Econometric Theory, **26**, 2010, 1180–1200.
doi:10.1017/S0266466609990491

CONFIDENCE BANDS IN QUANTILE REGRESSION

WOLFGANG K. HÄRDLE AND SONG SONG
Humboldt-Universität zu Berlin

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed random variables and let $l(x)$ be the unknown p-quantile regression curve of Y conditional on X. A quantile smoother $l_n(x)$ is a localized, nonlinear estimator of $l(x)$. The strong uniform consistency rate is established under general conditions. In many applications it is necessary to know the stochastic fluctuation of the process $\{l_n(x) - l(x)\}$. Using strong approximations of the empirical process and extreme value theory, we consider the asymptotic maximal deviation $\sup_{0 \leqslant x \leqslant 1} |l_n(x) - l(x)|$. The derived result helps in the construction of a uniform confidence band for the quantile curve $l(x)$. This confidence band can be applied as a econometric model check. An economic application considers the relation between age and earnings in the labor market by means of parametric model specification tests, which presents a new framework to describe trends in the entire wage distribution in a parsimonious way.

1. INTRODUCTION

In standard regression function estimation, most investigations are concerned with the conditional mean regression. However, new insights about the underlying structures can be gained by considering other aspects of the conditional distribution. The quantile curves are key aspects of inference in various economic problems and are of great interest in practice. These describe the conditional behavior of a response variable (e.g., wage of workers) given the value of an explanatory variable (e.g., education level, experience, occupation of workers) and investigate changes in both tails of the distribution, other than just the mean.

When examining labor markets, economists are concerned with whether discrimination exists, e.g., for different genders, nationalities, union status, etc. To study this question, we need to separate out other effects first, e.g., age, education, etc. The crucial relation between age and earnings or salaries belongs to the most carefully studied subjects in labor economics. The fundamental work in mean regression can be found in Murphy and Welch (1990). Quantile regression estimates could provide more accurate measures. Koenker and Hallock (2001) present a group of important economic applications, including quantile

Financial support from the Deutsche Forschungsgemeinschaft via SFB 649 "Ökonomisches Risiko," Humboldt-Universität zu Berlin, is gratefully acknowledged. We thank the editor and two referees for concrete suggestions on improving the manuscript and restructuring the paper. Their valuable comments and suggestions are gratefully acknowledged. Address correspondence to Song Song, Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany; e-mail: songsong@cms.hu-berlin.de.

 © Cambridge University Press 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

Engel curves, and claim that "quantile regression is gradually developing into a comprehensive strategy for completing the regression prediction." Besides this, it is also well known that a quantile regression model (e.g., the conditional median curve) is more robust to outliers, especially for fat-tailed distributions. For symmetric conditional distributions the quantile regression generates the nonparametric mean regression analysis because the $p = 0.5$ (median) quantile curve coincides with the mean regression.

As first introduced by Koenker and Bassett (1978), one may assume a parametric model for the p-quantile curve and estimate parameters by the interior point method discussed by Koenker and Park (1996) and Portnoy and Koenker (1997). Similarly, we can also adopt nonparametric methods to estimate conditional quantiles. The first one, a more direct approach using a check function such as a robustified local linear smoother, is provided by Fan, Hu, and Troung (1994) and further extended by Yu and Jones (1997, 1998). An alternative procedure is first to estimate the conditional distribution function using the double-kernel local linear technique of Fan, Yao, and Tong (1996) and then to invert the conditional distribution estimator to produce an estimator of a conditional quantile by Yu and Jones (1997, 1998). Beside these, Hall, Wolff, and Yao (1999) proposed a weighted version of the Nadaraya–Watson estimator, which was further studied by Cai (2002). Recently Jeong and Härdle (2008) have developed the conditional quantile causality test. More generally, for an M-regression function that involves quantile regression as a special case, the uniform Bahadur representation and application to the additive model are studied by Kong, Linton, and Xia (2010). An interesting question for parametric fitting, especially from labor economists, would be how well these models fit the data, when compared with the nonparametric estimation method.

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a sequence of independent and identically distributed (i.i.d.) bivariate random variables with joint probability density function (pdf) $f(x, y)$, joint cumulative distribution function (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 where $x \in J$ and J is a possibly infinite interval in $\mathbb{R}^d$ and $y \in \mathbb{R}$. In general, X may be a multivariate covariate, although here we restrict attention to the univariate case and $J = [0, 1]$ for convenience. Let $l(x)$ denote the p-quantile curve, i.e., $l(x) = F^{-1}_{Y|x}(p)$.

Under a "check function," the quantile regression curve $l(x)$ can be viewed as the minimizer of $L(\theta) \stackrel{\text{def}}{=} \mathsf{E}\{\rho_p(y - \theta)|X = x\}$ (with respect to θ) with $\rho_p(u) = pu\mathbf{1}\{u \in (0, \infty)\} - (1 - p)u\mathbf{1}\{u \in (-\infty, 0)\}$, which was originally motivated by an exercise in Ferguson (1967, p. 51) in the literature.

A kernel-based p-quantile curve estimator $l_n(x)$ can naturally be constructed by minimizing:

$$L_n(\theta) = n^{-1} \sum_{i=1}^{n} \rho_p(Y_i - \theta) K_h(x - X_i) \tag{1}$$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

with respect to $\theta \in I$ where I is a possibly infinite, or possibly degenerate, interval in $\mathbb{R}$ and $K_h(u) = h^{-1}K(u/h)$ is a kernel with bandwidth h. The numerical solution of (1) may be found iteratively as in Lejeune and Sarda (1988) and Yu, Lu, and Stander (2003).

In light of the concepts of M-estimation as in Huber (1981), if we define $\psi(u)$ as

$$\psi_p(u) = p\mathbf{1}\{u \in (0,\infty)\} - (1-p)\mathbf{1}\{u \in (-\infty,0)\}$$
$$= p - \mathbf{1}\{u \in (-\infty,0)\},$$

$l_n(x)$ and $l(x)$ can be treated as a zero (with respect to θ) of the function

$$\widetilde{H}_n(\theta,x) \stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^{n} K_h(x-X_i)\psi(Y_i-\theta), \tag{2}$$

$$\widetilde{H}(\theta,x) \stackrel{\text{def}}{=} \int_{\mathbb{R}} f(x,y)\psi(y-\theta)\,dy, \tag{3}$$

correspondingly.

To show the uniform consistency of the quantile smoother, we shall reduce the problem of strong convergence of $l_n(x) - l(x)$, uniformly in x, to an application of the strong convergence of $\widetilde{H}_n(\theta,x)$ to $\widetilde{H}(\theta,x)$, uniformly in x and θ, as given by Theorem 2.2 in Härdle, Janssen, and Serfling (1988). It is shown that under general conditions almost surely (a.s.)

$$\sup_{x\in J}|l_n(x)-l(x)| \leqslant B^* \max\left\{(nh/(\log n))^{-1/2}, h^{\tilde{\alpha}}\right\}, \quad \text{as } n\to\infty,$$

where B^* and $\tilde{\alpha}$ are parameters defined more precisely in Section 2.

Note that without assuming K has compact support (as we do here) under similar assumptions Franke and Mwita (2003) obtain

$$l_n(x) = \hat{F}^{-1}_{Y|x}(p),$$

$$\hat{F}(y|x) = \frac{\sum_{i=1}^n K_h(x-X_i)\mathbf{1}(Y_i<y)}{\sum_{i=1}^n K_h(x-X_i)},$$

$$\sup_{x\in J}|l_n(x)-l(x)| \leqslant B^{**}\left\{(nh/(s_n\log n))^{-1/2}+h^2\right\}, \quad \text{as } n\to\infty$$

for α-mixing data where B^{**} is some constant and $s_n, n \geqslant 1$ is an increasing sequence of positive integers satisfying $1 \leqslant s_n \leqslant n/2$ and some other criteria. Thus $\{nh/(\log n)\}^{-1/2} \leqslant \{nh/(s_n\log n)\}^{-1/2}$.

By employing similar methods to those developed in Härdle (1989) it is shown in this paper that

$$\mathrm{P}\left((2\delta\log n)^{1/2}\left[\sup_{x\in J} r(x)|\{l_n(x)-l(x)\}|/\lambda(K)^{1/2} - d_n\right] < z\right)$$
$$\to \exp\{-2\exp(-z)\}, \quad \text{as } n\to\infty \tag{4}$$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

from the asymptotic Gumbel distribution where $r(x)$, δ, $\lambda(K)$, d_n are suitable scaling parameters. The asymptotic result (4) therefore allows the construction of (asymptotic) uniform confidence bands for $l(x)$ based on specifications of the stochastic fluctuation of $l_n(x)$. The strong approximation with Brownian bridge techniques that we use in this paper is available only for the approximation of the two-dimensional empirical process. The extension to the multivariate covariable can be done by partial linear modeling, which deserves further research.

The plan of the paper is as follows. In Section 2, the stochastic fluctuation of the process $\{l_n(x) - l(x)\}$ and the uniform confidence band are presented through the equivalence of several stochastic processes, with a strong uniform consistency rate of $\{l_n(x) - l(x)\}$ also shown. In Section 3, in a small Monte Carlo study we investigate the behavior of $l_n(x)$ when the data are generated by fat-tailed conditional distributions of $(Y|X = x)$. In Section 4, an application considers a wage-earning relation in the labor market. All proofs are sketched in the Appendix.

2. RESULTS

The following assumptions will be convenient. To make x and X clearly distinguishable, we replace x by t sometimes, but they are essentially the same.

(A1) The kernel $K(\cdot)$ is positive and symmetric, has compact support $[-A, A]$, and is Lipschitz continuously differentiable with bounded derivatives.

(A2) $(nh)^{-1/2}(\log n)^{3/2} \to 0$, $(n \log n)^{1/2} h^{5/2} \to 0$, $(nh^3)^{-1}(\log n)^2 \leqslant M$, where M is a constant.

(A3) $h^{-3}(\log n) \int_{|y|>a_n} f_Y(y)dy = \mathcal{O}(1)$, where $f_Y(y)$ is the marginal density of Y and $\{a_n\}_{n=1}^{\infty}$ is a sequence of constants tending to infinity as $n \to \infty$.

(A4) $\inf_{t \in J} |q(t)| \geqslant q_0 > 0$, where $q(t) = \partial \mathsf{E}\{\psi(Y - \theta)|t\}/\partial\theta|_{\theta=l(t)} \cdot f_X(t) = f\{l(t)|t\} f_X(t)$.

(A5) The quantile function $l(t)$ is Lipschitz twice continuously differentiable for all $t \in J$.

(A6) $0 < m_1 \leqslant f_X(t) \leqslant M_1 < \infty$, $t \in J$; the conditional densities $f(\cdot|y)$, $y \in \mathbb{R}$, are uniform local Lipschitz continuous of order $\tilde{\alpha}$ (ulL-$\tilde{\alpha}$) on J, uniformly in $y \in \mathbb{R}$, with $0 < \tilde{\alpha} \leqslant 1$.

Define also

$$\sigma^2(t) = \mathsf{E}[\psi^2\{Y - l(t)\}|t] = p(1-p),$$

$$H_n(t) = (nh)^{-1} \sum_{i=1}^{n} K\{(t - X_i)/h\} \psi\{Y_i - l(t)\},$$

$$D_n(t) = \partial(nh)^{-1} \sum_{i=1}^{n} K\{(t - X_i)/h\} \psi\{Y_i - \theta\}/\partial\theta|_{\theta=l(t)}$$

and assume that $\sigma^2(t)$ and $f_X(t)$ are differentiable.

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

Assumption (A1) on the compact support of the kernel could possibly be relaxed by introducing a cutoff technique as in Csörgö and Hall (1982) for density estimators. Assumption (A2) has purely technical reasons: to keep the bias at a lower rate than the variance and to ensure the vanishing of some nonlinear remainder terms. Assumption (A3) appears in a somewhat modified form also in Johnston (1982). Assumptions (A5) and (A6) are common assumptions in robust estimation as in Huber (1981) and Härdle et al. (1988) that are satisfied by exponential and generalized hyperbolic distributions.

For the uniform strong consistency rate of $l_n(x) - l(x)$, we apply the result of Härdle et al. (1988) by taking $\beta(y) = \psi(y-\theta)$, $y \in \mathbb{R}$, for $\theta \in I = \mathbb{R}$, $q_1 = q_2 = -1$, $\gamma_1(y) = \max\{0, -\psi(y-\theta)\}$, $\gamma_2(y) = \min\{0, -\psi(y-\theta)\}$, and $\lambda = \infty$ to satisfy the representations for the parameters there. Thus from Härdle et al.'s Theorem 2.2 and Remark 2.3(v), we immediately have the following lemma.

LEMMA 2.1. *Let $\widetilde{H}_n(\theta, x)$ and $\widetilde{H}(\theta, x)$ be given by (2) and (3). Under Assumption (A6) and $(nh/\log n)^{-1/2} \to \infty$ through Assumption (A2), for some constant A^* not depending on n, we have a.s. as $n \to \infty$*

$$\sup_{\theta \in I} \sup_{x \in J} \left| \widetilde{H}_n(\theta, x) - \widetilde{H}(\theta, x) \right| \leq A^* \max \left\{ (nh/\log n)^{-1/2}, h^{\tilde{\alpha}} \right\}. \tag{5}$$

For our result on $l_n(\cdot)$, we shall also require

$$\inf_{x \in J} \left| \int \psi\{y - l(x) + \varepsilon\} \, dF(y|x) \right| \geqslant \tilde{q}|\varepsilon|, \quad \text{for } |\varepsilon| \leqslant \delta_1, \tag{6}$$

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 $f(l(x)|x) > \tilde{q}/p$, $x \in J$.

THEOREM 2.1. *Under the conditions of Lemma 2.1 and also assuming (6), we have a.s. as $n \to \infty$*

$$\sup_{x \in J} \left| l_n(x) - l(x) \right| \leq B^* \max \left\{ (nh/\log n)^{-1/2}, h^{\tilde{\alpha}} \right\} \tag{7}$$

with $B^* = A^*/m_1\tilde{q}$ not depending on n and m_1 a lower bound of $f_X(t)$. If additionally $\tilde{\alpha} \geqslant \{\log(\sqrt{\log n}) - \log(\sqrt{nh})\}/\log h$, it can be further simplified to

$$\sup_{x \in J} |l_n(x) - l(x)| \leq B^* \left\{ (nh/\log n)^{-1/2} \right\}.$$

THEOREM 2.2. *Let $h = n^{-\delta}$, $\frac{1}{5} < \delta < \frac{1}{3}$, $\lambda(K) = \int_{-A}^{A} K^2(u)\,du$, and*

$$d_n = (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} \left[\log \left\{ c_1(K)/\pi^{1/2} \right\} + \frac{1}{2} \left\{ \log \delta + \log\log n \right\} \right],$$

if $c_1(K) = \{K^2(A) + K^2(-A)\}/\{2\lambda(K)\} > 0$;

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

$d_n = (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} \log\{c_2(K)/2\pi\}$

otherwise with $c_2(K) = \int_{-A}^{A} \{K'(u)\}^2 \, du/\{2\lambda(K)\}$. *Then (4) holds with*

$r(x) = (nh)^{1/2} f\{l(x)|x\}\{f_X(x)/p(1-p)\}^{1/2}$.

This theorem can be used to construct uniform confidence intervals for the regression function as stated in the following corollary.

COROLLARY 2.1. *Under the assumptions of Theorem 2.2, an approximate* $(1-\alpha) \times 100\%$ *confidence band over* $[0, 1]$ *is*

$$l_n(t) \pm (nh)^{-1/2} \left\{ p(1-p)\lambda(K)/\hat{f}_X(t) \right\}^{1/2} \hat{f}^{-1}\{l(t)|t\} \left\{ d_n + c(\alpha)(2\delta \log n)^{-1/2} \right\},$$

where $c(\alpha) = \log 2 - \log|\log(1-\alpha)|$ *and* $\hat{f}_X(t)$, $\hat{f}\{l(t)|t\}$ *are consistent estimates for* $f_X(t)$, $f\{l(t)|t\}$.

In the literature, according to Fan et al. (1994, 1996), Yu and Jones (1997, 1998), Hall et al. (1999), Cai (2002), and others, asymptotic normality at interior points for various nonparametric smoothers, e.g., local constant, local linear, reweighted Nadaraya–Watson methods, etc., has been shown:

$$\sqrt{nh}\{l_n(t) - l(t)\} \sim \mathrm{N}\big(0, \tau^2(t)\big)$$

with $\tau^2(t) = \lambda(K)p(1-p)/[f_X(t)f^2\{l(t)|t\}]$. Note that the bias term vanishes here as we adjust h. With $\tau(t)$ introduced, we can further write Corollary 2.1 as

$$l_n(t) \pm (nh)^{-1/2} \left\{ d_n + c(\alpha)(2\delta \log n)^{-1/2} \right\} \hat{\tau}(t).$$

Through minimizing the approximation of asymptotic mean square error, the optimal bandwidth h_p can be computed. In practice, the rule of thumb for h_p is given by Yu and Jones (1998):

1. Use ready-made and sophisticated methods to select optimal bandwidth h_{mean} from conditional mean regression, e.g., Ruppert, Sheather, and Wand (1995);
2. $h_p = [p(1-p)/\varphi^2\{\Phi^{-1}(p)\}]^{1/5} \cdot h_{\text{mean}}$ with φ, Φ as the pdf and cdf of a standard normal distribution

Obviously the further p lies from 0.5, the more smoothing is necessary.

The proof is essentially based on a linearization argument after a Taylor series expansion. The leading linear term will then be approximated in a similar way as in Johnston (1982) and Bickel and Rosenblatt (1973). The main idea behind the proof is a strong approximation of the empirical process of $\{(X_i, Y_i)_{i=1}^n\}$ by a sequence of Brownian bridges as proved by Tusnady (1977).

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

As $l_n(t)$ is the zero (with respect to θ) of $\widetilde{H}_n(\theta, t)$, it follows by applying second-order Taylor expansions to $\widetilde{H}_n(\theta, t)$ around $l(t)$ that

$$l_n(t) - l(t) = \{H_n(t) - \mathsf{E}\, H_n(t)\}/q(t) + R_n(t), \tag{8}$$

where $\{H_n(t) - \mathsf{E}\, H_n(t)\}/q(t)$ is the leading linear term and

$$R_n(t) = H_n(t)\{q(t) - D_n(t)\}/\{D_n(t) \cdot q(t)\} + \mathsf{E}\, H_n(t)/q(t)$$

$$+ \frac{1}{2}\{l_n(t) - l(t)\}^2 \cdot \{D_n(t)\}^{-1} \tag{9}$$

$$\cdot (nh)^{-1} \sum_{i=1}^{n} K\{(x - X_i)/h\}\psi''\{Y_i - l(t) + r_n(t)\}, \tag{10}$$

$|r_n(t)| < |l_n(t) - l(t)|$

is the remainder term. In the Appendix it is shown (Lemma A.1) that $\|R_n\| = \sup_{t \in J} |R_n(t)| = \mathcal{O}_p\left\{(nh \log n)^{-1/2}\right\}$.

Furthermore, the rescaled linear part

$$Y_n(t) = (nh)^{1/2}\left\{\sigma^2(t) f_X(t)\right\}^{-1/2}\{H_n(t) - \mathsf{E}\, H_n(t)\}$$

is approximated by a sequence of Gaussian processes, leading finally to the Gaussian process

$$Y_{5,n}(t) = h^{-1/2} \int K\{(t - x)/h\}\, dW(x). \tag{11}$$

Drawing upon the result of Bickel and Rosenblatt (1973), we finally obtain asymptotically the Gumbel distribution.

We also need the Rosenblatt (1952) transformation,

$$T(x, y) = \{F_{X|y}(x|y), F_Y(y)\},$$

which transforms (X_i, Y_i) into $T(X_i, Y_i) = (X_i', Y_i')$ mutually independent uniform random variables. In the event that x is a d-dimensional covariate, the transformation becomes

$$T(x_1, x_2, \ldots, x_d, y) = \{F_{X_1|y}(x_1|y), F_{X_2|y}(x_2|x_1, y), \ldots, F_{X_k|x_{d-1},\ldots,x_1,y}$$

$$(x_k|x_{d-1}, \ldots, x_1, y), F_Y(y)\}. \tag{12}$$

With the aid of this transformation, Theorem 1 of Tusnady (1977) may be applied to obtain the following lemma.

LEMMA 2.2. *On a suitable probability space a sequence of Brownian bridges B_n exists such that*

$$\sup_{x \in J, y \in \mathbb{R}} |Z_n(x, y) - B_n\{T(x, y)\}| = \mathcal{O}\left\{n^{-1/2}(\log n)^2\right\} \quad a.s.,$$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

where $Z_n(x,y) = n^{1/2}\{F_n(x,y) - F(x,y)\}$ *denotes the empirical process of* $\{(X_i, Y_i)\}_{i=1}^n$.

For $d > 2$, it is still an open problem that deserves further research.

Before we define the different approximating processes, let us first rewrite (11) as a stochastic integral with respect to the empirical process $Z_n(x,y)$:

$$Y_n(t) = \{hg'(t)\}^{-1/2} \iint K\{(t-x)/h\}\psi\{y - l(t)\}\, dZ_n(x,y),$$

$$g'(t) = \sigma^2(t) f_X(t).$$

The approximating processes are now

$$Y_{0,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\}\psi\{y - l(t)\}\, dZ_n(x,y), \tag{13}$$

where $\Gamma_n = \{|y| \leqslant a_n\}$, $g(t) = \mathsf{E}[\psi^2\{y - l(t)\} \cdot \mathbf{1}(|y| \leqslant a_n)|X = t] \cdot f_X(t)$

$$Y_{1,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\}\psi\{y - l(t)\}\, dB_n\{T(x,y)\}, \tag{14}$$

$\{B_n\}$ being the sequence of Brownian bridges from Lemma 2.2.

$$Y_{2,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\}\psi\{y - l(t)\}\, dW_n\{T(x,y)\}, \tag{15}$$

$\{W_n\}$ being the sequence of Wiener processes satisfying

$$B_n(x', y') = W_n(x', y') - x'y'W_n(1,1),$$

$$Y_{3,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\}\psi\{y - l(x)\}\, dW_n\{T(x,y)\}, \tag{16}$$

$$Y_{4,n}(t) = \{hg(t)\}^{-1/2} \int g(x)^{1/2} K\{(t-x)/h\}\, dW(x), \tag{17}$$

$$Y_{5,n}(t) = h^{-1/2} \int K\{(t-x)/h\}\, dW(x), \tag{18}$$

$\{W(\cdot)\}$ being the Wiener process.

Lemmas A.2–A.7 in the Appendix ensure that all these processes have the same limit distributions. The result then follows from the next lemma.

LEMMA 2.3 (Theorem 3.1 in Bickel and Rosenblatt, 1973). *Let* d_n, $\lambda(K)$, δ *as in Theorem 2.2. Let*

$$Y_{5,n}(t) = h^{-1/2} \int K\{(t-x)/h\}\, dW(x).$$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

Then, as $n \to \infty$, the supremum of $Y_{5,n}(t)$ has a Gumbel distribution:

$$\mathrm{P}\left\{(2\delta \log n)^{1/2}\left[\sup_{t\in J}|Y_{5,n}(t)|/\{\lambda(K)\}^{1/2} - d_n\right] < z\right\} \to \exp\{-2\exp(-z)\}.$$

3. A MONTE CARLO STUDY

We generate bivariate data $\{(X_i, Y_i)\}_{i=1}^n, n = 500$ with joint pdf:

$$f(x,y) = g\left(y - \sqrt{x+2.5}\right)\mathbf{1}(x \in [-2.5, 2.5]), \tag{19}$$

$$g(u) = \frac{9}{10}\varphi(u) + \frac{1}{90}\varphi(u/9).$$

The p-quantile curve $l(x)$ can be obtained from a zero (with respect to θ) of

$$9\Phi(\theta) + \Phi(\theta/9) = 10p,$$

with Φ as the cdf of a standard normal distribution. Solving it numerically gives the 0.5-quantile curve $l(x) = \sqrt{x+2.5}$ and the 0.9-quantile curve $l(x) = 1.5296 + \sqrt{x+2.5}$. We use the quartic kernel:

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

$$= 0, \qquad |u| > 1.$$

In Figure 1 the raw data, together with the 0.5-quantile curve, are displayed. The random variables generated with probability $\frac{1}{10}$ from the fat-tailed pdf $\frac{1}{9}\varphi(u/9)$ (see eqn. (19)) are marked as squares whereas the standard normal random variables are shown as stars. We then compute both the Nadaraya–Watson estimator $m_n^*(x)$ and the 0.5-quantile smoother $l_n(x)$. The bandwidth is set to 1.25, which is equivalent to 0.25 after rescaling x to $[0, 1]$ and fulfills the requirements of Theorem 2.2.

In Figure 1 $l(x)$, $m_n^*(x)$, and $l_n(x)$ are shown as a dotted line, dashed-dot line, and solid line, respectively. At first sight $m_n^*(x)$ has clearly more variation and has the expected sensitivity to the fat tails of $f(x, y)$. A closer look reveals that $m_n^*(x)$ for $x \approx 0$ apparently even leaves the 0.5-quantile curve. It may be surprising that this happens at $x \approx 0$ where no outlier is placed, but a closer look at Figure 1 shows that the large negative data values at both $x \approx -0.1$ and $x \approx 0.25$ cause the problem. This data value is inside the window ($h = 1.10$) and therefore distorts $m_n^*(x)$ for $x \approx 0$. The quantile smoother $l_n(x)$ (solid line) is unaffected and stays fairly close to the 0.5-quantile curve. Similar results can be obtained in Figure 2 corresponding to the 0.9 quantile ($h = 1.25$) with the 95% confidence band.

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

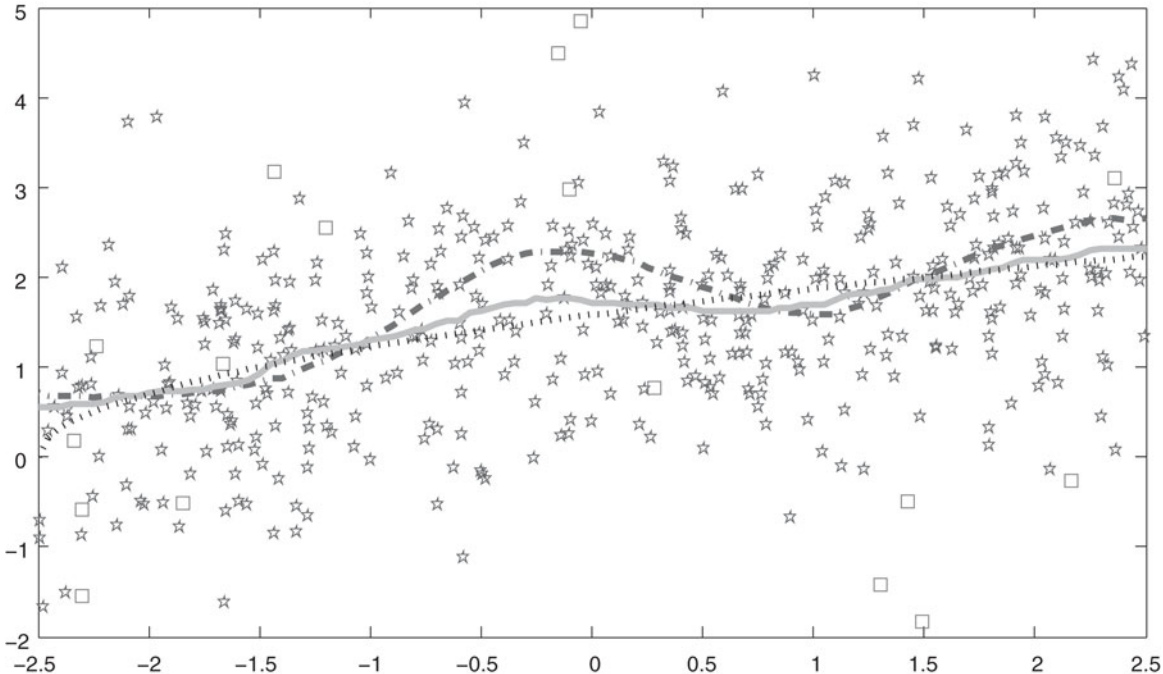


FIGURE 1. The 0.5-quantile curve, the Nadaraya–Watson estimator $m_n^*(x)$, and the 0.5-quantile smoother $l_n(x)$.

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

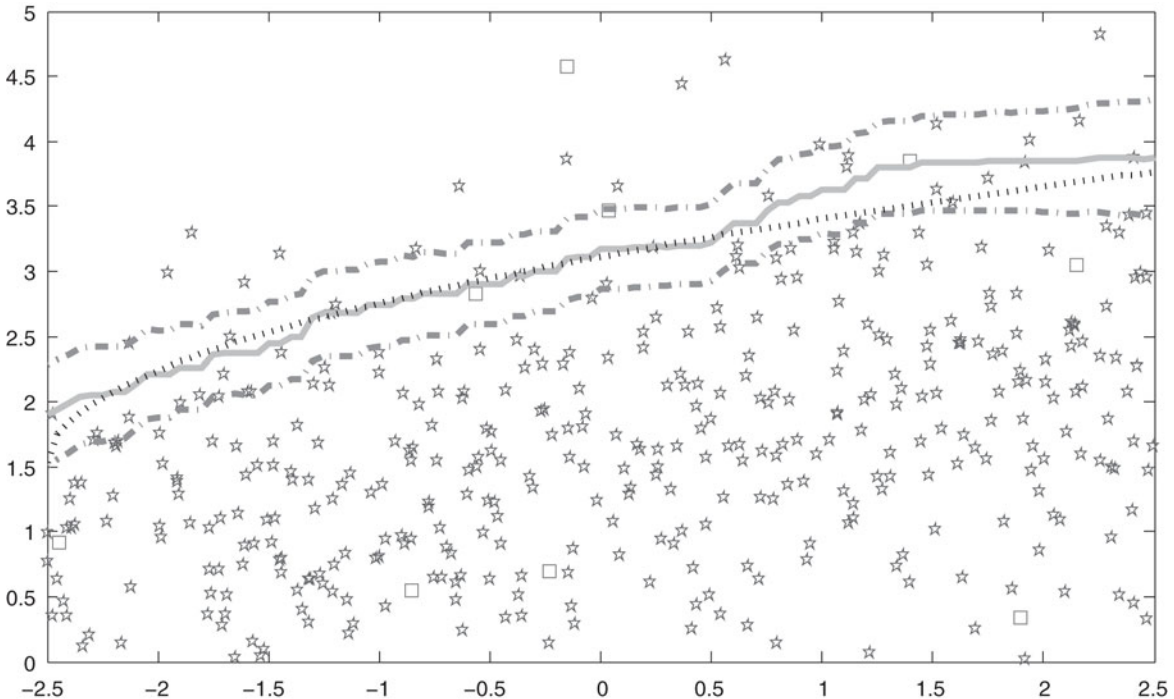


FIGURE 2. The 0.9-quantile curve, the 0.9-quantile smoother, and 95% confidence band.

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

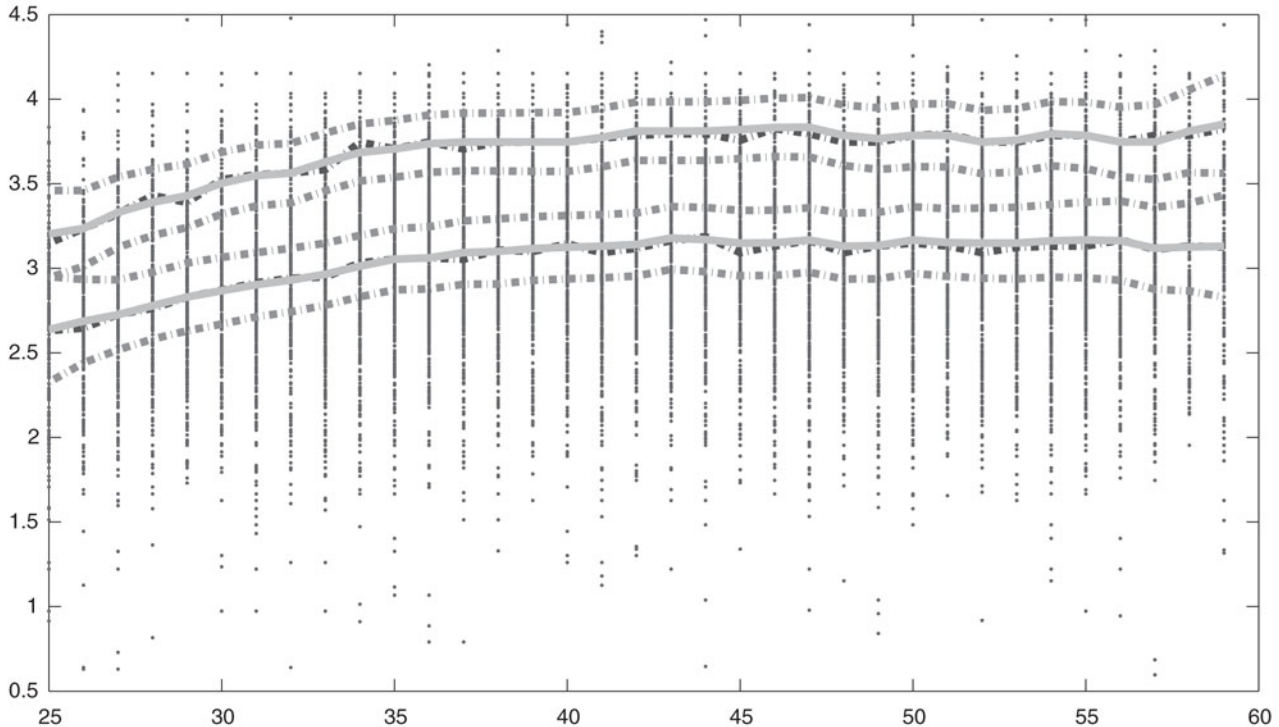


FIGURE 3. The original observations, local quantiles, 0.5- and 0.9-quantile smoothers, and corresponding 95% confidence bands.

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

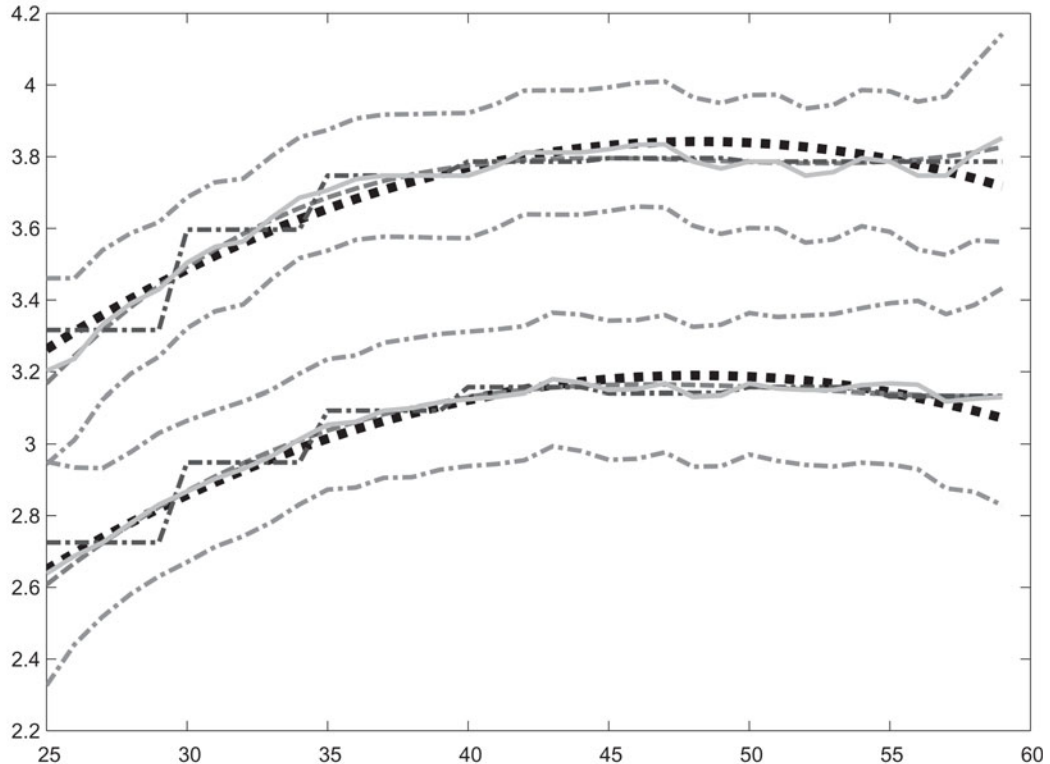


FIGURE 4. Quadratic, quartic, set of dummies (for age groups) estimates, 0.5- and 0.9-quantile smoothers, and their corresponding 95% confidence bands.

4. APPLICATION

Recently there has been great interest in finding out how the financial returns of a job depend on the age of the employee. We use the Current Population Survey (CPS) data from 2005 for the following group: male aged 25–59, full-time employed, and college graduate containing 16,731 observations, for the age-earning estimation. As is usual for wage data, a log transformation to hourly real wages (unit: U.S. dollar) is carried out first. In the CPS all ages (25–59) are reported as integers. We rescaled them into $[0, 1]$ by dividing 40 by bandwidth 0.059 for nonparametric quantile smoothers. This is equivalent to setting bandwidth 2 for the original age data.

In Figure 3 the original observations are displayed as small stars. The local 0.5 and 0.9 quantiles at the integer points of age are shown as dashed lines, whereas the corresponding nonparametric quantile smoothers are displayed as solid lines with corresponding 95% uniform confidence bands shown as dashed-dot lines. A closer look reveals a quadratic relation between age and logged hourly real wages. We use several popular parametric methods to estimate the 0.5 and 0.9 conditional quantiles, e.g., quadratic, quartic, and set of dummies (a dummy variable for each 5-year age group) models; the results are displayed in Figure 4. With the help of the 95% uniform confidence bands, we can conduct the parametric model specification test. At the 5% significance level, we could not reject any model. However, when the confidence level further decreases and the uniform confidence bands get narrower, the "set of dummies" parametric model will be the first one to be rejected. At the 10% significance level, the set of dummies (for age groups) model is rejected whereas the other two are not. As the quadratic model performs quite similarly to the quartic one, for simplicity it is suggested in practice to measure the log(wage)-earning relation in mean regression, which coincides with the approach of Murphy and Welch (1990).

REFERENCES

Bickel, P. & M. Rosenblatt (1973) On some global measures of the deviation of density function estimatiors. *Annals of Statistics* 1, 1071–1095.

Cai, Z.W. (2002) Regression quantiles for time series. *Econometric Theory* 18, 169–192.

Csörgö, S. & P. Hall (1982) Upper and lower classes for triangular arrays. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 61, 207–222.

Fan, J., T.C. Hu, & Y.K. Troung (1994) Robust nonparametric function estimation. *Scandinavian Journal of Statistics* 21, 433–446.

Fan, J., Q. Yao, & H. Tong (1996) Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* 83, 189–206.

Ferguson, T.S. (1967) *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press.

Franke, J. & P. Mwita (2003) Nonparametric Estimates for Conditional Quantiles of Time Series. Report in Wirtschaftsmathematik 87, University of Kaiserslautern.

Hall, P., R. Wolff, & Q. Yao (1999) Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* 94, 154–163.

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

Härdle, W. (1989) Asymptotic maximal deviation of M-smoothers. *Journal of Multivariate Analysis* 29, 163–179.

Härdle, W., P. Janssen & R. Serfling (1988) Strong uniform consistency rates for estimators of conditional functionals. *Annals of Statistics* 16, 1428–1429.

Härdle, W. & S. Luckhaus (1984) Uniform consistency of a class of regression function estimators. *Annals of Statistics* 12, 612–623.

Huber, P. (1981) *Robust Statistics*. Wiley.

Jeong, K. & W. Härdle. (2008) A Consistent Nonparametric Test for Causality in Quantile. SFB 649 Discussion Paper.

Johnston, G. (1982) Probabilities of maximal deviations of nonparametric regression function estimates. *Journal of Multivariate Analysis* 12, 402–414.

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

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

Koenker, R. & B.J. Park (1996) An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics* 71, 265–283.

Kong, E., O. Linton, & Y. Xia (2010) Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory*, forthcoming.

Lejeune, M.G. & P. Sarda (1988) Quantile regression: A nonparametric approach. *Computational Statistics and Data Analysis* 6, 229–239.

Murphy, K. & F. Welch (1990) Empirical age-earnings profiles. *Journal of Labor Economics* 8, 202–229.

Parzen, M. (1962) On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 32, 1065–1076.

Portnoy, S. & R. Koenker (1997) The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators (with discussion). *Statistical Sciences* 12, 279–300.

Rosenblatt, M. (1952) Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23, 470–472.

Ruppert, D., S.J. Sheather, & M.P. Wand (1995) An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* 90, 1257–1270.

Tusnady, G. (1977) A remark on the approximation of the sample distribution function in the multidimensional case. *Periodica Mathematica Hungarica* 8, 53–55.

Yu, K. & M.C. Jones (1997) A comparison of local constant and local linear regression quantile estimation. *Computational Statistics and Data Analysis* 25, 159–166.

Yu, K. & M.C. Jones (1998) Local linear quantile regression. *Journal of the American Statistical Association* 93, 228–237.

Yu, K., Z. Lu, & J. Stander (2003) Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society, Series D* 52, 331–350.

APPENDIX

Proof of Theorem 2.1 . By the definition of $l_n(x)$ as a zero of (2), we have, for $\varepsilon > 0$,

$$\text{if} \quad l_n(x) > l(x)+\varepsilon, \quad \text{then} \quad \widetilde{H}_n\{l(x)+\varepsilon, x\} > 0. \tag{A.1}$$

Now

$$\widetilde{H}_n\{l(x)+\varepsilon, x\} \leqslant \widetilde{H}\{l(x)+\varepsilon, x\} + \sup_{\theta\in I} \left|\widetilde{H}_n(\theta, x) - \widetilde{H}(\theta, x)\right|. \tag{A.2}$$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

Also, by the identity $\widetilde{H}\{l(x),x\}=0$, the function $\widetilde{H}\{l(x)+\varepsilon,x\}$ is not positive and has a magnitude $\geqslant m_1\tilde{q}\varepsilon$ by Assumption (A6) and (6), for $0<\varepsilon<\delta_1$. That is, for $0<\varepsilon<\delta_1$,

$$\widetilde{H}\{l(x)+\varepsilon,x\}\leqslant -m_1\tilde{q}\varepsilon. \tag{A.3}$$

Combining (A.1)–(A.3), we have, for $0<\varepsilon<\delta_1$,

$$\text{if}\quad l_n(x)>l(x)+\varepsilon,\quad \text{then}\quad \sup_{\theta\in I}\sup_{x\in J}\left|\widetilde{H}_n(\theta,x)-\widetilde{H}(\theta,x)\right|>m_1\tilde{q}\varepsilon.$$

With a similar inequality proved for the case $l_n(x)<l(x)+\varepsilon$, we obtain, for $0<\varepsilon<\delta_1$,

$$\text{if}\quad \sup_{x\in J}|l_n(x)-l(x)|>\varepsilon,\quad \text{then}\quad \sup_{\theta\in I}\sup_{x\in J}\left|\widetilde{H}_n(\theta,x)-\widetilde{H}(\theta,x)\right|>m_1\tilde{q}\varepsilon. \tag{A.4}$$

It readily follows that (A.4) and (5) imply (7). ■

Subsequently we first show that $\|R_n\|_\infty=\sup_{t\in J}|R_n(t)|$ vanishes asymptotically faster than the rate $(nh\log n)^{-1/2}$; for simplicity we will just use $\|\cdot\|$ to indicate the sup-norm.

LEMMA A.1. *For the remainder term $R_n(t)$ defined in (9) we have*

$$\|R_n\|=\mathcal{O}_p\{(nh\log n)^{-1/2}\}. \tag{A.5}$$

Proof. First we have by the positivity of the kernel K,

$$\|R_n\|\leqslant\left[\inf_{0\leqslant t\leqslant 1}\{|D_n(t)|\cdot q(t)\}\right]^{-1}\{\|H_n\|\cdot\|q-D_n\|+\|D_n\|\cdot\|\mathsf{E}\,H_n\|\}$$

$$+C_1\cdot\|l_n-l\|^2\cdot\left\{\inf_{0\leqslant t\leqslant 1}|D_n(t)|\right\}^{-1}\cdot\|f_n\|_\infty,$$

where $f_n(x)=(nh)^{-1}\sum_{i=1}^n K\{(x-X_i)/h\}$.

The desired result, Lemma A.1, will then follow if we prove

$$\|H_n\|=\mathcal{O}_p\left\{(nh)^{-1/2}(\log n)^{1/2}\right\}, \tag{A.6}$$

$$\|q-D_n\|=\mathcal{O}_p\left\{(nh)^{-1/4}(\log n)^{-1/2}\right\}, \tag{A.7}$$

$$\|\mathsf{E}\,H_n\|=\mathcal{O}\left(h^2\right), \tag{A.8}$$

$$\|l_n-l\|^2=\mathcal{O}_p\left\{(nh)^{-1/2}(\log n)^{-1/2}\right\}. \tag{A.9}$$

Because (A.8) follows from the well-known bias calculation

$$\mathsf{E}\,H_n(t)=h^{-1}\int K\{(t-u)/h\}\,\mathsf{E}[\psi\{y-l(t)\}|X=u]f_X(u)\,du=\mathcal{O}(h^2),$$

where $\mathcal{O}(h^2)$ is independent of t in Parzen (1962), we have from Assumption (A2) that $\|\mathsf{E}\,H_n\|=\mathcal{O}_p\{(nh)^{-1/2}(\log n)^{-1/2}\}$.

According to Lemma A.3 in Franke and Mwita (2003),

$$\sup_{t\in J}|H_n(t)-\mathsf{E}\,H_n(t)|=\mathcal{O}\left\{(nh)^{-1/2}(\log n)^{1/2}\right\}$$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

and the following inequality

$$\begin{aligned}\|H_n\| &\leqslant \|H_n - \mathsf{E}\, H_n\| + \|\mathsf{E}\, H_n\| \\ &= \mathcal{O}\left\{(nh)^{-1/2}(\log n)^{1/2}\right\} + \mathcal{O}_p\left\{(nh)^{-1/2}(\log n)^{-1/2}\right\} \\ &= \mathcal{O}\left\{(nh)^{-1/2}(\log n)^{1/2}\right\},\end{aligned}$$

statement (A.6) thus is obtained.

Statement (A.7) follows in the same way as (A.6) using Assumption (A2) and the Lipschitz continuity properties of K, ψ', l.

According to the uniform consistency of $l_n(t) - l(t)$ shown before, we have

$$\|l_n - l\| = \mathcal{O}_p\{(nh)^{-1/2}(\log n)^{1/2}\},$$

which implies (A.9).

Now the assertion of the lemma follows, because by tightness of $D_n(t)$, $\inf_{0\leqslant t\leqslant 1} |D_n(t)| \geqslant q_0$ a.s. and thus

$$\|R_n\| = \mathcal{O}_p\{(nh\log n)^{-1/2}\}(1 + \|f_n\|).$$

Finally, by Theorem 3.1 of Bickel and Rosenblatt (1973), $\|f_n\| = \mathcal{O}_p(1)$; thus the desired result $\|R_n\| = \mathcal{O}_p\{(nh\log n)^{-1/2}\}$ follows. ■

We now begin with the subsequent approximations of the processes $Y_{0,n}$–$Y_{5,n}$.

LEMMA A.2.

$$\|Y_{0,n} - Y_{1,n}\| = \mathcal{O}\left\{(nh)^{-1/2}(\log n)^2\right\} \quad a.s.$$

Proof. Let t be fixed and put $L(y) = \psi\{y - l(t)\}$ still depending on t. Using integration by parts, we obtain

$$\begin{aligned}&\iint_{\Gamma_n} L(y)K\{(t-x)/h\}\,dZ_n(x,y) \\ &\quad= \int_{u=-A}^{A}\int_{y=-a_n}^{a_n} L(y)K(u)\,dZ_n(t-h\cdot u, y) \\ &\quad= -\int_{-A}^{A}\int_{-a_n}^{a_n} Z_n(t-h\cdot u, y)\,d\{L(y)K(u)\} \\ &\qquad + L(a_n)(a_n)\int_{-A}^{A} Z_n(t-h\cdot u, a_n)\,dK(u) \\ &\qquad - L(-a_n)(-a_n)\int_{-A}^{A} Z_n(t-h\cdot u, -a_n)\,dK(u) \\ &\qquad + K(A)\left\{\int_{-a_n}^{a_n} Z_n(t-h\cdot A, y)\,dL(y)\right. \\ &\qquad\qquad \left. + L(a_n)(a_n)Z_{n_a}(t-h\cdot A, a_n) - L(-a_n)(-a_n)Z_n(t-h\cdot A, -a_n)\right\}\end{aligned}$$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

$$- K(-A)\left\{\int_{-a_n}^{a_n} Z_n(t+h\cdot A, y)\,dL(y) + L(a_n)(a_n)Z_n(t+h\cdot A, a_n)\right.$$
$$\left. - L(-a_n)(-a_n)Z_n(t+h\cdot A, -a_n)\right\}.$$

If we apply the same operation to $Y_{1,n}$ with $B_n\{T(x,y)\}$ instead of $Z_n(x,y)$ and use Lemma 2.2, we finally obtain

$$\sup_{0\leqslant t\leqslant 1} h^{1/2} g(t)^{1/2} |Y_{0,n}(t) - Y_{1,n}(t)| = \mathcal{O}\left\{n^{-1/2}(\log n)^2\right\} \quad \text{a.s.} \qquad \blacksquare$$

LEMMA A.3. $\|Y_{1,n} - Y_{2,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof. Note that the Jacobian of $T(x,y)$ is $f(x,y)$. Hence

$$Y_{1,n}(t) - Y_{2,n}(t) = \left|\{g(t)h\}^{-1/2} \iint_{\Gamma_n} \psi\{y - l(t)\} K\{(t-x)/h\} f(x,y)\,dx\,dy\right| \cdot |W_n(1,1)|.$$

It follows that

$$h^{-1/2}\|Y_{1,n} - Y_{2,n}\| \leqslant |W_n(1,1)| \cdot \left\|g^{-1/2}\right\|$$
$$\cdot \sup_{0\leqslant t\leqslant 1} h^{-1} \iint_{\Gamma_n} |\psi\{y - l(t)\} K\{(t-x)/h\}| f(x,y)\,dx\,dy.$$

Because $\|g^{-1/2}\|$ is bounded by assumption, we have

$$h^{-1/2}\|Y_{1,n} - Y_{2,n}\| \leqslant |W_n(1,1)| \cdot C_4 \cdot h^{-1} \int K\{(t-x)/h\}\,dx = \mathcal{O}_p(1). \qquad \blacksquare$$

LEMMA A.4. $\|Y_{2,n} - Y_{3,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof. The difference $|Y_{2,n}(t) - Y_{3,n}(t)|$ may be written as

$$\left|\{g(t)h\}^{-1/2} \iint_{\Gamma_n} [\psi\{y - l(t)\} - \psi\{y - l(x)\}] K\{(t-x)/h\}\,dW_n\{T(x,y)\}\right|.$$

If we use the fact that l is uniformly continuous, this is smaller than

$$h^{-1/2}|g(t)|^{-1/2} \cdot \mathcal{O}_p(h),$$

and the lemma thus follows. $\blacksquare$

LEMMA A.5. $\|Y_{4,n} - Y_{5,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof.

$$|Y_{4,n}(t) - Y_{5,n}(t)| = h^{-1/2}\left|\int \left[\left\{\frac{g(x)}{g(t)}\right\}^{1/2} - 1\right] K\{(t-x)/h\}\,dW(x)\right|$$
$$\leqslant h^{-1/2}\left|\int_{-A}^{A} W(t-hu)\frac{\partial}{\partial u}\left[\left\{\frac{g(t-hu)}{g(t)}\right\}^{1/2} - 1\right] K(u)\,du\right|$$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

$$+h^{-1/2}\left|K(A)W(t-hA)\left[\left\{\frac{g(t-Ah)}{g(t)}\right\}^{1/2}-1\right]\right|$$

$$+h^{-1/2}\left|K(-A)W(t+hA)\left[\left\{\frac{g(t+Ah)}{g(t)}\right\}^{1/2}-1\right]\right|$$

$$S_{1,n}(t)+S_{2,n}(t)+S_{3,n}(t), \quad \text{say.}$$

The second term can be estimated by

$$h^{-1/2}\|S_{2,n}\| \leqslant K(A)\cdot \sup_{0\leqslant t\leqslant 1}|W(t-Ah)|\cdot \sup_{0\leqslant t\leqslant 1} h^{-1}\left|\left[\left\{\frac{g(t-Ah)}{g(t)}\right\}^{1/2}-1\right]\right|.$$

By the mean value theorem it follows that

$$h^{-1/2}\|S_{2,n}\| = \mathcal{O}_p(1).$$

The first term $S_{1,n}$ is estimated as

$$h^{-1/2}S_{1,n}(t) = \left|h^{-1}\int_{-A}^{A} W(t-uh)K'(u)\left[\left\{\frac{g(t-uh)}{g(t)}\right\}^{1/2}-1\right]du\right.$$

$$\left.\cdot\frac{1}{2}\int_{-A}^{A} W(t-uh)K(u)\left\{\frac{g(t-uh)}{g(t)}\right\}^{1/2}\left\{\frac{g'(t-uh)}{g(t)}\right\}du\right|$$

$$= |T_{1,n}(t)-T_{2,n}(t)|, \quad \text{say;}$$

$\|T_{2,n}\| \leqslant C_5\cdot\int_{-A}^{A}|W(t-hu)|du = \mathcal{O}_p(1)$ by assumption on $g(t)=\sigma^2(t)\cdot f_X(t)$. To estimate $T_{1,n}$ we again use the mean value theorem to conclude that

$$\sup_{0\leqslant t\leqslant 1} h^{-1}\left|\left\{\frac{g(t-uh)}{g(t)}\right\}^{1/2}-1\right| < C_6\cdot|u|;$$

hence

$$\|T_{1,n}\| \leqslant C_6\cdot\sup_{0\leqslant t\leqslant 1}\int_{-A}^{A}|W(t-hu)|K'(u)u/du = \mathcal{O}_p(1).$$

Because $S_{3,n}(t)$ is estimated as $S_{2,n}(t)$, we finally obtain the desired result. ■

The next lemma shows that the truncation introduced through $\{a_n\}$ does not affect the limiting distribution.

LEMMA A.6. $\|Y_n - Y_{0,n}\| = \mathcal{O}_p\{(\log n)^{-1/2}\}$.

Proof. We shall only show that $g'(t)^{-1/2}h^{-1/2}\iint_{\mathbb{R}-\Gamma_n}\psi\{y-l(t)\}K\{(t-x)/h\}dZ_n(x,y)$ fulfills the lemma. The replacement of $g'(t)$ by $g(t)$ may be proved as in Lemma A.4 of Johnston (1982). The preceding quantity is less than $h^{-1/2}\|g^{-1/2}\|\cdot\|\iint_{\{|y|>a_n\}}\psi\{y-$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

$l(\cdot)\}K\{(\cdot - x)/h\}dZ(x, y)\|$. It remains to be shown that the last factor tends to zero at a rate $\mathcal{O}_p\{(\log n)^{-1/2}\}$. We show first that

$$V_n(t) = (\log n)^{1/2} h^{-1/2} \iint_{\{|y|>a_n\}} \psi\{y - l(t)\}K\{(t-x)/h\}dZ_n(x, y)$$

$$\xrightarrow{p} 0 \quad \text{for all } t,$$

and then we show tightness of $V_n(t)$. The result then follows:

$$V_n(t) = (\log n)^{1/2}(nh)^{-1/2} \sum_{i=1}^{n} [\psi\{Y_i - l(t)\}\mathbf{1}(|Y_i| > a_n)K\{(t - X_i)/h\}$$

$$- \mathsf{E}\,\psi\{Y_i - l(t)\}\mathbf{1}(|Y_i| > a_n)K\{(t - X_i)/h\}]$$

$$= \sum_{i=1}^{n} X_{n,t}(t),$$

where $\{X_{n,t}(t)\}_{i=1}^{n}$ are i.i.d. for each n with $\mathsf{E}\,X_{n,t}(t) = 0$ for all $t \in [0, 1]$. We then have

$$\mathsf{E}\,X_{n,t}^2(t) \leqslant (\log n)(nh)^{-1}\,\mathsf{E}\,\psi^2\{Y_i - l(t)\}\mathbf{1}(|Y_i| > a_n)K^2\{(t - X_i)/h\}$$

$$\leqslant \sup_{-A\leqslant u\leqslant A} K^2(u) \cdot (\log n)(nh)^{-1}\,\mathsf{E}\,\psi^2\{Y_i - l(t)\}\mathbf{1}(|Y_i| > a_n).$$

Hence

$$\mathrm{Var}\{V_n(t)\} = \mathsf{E}\left\{\sum_{i=1}^{n} X_{n,t}(t)\right\}^2 = n \cdot \mathsf{E}\,X_{n,t}^2(t)$$

$$\leqslant \sup_{-A\leqslant u\leqslant A} K^2(u)h^{-1}(\log n) \int_{\{|y|>a_n\}} f_y(y)\,dy \cdot M_\psi,$$

where M_ψ denotes an upper bound for ψ^2. This term tends to zero by Assumption (A3). Thus by Markov's inequality we conclude that

$$V_n(t) \xrightarrow{p} 0 \quad \text{for all } t \in [0, 1].$$

To prove tightness of $\{V_n(t)\}$ we refer again to the following moment condition as stated in Lemma A.1:

$$\mathsf{E}\{|V_n(t) - V_n(t_1)| \cdot |V_n(t_2) - V_n(t)|\} \leqslant C' \cdot (t_2 - t_1)^2$$

C' denoting a constant, $\quad t \in [t_1, t_2]$.

We again estimate the left-hand side by Schwarz's inequality and estimate each factor separately:

$$\mathsf{E}\{V_n(t) - V_n(t_1)\}^2 = (\log n)(nh)^{-1}\,\mathsf{E}\left[\sum_{i=1}^{n} \Psi_n(t, t_1, X_i, Y_i) \cdot \mathbf{1}(|Y_i| > a_n)\right.$$

$$\left. - \mathsf{E}\{\Psi_n(t, t_1, X_i, Y_i) \cdot \mathbf{1}(|Y_i| > a_n)\}\right]^2,$$

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press

where $\Psi_n(t, t_1, X_i, Y_i) = \psi\{Y_i - l(t)\}K\{(t - X_i)/h\} - \psi\{Y_i - l(t_1)\}K\{(t_1 - X_1)/h\}$. Because ψ, K are Lipschitz continuous except at one point and the expectation is taken afterward, it follows that

$$[\mathsf{E}\{V_n(t) - V_n(t_1)\}^2]^{1/2}$$

$$\leqslant C_7 \cdot (\log n)^{1/2} h^{-3/2} |t - t_1| \cdot \left\{ \int_{\{|y| > a_n\}} f_y(y)\, dy \right\}^{1/2}.$$

If we apply the same estimation to $V_n(t_2) - V_n(t_1)$ we finally have

$$\mathsf{E}\{|V_n(t) - V_n(t_1)| \cdot |V_n(t_2) - V_n(t)|\}$$

$$\leqslant C_7^2 (\log n) h^{-3} |t - t_1||t_2 - t| \times \int_{\{|y| > a_n\}} f_y(y)\, dy$$

$$\leqslant C' \cdot |t_2 - t_1|^2 \quad \text{because } t \in [t_1, t_2] \quad \text{by Assumption (A3).}$$ ■

LEMMA A.7. *Let $\lambda(K) = \int K^2(u)\, du$ and let $\{d_n\}$ be as in Theorem 2.2. Then*

$$(2\delta \log n)^{1/2} [\|Y_{3,n}\| / \{\lambda(K)\}^{1/2} - d_n]$$

has the same asymptotic distribution as

$$(2\delta \log n)^{1/2} [\|Y_{4,n}\| / \{\lambda(K)\}^{1/2} - d_n].$$

Proof. $Y_{3,n}(t)$ is a Gaussian process with

$$\mathsf{E}\, Y_{3,n}(t) = 0$$

and covariance function

$$r_3(t_1, t_2) = \mathsf{E}\, Y_{3,n}(t_1) Y_{3,n}(t_2)$$

$$= \{g(t_1) g(t_2)\}^{-1/2} h^{-1} \iint_{\Gamma_n} \psi^2\{y - l(x)\} K\{(t_1 - x)/h\}$$

$$\times K\{(t_2 - x)/h\} f(x, y)\, dx\, dy$$

$$= \{g(t_1) g(t_2)\}^{-1/2} h^{-1} \iint_{\Gamma_n} \psi^2\{y - l(x)\} f(y|x)\, dy K\{(t_1 - x)/h\}$$

$$\times K\{(t_2 - x)/h\} f_X(x)\, dx$$

$$= \{g(t_1) g(t_2)\}^{-1/2} h^{-1} \int g(x) K\{(t_1 - x)/h\} K\{(t_2 - x)/h\}\, dx$$

$$= r_4(t_1, t_2),$$

where $r_4(t_1, t_2)$ is the covariance function of the Gaussian process $Y_{4,n}(t)$, which proves the lemma. ■

https://doi.org/10.1017/S0266466609990491 Published online by Cambridge University Press