SFB 649 Discussion Paper 2008-027

The Stochastic Fluctuation of the Quantile Regression Curve

Wolfgang Härdle*
Song Song*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

The Stochastic Fluctuation of the Quantile Regression Curve *

Wolfgang K. Härdle †, Song Song ‡

Abstract

Let (X_1, Y_1), ..., (X_n, Y_n) be i.i.d. rvs and let $l(x)$ be the unknown p-quantile regression curve of Y on X. A quantile-smoother $l_n(x)$ is a localised, 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 allows us to 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 model check, e.g. in econometrics. An application considers a labour market discrimination effect.

Keywords: Quantile Regression; Consistency Rate; Confidence Band; Check Function; Kernel Smoothing; Nonparametric Fitting

JEL classification: C00; C14; J01; J31

1 Introduction

In 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

*The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin is gratefully acknowledged.

†Professor at Humboldt-Universität zu Berlin and Director of CASE - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany.

‡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.

distribution. The quantile curves are key aspects of inference in various statistical problems and are of great interest in practice. These describe the conditional behaviour of a response variable given the value of an explanatory variable, and investigate changes in both tails of the distribution, other than just the mean. 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 since the $p = 0.5$ (median) quantile curve coincides with the mean regression.

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 where $x \in J$, 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. Special treatment of the multivariate case will be indicated when necessary. Let $l(x)$ denote the p-quantile curve, i.e. $l(x) = F^{-1}_{Y|x}(p)$.

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 et al. (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 et al. (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 et al. (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 M-regression function which involves quantile regression as a special case, its uniform Bahadur representation and application to the additive model is studied by Kong et al. (2008).

Under a "check function", $l(x)$ can be viewed as minimiser of $L(\theta) \stackrel{\text{def}}{=} \mathsf{E}\{\rho_p(y - \theta)|X = x\}$ (w.r.t. θ) with $\rho_p(u) = pu\mathbf{1}\{u \in (0, \infty)\} - (1-p)u\mathbf{1}\{u \in (-\infty, 0)\}$. Interestingly there is only one reference to an exercise in Ferguson (1967)[p.51] in the literature, but no detailed proof. It is however easy to see that:

$$\begin{aligned} L(\theta) &= \mathsf{E}\{\rho_p(y-\theta)|X=x\} \\ &= \int_{-\infty}^{+\infty} \{p(y-\theta)\mathbf{1}(y>\theta) + (1-p)(\theta-y)\mathbf{1}(y<\theta)\}f(y|x)dy \\ &= p\int_{\theta}^{+\infty}(y-\theta)f(y|x)dy + (1-p)\int_{-\infty}^{\theta}(\theta-y)f(y|x)dy \\ &= p\int_{-\infty}^{+\infty} yf(y|x)dy - p\theta\int_{-\infty}^{+\infty} f(y|x)dy \\ &\quad -\int_{-\infty}^{\theta} yf(y|x)dy + \theta\int_{-\infty}^{\theta} f(y|x)dy \end{aligned}$$

With differentiation under the integral the FOC (w.r.t. θ) is:

$$\begin{aligned} -p - \theta f(y|x) + \int_{-\infty}^{\theta} f(y|x)dy + \theta f(y|x) &= 0 \\ \int_{-\infty}^{\theta} f(y|x)dy - p &= 0. \end{aligned}$$

The SOC (w.r.t. θ): $f(\theta|x) \geqslant 0$, guarantees that the p-quantile θ minimizes $L(\theta)$ according to the definition of quantile.

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

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

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.

Numerically the minimisation of (1) through the check function is not a trivial task. Consequently Lejeune and Sarda (1988) and Yu et al. (2003) proposed to employ iterative methods to compute $l_n(x)$. This is based on writing (1) as:

$$L_n(\theta) = n^{-1}\sum_{i=1}^{n}(Y_i-\theta)^2\Big\{\frac{\rho_p(Y_i-\theta)}{(Y_i-\theta)^2}\Big\}K_h(x-X_i). \tag{2}$$

Note that if we define $(Y_i;\theta)$ as:

$$\begin{aligned} w_p(Y_i;\theta) &= \frac{\rho_p(Y_i-\theta)}{(Y_i-\theta)^2} = \frac{p}{Y_i-\theta}\mathbf{1}\{(Y_i-\theta)\in(0,\infty)\} \\ &\quad + \frac{p-1}{Y_i-\theta}\mathbf{1}\{(Y_i-\theta)\in(-\infty,0)\}, \end{aligned}$$

and integrate w_p into K_h we can rewrite (2) as a reweighted sum of squares:

$$\begin{aligned} L_n(\theta) &= n^{-1}\sum_{i=1}^{n}(Y_i-\theta)^2 w_p(Y_i;\theta)K_h(x-X_i) \\ &= n^{-1}\sum_{i=1}^{n}(Y_i-\theta)^2 K_p(x;X_i;Y_i;\theta) \end{aligned} \tag{3}$$

An algorithm for finding the minimiser of (2) is given in Yu et al. (2003). This algorithm is an iteratively reweighted least squares procedure for finding l_n. Convergence of this algorithm is guaranteed for any initial value of $l_{n,k}$. At step $k=1$, we simply take $l_{n,1}$ as the global p-quantile. At each iteration from $k \to k+1$, new weight $w_p(Y_i;l_{n,k})$ is defined to compute $K_p(x;X_i;Y_i;l_{n,k})$. A new estimator of l is obtained as:

$$\begin{aligned} l_{n,k+1} &= \operatorname*{argmin}_{\theta}\, n^{-1}\sum_{i=1}^{n}(Y_i-\theta)^2 K_p(x;X_i;Y_i;l_{n,k}) \\ &= \frac{\sum_{i=1}^{n} K_p(x;X_i;Y_i;l_{n,k})Y_i}{\sum_{i=1}^{n} K_p(x;X_i;Y_i;l_{n,k})}. \end{aligned} \tag{4}$$

Equivalently, $l_{n,k}(x)$ can be viewed as a local average of those Y-observations with corresponding X-observations in a reweighed neighbourhood of x. The size and weight of that neighbourhood is regulated by the bandwidth h and p. In Theorem 2.1 we show that $\exists$ some k_1, s.t. $l_{n,k}(x) = l_n(x), \forall k \geqslant k_1$.

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

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

$l_n(x)$ and $l(x)$ can be treated as a zero (w.r.t. θ) of the function:

$$\widetilde{H}_n(\theta,x) \stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^{n}K_h(x-X_i)\psi(Y_i-\theta) \tag{5}$$

$$\widetilde{H}(\theta,x) \stackrel{\text{def}}{=} \int_{\mathbb{R}} f(x,y)\psi(y-\theta)dy \tag{6}$$

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 et al. (1988). It is shown that under general conditions almost surely (a.s.)

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

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

Please note that without assuming K has the compact support (as we do here) under similar assumptions Franke and Mwita (2003) get:

$$\begin{aligned} l_n(x) &= \hat{F}^{-1}_{Y|x}(p) \\ \hat{F}(y|x) &= \frac{\sum_{i=1}^n K_h(x - X_i)\mathbf{1}(Y_i < y)}{\sum_{i=1}^n K_h(x - X_i)} \\ \sup_{x\in J} |l_n(x) - l(x)| &\leqslant B^{**}\{(nh/(s_n \log n))^{-1/2} + h^2\}, \quad \text{as } n \to \infty. \end{aligned}$$

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

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

$$\begin{aligned} &\mathrm{P}\left((2\delta \log n)^{1/2}\left[\sup_{x\in J} r(x)|\{l_n(x) - l(x)\}|\lambda(K)^{1/2} - d_n\right] < z\right) \\ &\longrightarrow \exp\{-2\exp(-z)\}, \quad \text{as } n \to \infty. \end{aligned} \tag{7}$$

where $r(x)$, δ, $\lambda(K)$, d_n are suitable scaling parameters. This result allows the construction of (asymptotic) uniform confidence bands for $l(x)$ based on specifications of the stochastic fluctuation of $l_n(x)$.

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 present through the equivalence of several stochastic processes with a strong uniform consistency rate of $\{l_n(x) - l(x)\}$ also shown, the proof of which are sketched in Section 3. In Section 4, in a small Monte Carlo study we investigate the behaviour of $l_n(x)$ when the data is generated by fat-tailed conditional distributions of $(Y|X = x)$. In Section 5, an application considers a wage-earning relation in labour market.

2 Results

THEOREM 2.1 *For $l_n(x)$ as the minimiser of $L_n(\theta)$, and $l_{n,k}$ defined in (4), $\exists$ some k_1, s.t. $l_{n,k}(x) = l_n(x), \forall k \geqslant k_1$.*

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, symmetric, has compact support $[-A, A]$ and is Lipschitz continuously differentiable with bounded derivatives;
$(A2)$ $(nh)^{-1/2}(\log n)^{3/2} \to 0$, $(n \log n)^{1/2} h^{5/2} \to 0$, $(nh^3)^{-1}(\log n)^2 \leqslant M$, M a constant;
$(A3)$ $h^{-3}(\log n) \int_{|y|>a_n} f_Y(y) dy = \mathcal{O}(1)$, $f_Y(y)$ the marginal density of Y, $\{a_n\}_{n=1}^{\infty}$ a sequence of constants tending to infinity as $n \to \infty$;
$(A4)$ $\inf_{t \in J} |q(t)| \geqslant q_0 > 0$, where $q(t) = \partial \, \mathsf{E}\{\psi(Y - \theta)|t\}/\partial\theta \; \{l(t)\} \cdot f_X(t) = f\{l(t)|t\} f_X(t)$;
$(A5)$ 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 uniformly locally Lipschitz of order $\tilde{\alpha}$ (ulL-$\tilde{\alpha}$) on J, uniformly in $y \in \mathbb{R}$, with $0 < \tilde{\alpha} \leqslant 1$.

Define also

$$
\begin{aligned}
\sigma^2(t) &= \mathsf{E}[\psi^2\{Y - l(t)\}|t] = p(1-p) \\
H_n(t) &= (nh)^{-1} \sum_{i=1}^{n} K\{(t - X_i)/h\} \psi\{Y_i - l(t)\} \\
D_n(t) &= \partial (nh)^{-1} \sum_{i=1}^{n} K\{(t - X_i)/h\} \psi\{Y_i - \theta\}/\partial\theta \; \{l(t)\}
\end{aligned}
$$

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

Assumption $(A1)$ on the compact support of the kernel could possibly be relaxed introducing a cutoff technique as 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 non-linear remainder terms. Assumption $(A3)$ appears in a somewhat modified form also in Johnston (1982). Assumption $(A5, A6)$ are common assumptions in robust estimation as in Huber (1981), Härdle et al. (1988) which exponential, and generalised hyperbolic distributions satisfy.

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 Theorem 2.2 and Remark 2.3(v) there we immediately have the following lemma.

LEMMA 2.1 *Let $\widetilde{H}_n(\theta, x)$ and $\widetilde{H}(\theta, x)$ be given by (5) and (6). Under assumption (A6) and $(nh/\log n)^{-1/2} \to \infty$ through (A2), for some constant A^* not depending on n, we have a.s. as $n \to \infty$*

$$\sup_{\theta \in I} \sup_{x \in J} |\widetilde{H}_n(\theta, x) - \widetilde{H}(\theta, x)| \quad \leq \quad A^* \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{8}$$

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

$$\inf_{x \in J} \Big| \int \psi\{y - l(x) + \varepsilon\} dF(y|x) \Big| \geqslant \tilde{q}|\varepsilon|, \quad \text{for } |\varepsilon| \leqslant \delta_1, \tag{9}$$

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 2.2 *Under the conditions of Lemma 2.1 and also assuming (9), we have a.s. as $n \to \infty$*

$$\sup_{x \in J} |l_n(x) - l(x)| \quad \leq \quad B^* \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{10}$$

with $B^* = A^*/m_1\tilde{q}$ not depending on n and m_1 a lower bound of $f_X(t)$.

THEOREM 2.3 *Let $h = n^{-\delta}$, $\frac{1}{5} < \delta < \frac{1}{3}$, $\lambda(K) = \int_{-A}^{A} K^2(u)du$ and*

$$\begin{aligned} d_n &= (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2}[\log\{c_1(K)/\pi^{1/2}\} + \frac{1}{2}\{\log \delta + \log\log n\}], \\ & \quad \textit{if } c_1(K) = \{K^2(A) + K^2(-A)\}/\{2\lambda(K)\} > 0 \\ d_n &= (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} \log\{c_2(K)/2\pi\} \\ & \quad \textit{otherwise with } c_2(K) = \int_{-A}^{A} \{K'(u)\}^2 du/\{2\lambda(K)\}. \end{aligned}$$

Then (7) 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 the theorem above, an approximate* $(1-\alpha)\times 100\%$ *confidence band over* $[0,1]$ *is*

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

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 NW methods, etc. has been shown:

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

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

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

Through minimising the approximation of AMSE (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. Select optimal bandwidth h_{mean} from conditional mean regression
2. $h_p = [p(1-p)/\varphi^2\{\Phi^{-1}(p)\}]^{1/5}\cdot h_{\mathrm{mean}}$
   with φ, Φ as the pdf and cdf of a standard normal distribution

Obviously more p deviates from 0.5, means more smoothing is necessary.

The proof is essentially based on a linearisation argument after a Taylor series expansion. The leading linear term will then be approximated in a similar way as in Johnston (1982), 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).

As $l_n(t)$ is the zero (w.r.t. θ) of $\widetilde{H}_n(\theta, t)$, it follows by applying 2nd-order Taylor expansions to $\widetilde{H}_n(\theta, t)$ around $l(t)$ that

$$l_n(t) - l(t) = \{H_n(t) - \mathsf{E}\,H_n(t)\}/q(t) + R_n(t) \tag{11}$$

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

$$\begin{aligned} R_n(t) \;=\;& H_n(t)\{q(t) - D_n(t)\}/\{D_n(t)\cdot q(t)\} + \mathsf{E}\,H_n(t)/q(t) \\ &+\frac{1}{2}\{l_n(t) - l(t)\}^2\cdot\{D_n(t)\}^{-1} & (12)\\ &\cdot(nh)^{-1}\sum_{i=1}^{n} K\{(x - X_i)/h\}\psi''\{Y_i - l(t) + r_n(t)\}, & (13) \end{aligned}$$

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

is the remainder term. In Section 3 it is shown (Lemma 3.1) that $\|R_n\| = \sup_{t\in J} |R_n(t)| = \mathcal{o}_p\{(nh\log n)^{-1/2}\}$.

Furthermore, the rescaled linear part

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

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

$$Y_{5,n}(t) = h^{-1/2}\int K\{(t-x)/h\}dW(x). \tag{14}$$

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 rv's. In the event that x is a d-dimension covariate, the transformation becomes:

$$\begin{aligned} T(x_1, x_2, \ldots, x_d, y) &= \{F_{X_1|y}(x_1|y), F_{X_2|y}(x_2|x_1,y), \ldots, \\ &\quad F_{X_k|x_{d-1},\ldots,x_1,y}(x_k|x_{d-1},\ldots,x_1,y), F_Y(y)\}. \end{aligned} \tag{15}$$

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 that*

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

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 which deserves further research, while the current best result is given in Rio (1996).

Before we define the different approximating processes let us first rewrite (14) as a stochastic integral w.r.t. 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:

$$Y_{0,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\}\psi\{y - l(t)\} dZ_n(x,y) \tag{16}$$

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

$$Y_{1,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\}\psi\{y - l(t)\} dB_n\{T(x,y)\} \tag{17}$$

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

$$Y_{2,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\}\psi\{y - l(t)\} dW_n(T(x,y)) \tag{18}$$

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

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

$$Y_{3,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\}\psi\{y - l(x)\} dW_n(T(x,y)) \tag{19}$$

$$Y_{4,n}(t) = \{hg(t)\}^{-1/2} \int g(x)^{1/2} K\{(t-x)/h\} dW(x) \tag{20}$$

$$Y_{5,n}(t) = h^{-1/2} \int K\{(t-x)/h\} dW(x) \tag{21}$$

$\{W(\cdot)\}$ being the Wiener process on $(-\infty, \infty)$.

Lemma 3.2 to 3.7 ensure that all these processes have the same limit distributions. The result then follows from

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

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

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

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

3 Proof

Proof of Theorem 2.1. For $l_{n,k}$ and l_n there are only 2 cases:

1. $\nexists$ some k_1, s.t. $l_{n,k_1} = l_n$
2. $\exists$ some k_1, s.t. $l_{n,k_1} = l_n$

W.L.G. we assume $l_n \in (Y_m, Y_{m+1})$ for some m. According to the convergence of $l_{n,k} \to l_n$ shown in Lejeune and Sarda (1988), $\exists$ some k_0, s.t. $|l_{n,k} - l_n| \leqslant \min\{(l_n - y_m), (y_{m+1} - l_n)\}, \forall k \geqslant k_0$. Using same transformation proposed by Lejeune and Sarda (1988), K_h in (4) can be seen as not relevant and discarded here. For simplicity of notation, if we write $p\mathbf{1}\{(Y_i - l_{n,k}) \in (0, \infty)\} - (1 - p)\mathbf{1}\{(Y_i - l_{n,k}) \in (-\infty, 0)\}$ as $\psi_p(Y_i - l_{n,k})$, therefore

$$\begin{aligned} l_{n,k+1} &= \frac{\sum_{i=1}^n \psi_p(Y_i - l_{n,k})/(Y_i - l_{n,k}) \cdot (Y_i - l_{n,k} + l_{n,k})}{\sum_{i=1}^n \psi_p(Y_i - l_{n,k})/(Y_i - l_{n,k})} \\ &= \frac{\sum_{i=1}^n \psi_p(Y_i - l_{n,k})}{\sum_{i=1}^n \psi_p(Y_i - l_{n,k})/(Y_i - l_{n,k})} + l_{n,k}. \end{aligned}$$

As $|l_{n,k} - l_n| \leqslant \min\{(l_n - y_m), (y_{m+1} - l_n)\}, \forall k \geqslant k_0$, $\sum_{i=1}^n p\mathbf{1}\{(Y_i - l_{n,k}) \in (0, \infty)\} - (1-p)\mathbf{1}\{(Y_i - l_{n,k}) \in (-\infty, 0)\} = pn(1-p) + (p-1)np = 0, \forall\, l_{n,k}$ with $k \geqslant k_0$.

If case 1 holds, it means $l_{n,k} \neq l_n, \forall k \geqslant k_0$. Just pick anyone of them, namely $l_{n,k_0^\star} = l_n + \varepsilon \neq 0$, with $\varepsilon \neq 0, k_0^\star \geqslant k_0$. Thus:

$$l_{n,k_0^\star+1} = \frac{0}{\sum_{i=1}^n \psi_p(Y_i - l_{n,k_0^\star})/(Y_i - l_{n,k_0^\star})} + l_{n,k_0^\star} = l_{n,k_0^\star}.$$

So $l_{n,k} = l_{n,k_0^\star} = l_n + \varepsilon, \forall k \geqslant k_0^\star$, however, this contradicts to $l_{n,k} \to l_n$ as $\varepsilon \neq 0$. So $\exists$ some k_1, s.t. $l_{n,k_1} = l_n$. Similarly we get:

$$l_{n,k_1+1} = \frac{0}{\sum_{i=1}^n \psi_p(Y_i - l_{n,k_1})/(Y_i - l_{n,k_1})} + l_{n,k_1} = l_{n,k_1}.$$

So $l_{n,k} = l_{n,k_1} = l_n, \forall k \geqslant k_1$. $\square$

Proof of Theorem 2.2. By the definition of $l_n(x)$ as a zero of (5), we have, for $\varepsilon > 0$,

$$\text{if } l_n(x) > l(x) + \varepsilon, \text{and then } \widetilde{H}_n\{l(x) + \varepsilon, x\} > 0. \tag{22}$$

Now

$$\widetilde{H}_n\{l(x) + \varepsilon, x\} \leqslant \widetilde{H}\{l(x) + \varepsilon, x\} + \sup_{\theta \in I} |\widetilde{H}_n(\theta, x) - \widetilde{H}(\theta, x)|. \tag{23}$$

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 (9), for $0 < \varepsilon < \delta_1$. That is, for $0 < \varepsilon < \delta_1$,

$$\widetilde{H}\{l(x)+\varepsilon, x\} \leqslant -m_1\tilde{q}\varepsilon. \tag{24}$$

Combining (22), (23) and (24), we have, for $0 < \varepsilon < \delta_1$:

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

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

$$\text{if } \sup_{x\in J} |l_n(x) - l(x)| > \varepsilon, \text{and then } \sup_{\theta\in I} |\widetilde{H}_n(\theta, x) - \widetilde{H}(\theta, x)| > m_1\tilde{q}\varepsilon. \tag{25}$$

It readily follows that (25), and (8) imply (10). □

Below 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 3.1 *For the remainder term $R_n(t)$ defined in (12) we have*

$$\|R_n\| = \mathcal{o}_p\{(nh\log n)^{-1/2}\}. \tag{26}$$

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

$$\begin{aligned} \|R_n\| &\leqslant \left[\inf_{0\leqslant t\leqslant 1}\{|D_n(t)|\cdot q(t)\}\right]^{-1}\{\|H_n\|\cdot\|q-D_n\| + \|D_n\|\cdot\|\mathsf{E}\, H_n\|\} \\ &\quad + C_1\cdot\|l_n - l\|^2\cdot\left\{\inf_{0\leqslant t\leqslant 1}|D_n(t)|\right\}^{-1}\cdot\|f_n\|_\infty, \end{aligned}$$

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

The desired result (3.1) will then follow if we prove

$$\begin{aligned} \|H_n\| &= \mathcal{o}_p\{(nh)^{-1/2}(\log n)^{1/2}\} && (27)\\ \|q-D_n\| &= \mathcal{o}_p\{(nh)^{-1/4}(\log n)^{-1/2}\} && (28)\\ \|\mathsf{E}\, H_n\| &= \mathcal{O}(h^2) && (29)\\ \|l_n - l\|^2 &= \mathcal{o}_p\{(nh)^{-1/2}(\log n)^{-1/2}\} && (30) \end{aligned}$$

Since (29) 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 $\| \mathsf{E}\, H_n\| = o_p\{(nh)^{-1/2}(\log n)^{-1/2}\}$.

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

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

and the following inequality

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

Statement (27) thus is obtained.

Statement (28) follows in the same way as (27) 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 (30).

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

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

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

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

LEMMA 3.2

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

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

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

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

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

□

LEMMA 3.3 $\|Y_{1,n} - Y_{2,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof. Note that the Jacobi of $T(x,y)$ is $f(x,y)$ hence

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

It follows that

$$\begin{aligned}
h^{-1/2}\|Y_{1,n} - Y_{2,n}\| &\leqslant |W_n(1,1)|\cdot\|g^{-1/2}\| \\
&\quad \cdot \sup_{0\leqslant t\leqslant 1} h^{-1}\iint_{\Gamma_n}|\psi\{y-l(t)\}K\{(t-x)/h\}|f(x,y)dxdy.
\end{aligned}$$

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

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

$\square$

LEMMA 3.4 $\|Y_{2,n} - Y_{3,n}\| = \mathcal{O}_p(h^{1/2})$.

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 are uniformly continuous this is smaller than

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

and the lemma thus follows. $\square$

LEMMA 3.5 $\|Y_{4,n} - Y_{5,n}\| = \mathcal{O}_p(h^{1/2})$.

Proof.

$$\begin{aligned}
|Y_{4,n}(t) - Y_{5,n}(t)| &= h^{-1/2}\left|\int \left[\left\{\frac{g(x)}{g(t)}\right\}^{1/2} - 1\right] K\{(t-x)/h\}dW(x)\right| \\
&\leqslant h^{-1/2}\left|\int_{-A}^{A} W(t-hu)\frac{\partial}{\partial u}\left[\left\{\frac{g(t-hu)}{g(t)}\right\}^{1/2} - 1\right]K(u)du\right| \\
&\quad + h^{-1/2}\left|K(A)W(t-hA)\left[\left\{\frac{g(t-Ah)}{g(t)}\right\}^{1/2} - 1\right]\right| \\
&\quad + h^{-1/2}\left|K(-A)W(t+hA)\left[\left\{\frac{g(t+Ah)}{g(t)}\right\}^{1/2} - 1\right]\right| \\
&\qquad S_{1,n}(t) + S_{2,n}(t) + S_{3,n}(t), \quad \textit{say}.
\end{aligned}$$

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

$$\begin{aligned} h^{-1/2}S_{1,n}(t) &= \Big|h^{-1}\int_{-A}^{A} W(t-uh)K'(u)\Big[\Big\{\frac{g(t-uh)}{g(t)}\Big\}^{1/2}-1\Big]du \\ &\quad \frac{1}{2}\int_{-A}^{A} W(t-uh)K(u)\Big\{\frac{g(t-uh)}{g(t)}\Big\}^{1/2}\Big\{\frac{g'(t-uh)}{g(t)}\Big\}du\Big| \\ &= |T_{1,n}(t)-T_{2,n}(t)|, \quad say; \end{aligned}$$

$\|T_{2,n}\| \leqslant C_5 \cdot \int_{-A}^{A}|W(t-hu)|du = \mathcal{O}_p(1)$ 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

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

hence

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

Since $S_{3,n}(t)$ 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 3.6 $\|Y_n - Y_{0,n}\| = \mathcal{O}_p\{(\log n)^{-1/2}\}$.

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' by $g(t)$ may be proved as in Johnston (1982). The quantity above is less than $h^{-1/2}\|g^{-1/2}\|\cdot\|\iint_{\{|y|>a_n\}}\psi\{y-l(\cdot)\}K\{(\cdot-x)/h\}dZ(x,y)\|$. 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

$$\begin{aligned} V_n(t) &= (\log n)^{1/2}h^{-1/2}\iint_{\{|y|>a_n\}}\psi\{y-l(t)\}K\{(t-x)/h\}dZ_n(x,y) \\ &\overset{p}{\to} 0 \quad \text{for all} \quad t \end{aligned}$$

and then we show tightness of $V_n(t)$, the result then follows:

$$\begin{aligned} V_n(t) &= (\log n)^{1/2}(nh)^{-1/2}\sum_{i=1}^{n}[\psi\{Y_i-l(t)\}\mathbf{1}(|Y_i|>a_n)K\{(t-X_i)/h\} \\ &\quad -\mathsf{E}\,\psi\{Y_i-l(t)\}\mathbf{1}(|Y_i|>a_n)K\{(t-X_i)/h\}] \\ &= \sum_{i=1}^{n}X_{n,t}(t), \end{aligned}$$

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

$$\begin{aligned} \mathsf{E}\, X_{n,t}^2(t) &\leqslant (\log n)(nh)^{-1}\, \mathsf{E}\, \psi^2\{Y_i - l(t)\}\mathbf{1}(|Y_i| > a_n) K^2\{(t - X_i)/h\} \\ &\leqslant \sup_{-A \leqslant u \leqslant A} K^2(u) \cdot (\log n)(nh)^{-1}\, \mathsf{E}\, \psi^2\{Y_i - l(t)\}\mathbf{1}(|Y_i| > a_n); \end{aligned}$$

hence

$$\begin{aligned} \mathrm{Var}\{V_n(t)\} &= \mathsf{E}\Big\{\sum_{i=1}^n X_{n,t}(t)\Big\}^2 = n \cdot \mathsf{E}\, X_{n,t}^2(t) \\ &\leqslant \sup_{-A \leqslant u \leqslant A} K^2(u) h^{-1} (\log n) \int_{\{|y| > a_n\}} f_y(y) dy \cdot M_\psi. \end{aligned}$$

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 3.1:

$$\begin{gathered} \mathsf{E}\{|V_n(t) - V_n(t_1)| \cdot |V_n(t_2) - V_n(t)|\} \leqslant C' \cdot (t_2 - t_1)^2 \\ C' \text{ denoting a constant}, \quad t \in [t_1, t_2]. \end{gathered}$$

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

$$\begin{aligned} \mathsf{E}\{V_n(t) - V_n(t_1)\}^2 &= (\log n)(nh)^{-1}\, \mathsf{E}\Big[\sum_{i=1}^n \Psi_n(t, t_1, X_i, Y_i) \cdot \mathbf{1}(|Y_i| > a_n) \\ &\qquad - \mathsf{E}\{\Psi_n(t, t_1, X_i, Y_i) \cdot \mathbf{1}(|Y_i| > a_n)\}\Big]^2, \end{aligned}$$

where $\Psi_n(t, t_1, X_i, Y_i) = \psi\{Y_i - l(t)\}K\{(t - X_i)/h\} - \psi\{Y_i - l(t_1)\}K\{(t_1 - X_1)/h\}$. Since m, K are Lipschitz continuous, it follows that

$$\begin{aligned} &[\mathsf{E}\{V_n(t) - V_n(t_1)\}^2]^{1/2} \\ &\leqslant C_7 \cdot (\log n)^{1/2} h^{-3/2} |t - t_1| \cdot \Big\{\int_{\{|y| > a_n\}} f_y(y) dy\Big\}^{1/2}. \end{aligned}$$

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

$$\begin{aligned} &\mathsf{E}\{|V_n(t) - V_n(t_1)| \cdot |V_n(t_2) - V_n(t)|\} \\ &\leqslant C_7^2 (\log n) h^{-3} |t - t_1||t_2 - t| \times \int_{\{|y| > a_n\}} f_y(y) dy \\ &\leqslant C' \cdot |t_2 - t_1|^2 \text{ since } t \in [t_1, t_2] \quad \text{by (A3)}. \end{aligned}$$

□

LEMMA 3.7 *Let* $\lambda(K) = \int K^2(u)du$ *and let* $\{d_n\}$ *be as in the theorem. 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

$$\begin{aligned}
r_3(t_1, t_2) &= \mathsf{E}\, Y_{3,n}(t_1)Y_{3,n}(t_2) \\
&= \{g(t_1)g(t_2)\}^{-1/2}h^{-1} \iint_{\Gamma_n} \psi^2\{y - l(t)\}K\{(t_1 - x)/h\} \\
&\quad \times K\{(t_2 - x)/h\}f(x, y)dxdy \\
&= \{g(t_1)g(t_2)\}^{-1/2}h^{-1} \iint_{\Gamma_n} \psi^2\{y - l(t)\}f(y|x)dyK\{(t_1 - x)/h\} \\
&\quad \times K\{(t_2 - x)/h\}f_X(x)dx \\
&= \{g(t_1)g(t_2)\}^{-1/2}h^{-1} \int g(x)K\{(t_1 - x)/h\}K\{(t_2 - x)/h\}dx \\
&= r_4(t_1, t_2)
\end{aligned}$$

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

4 A Monte Carlo Study

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

$$\begin{aligned}
f(x, y) &= g(y - \sqrt{x + 2.5})\mathbf{1}(x \in [-2.5, 2.5]) \qquad (31) \\
g(u) &= \frac{9}{10}\varphi(u) + \frac{1}{90}\varphi(u/9).
\end{aligned}$$

The p-quantile curve $l(x)$ can be obtained from a zero (w.r.t. θ) of:

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

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

$$\begin{aligned} K(u) &= \frac{15}{16}(1-u^2)^2, \quad |u| \leqslant 1, \\ &= 0, \qquad\qquad\qquad |u| > 1. \end{aligned}$$

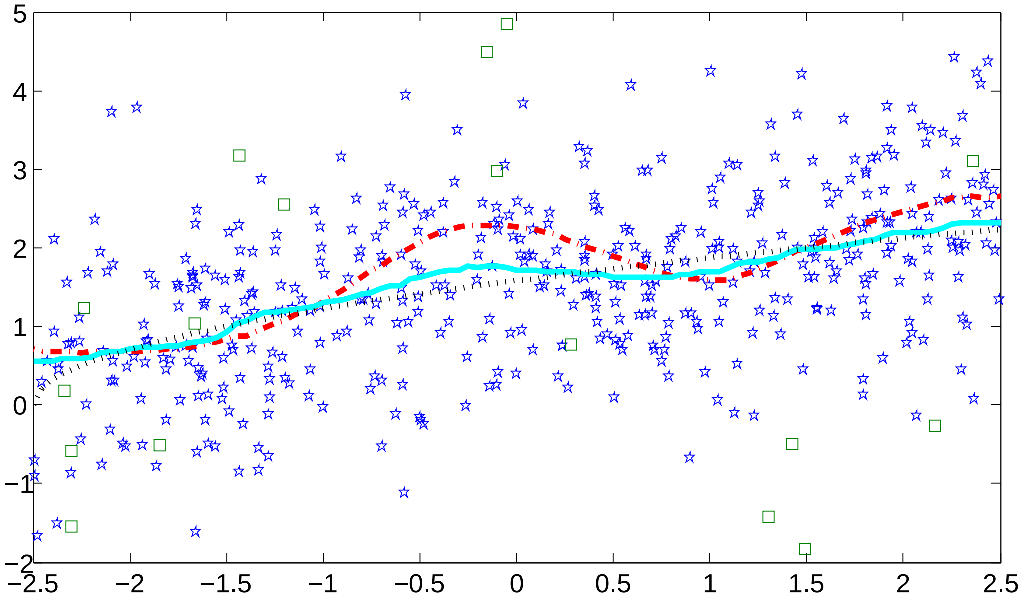

Figure 1: The 0.5-quantile curve, the Nadaraya-Watson estimator $m_n^*(x)$, and the 0.5-quantile smoother $l_n(x)$.

In Fig. 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 (31), are marked as squares whereas the standard normal rv's are shown as stars. We then compute both Nadaraya-Watson estimator $m_n^*(x)$ and 0.5-quantile smoother $l_n(x)$ using an iterative method as in (4). The iteration step and bandwidth are set to 150 and 1.25 which is equivalent to 0.25 after rescalling x to $[0, 1]$ and fullfills the requirement of Theorem 2.3. Fig. 2 displays the simulation result for $l_{n,k}(x)$ with different starting values and different x values. $l_{n,k}(-1)$ are shown as solid and dashed lines, while $l_{n,k}(1)$ are shown as dashed-dotted and dotted lines. As we can see, convergence of $l_{n,k}(x)$ is guaranteed no matter that the initial value is 5 or -5. After around 35 steps, $l_{n,k}(x)$ arrives at $l_n(x)$ and never leaves again; this coincides with the result of Theorem 2.1.

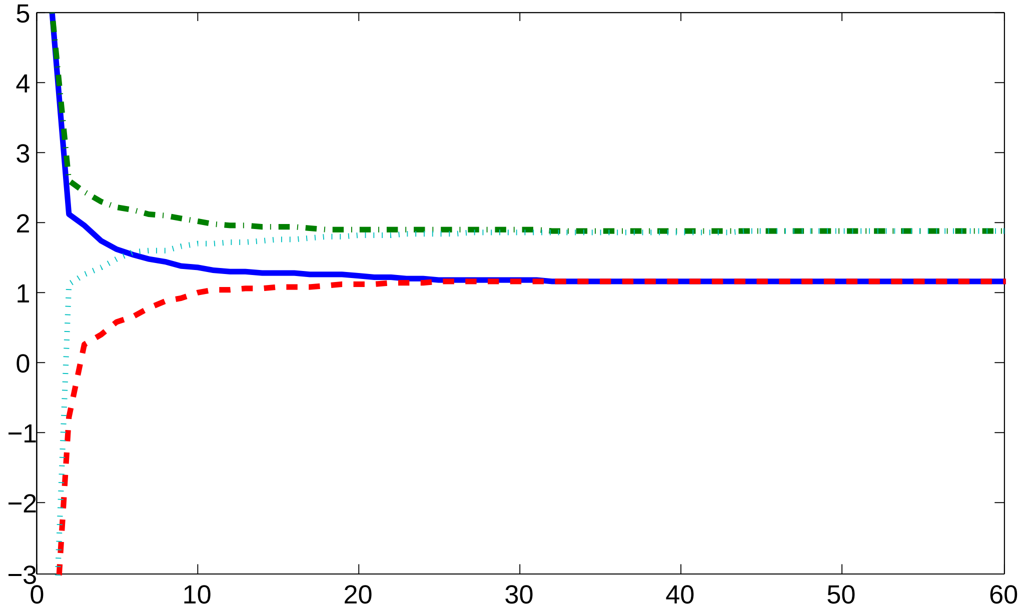

Figure 2: Convergence of $l_{n,k}(x)$ to 0.5-quantile smoother $l_n(x)$ with starting value 5, -5 and x value -1, 1.

In Fig. 1 the $l(x)$, $m_n^*(x)$ 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 fat-tails of $f(x, y)$. A closer look reveals that $m_n^*(x)$ for $x \approx 0$ even leaves apparently from 0.5-quantile curve. It may be surprising that this happens at $x \approx 0$ where no outliers is placed, but a closer look at Fig. 1 shows that the large negative data values at both $x \approx -0.1$ and $x \approx 0.25$ causes 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 Fig. 3 corresponding to 0.9 quantile ($h = 1.25$) with the 95% confidence band.

5 Application

In labour markets economists are concerned with whether discrimination exists, for example for different genders, nationalities, union status and so on. To study this, we need to separate out other effects first, e.g. age, education, etc. Recently there has been great interest in finding out how the financial returns of a job depends on the age of the employee. We use the

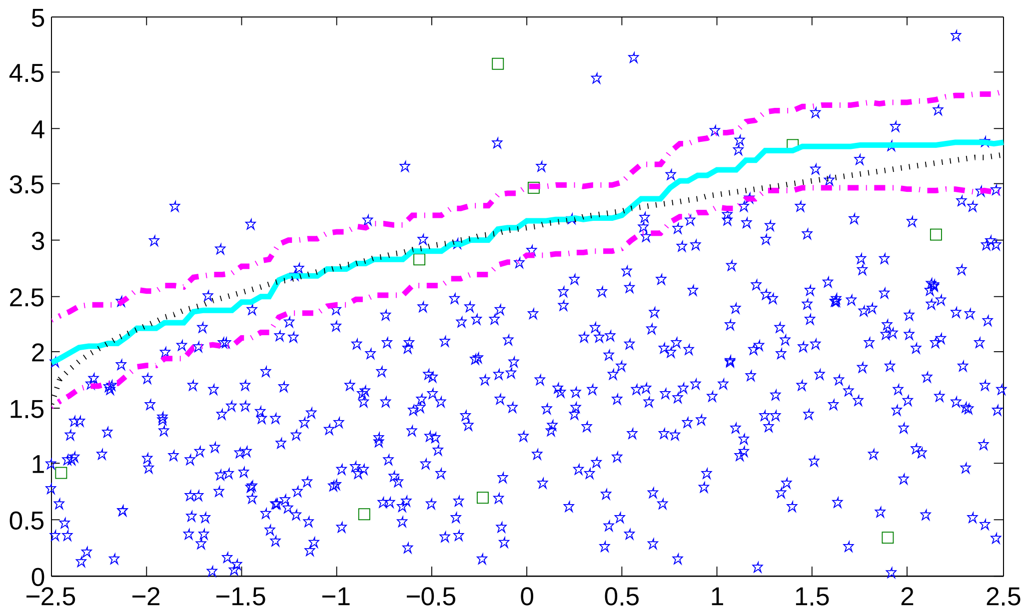

Figure 3: The 0.9-quantile curve, the 0.9-quantile smoother and 95% confidence band.

Current Population Survey (CPS) data from 2005 for the following group: male, $25 \sim 59$ for age, full-time employed, and college graduate containing $16,731$ observations, for the age-earning estimation. As is usual for wage data, log transformation to hourly real wages (unit: US dollar) is carried out first. In CPS all ages ($25 \sim 59$) are reported as integers, we rescale them into $[0,1]$ through dividing 40 with bandwidth 0.059 for nonparametric quantile-smoothers. This is equivalent to set bandwidth 2 for the original age data.

In Fig. 4 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. If we use several popular parametric methods to estimate the 0.5 and 0.9 conditional quantiles, e.g. quadratic, quartic and set of dummies (for ages groups) models as in Fig. 5, with help of the 95% uniform confidence bands, we can do the parametric model specification test. At the 5% significance level, we could not reject any model. However, at the 10% significance level, when the uniform confidence bands get narrower, the set of dummies (for age groups) model is rejected while the other two could not. As the quadratic model performs quite similar to the quartic one, for simplicity, it is suggested in practice for measuring the log(wage)-earing relation.

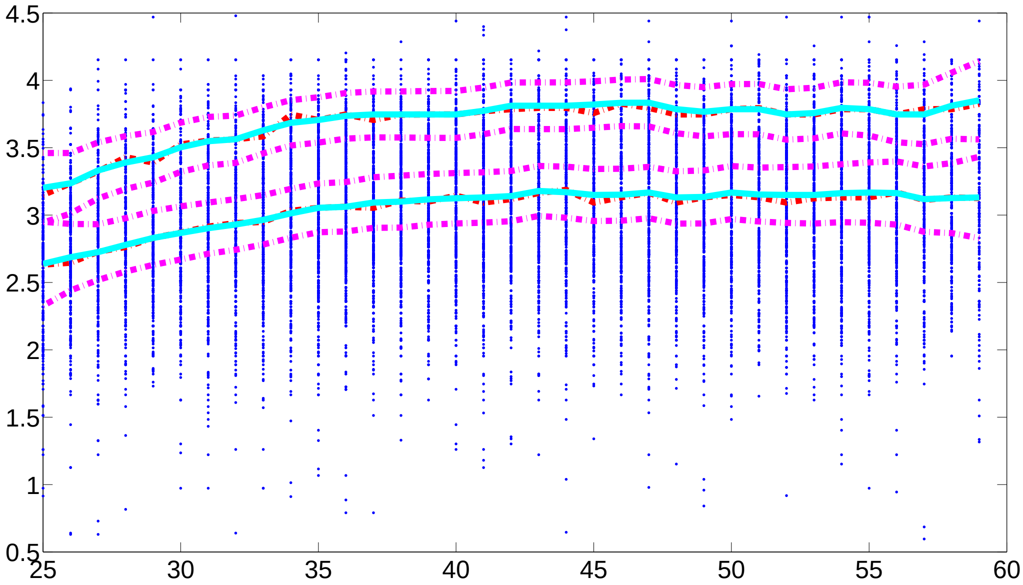

Figure 4: The original observations, local quantiles, 0.5, 0.9-quantile smoothers and corresponding 95% confidence bands.

References

Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviation of density function estimations. *Ann. Statist.*, 1:1071–1095.

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

Csörgö, S. and Hall, P. (1982). Upper and lower classes for triangular arrays. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete*, 61:207–222.

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

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

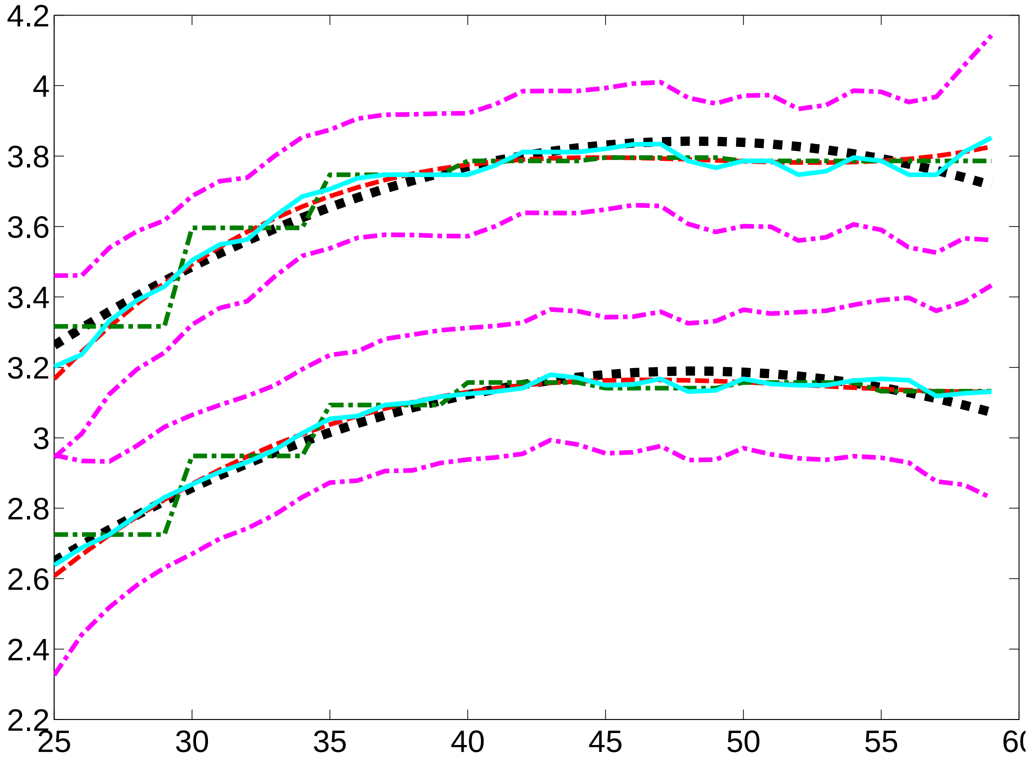

Figure 5: Quadratic, quartic, set of dummies (for age groups) estimates, 0.5, 0.9-quantile smoothers and their corresponding 95% confidence bands.

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

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

Hall, P., Wolff, R., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.*, 94:154–163.

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

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 cnsistency of a class of regression function estimators. *Ann. Statist.*, 12:612–623.

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

Jeong, K. and Härdle, W. (2008). A consistent nonparametric test for causality in quantile. *SFB* 649 *Discussion Paper*.

Johnston, G. (1982). Probabilities of maximal deviations of nonparametric regression function estimation. *J. Multivariate Anal.*, 12:402–414.

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

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

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.*

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

Parzen, M. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 32:1065–1076.

Portnoy, S. and Koenker, R. (1997). The gaussian hare and the laplacian tortoise: computability of squared-error versus absolute -error estimatiors (with discussion). *Statistical Sciences*, 12:279–300.

Rio, E. (1996). Vitesses de convergence dans le principe d'invariance faible pour la fonction de répartition empirique multivariée (rates of convergence in the invariance principle for the multivariate empirical distribution function). *C. r. Acad. sci., Sér. 1, Math.*, 322(2):169–172.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.*, 23:470–472.

Tusnady, G. (1977). A remark on the approximation of the sample distribution function in the multidimensional case. *Period. Math. Hungar.*, 8:53–55.

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

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

Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: applications and current research areas. *J. Roy. Statistical Society : Series D (The Statistician)*, 52:331–350.

SFB 649 Discussion Paper Series 2008

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

001 "Testing Monotonicity of Pricing Kernels" by Yuri Golubev, Wolfgang Härdle and Roman Timonfeev, January 2008.
002 "Adaptive pointwise estimation in time-inhomogeneous time-series models" by Pavel Cizek, Wolfgang Härdle and Vladimir Spokoiny, January 2008.
003 "The Bayesian Additive Classification Tree Applied to Credit Risk Modelling" by Junni L. Zhang and Wolfgang Härdle, January 2008.
004 "Independent Component Analysis Via Copula Techniques" by Ray-Bing Chen, Meihui Guo, Wolfgang Härdle and Shih-Feng Huang, January 2008.
005 "The Default Risk of Firms Examined with Smooth Support Vector Machines" by Wolfgang Härdle, Yuh-Jye Lee, Dorothea Schäfer and Yi-Ren Yeh, January 2008.
006 "Value-at-Risk and Expected Shortfall when there is long range dependence" by Wolfgang Härdle and Julius Mungo, Januray 2008.
007 "A Consistent Nonparametric Test for Causality in Quantile" by Kiho Jeong and Wolfgang Härdle, January 2008.
008 "Do Legal Standards Affect Ethical Concerns of Consumers?" by Dirk Engelmann and Dorothea Kübler, January 2008.
009 "Recursive Portfolio Selection with Decision Trees" by Anton Andriyashin, Wolfgang Härdle and Roman Timofeev, January 2008.
010 "Do Public Banks have a Competitive Advantage?" by Astrid Matthey, January 2008.
011 "Don't aim too high: the potential costs of high aspirations" by Astrid Matthey and Nadja Dwenger, January 2008.
012 "Visualizing exploratory factor analysis models" by Sigbert Klinke and Cornelia Wagner, January 2008.
013 "House Prices and Replacement Cost: A Micro-Level Analysis" by Rainer Schulz and Axel Werwatz, January 2008.
014 "Support Vector Regression Based GARCH Model with Application to Forecasting Volatility of Financial Returns" by Shiyi Chen, Kiho Jeong and Wolfgang Härdle, January 2008.
015 "Structural Constant Conditional Correlation" by Enzo Weber, January 2008.
016 "Estimating Investment Equations in Imperfect Capital Markets" by Silke Hüttel, Oliver Mußhoff, Martin Odening and Nataliya Zinych, January 2008.
017 "Adaptive Forecasting of the EURIBOR Swap Term Structure" by Oliver Blaskowitz and Helmut Herwatz, January 2008.
018 "Solving, Estimating and Selecting Nonlinear Dynamic Models without the Curse of Dimensionality" by Viktor Winschel and Markus Krätzig, February 2008.
019 "The Accuracy of Long-term Real Estate Valuations" by Rainer Schulz, Markus Staiber, Martin Wersing and Axel Werwatz, February 2008.
020 "The Impact of International Outsourcing on Labour Market Dynamics in Germany" by Ronald Bachmann and Sebastian Braun, February 2008.
021 "Preferences for Collective versus Individualised Wage Setting" by Tito Boeri and Michael C. Burda, February 2008.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

022 "Lumpy Labor Adjustment as a Propagation Mechanism of Business Cycles" by Fang Yao, February 2008.

023 "Family Management, Family Ownership and Downsizing: Evidence from S&P 500 Firms" by Jörn Hendrich Block, February 2008.

024 "Skill Specific Unemployment with Imperfect Substitution of Skills" by Runli Xie, March 2008.

025 "Price Adjustment to News with Uncertain Precision" by Nikolaus Hautsch, Dieter Hess and Christoph Müller, March 2008.

026 "Information and Beliefs in a Repeated Normal-form Game" by Dietmar Fehr, Dorothea Kübler and David Danz, March 2008.

027 "The Stochastic Fluctuation of the Quantile Regression Curve" by Wolfgang Härdle and Song Song, March 2008.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".