alpha_n(X_j - X_i)\}$ and $\alpha_n \approx h^{-1}$, i.e.

$$\hat{F}(\cdot|X_i) = \frac{\sum_{j=1}^n K_h(X_j - X_i)\mathbf{1}\{Y_j - l_h(X_i) \leq \cdot\}}{\sum_{j=1}^n K_h(X_j - X_i)}$$

We additionally 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 (positive) $\varepsilon \to 0$,*

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

Note that the result in Lemma 2.1 is natural, since by definition, there is no error at $t = 0$, since $\hat{F}(0|X_i) \equiv p \equiv F(0|X_i)$. For $t \in (0, \varepsilon)$, $\hat{F}(t|X_i)$, 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 $\mathcal{O}_p(\varepsilon h^2) = o_p(\delta_n)$. The S_n term takes care of the maximisation.

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

$$Y_i^* = l_g(X_i) + \hat{F}^{-1}(U_i|X_i), \quad i = 1, \ldots, n \qquad (5)$$

be the bootstrap sample. We couple this sample to an unobserved hypothetical sample from the true conditional distribution:

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

Note that the vectors $(Y_1, \ldots, Y_n)$ and $(Y_1^{\#}, \ldots, Y_n^{\#})$ are equally distributed given $X_1, \ldots, X_n$. We are really interested in the exact values of $Y_i^{\#}$ and Y_i^* only when they are 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:

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

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, i.e. $q_{hi}(Y_1, \ldots, Y_n) \stackrel{\text{def}}{=} l_h(X_i)$ for data set $\{(X_i, Y_i)\}_{i=1}^n$. Note that by (3), if $|X_i - X_j| = \mathcal{O}(h)$, then

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

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

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

while

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

From (7) – (10) we conclude that

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

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.

Although Theorem 2.1 is stated with a fixed bandwidth, in practice, to take care of the heteroscedasticity effect, we construct confidence bands with the width depending on the densities, which is motivated by the counterpart based on the asymptotic theory as in Härdle and Song (2010). Thus we have the following corollary:

COROLLARY 2.1 *Under the assumptions (A1) - (A8), an approximate $(1 - \alpha) \times 100\%$ confidence band over $\mathbb{R}$ is*

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

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

Below is the summary of the basic steps for the bootstrap procedure:

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

2) Compute the conditional edf:

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

3) For each $i = 1, \ldots, n$, generate random variables $\varepsilon_{i,b}^* \sim \hat{F}(t|x)$, $b = 1, \ldots, B$ and construct the bootstrap sample $Y_{i,b}^*$, $i = 1, \ldots, n$, $b = 1, \ldots, B$ as follows:

$$Y_{i,b}^* = l_g(X_i) + \varepsilon_{i,b}^*.$$

4) For each bootstrap sample $\{(X_i, Y_{i,b}^*)\}_{i=1}^{n}$, compute l_h^* and the random variable

$$d_b \stackrel{\text{def}}{=} \sup_{x \in J^*} \left[\hat{f}\{l_h^*(x)|x\}\sqrt{\hat{f}_X(x)}|l_h^*(x) - l_g(x)| \right]. \tag{12}$$

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

5) Calculate the $(1 - \alpha)$ quantile d_α^* of $d_1, \ldots, d_B$.

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

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 bandwidth is crucial here. In our experience the bootstrap works well with a rather crude choice of g, one may, however, specify g more precisely. 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. 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

$$\hat{b}_{h,g}(x) = \mathsf{E}\, l_h^*(x) - l_g(x). \tag{13}$$

Note that in (13) the expected value is computed under the bootstrap estimation. 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 minimise 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:

8

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

(A6). K is twice continuously differentiable.

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

$$\mathsf{E}\left[\left\{\hat{b}_{h,g}(x) - b_h(x)\right\}^2 | X_1, \ldots, X_n\right] \sim h^4(C_1 g^4 + C_2 n^{-1} g^{-5}) \qquad (14)$$

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 $n^{-1/9}$, see also Härdle and Marron (1991). 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, (14) 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

$$
\begin{aligned}
y &= u^\top \beta + l(v) + \varepsilon \qquad (15) \\
&= \tilde{l}(x) + \varepsilon
\end{aligned}
$$

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 considered as being approximately constant, and hence (15) can be considered as a linear model. This observation motivates the following two stage estimation procedure:

1) A 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 $u, \mathbf{1}(v \in [0, b_n)), \mathbf{1}(v \in [b_n, 2b_n)), \ldots, \mathbf{1}(v \in [1 - b_n, 1])$. That is, let

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

Then let

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

2) Calculate the smooth quantile estimate as in (2) from $(V_i, Y_i - U_i^{\top} \hat{\beta})_{i=1}^{n}$, and name it as $\tilde{\tilde{l}}_h(v)$.

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

THEOREM 3.1 *There exist positive definite matrices D, C (defined in the appendix), such that*

$$\sqrt{n}(\hat{\beta} - \beta) \overset{\mathcal{L}}{\to} N\{0, p(1 - p)D^{-1}CD^{-1}\} \ as \ n \to \infty.$$

Note that $l(v), \tilde{l}_h(v)$ (quantile smoother based on $(v, y - u^{\top}\beta)$) and $\tilde{\tilde{l}}_h(v)$ 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

$$\widetilde{H}(\theta, v) \overset{\text{def}}{=} \int_{\mathbb{R}} f(v, \tilde{y})\psi(\tilde{y} - \theta)d\tilde{y}, \tag{16}$$

$$\widetilde{H}_n(\theta, v) \overset{\text{def}}{=} n^{-1} \sum_{i=1}^{n} K_h(v - V_i)\psi(\widetilde{Y}_i - \theta), \tag{17}$$

$$\widetilde{\tilde{H}}_n(\theta, v) \overset{\text{def}}{=} n^{-1} \sum_{i=1}^{n} K_h(v - V_i)\psi(\widetilde{\tilde{Y}}_i - \theta), \tag{18}$$

10

where

$$
\begin{aligned}
\widetilde{Y}_i &\overset{\mathrm{def}}{=} Y_i - U_i^\top \beta \\
\widetilde{\widetilde{Y}}_i &\overset{\mathrm{def}}{=} Y_i - U_i^\top \hat{\beta} = Y_i - U_i^\top \beta + U_i^\top(\beta - \hat{\beta}) \overset{\mathrm{def}}{=} \widetilde{Y}_i + Z_i.
\end{aligned}
$$

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, which are satisfied by exponential, and generalised hyperbolic distributions, also used in Härdle et al. (1988):

(A7). The conditional densities $f(\cdot|\tilde{y})$, $\tilde{y} \in \mathbb{R}$, are uniformly local Lipschitz continuous of order $\tilde{\alpha}$ (ulL-$\tilde{\alpha}$) on J, uniformly in $\tilde{y} \in \mathbb{R}$, 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 (2010) shows a.s. as $n \to \infty$:

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

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

$$
\begin{aligned}
\sup_{\theta \in I} \sup_{v \in J^*} |\widetilde{\widetilde{H}}_n(\theta, v) - \widetilde{H}(\theta, v)| &\leq \sup_{\theta \in I} \sup_{v \in J^*} |\widetilde{H}_n(\theta, v) - \widetilde{H}(\theta, v)| \\
&\quad + \underbrace{\sup_{\theta \in I} \sup_{v \in J^*} |\widetilde{H}_n(\theta, v) - \widetilde{\widetilde{H}}_n(\theta, v)|}_{\leq \mathcal{O}_p(1/\sqrt{n}) \sup_{v \in J} |n^{-1} \sum K_h|} \\
&\leq C_4 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \qquad (19)
\end{aligned}
$$

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 $\widetilde{\widetilde{l}}_h(v) - l(v)$, uniformly in v, to an application of the strong convergence of $\widetilde{\widetilde{H}}_n(\theta, v)$ to $\widetilde{H}(\theta, v)$, uniformly in v and θ. For our result on $\widetilde{\widetilde{l}}_h(\cdot)$, we shall also require

(A8). $\inf_{v \in J^*} \left| \int \psi\{y - l(v) + \varepsilon\} dF(y|v) \right| \geqslant \tilde{q}|\varepsilon|$, 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 a constant $\tilde{q}$ exists giving $f\{l(v)|v\} > \tilde{q}/p$, $x \in J$. Härdle and Song (2010) showed:

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

$$\sup_{v \in J^*} |\tilde{\tilde{l}}_h(v) - l(v)| \leq C_5 \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{20}$$

with another constant C_5 not depending on n. If additionally $\tilde{\alpha} \geqslant \{\log(\sqrt{\log n}) - \log(\sqrt{nh})\}/\log h$, (20) can be further simplified to:

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

Since the proof is essentially the same as Theorem 2.1 of the above mentioned reference, it is omitted here.

The convergence rate for the parametric part $\mathcal{O}_p(n^{-1/2})$ (Theorem 3.1) is smaller than the bootstrap approximation error for the nonparametric part $\mathcal{O}_p(n^{-2/5})$ as shown in Theorem 2.1. This makes the construction of uniform confidence bands for multivariate $x \in \mathbb{R}^d$ with a partial linear model possible.

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

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

where $\hat{f}\{l_h(x)|x\}$, $\hat{f}_X(x)$ are consistent estimators of $f\{l(x)|x\}$, $f_X(x)$.

4 A Monte Carlo study

This section is divided into two parts. First we concentrate on a univariate regressor variable x, check the validity of the bootstrap procedure together with settings in the specific example, and compare it with asymptotic uniform bands. Secondly 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:

   $$f(x, y) = f_{y|x}(y - \sin x)\mathbf{1}(x \in [0, 1]), \tag{21}$$

   where $f_{y|x}(x)$ is the pdf of $N(0, x)$ 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 $\Phi^{-1}(p) = 1.2816$, $l(x) = \sin(x) + 1.2816\sqrt{x}$. Set $h = 0.05$.

3) Compute the conditional edf:

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

   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 $\varepsilon_{i,b}^* \sim \hat{F}(t|x)$, $b = 1, \ldots, B$ and construct the bootstrap sample $Y_{i,b}^*, i = 1, \ldots, n, b = 1, \ldots, B$ as follows:
$$Y_{i,b}^* = l_g(X_i) + \varepsilon_{i,b}^*,$$
   with $g = 0.2$.

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

$$d_b \stackrel{\text{def}}{=} \sup_{x \in J^*} \left[\hat{f}\{l_h^*(x)|x\} \sqrt{\hat{f}_X(x)} |l_h^*(x) - l_g(x)| \right]. \tag{22}$$

   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. $l_h(x) \pm \left[\hat{f}\{l_h(x)|x\} \sqrt{\hat{f}_X(x)} \right]^{-1} d_\alpha^*$.

Figure 1 shows the theoretical 0.9 quantile curve, 0.9 quantile estimate with corresponding 95% uniform confidence band from the asymptotic theory and the confidence band from the bootstrap. 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 (2010). The 95% confidence band from the bootstrap is displayed as
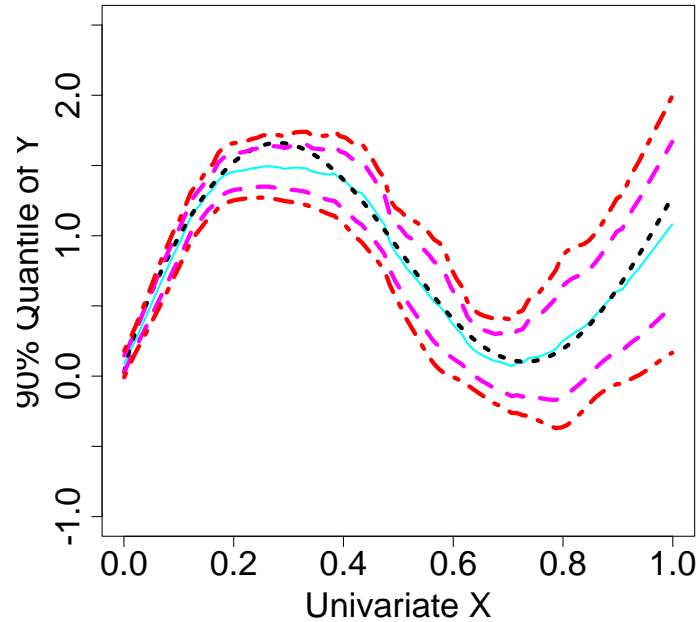
Figure 1: The real 0.9 quantile curve, 0.9 quantile estimate with corresponding 95% uniform confidence band from asymptotic theory and confidence band from bootstrapping.

red dashed-dot lines. At first sight, the quantile smoother, together with two corresponding bands, all capture the heteroscedastic structure quite well, and the width of the bootstrap confidence band is similar to the one based on asymptotic theory in Härdle and Song (2010).

To compare the small sample performance and convergence rate of both methods, Table 1 presents the simulated coverage probabilities together with the calculated area of the 95% confidence band of the quantile smoother, for three sample sizes, $n = 50$, 100 and 200. 500 simulation runs are carried out and for each simulation, 500 bootstrap samples are generated. From Table 1 we observe that, for the asymptotic method, coverage probabilities improve with increasing sample size and the bootstrap method (shown in brackets) obtains a significantly larger coverage probability than the asymptotic one, though still smaller than the nominal coverage, which results from the fact that quantile regression usually needs a larger sample size than mean regression and n here is quite moderate. It is also observed that the size of the bands decrease with increasing sample size. Overall, the bootstrap method displays a better convergence rate, while not sacrificing much on the width of the bands.

n	Cov. Prob.	Area
50	0.144 (0.642)	0.58 (1.01)
100	0.178 (0.742)	0.42 (0.58)
200	0.244 (0.862)	0.31 (0.36)

Table 1: Simulated coverage probabilities & areas of nominal asymptotic (bootstrap) 95% confidence bands with 500 repetition.

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:

$$y = 2u + v^2 + \varepsilon - 1.2816, \tag{23}$$

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 2. 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 $a_n = n^{1/3}$. Certainly this issue is far from settled and needs further investigations.

Thus for the specific model (23), 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.

15

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

5 A labour market application

Our intuition of the effect of education on income is summarised by Day and Newburger (2002)'s basic claim: "At most ages, more education equates with higher earnings, and the payoff is most notable at the highest educational levels", which is actually from the point of view of mean regression. However, whether this difference is significant or not is still questionable, especially for different ends of the (conditionally) income distribution. To this end, a careful investigation of quantile regression is necessary. Since different education levels may reflect different productivity, which is unobservable and may also results from different ages, abilities etc, to study the labour market differential effect with respect to different education levels, a semiparametric partial linear quantile model is preferred, which can retain the flexibility of the nonparametric models for the age and other unobservable factors and ease the interpretation of the education factor.

We use the administrative data from the German National Pension Office (Deutsche Rentenversicherung Bund) for the following group: West Germany part, males aged $25 - 59$, born between 1939 and 1942 who began receiving a pension in 2004 or 2005, with at least 30 yearly uncensored observations, and thus in total, $n = 128429$ observations are available. We have the following three education categories: "low education", "apprenticeship" and "university" for the variable u (assign them the numerical values 1, 2 and 3 respectively); the variable v is the age of the employee. "Low education" means without post-secondary education in Germany. "Apprenticeship" are part of Germany's dual education system. Depending on the profession, they may work for three to four days a week in the company and then spend one or two days at a vocational school (Berufsschule). "University" in Germany also includes the technical colleges (applied universities). Since the level and

structure of wages differs substantially between East and West Germany, we concentrate on West Germany only here (which we usually refer to simply as Germany). Our data have several advantages over the most often used German Socio-Economics Panel (GSOEP) data to analyze wages in Germany. Firstly, it is available for a much longer period, as opposed to from 1984 only for the GSOEP data. Secondly, more importantly, it has a much larger sample size. Thirdly, wages are likely to be measured much more precisely. Fourthly, we observe a complete earnings history from the individual's first job until his retirement, therefore this is a true panel, not a pseudo-panel. There are also several drawbacks. For example, some very wealthy individuals are not registered in the German pension system, e.g. if the monthly income is more than some threshold (which may vary for different years due to the inflation effect), the individual has the right not to be included in the public pension system, and thus not recorded. Besides this, it is also right-censored at the highest level of earnings that are subject to social security contributions, so the censored observations in the data are only for those who actually decided to stay within the public system. Because of the combination of truncation and censoring, this paper focuses on the uncensored data only, and we should not draw inferences from the very high quantile. Recently, similar data is also used to investigate the German wage structure as in Dustmann et al. (2009).

Following from Becker (1994)'s human capital mode, a log transformation is performed first on the hourly real wages (unit: EUR, in year 2000 prices). Figure 4 displays the boxplots for the "low education", "apprenticeship" and "university" groups corresponding to different ages. In the data all ages $(25 \sim 59)$ are reported as integers and are categorised in one-year groups. We rescaled them to the interval $[0, 1]$ by dividing by 40, with a corresponding bandwidth of 0.059 for the nonparametric quantile smoothers. This is equivalent to setting a bandwidth 2 in the original age data. This makes sense, because to detect whether a differential effect for different education levels exists, we compare the corresponding uniform confidence bands, i.e. differences indicate that the differential 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 5 displays $\hat{\beta}$ with respect to different quantiles for 6, 13, 25 partitions, respectively. At first, 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. $\hat{\beta}(0.20) = 0.026$, $\hat{\beta}(0.50) = 0.057$ and $\hat{\beta}(0.80) = 0.061$. Furthermore, it is robust to different numbers of partitions. It seems that the differences between the "low education" and "university" groups are different for different tails of the wage distribution. To judge whether these differences are significant, we

17

use the uniform confidence band techniques discussed in Section 2 which are displayed in Figure 6 - 8 corresponding to the 0.20, 0.50 and 0.80 quantiles respectively.

The 95% uniform confidence bands from bootstrapping for the "low education" group are marked as red dashed lines, while the ones for "apprenticeship" and "university" are displayed as blue dotted and brown dashed-dot lines, respectively. For the 0.20 quantile in Figure 6, the bands for "university", "apprenticeship" and "low education" do not differ significantly from one another although they become progressively lower, which indicates that high education does not equate to higher earnings significantly for the lower tails of wages, while increasing age seems the main driving force. For the 0.50 quantile in Figure 7, the bands for "university" and "low education" differ significantly from one another although not from "apprenticeship"'s. However, for the 0.80-quantiles in Figure 8, all the bands differ significantly (except on the right boundary because of the nonparametric method's boundary effect) resulting from the relatively large $\hat{\beta}(0.80) = 0.061$, which indicates that high education is significantly associated with higher earnings for the uppers tails of wages.

If we investigate the explanations for the differences in different tails of the income distribution, maybe the most prominent reason is the rapid development of technology, which has been extensively studied. The point is technology does not simply increase the demand for upper-end labour realtive to that of lower-end labour, but instead asymmetrically affects the bottom and the top of the wage distribution, resulting in its strong asymmetry.

Conclusions from the point of view of quantile regression are consistent with the (grouped) mean regression's, but in a careful way, i.e. we provide formal statistical tools to judge these uniformly. Partial linear quantile regression techniques, together with confidence bands, as developed in this paper, display very interesting findings compared with classic (mean) methods. Motivated by several key observations like the average income for female employees increase more than men's during the past few decades, partially because a better social welfare system means women can be more and more selective; and the "hollowing out" effect of employment, i.e. job growth in U.S., U.K. and continental Europe has increasingly been concentrated in the tails of the skill distribution over the last two decades, with disproportionate employment gains in high-wage, high-education occupations and low-wage, low-education occupations, further applications, for example to different genders, labour union status, nationalities and inequality analysis amongst other things will definitely bring more contributions to the differential analysis of the labour market.
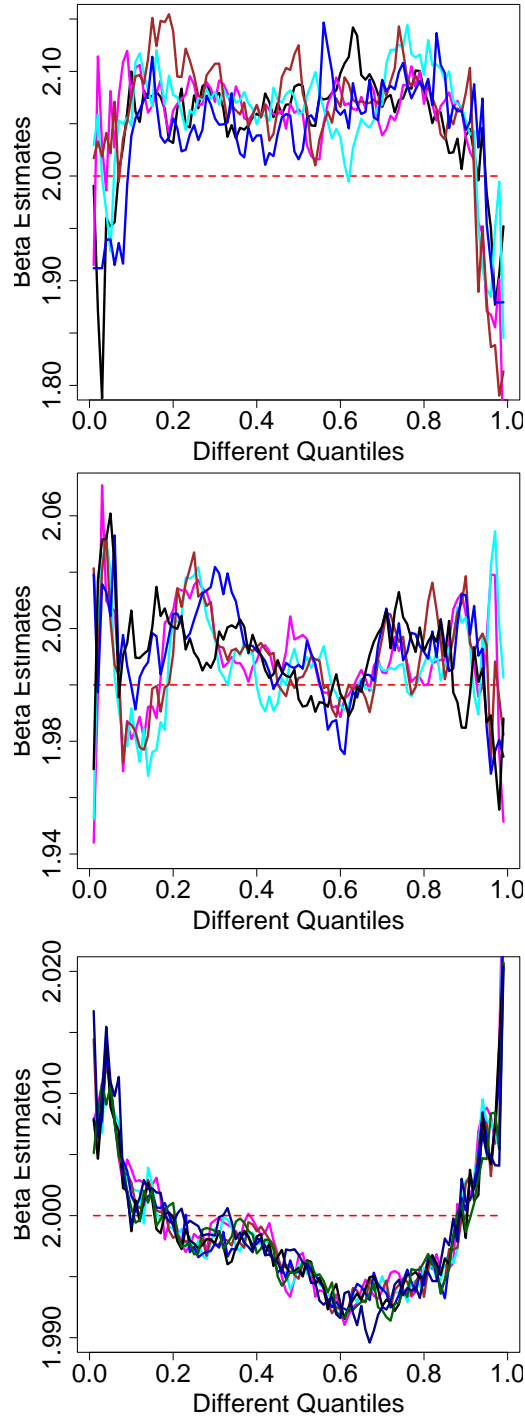
Figure 2: $\hat{\beta}$ with respect to different quantiles for different numbers of observations, i.e. $n = 1000$, $n = 8000$, $n = 261148$.
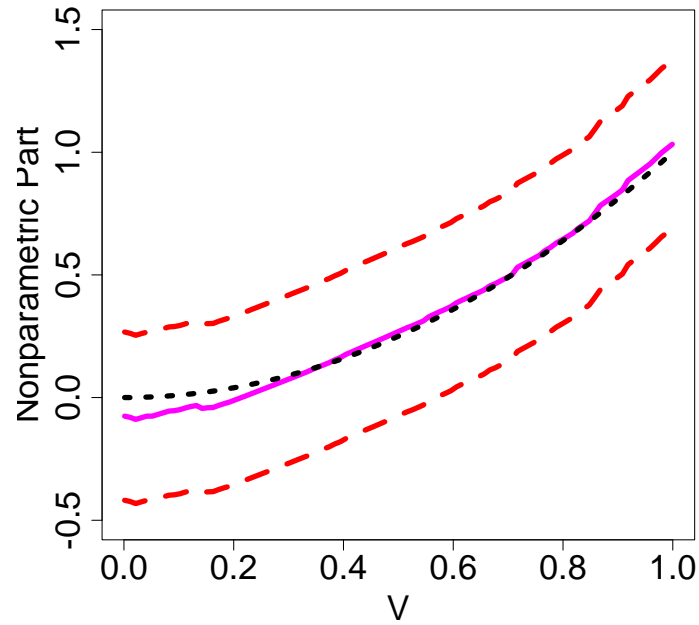
Figure 3: Nonparametric part smoothing, real 0.9 quantile curve with respect to v, 0.9 quantile smoother with corresponding 95% bootstrap uniform confidence band.
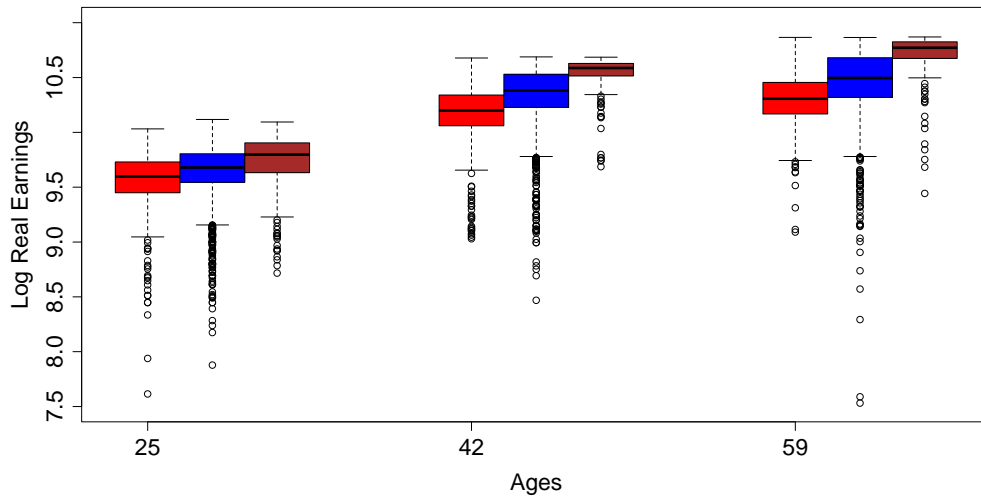

Figure 4: Boxplots for "low education", "apprenticeship" & "university" groups corresponding to different ages.
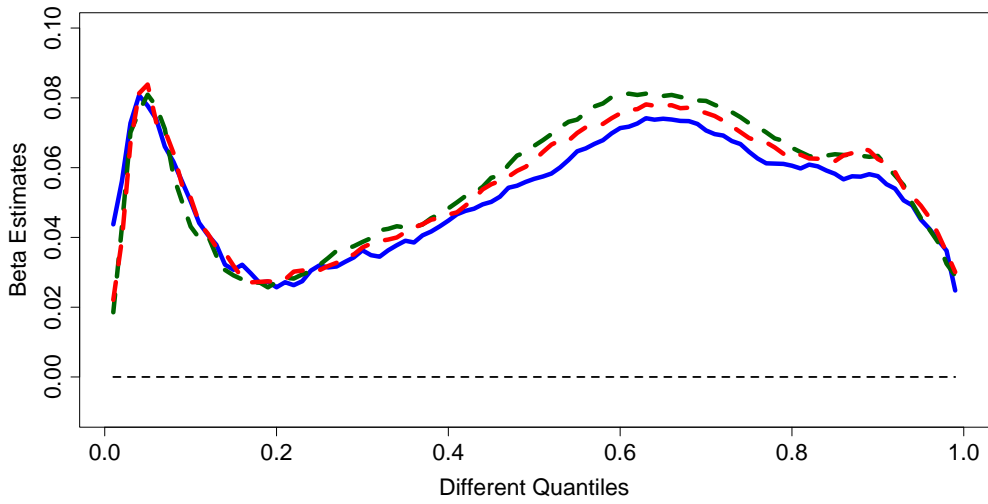
Figure 5: $\hat{\beta}$ corresponding to different quantiles with 6, 13, 25 partitions.
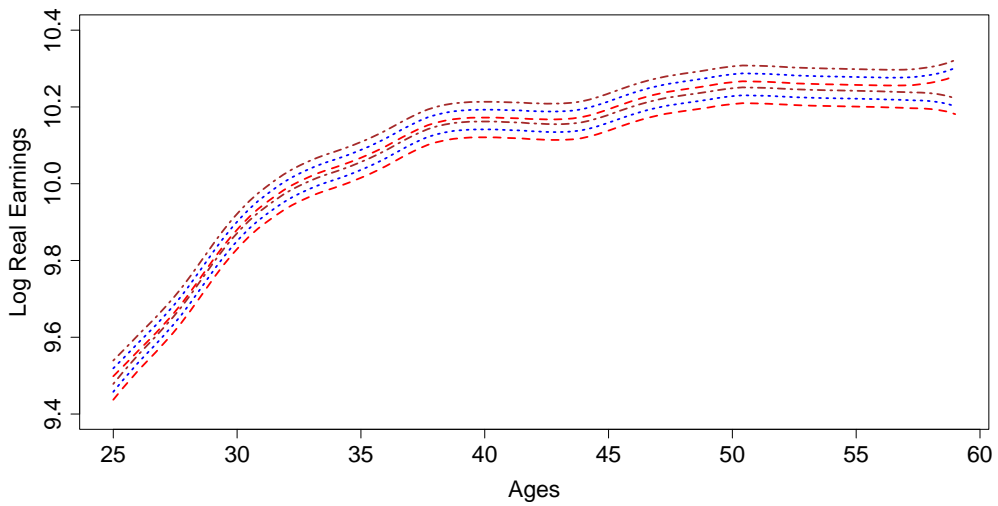

Figure 6: 95% uniform confidence bands for 0.05-quantile smoothers with 3 different education levels
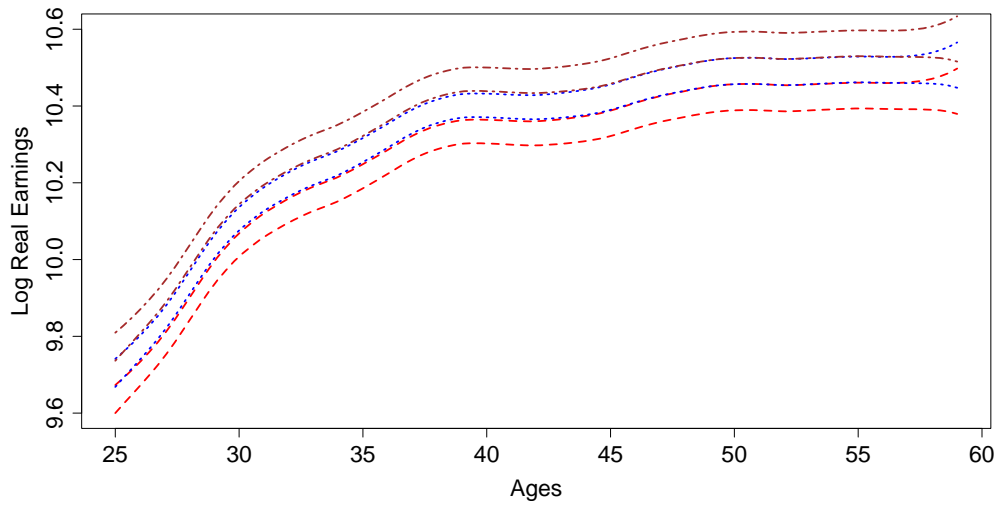
Figure 7: 95% uniform confidence bands for 0.50-quantile smoothers with 3 different education levels
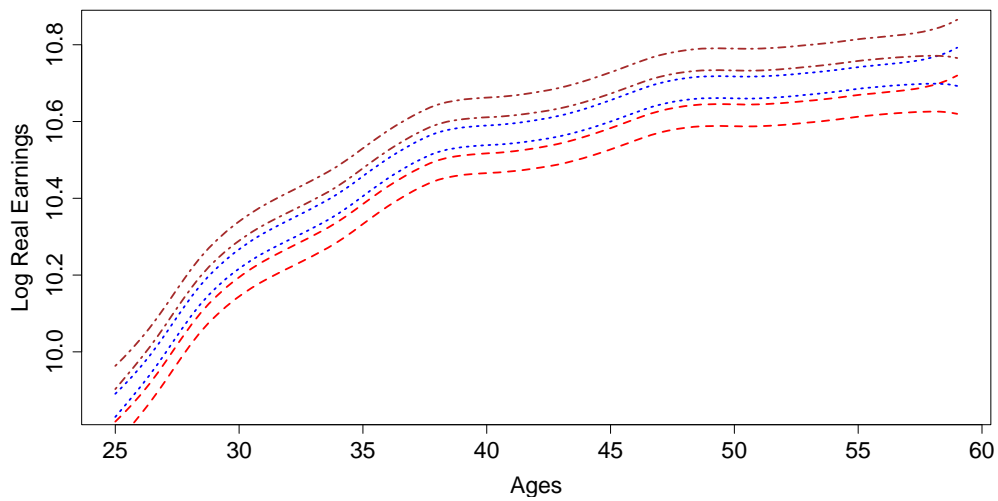

Figure 8: 95% uniform confidence bands for 0.99-quantile smoothers with 3 different education levels

6 Appendix

Proof of Theorem 2.1 We start by proving equation (7). Write first $\hat{F}^{-1}(U_i|X_i) = F^{-1}(U_i|X_i) + \Delta_i$. Fix any i such that $|F^{-1}(U_i|X_i)| \leq S_n\delta_n$, which, by equation (1), implies that $|U_i - p| < S_n\delta_n$. Lemma 2.1 gives:

$$\max_i |\hat{F}(S_n^2\delta_n|X_i) - F(S_n^2\delta_n|X_i)| = o_p(\delta_n). \tag{24}$$

Together with $F(\pm S_n^2\delta_n|X_i) = p \pm \mathcal{O}(S_n^2\delta_n)$ again by equation (1), we have $\hat{F}(\pm S_n^2\delta_n|X_i) = p \pm \mathcal{O}_p(S_n^2\delta_n)$ and thus

$$\hat{F}(-S_n^2\delta_n|X_i) = p - \mathcal{O}_p(S_n^2\delta_n)$$
$$< \quad p - S_n\delta_n < U_i < p + S_n\delta_n$$
$$< \quad p + \mathcal{O}_p(S_n^2\delta_n) = \hat{F}(S_n^2\delta_n|X_i).$$

Since $\hat{F}(\cdot|X_i)$ is monotone non-decreasing, $|\hat{F}^{-1}(U_i|X_i)| \leq S_n^2\delta_n$, which means, by $S_n^2 = S_n$,

$$|\hat{F}^{-1}(U_i|X_i)| \leq S_n\delta_n. \tag{25}$$

Apply now Lemma 2.1 again to equation (25), and obtain:

$$S_n\delta^{3/2} \quad \geq \quad |\hat{F}_i\{\hat{F}^{-1}(U_i|X_i)\} - F\{\hat{F}^{-1}(U_i|X_i)|X_i\}|$$
$$= \quad |U_i - F\{F^{-1}(U_i|X_i) + \Delta_i|X_i\}|$$
$$= \quad |F\{F^{-1}(U_i|X_i)|X_i\} - F\{F^{-1}(U_i|X_i) + \Delta_i|X_i\}|$$
$$\geq \quad f_0(X_i)|\Delta_i| \tag{26}$$

Hence $|\Delta_i| < S_n\delta_n^{3/2}$, and we summarise it as:

$$\max_{i:|F^{-1}(U_i|X_i)-F^{-1}(p)|<S_n\delta_n} |F^{-1}(U_i|X_i) - \widehat{F}^{-1}(U_i|X_i)| = \mathcal{O}_p\{S_n\delta_n^{3/2}\}.$$

Beside the above approach, there is an alternative way. Note that $|\hat{F}^{-1}(U_i|X_i)| \leq |F^{-1}(U_i|X_i)| + |\Delta_i| \leq S_n\delta_n + |\Delta_i|$. Similar to inequality (26), by applying Lemma 2.1, we have $S_n\delta_n(|\Delta_i| + S_n\delta_n)^{1/2} \geq f_0(X_i)|\Delta_i|$. Solving this inequality w.r.t. $|\Delta_i|$ gives:

$$|\Delta_i| < \{S_n\delta_n^2 + (S_n\delta_n^2 + 4S_n\delta_n^3)^{1/2}\}/2 = O_p(S_n\delta_n^{3/2}),$$

which leads to the same conclusion.

To show equation (11), define

$$Z_{1j} \quad \overset{\text{def}}{=} \quad Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)$$
$$Z_{2j} \quad \overset{\text{def}}{=} \quad Y_j^\# - l(X_j) + l(X_j) - l(X_i).$$

23

Thus $q_{hi}[\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\})_{j=1}^n]$ and $q_{hi}[\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n]$ can be seen as $l_h(X_i)$ for data sets $\{(X_i, Z_{1i})\}_{i=1}^n$ and $\{(X_i, Z_{2i})\}_{i=1}^n$ respectively. Similar to Härdle and Song (2010), they 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

$$\widetilde{G}_n(\theta, X_i) \stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^n K_h(X_i - X_j)\psi(Z_{1j} - \theta), \qquad (27)$$

$$\widetilde{\widetilde{G}}_n(\theta, X_i) \stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^n K_h(X_i - X_j)\psi(Z_{2j} - \theta). \qquad (28)$$

From (7) and (8), we have

$$\max_i \left| [\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\})_{j=1}^n] - [\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n] \right|$$
$$= \mathcal{O}_p\{S_n \delta_n^{3/2}\} + \mathcal{O}_p(\delta_n) = \mathcal{O}_p(\delta_n) \qquad (29)$$

Thus

$$\sup_{\theta \in I} \max_i |\widetilde{G}_n(\theta, X_i) - \widetilde{\widetilde{G}}_n(\theta, X_i)| \le \mathcal{O}_p(\delta_n) \max |n^{-1} \sum K_h| = \mathcal{O}_p(\delta_n)$$

To show the difference of the two quantile smoothers, we shall reduce the strong convergence of $q_{hi}[\{Y_j^* - l_g(X_j) + l_g(X_j) - l_g(X_i)\})_{j=1}^n] - q_{hi}[\{Y_j^\# - l(X_j) + l(X_j) - l(X_i)\}_{j=1}^n]$, for any i, to an application of the strong convergence of $\widetilde{G}(\theta, X_i)$ to $\widetilde{\widetilde{G}}_n(\theta, X_i)$, uniformly in θ, for any i. Under assumptions (A7) and (A8), in a similar spirit of Härdle and Song (2010), we get

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

To show the supremum of the bootstrap approximation error, without loss of generality, based on assumption (A1), we reorder the original observations $\{X_i, Y_i\}_{i=1}^n$, such that $X_1 \le X_2 \le \ldots, \le X_n$. First decompose:

$$\sup_{x \in J^*} |l_h^*(x) - l_g(x) - l_h^\#(x) - l(x)| = \max_i |l_h^*(X_i) - l_g(X_i) - l_h^\#(X_i) - l(X_i)|$$
$$+ \max_i \sup_{x \in [X_i, X_{i+1}]} |l_h^*(x) - l_g(x) - l_h^\#(x) - l(x)|. (30)$$

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 (30) is of a lower order than the first term. Together with equation (11) we have

$$\sup_{x \in J^*} |l_h^*(x) - l_g(x) - l_h^{\#}(x) - l(x)|$$

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

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. $\quad \square$

Proof of Theorem 2.2. The proof of (14) 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,

$$\mathsf{E}\left[\left\{\hat{b}_{h,g}(x) - b_h(x)\right\}^2 |X_1, ... X_n\right] = \mathcal{V}_n + \mathcal{B}_n^2 \tag{31}$$

where

$$\mathcal{V}_n = \mathsf{Var}\left[\hat{b}_{h,g}(x)|X_1, ... X_n\right],$$

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

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

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

Then we have

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

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

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

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

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

Collecting terms, we get

$$\mathcal{U}_h'(x) = \int K_h(t)\{\psi'l''(x)f_X'(x)t^2 + \psi'l'''f_X(x)t^2$$
$$+ af'''(x)t^2 + o(t^2)\}dt$$
$$= \int K_h(t)\{C_0 t^2 + o(t^2)\}dt = h^2 d_K \cdot C_0 + o(h^2),$$

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,

$$\mathcal{B}_n^2 = C_1 g^4 h^4 + o(g^4 h^4) \tag{32}$$

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

$$\mathcal{V}_n = n^{-1} h^4 g^{-5} C_2 + o(n^{-1} h^4 g^{-5}) \tag{33}$$

for a constant C_2. This, together with (31) and (32) completes the proof of Theorem 2.2. $\square$

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

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

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 $|l(V_i) - \alpha_i| \leq \lambda_1 b_n < \infty$. If we define our first step estimate $\hat{\beta}_i$ inside each small interval as

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

$|\{Y_i - l(V_i) - U_i^\top \beta\} - (Y_i - \alpha - U_i^\top \beta)| \le \lambda_1 b_n < \infty$ 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

$$\sqrt{d_i}(\hat{\beta}_i - \beta) \overset{\mathcal{L}}{\to} \mathrm{N}\{0, p(1-p)D_i'^{-1}(p)C_i'D_i'^{-1}(p)\} \qquad (34)$$

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

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

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

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 condition for the central limit theorem. To this end, we use s_n^2 to denote $\mathsf{Var}(\sum_{i=1}^{a_n} d_i \hat{\beta}_i / n)$, and we need to further check whether the following "Lindeberg condition" holds:

$$\lim_{a_n \to \infty} \frac{1}{s_n^2} \sum_{i=1}^{a_n} \int_{(|d_i \hat{\beta}_i / n - \beta| > \varepsilon s_n)} (\hat{\beta}_i - \beta)^2 \, dF = 0, \quad \text{for all } \varepsilon > 0. \qquad (35)$$

Since

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

28

where $D_n = \frac{1}{n} \sum_{j=1}^{n} f\{l(V_j)|V_i\}U_j^\top U_j$ 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 (34) and properties of the tail of the normal distribution, e.g. Exe. $14.3 - 14.4$ of Borak et al. (2010), (35) follows.

Thus as $n, a_n \to \infty$ (although at a lower rate than n), together with $C = \text{plim}_{n\to\infty} C_n$, $D = \text{plim}_{n\to\infty} D_n$, we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} \text{N}\{0, p(1-p)D^{-1}CD^{-1}\}. \tag{36}$$

References

Becker, G. (1994). Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, 3rd edition. The University of Chicago Press.

Belloni, A. and Chernozhukov, V. (2009). l_1-penalized quantile regression in high-dimensional sparse models. CEMMAP Working Paper, 10/09.

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

Buchinsky, M. (1995). Quantile regression, box-cox transformation model, and the u.s. wage structure, 1963-1987. Journal of Econometrics, 65:109–154.

Chernozhukov, V., Lee, S., and Rosen, A. M. (2009). Intersection bounds: estimation and inference. CEMMAP Working Paper, 19/09.

Day, J. C. and Newburger, E. C. (2002). The big payoff: Educational attainment and synthetic estimates of work-life earnings. special studies. current population reports. Statistical report p23-210, U.S. Department of Commerce Economics and Statistics Administration, U.S. CENSUS BUREAU.

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

Dustmann, C., Ludsteck, J., and Schönberg, U. (2009). Revisitng the german wage structure. Quartely Journal of Economics, forthcoming.

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., Janssen, P., and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. Ann. Statist., 16:1428–1429.

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. (2010). Confidence bands in quantile regression. Econometric Theory, 26:1–22.

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

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.

30

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.

SFB 649 Discussion Paper Series 2010

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.