Local Quantile Regression

Wolfgang K. Härdle
Vladimir Spokoiny
Weining Wang

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
http://ise.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

Weierstrass Institute
for Applied Analysis and Stochastics

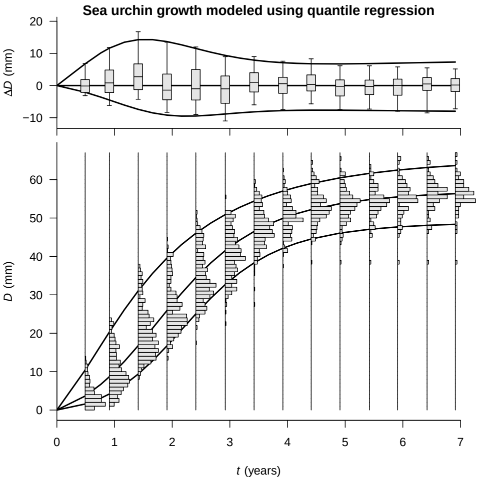


Figure 1: Individual growth over time in a population, Grosjean, PH, Ch. Spirlet M. Jangoux, 2003

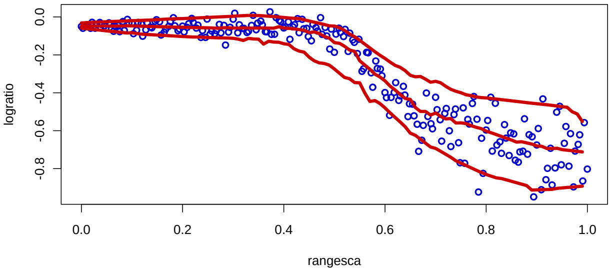


Figure 2: 221 observations, a light detection and ranging (LIDAR) experiment, X: range distance travelled before the light is reflected back to its source, Y: logarithm of the ratio of received light from two laser sources.

Quantile Estimation

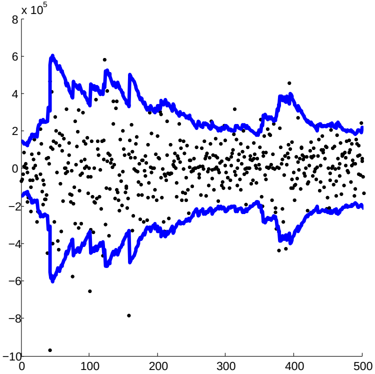


Figure 3: Forecasts of Value at Risk

Conditional Quantile Regression

- Median regression = mean regression (symmetric)
- Describes the conditional behavior of a response Y in a broader spectrum
- ...

Example

- Financial Market & Econometrics
  - VaR (Value at Risk) tool
  - Detect conditional heteroskedasticity
  - ...

Example

- Labor Market
  - Analyze income of football players w.r.t. different ages, years, and countries, etc.
  - To detect discrimination effects, need split other effects at first.

$$
\begin{aligned}
\log(\text{Income}) &= A(\text{year, age, education, etc}) \\
&\quad + \beta\, B(\text{gender, nationality, union status, etc}) + \varepsilon
\end{aligned}
$$

  - ...

- Parametric e.g. polynomial model
  - Interior point method, Koenker and Bassett (1978), Koenker and Park (1996), Portnoy and Koenker (1996).
- Nonparametric model
  - Check function approach, Fan et al. (1994), Yu and Jones (1997, 1998).
  - Double-kernel technique, Fan et al. (1996).
  - Weighted NW estimation, Hall et al. (1999), Cai (2002).

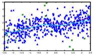
- Global Smoothing Parameter is not sensible to local variations of quantile curve
- Adaptive approach to choose local smoothing parameter
  - Local changing point method for Volatility model , Spokoiny (2009).
  - Local model selection method, Katkovnik and Spokoiny (2007) .

Outline

1. Motivation ✓
2. Basic Setup
3. Parametric Exponential Bounds
4. Calibration by "Propagation" Condition
5. "Small Modeling Bias" and Propagation Properties
6. "Stability" and "Oracle" Property
7. Simulation
8. Application
9. Further work

Basic Setup

- $\{(X_i, Y_i)\}_{i=1}^n$ i.i.d. rv's, pdf $f(x,y)$, $f(y|x)$, cdf $F(x,y)$, $F(y|x)$, and marginal f_X, f_Y, $x \in J = [0,1] \subseteq \mathbb{R}^d$, $y \in \mathbb{R}$
- $l(x) = F^{-1}_{Y|x}(\tau)$ τ-quantile regression curve
- $l_n(x)$ quantile-smoother



Quantile & ALD Location Model

$$Y_i \sim F(y), l_\tau = F^{-1}(\tau) = \theta^* = 0$$

Asymmetric Laplace Distribution (ALD)

$$\begin{aligned}
Y_i &= \theta^* + \varepsilon_i, \varepsilon_i \sim ALD_\tau(0,1) \\
p(u,0) &= \tau(1-\tau)\exp\{-\rho_\tau(u)\} \\
\ell(u) &= \log p(u,0) = \log\{\tau(1-\tau)\} - \rho_\tau(u) \\
\rho_\tau(u) &\overset{\text{def}}{=} \tau u \mathbf{1}\{u \in (0,\infty)\} - (1-\tau) u \mathbf{1}\{u \in (-\infty,0)\}
\end{aligned}$$

Check Function

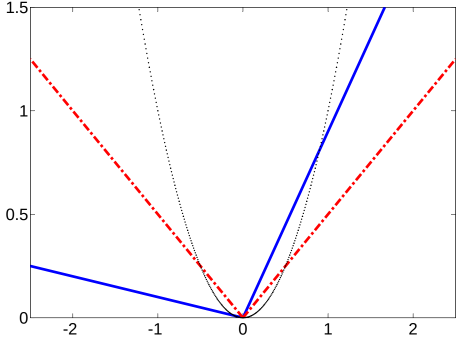


Figure 4: Check function for τ=0.9, τ=0.5 and weight function in conditional mean regression

Quasi-likelihood Approach

$$
\begin{aligned}
\theta^* &= \arg\min_\theta \mathsf{E}_{\theta^*,\lambda_0(.)}\, \rho_\tau(Y_i - \theta) \\
\tilde{\theta} &= \arg\min_\theta L(\theta) \stackrel{\text{def}}{=} \arg\min_\theta \sum_{i=1}^{n} \{-\log \tau(1-\tau) + \rho_\tau(Y_i - \theta)\}
\end{aligned}
$$

$$
\begin{aligned}
\lambda^+(y) &= (\tau^{-1} y)^{-1} \log\{\tau^{-1}\,\mathsf{P}(Y_1 > y)\}, y > 0 \\
\lambda^-(y) &= \{(1-\tau)^{-1} y\}^{-1} \log\{(1-\tau)^{-1}\,\mathsf{P}(Y_1 < y)\}, y < 0
\end{aligned}
$$

where we assume $\lambda^{+(-)}(y) \to 0$ as $y \to \infty$ (heavy tail distribution), $\lambda_0(.)$ tail function of true P_0.

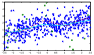
- $l(x)$ minimizes (w.r.t. θ)

$$L(\theta) = \mathsf{E}_{l(.),\lambda_0(.)}\{\rho_\tau(Y-\theta)|X=x\}$$

- $l_n(x)$ minimizes (w.r.t. θ)

$$L_n(\theta) = n^{-1}\sum_{i=1}^{n}\rho_\tau(Y_i-\theta)K_{h(x)}(x-X_i) \tag{1}$$

- $K_{h(x)}(u) = h(x)^{-1}K\{u/h(x)\}$ is a kernel with bandwidth $h(x)$, which is determined by a local adaptive algorithm.



First Step to Happiness

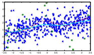
Parametric Exponential Bounds

Define rate function:

$$\mathfrak{M}(\theta, \theta^*) \stackrel{\text{def}}{=} -\log \mathsf{E}_{\theta^*, \lambda(.)} \exp\{\mu(\theta) L(\theta, \theta^*)\} \tag{2}$$

Let $\theta^* = 0$, and $\theta > 0$, we can derive, for $\mu(\theta) = \frac{1}{2\tau}\lambda^+(\theta)$:

$$\mathfrak{M}(\theta, \theta^*) \geq -\log\{1 + \theta\lambda^+(\theta)\} + \frac{4\tau - 1}{2\tau}\theta\lambda^+(\theta) \tag{3}$$

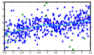


Parametric Exponential Bounds

Golubev Y., Spokoiny V. (2009)

$$
\begin{aligned}
V(\theta) &\stackrel{\mathrm{def}}{=} \mathsf{E}_{\theta^*} \nabla\zeta(\theta)[\nabla\zeta(\theta)]^\top \\
H(\lambda,\gamma,\theta) &\stackrel{\mathrm{def}}{=} \log \mathsf{E}_{\theta^*,\lambda(.)} \exp\{2\lambda \frac{\gamma^\top \nabla\zeta(\theta)}{\sqrt{\gamma^\top V(\theta)\gamma}}\},
\end{aligned}
$$

where $\zeta(\theta) \stackrel{\mathrm{def}}{=} \ell(\theta,\theta^*) - \mathsf{E}_{\theta^*,\lambda(.)}\, \ell(\theta,\theta^*)$.



Parametric Exponential Bounds

(ED) There exists $\overline{\lambda} > 0$ such that for some $\nu_0 \geq 1$ uniformly in $\theta \in \Theta$

$$\sup_{|\lambda| \leq \overline{\lambda}} \sup_{\gamma \in S^p} \lambda^{-2} H(\lambda, \gamma, \theta) \leq 2\nu_0^2$$

Theorem

Under (ED), we have, for prescribed $\rho < 1$, $s < 1$,

$$E_{\theta^*,\lambda(.)} \exp[\rho\{\mu(\tilde{\theta}) L(\tilde{\theta}, \theta^*) + s\mathfrak{M}(\tilde{\theta}, \theta^*)\}] \leq C(1-\rho)^{-1/2}(1-s)^{-1/2}$$

with fixed constant C provided that n exceeds some minimal sample size.

Parametric Exponential Bounds

Obtain exponential risk bound for the likelihood function:

$$\mathsf{E}_{\theta^*,\lambda(.)} \exp\{\rho\mu(\tilde{\theta}) L(\tilde{\theta}, \theta^*)\} \le C(1-\rho)^{-0.5}$$

$$L(\tilde{\theta}, \theta^*) = n\frac{2\tau - 1}{2}(\tilde{\theta} - \theta^*) + \frac{1}{2}\sum_{i=1}^{n}\{|Y_i - \theta^*| - |Y_i - \tilde{\theta}|\}$$

Parametric Exponential Bounds

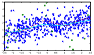
This further implies:

$$\mathsf{E}_{\theta^*,\lambda(.)} \left| L(\tilde{\theta}, \theta^*) \right|^r \le \tau_r$$

Adaptation Scale

Fix x, we have a sequence of ordered weights
$W^{(k)} = (w_1^{(k)}, w_2^{(k)}, \dots, w_n^{(k)})^\top$.
Define $w_i^{(k)} = K_{h_k}(x - X_i), (h_1 < h_2 < \ldots < h_K)$.
LPA: $l(X_i) \approx l_\theta(X_i)$

$$\begin{aligned}\tilde{\theta}_k &= \operatorname*{argmax}_\theta \sum_{i=1}^n \ell\{Y_i, l_\theta(X_i)\} w_i^{(k)} \\ &\stackrel{\text{def}}{=} \operatorname*{argmax}_\theta L(W^{(k)}, \theta)\end{aligned}$$



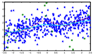
LMS Procedure for Quantile

Construct an estimate $\hat{\theta} = \hat{\theta}(x)$, on the base of $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_K$.

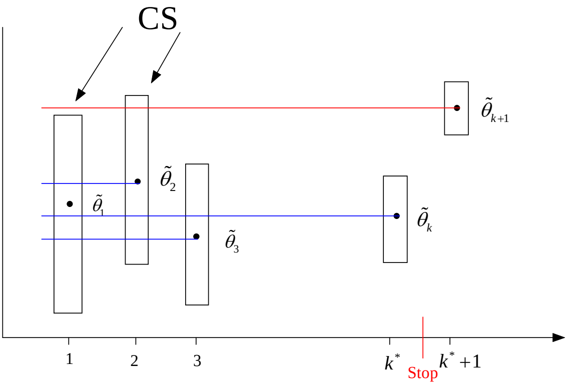
- Start with $\hat{\theta}_1 = \tilde{\theta}_1$.
- For $k \geq 2$, $\tilde{\theta}_k$ is accepted and $\hat{\theta}_k = \tilde{\theta}_k$ if $\tilde{\theta}_{k-1}$ was accepted and

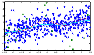
$$L(W^{(\ell)}, \tilde{\theta}_\ell, \tilde{\theta}_k) \leq \mathfrak{z}_\ell, \ell = 1, \ldots, k-1$$

$\hat{\theta}_k$ is the the latest accepted estimate after the first k steps.



Illustration





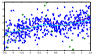
"Propagation" Condition

Theorem

Suppose $\Delta(W^{(k)}, \theta) \leq \Delta$ *for* $k \leq k^*$. *Then*

$$E_{\theta^*,\lambda(.)} \frac{|L(W^{(k)}, \tilde{\theta}_k, \hat{\theta}_k)|^r}{\tau_r(W^{(k)})} \leq \alpha,$$

where k^**, the "oracle" choice.*



Sequential Choice of Critical Values

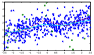
- Consider first only $\mathfrak{z}_1$ letting $\mathfrak{z}_2 = \ldots = \mathfrak{z}_{K-1} = \infty$. Leads to the estimates $\hat{\theta}_k(\mathfrak{z}_1)$ for $k = 2, \ldots, K$.
- The value $\mathfrak{z}_1$ is selected as the minimal one for which

$$\sup_{\theta^*} \mathrm{E}_{\theta^*,\lambda(.)} \frac{|L\{W^{(k)}, \tilde{\theta}_k, \hat{\theta}_k(\mathfrak{z}_1)\}|^r}{\tau_r(W^{(k)})} \le \frac{\alpha}{K-1}, k = 2, \ldots, K.$$

- Set $\mathfrak{z}_{k+1} = \ldots = \mathfrak{z}_{K-1} = \infty$ and fix $\mathfrak{z}_k$ lead the set of parameters $\mathfrak{z}_1, \ldots, \mathfrak{z}_k, \infty, \ldots, \infty$ and the estimates $\hat{\theta}_m(\mathfrak{z}_1, \ldots, \mathfrak{z}_k)$ for $m = k+1, \ldots, K$. Select $\mathfrak{z}_k$ s.t.

$$\sup_{\theta^*} \mathrm{E}_{\theta^*,\lambda(.)} \frac{|L\{W^{(k)}, \tilde{\theta}_m, \hat{\theta}_m(\mathfrak{z}_1, \mathfrak{z}_2, \ldots, \mathfrak{z}_k)\}|^r}{\tau_r(W^{(k)})} \le \frac{k\alpha}{K-1},$$

$m = k+1, \ldots, K$.



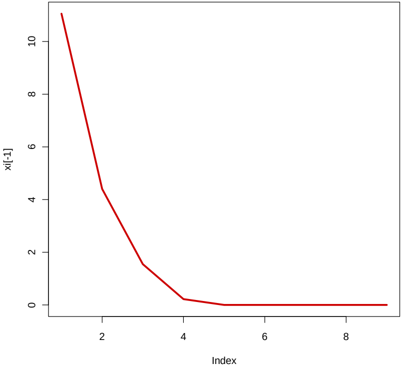


Figure 5: Critical value with $\lambda(.)$ corresponding to standard normal distribution , $\rho = 0.2$, $r = 0.5$, $\tau = 0.75$.

Bounds for Critical Values

a) For some constants $0 < u_0, u < 1$, it holds
$0 < u_0 \le h_{k-1}/h_k \le u < 1$.
In addition, we assume the parameter set Θ satisfy the condition:
b) For some constants a with $0 < a < 1$, and for any θ and $\theta^* \in \Theta$,

$$a^2 < \theta^*/\theta < a^{-2}$$

Theorem

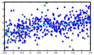
Suppose that $r > 0, \alpha > 0$. Under the assumptions of probability bounds, there are a_0, a_1, a_2, s.t. the propagation condition is fulfilled with the choice of
$\mathfrak{z}_k = a_0 r \log(\rho^{-1}) + a_1 r \log(h_K/h_k) + a_2 \log(nh_k)$.

"Small Modeling Bias" Condition

$$\Delta_{\lambda(.),\lambda_0(.)}(W^{(k)}, \theta) = \sum_{i=1}^{n} \mathcal{K}\{P_{I(X_i),\lambda_0(.)}, P_{I_\theta(X_i),\lambda(.)}\}\mathbf{1}\{w_i^{(k)} > 0\}$$

"Small Modeling Bias" Condition:

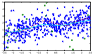
$$\Delta_{\lambda,\lambda_0(.)}(W^{(k)}, \theta) \leq \Delta, \forall k < k^*$$



"Small Modeling Bias" Property

Then, it holds for any estimate $\tilde{\theta}_k$ and θ satisfies "SMB":

$$\mathsf{E}_{l(.),\lambda_0(.)} \log\{1 + |L(W^{(k)}, \tilde{\theta}_k, \theta)|^r / \tau_r(W^{(k)})\} \leq \Delta + 1,$$



"Stability" Property

Stability: the attained quality of estimation during "propagation" can not get lost at further steps.

Theorem

$$L(W^{(k^*)}, \tilde{\theta}_{k^*}, \hat{\theta}_{\hat{k}})\mathbf{1}\{\hat{k} > k^*\} \leq \mathfrak{z}_{k^*}$$

"Oracle" Property

Combing the "propagation" and "stability" results yields

Theorem

Let $\Delta(W^{(k)}, \theta) \le \Delta$ for some $\theta \in \Theta$ and $k \le k^*$. Then

$$E_{l(.),\lambda_0(.)} \log \left\{ 1 + \frac{|L(W^{(k^*)}, \tilde{\theta}_{k^*}, \theta)|^r}{\tau_r(W^{(k^*)})} \right\} \le \Delta + 1$$

$$E_{l(.),\lambda_0(.)} \log \left\{ 1 + \frac{|L(W^{(k^*)}, \tilde{\theta}_{k^*}, \hat{\theta})|^r}{\tau_r(W^{(k^*)})} \right\} \le \Delta + \alpha$$

$$+ \log\{1 + \frac{\mathfrak{z} k^*}{\tau_r(W^{(k^*)})}\}$$

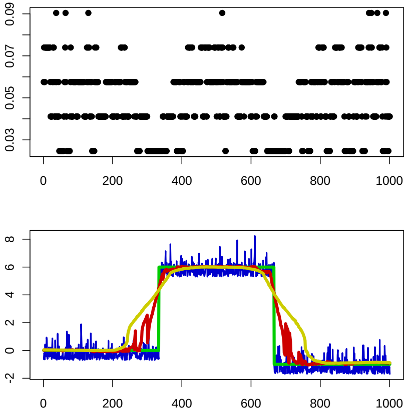


Figure 6: The bandwidth sequence (upper panel), the data with exp(2) noise (blue line), the adaptive estimation of 0.75 quantile (red line), the quantile smoother with fixed optimal bandwidth = 0.06 (yellow)

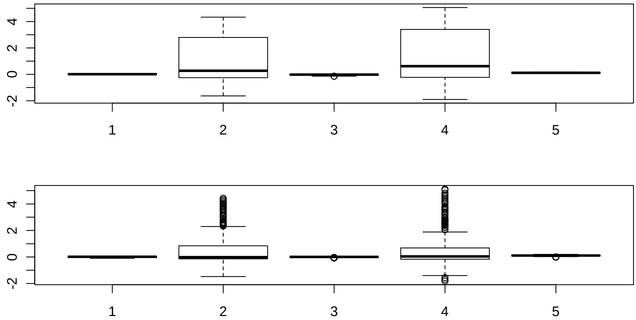


Figure 7: The block residuals of the fixed bandwidth estimation (upper panel)and the adaptive estimation (lower panel)

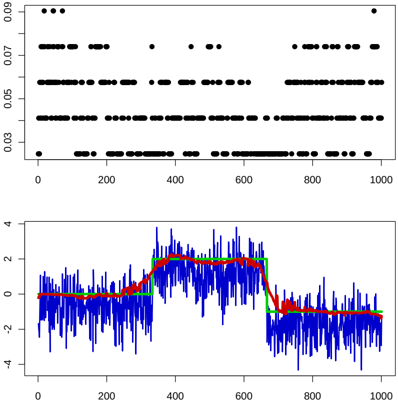


Figure 8: The bandwidth sequence (upper panel), the data with standard normal noise(blue line), the adaptive estimation of 0.75 quantile (red line)

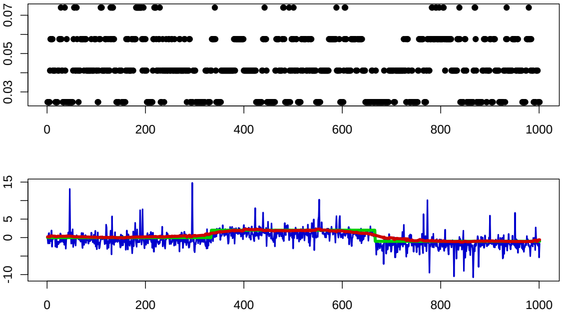
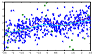


Figure 9: The bandwidth sequence (upper panel), the data with t(3) noise (blue line), the adaptive estimation of 0.75 quantile (red line)



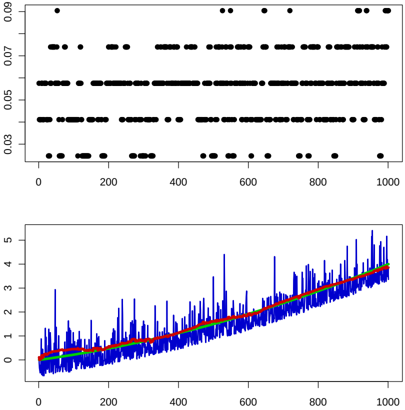
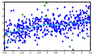


Figure 10: The bandwidth sequence (upper panel), the data with exp(2) noise (blue line), the adaptive estimation of 0.75 quantile (red line)



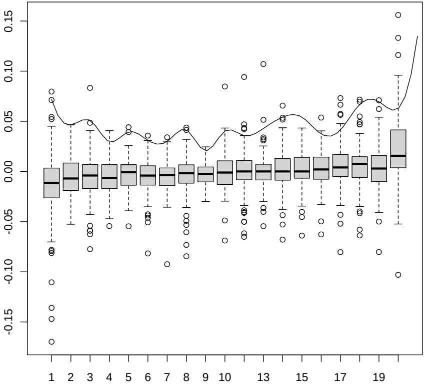
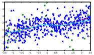


Figure 11: Y: HK stock cheung kong returns, X: HK stock clpholdings returns, daily 01/01/2003 – 04/26/2010



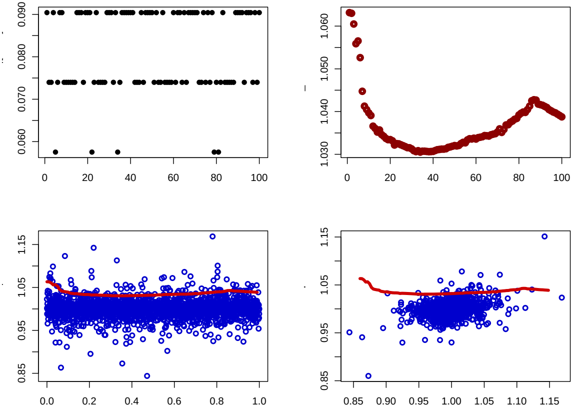
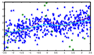


Figure 12: Plot of bandwidth sequence (upper left), plot of quantile smoother (upper left), scatter plot on scaled X (lower left), original scale (lower right)



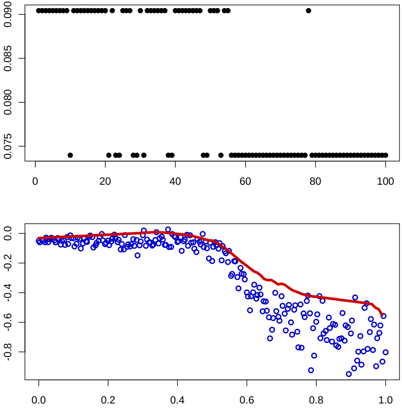


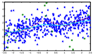
Figure 13: 221 observations, a light detection and ranging (LIDAR) experiment, X: range distance travelled before the light is reflected back to its source, Y: logarithm of the ratio of received light from two laser sources.

References

- P. Bickel and M. Rosenblatt
  On some global measures of the deviation of density function estimatiors
  Ann. Statist., 1:1071-1095, 1973.

- Z. W. Cai
  Regression quantiles for time series
  Econometric Theory, 18:169-192, 2002.

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



References

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

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

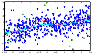
P. Hall and R.C.L. Wolff and Q. Yao
Methods for estimating a conditional distribution function
J. Amer. Statist. Assoc., 94:154-163, 1999.

References

W. Härdle, P. Janssen and R. Serfling
Strong uniform consistency rates for estimators of conditional functionals
Ann. Statist., 16:1428-1449 1988.

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

X. He and H. Liang
Quantile regression estimates for a class of linear and partially linear errors-in-variables model
Ann. Statist., 10:129-140 2000.



References

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

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

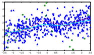
R. Koenker and G. W. Bassett
Regression quantiles
Econometrica, 46:33-50, 1978.

References

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

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

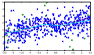
K. Murphy and F. Welch
Empirical age-earnings profiles
Journal of Labor Economics, 8(2):202-229 1990.



References

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

E. Rio
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, 1996.



References

V. Spokoiny
Local parametric estimation
Heildelberg: Springer Verlag, 2008

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

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

References

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

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

- L. X. Zhu, R. Q. Zhu and S. Song
  Diagnostic checking for multivariate regression models
  J. Multivariate Anal., accepted.