

# 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://lvb.wiwi.hu-berlin.de>

<http://www.case.hu-berlin.de>



**Weierstrass Institute**  
for Applied Analysis and Stochastics

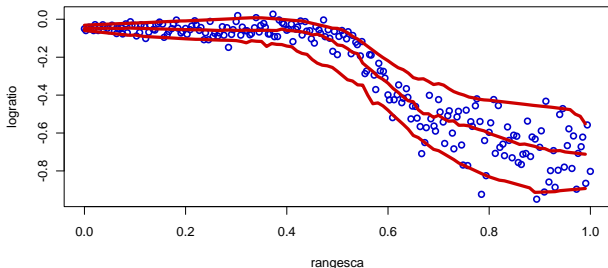
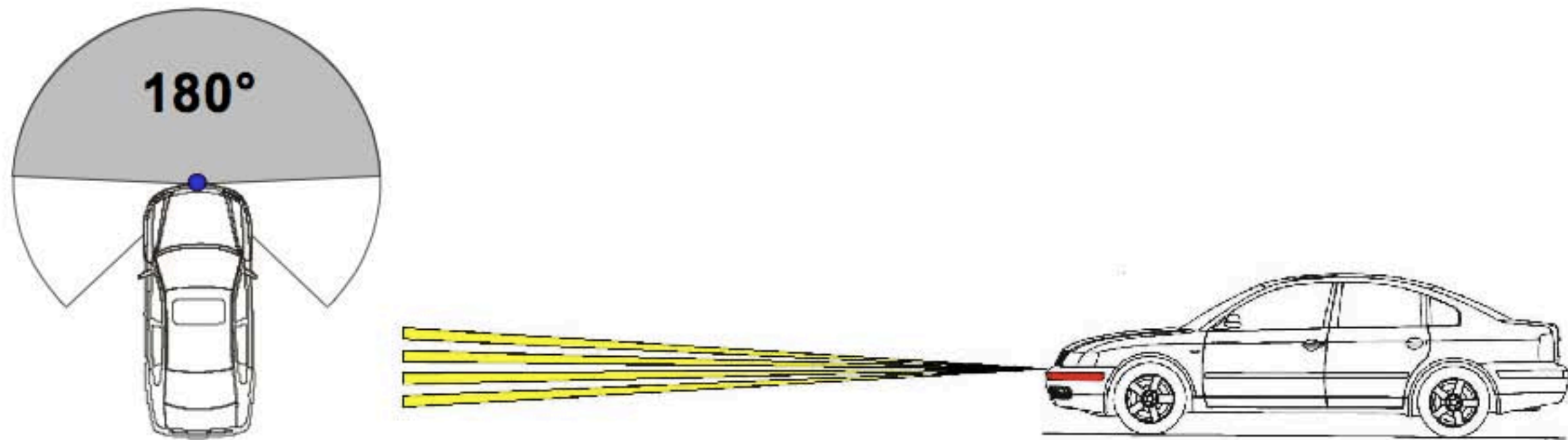
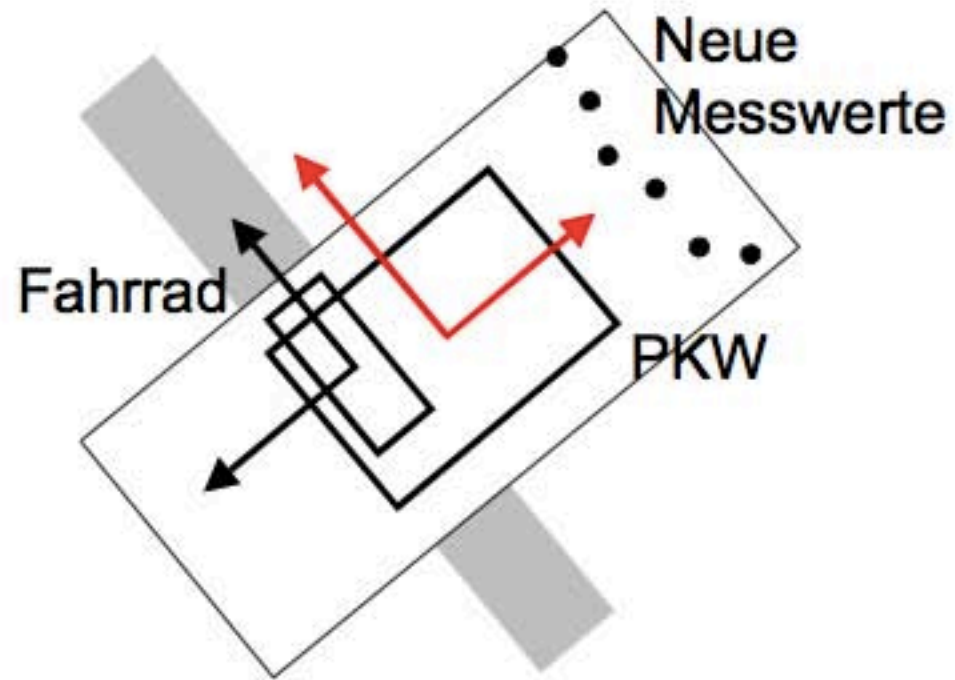
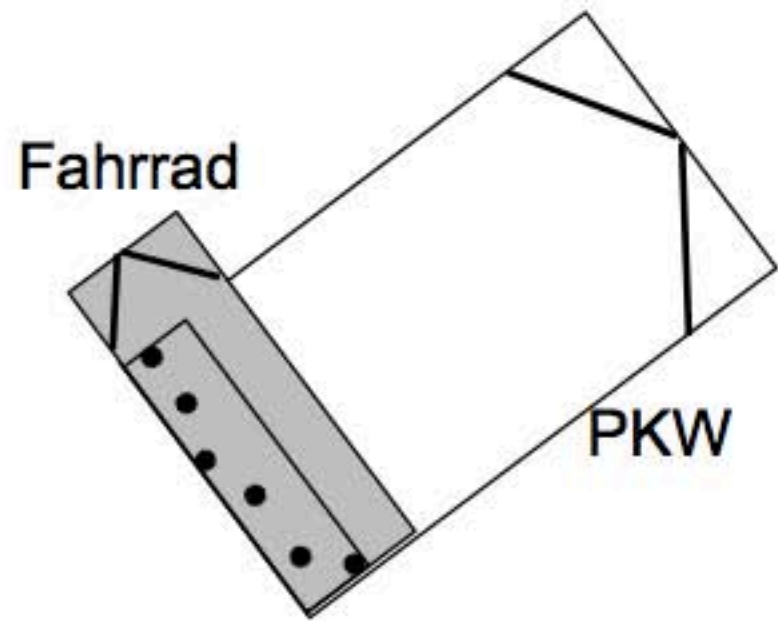


Figure 1: 221 observations, a light detection and ranging (LIDAR) experiment, X: range distance traveled before the light is reflected back to its source, Y: logarithm of the ratio of received light from two laser sources. 95% quantile.





**Bild 4:** Erfassungsbereiche des mehrzeiligen Sensors LD ML, der vertikale Erfassungsbereich beträgt ca.  $3,2^\circ$ , der horizontale nahezu  $180^\circ$ .



## Quantile Estimation

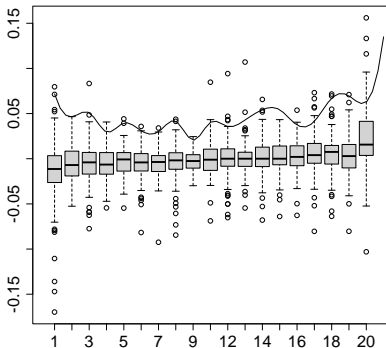


Figure 2: Y: HK stock cheung kong returns, X: HK stock clpholdings returns, daily 01/01/2003 – 04/26/2010, 90% quantile.



## Electricity Demand

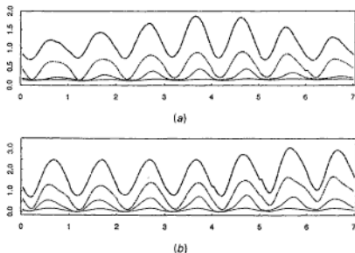


Figure 2. Estimated Conditional Quantiles of the Demand Distribution for Two Representative Households. The plotted curves indicate, in ascending order the .25, .50, .75 and .95 fitted quantiles of the weekly demand distribution measured in kilowatts. On the horizontal axis 0 represents 0:00 a.m. on Sunday; 7 represents 24:00 on Saturday; (a) Household 1 and (b) Household 2.

Figure 3: Electricity Demand over Time



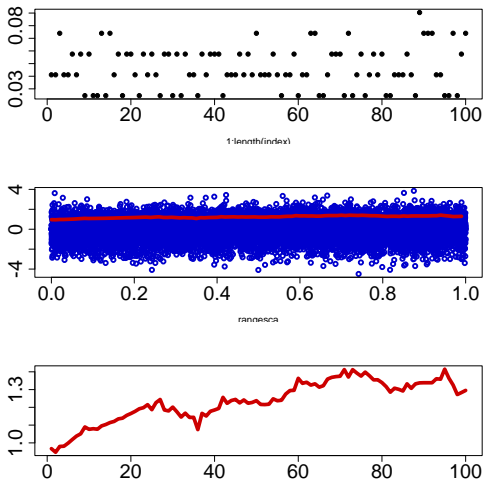


Figure 4: Plot of quantile for standardized weather residuals over 51 years at Berlin, 90% quantile .

Local Quantile Regression



## Weather Residuals

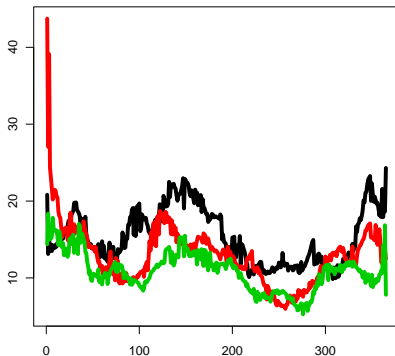


Figure 5: Estimated 90% quantile of variance functions, Berlin, average over 1995 – 1998, 1999 – 2002, 2003 – 2006





## 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

- ▶ To detect discrimination effects, need split other effects at first.

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



□ 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).
- ▶ LCP method, Spokoiny (2009).
- ▶ LMS , 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., pdf  $f(x, y)$ ,  $f(y|x)$
- $l(x) = F_{y|x}^{-1}(\tau)$ ,  $\tau$ -quantile regression curve
- $l_n(x)$  quantile smoother



# Asymmetric Laplace Distribution (ALD)

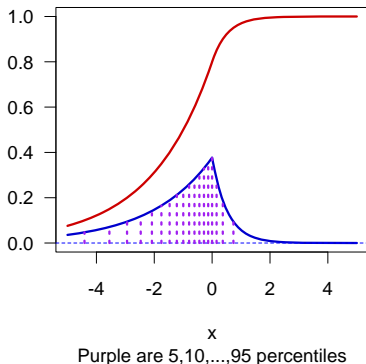


Figure 6: The **cdf** and **pdf** of an ALD with  $\mu = 0, \sigma = 1.5, \tau = 0.8$ .



## Quantile & ALD Location Model

$$Y_i = \theta^* + \varepsilon_i, \varepsilon_i \sim \text{ALD}_\tau(0, 1)$$

$$l = F^{-1}(\tau) = \theta^* = 0$$

$$f(u, \theta^*) = \tau(1 - \tau) \exp\{-\rho_\tau(u - \theta^*)\}$$

$$\ell(u, \theta^*) = \log\{\tau(1 - \tau)\} - \rho_\tau(u - \theta^*)$$

$$\rho_\tau(u) \stackrel{\text{def}}{=} \tau u \mathbf{1}\{u \in (0, \infty)\} - (1 - \tau)u \mathbf{1}\{u \in (-\infty, 0)\}$$



## Check Function

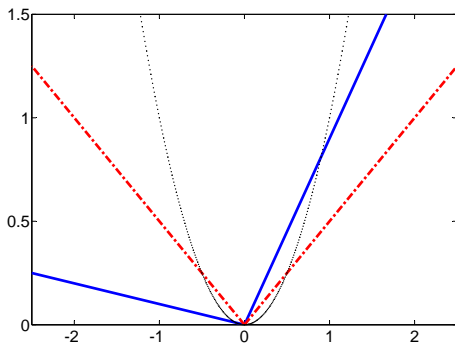


Figure 7: Check function  $\rho_\tau$  for  $\tau=0.9$ ,  $\tau=0.5$  and weight function in conditional mean regression





## Quasi-likelihood Approach

$$\theta^* = \arg \min_{\theta} E_{\lambda_0(\cdot)} \rho_{\tau}(Y - \theta)$$

$$\tilde{\theta} = \arg \min_{\theta} L(\theta) \stackrel{\text{def}}{=} \arg \min_{\theta} \sum_{i=1}^n \{-\log \tau(1 - \tau) + \rho_{\tau}(Y_i - \theta)\}$$

Tail probabilities determine  $F$ :

$$\lambda_0^+(y) = (\tau^{-1}y)^{-1} \log\{\tau^{-1} P(Y_1 > y)\}, y > 0$$

$$\lambda_0^-(y) = \{(1 - \tau)^{-1}y\}^{-1} \log\{(1 - \tau)^{-1} P(Y_1 < y)\}, y < 0$$

Assume  $\lambda_0^{+(-)}(y)\{\lambda_0(y)\} \rightarrow 0$  as  $y \rightarrow +/ - \infty$  (heavy tails).



## Conditional Quantile Regression

$$\square I(x) = \operatorname{argmin}_{\theta} E_{\lambda_0(\cdot)} \{ \rho_{\tau}(Y - \theta) | X = x \}$$

$\square$

$$I_n(\theta) = \operatorname{argmin}_{\theta} n^{-1} \sum_{i=1}^n \rho_{\tau}(Y_i - \theta) K_{h(x)}(x - X_i)$$

$\square K_{h(x)}(u) = h(x)^{-1} K\{u/h(x)\}$  is a kernel with bandwidth  $h(x)$ .

$\square L(\theta, \theta') \stackrel{\text{def}}{=} L(Y, \theta) - L(Y, \theta')$



## Local Polynomial Quantile Regression

$$l(u) \approx \theta_0(u) + \theta_1(u)(x - u) + \theta_2(u)(x - u)^2 + \dots + \theta_p(u)(x - u)^p,$$

$$\tilde{\theta}(u) = \{\tilde{\theta}_0(u), \tilde{\theta}_1(u), \dots, \tilde{\theta}_p(u)\}^\top \quad (1)$$

$$= \operatorname{argmax}_{\theta \in \Theta} \log \tau(1 - \tau) \sum_{i=1}^n w_i - \sum_{i=1}^n \rho_\tau(Y_i - \theta^\top \psi_i) w_i$$

$$= \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1}^n \rho_\tau(Y_i - \theta^\top \psi_i) w_i,$$

where  $\psi_i = \{(X_i - u), (X_i - u)^2, \dots, (X_i - u)^p\}^\top$  and  $w_i = K\{(x_i - u)/h\}$ , for  $i$  in  $1, 2, \dots, n$ .



## First Step to Happiness



## Parametric Exponential Bounds

**Theorem 1** Let  $V_0 = V(\theta^*)$ . Let  $R$  satisfy  $\lambda^* \geq qR/\mu$  for  $\mu = \alpha/(2\nu_0q^3)$ . We have for any  $r < R$ , if  $R \leq \mu\lambda^*/(\nu_0q)$ , then for any  $\beta > 0$  with  $2\beta/\alpha \leq R^2$ , it holds

$$P_{\lambda_0(\cdot)}\{L(\tilde{\theta}, \theta^*) > \beta\} \leq \exp\{-\beta\alpha/(\nu_0q^3) - \Omega_0(p, q)\} + \exp\{-g(R)\}, \quad (2)$$

where  $g(R)$  is defined as:

$$\begin{aligned} g(R) &\stackrel{\text{def}}{=} \inf_{\theta \in \Theta} \sup_{\mu \in M} \{\mu R + \mathfrak{C}(\mu, \theta, \theta^*)\} \\ &= \inf_{\theta \in \Theta} g(R, \theta, \theta^*) \end{aligned}$$

[Go to details](#)



## Parametric Exponential Bounds

**Theorem 2** Under technical assumptions,

$$E_{\lambda_0(\cdot)} |L(\theta, \theta^*)|^r \leq \mathfrak{R}_r,$$

where  $\mathfrak{R}_r > 0$  is a constant.



## Adaptation Scale

Fix  $x$ , and a sequence of ordered weights

$$W^{(k)} = (w_1^{(k)}, w_2^{(k)}, \dots, w_n^{(k)})^\top$$

Define

$$w_i^{(k)} = K_{h_k}(x - X_i), (h_1 < h_2 < \dots < h_K)$$

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

Construct an estimate  $\hat{\theta} = \hat{\theta}(x)$ , based on  $\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_K$ .

- 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

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

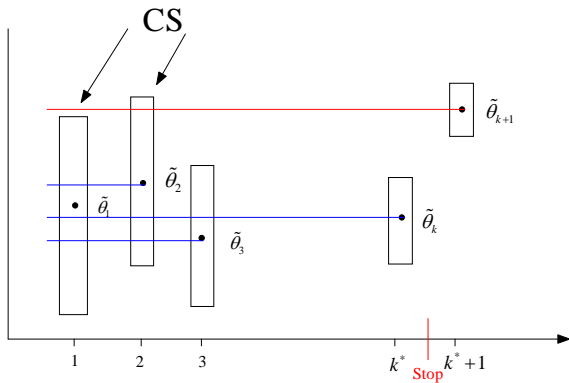
$\hat{\theta}_k$  is **the latest accepted estimate after the first  $k$  steps.**

- $L(W^{(k)}, \theta, \theta') = L(W^{(k)}, \theta) - L(W^{(k)}, \theta')$





# Illustration



## "Propagation" Condition

The critical value is determined via a false alarm consideration:

$$E_{\theta^*} \frac{|L(W^{(k)}, \tilde{\theta}_k, \hat{\theta}_k)|^r}{\mathfrak{R}_r(W^{(k)})} \leq \alpha,$$

where  $\mathfrak{R}_r(W^{(k)})$  is a risk bound depending on  $W^{(k)}$ .



## Critical Values

- Consider first  $\beta_1$  letting  $\beta_2 = \dots = \beta_{K-1} = \infty$ . Leads to the estimates  $\hat{\theta}_k(\beta_1)$  for  $k = 2, \dots, K$ .
- The value  $\beta_1$  is selected as the minimal one for which

$$\sup_{\theta^*} E_{\theta^*, \lambda(\cdot)} \frac{|L\{W^{(k)}, \tilde{\theta}_k, \hat{\theta}_k(\beta_1)\}|^r}{\mathfrak{R}_r(W^{(k)})} \leq \frac{\alpha}{K-1}, k = 2, \dots, K.$$

- Set  $\beta_{k+1} = \dots = \beta_{K-1} = \infty$  and fix  $\beta_k$  gives the set of parameters  $\beta_1, \dots, \beta_k, \infty, \dots, \infty$  and the estimates  $\hat{\theta}_m(\beta_1, \dots, \beta_k)$  for  $m = k+1, \dots, K$ . Select  $\beta_k$  s.t.

$$\sup_{\theta^*} E_{\theta^*, \lambda(\cdot)} \frac{|L\{W^{(k)}, \tilde{\theta}_m, \hat{\theta}_m(\beta_1, \beta_2, \dots, \beta_k)\}|^r}{\mathfrak{R}_r(W^{(k)})} \leq \frac{k\alpha}{K-1},$$

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



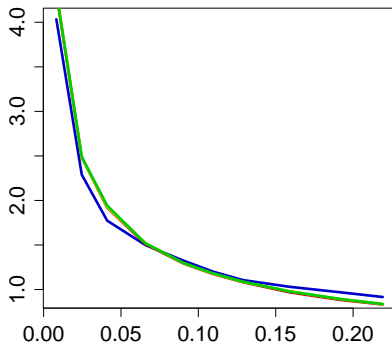


Figure 8: Critical value with  $\lambda(\cdot)$  corresponding to **standard normal distribution**, **t(3)**, **pareto(0.5,1)**  $\rho = 0.2$ ,  $r = 0.5$ ,  $\tau = 0.75$ .



## Bounds for Critical Values

**Theorem 3** 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  $\beta_k = a_0 r \log(\rho^{-1}) + a_1 r \log(h_K/h_k) + a_2 \log(nh_k)$ .



## Small Modeling Bias

$$\Delta_{\lambda_0(\cdot)}(W^{(k)}, \theta) = \sum_{i=1}^n \mathcal{K}\{P_{l(X_i), \lambda_0(\cdot)}, P_{l_\theta(X_i), \lambda_0(\cdot)}\} \mathbf{1}\{w_i^{(k)} > 0\}$$

"Small Modeling Bias" Condition:

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

$$E_{\lambda_0(\cdot)} \log\{1 + |L(W^{(k)}, \tilde{\theta}_k, \theta)|^r / \mathfrak{R}_r(W^{(k)})\} \leq \Delta + 1,$$



## Stability

The attained quality of estimation during "propagation" can not get lost at further steps.

**Theorem 4**

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



## Oracle

Combing the "propagation" and "stability" results yields

**Theorem 5** Let  $\Delta(W^{(k)}, \theta) \leq \Delta$  for some  $\theta \in \Theta$  and  $k \leq k^*$ .

Then

$$\begin{aligned}
 E_{\lambda_0(\cdot)} \log \left\{ 1 + \frac{|L(W^{(k^*)}, \tilde{\theta}_{k^*}, \theta)|^r}{\mathfrak{R}_r(W^{(k^*)})} \right\} &\leq \Delta + 1 \\
 E_{\lambda_0(\cdot)} \log \left\{ 1 + \frac{|L(W^{(k^*)}, \tilde{\theta}_{k^*}, \hat{\theta})|^r}{\mathfrak{R}_r(W^{(k^*)})} \right\} &\leq \Delta + \alpha \\
 + \log \left\{ 1 + \frac{\hat{\delta}^{k^*}}{\mathfrak{R}_r(W^{(k^*)})} \right\} &
 \end{aligned}$$





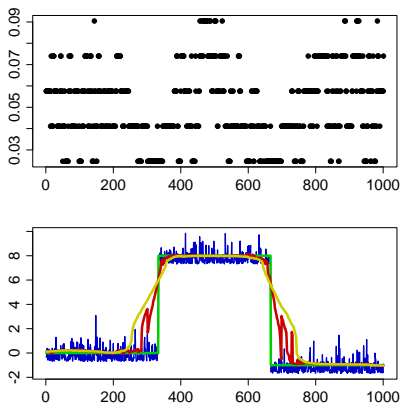


Figure 9: The bandwidth sequence (upper panel), the data with  $\exp(2)$  noise, the adaptive estimation of 0.75 quantile, the quantile smoother with fixed optimal bandwidth = 0.08 (yellow)



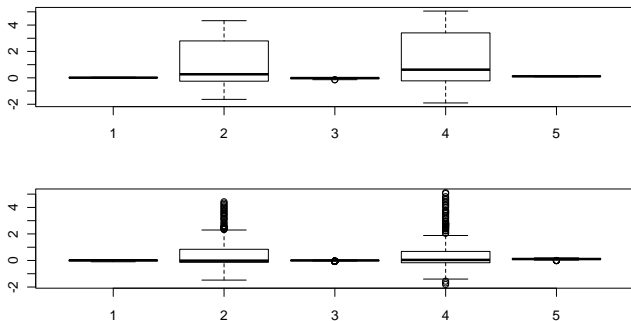


Figure 10: The block residuals of the fixed bandwidth estimation (upper panel) and the adaptive estimation (lower panel)



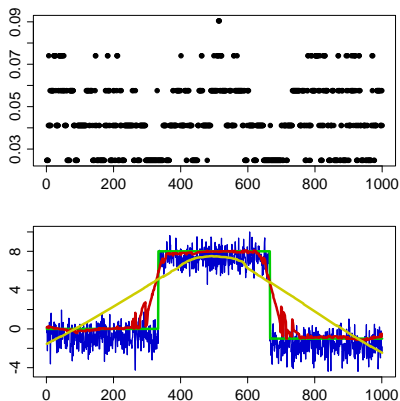


Figure 11: The bandwidth sequence (upper panel), the data with standard normal noise, the adaptive estimation of 0.75 quantile, the quantile smoother with fixed optimal bandwidth = 0.089 (yellow) ,



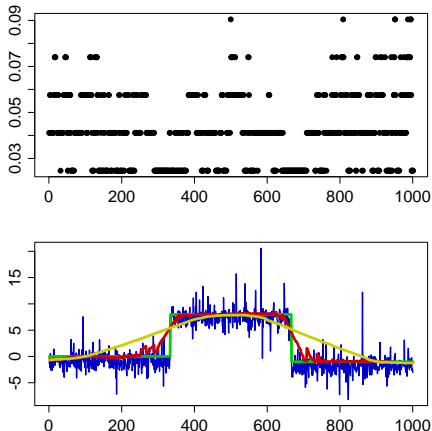


Figure 12: The bandwidth sequence (upper panel), the data with  $t(3)$  noise, the adaptive estimation of 0.75 quantile, the quantile smoother with fixed optimal bandwidth = 0.06 (yellow)



## Local linear

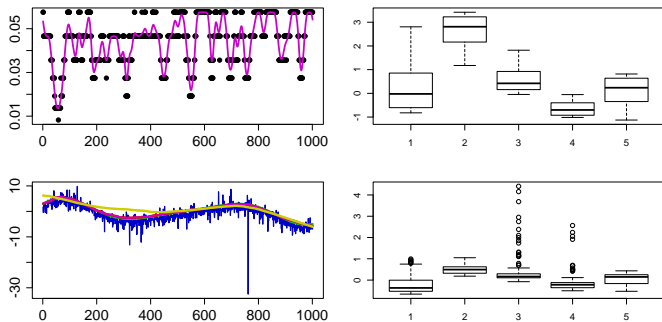


Figure 13: The bandwidth sequence (upper left panel), the data with  $t(3)$  noise, the adaptive estimation of 0.80 quantile, the quantile smoother with fixed optimal bandwidth = 0.06 (yellow); The blocked error fixed vs. adaptive(right panel).



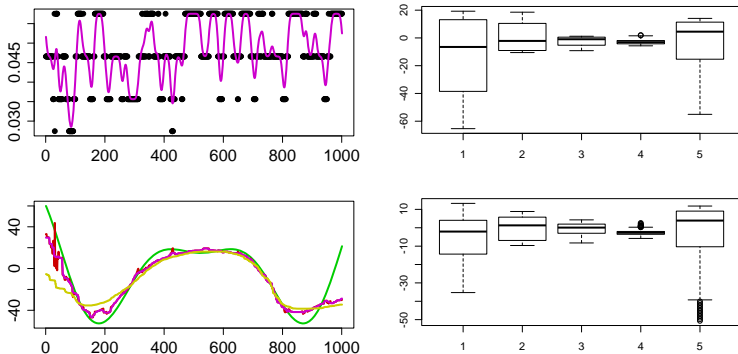


Figure 14: The bandwidth sequence (upper left panel), the true trend curve (lower left panel), the adaptive estimation of 0.80 quantile, the quantile smoother with fixed optimal bandwidth = 0.078 (yellow); The blocked error fixed vs. adaptive (right panel).



## Tail Dependency

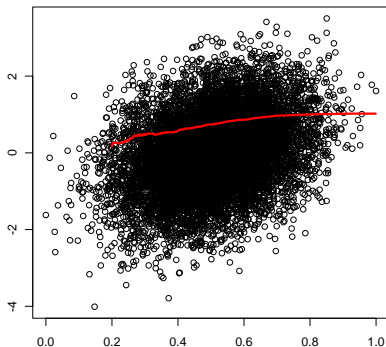


Figure 15: Plot of quantile curve for two normal random variables with  $\rho = 0.3204$



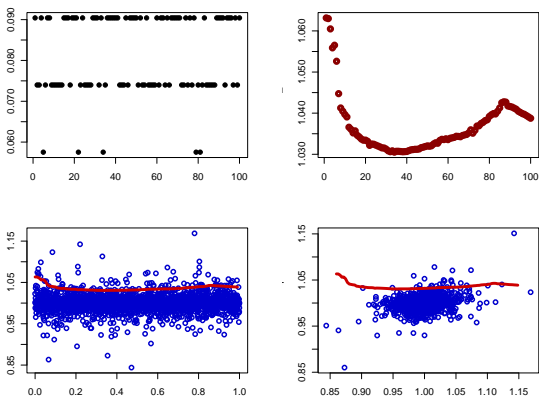


Figure 16: Plot of bandwidth sequence (upper left), plot of quantile smoother (upper left), scatter plot on scaled  $X$  (lower left), original scale (lower right)





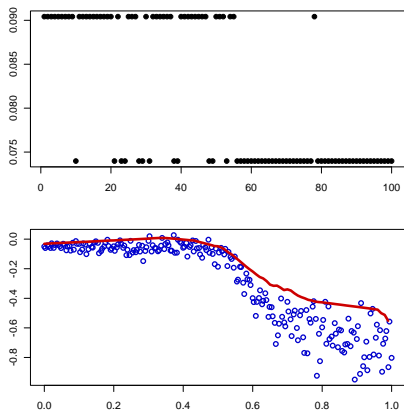


Figure 17: 221 observations, a light detection and ranging (LIDAR) experiment, X: range distance travelled before the light is reflected back to its source, Y: logarithm of the ratio of received light from two laser sources.





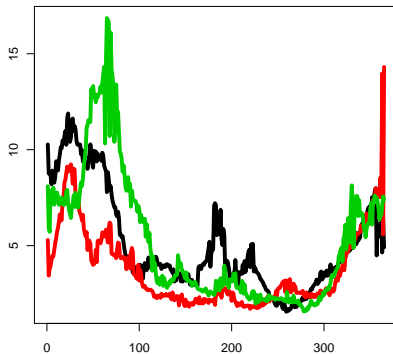


Figure 18: Estimated 90% quantile of variance functions, Kaoshiung, average over 1995 – 1998, 1999 – 2002, 2003 – 2006



# 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://lvb.wiwi.hu-berlin.de>

<http://www.case.hu-berlin.de>



**Weierstrass Institute**  
for Applied Analysis and Stochastics

## Assumptions and Definitions

$$\mathfrak{C}(\mu, \theta, \theta^*) = q^{-1} \mathcal{M}(q\mu, \theta, \theta^*) - (p+1) \log_+ \{ \epsilon^{-1} \mu^* \|V(\theta - \theta^*)\| \} - \mathfrak{Q}(p, q)$$

and

$$\mathfrak{Q}(p, q) = -\nu_0 q \epsilon^2 / 2 - \log(2 + 2p) - pC(q) - \log(1 - a^{-(p+1)})$$

## Assumptions and Definitions

$$\begin{aligned}\zeta(\theta) &\stackrel{\text{def}}{=} L(\theta) - \mathbb{E} L(\theta) \\ \zeta(\theta, \theta^*) &\stackrel{\text{def}}{=} \zeta(\theta) - \zeta(\theta^*) \\ \mathfrak{N}(\mu, \theta, \theta^*) &\stackrel{\text{def}}{=} -\log \mathbb{E} \exp\{\mu \zeta(\theta, \theta^*)\} \\ \mathfrak{M}(\mu, \theta, \theta^*) &\stackrel{\text{def}}{=} -\log \mathbb{E} \exp\{\mu L(\theta, \theta^*)\}\end{aligned}\tag{3}$$



## Assumptions and Definitions

$\nabla\zeta(\theta)$  satisfying the condition (ED):  $\exists\lambda^* > 0$  and a symmetric positive matrix  $V^2$  s.t.  $\forall 0 < \lambda \leq \lambda^*$ ,

$$\log E \exp\{\lambda\gamma^\top \nabla\zeta(\theta)\} \leq \nu_0\lambda^2\|V^2\gamma\|^2/2, \quad (4)$$

for some fixed  $\nu_0 \geq 1$  and  $\gamma \in R^{p+1}$ . Set

$$V^2(\theta) = \sum_{i=1}^n \psi_i \psi_i^\top p_i(\psi^\top \theta) W_i.$$

( $\mathcal{L}$ ) For  $R > 0$ , there exists constant  $a = a(R) > 0$  such that it holds on the set  $\Theta_0(R) = \{\theta : \|V(\theta - \theta^*)\| \leq R\}$ :

$$-E L(\theta, \theta^*) \geq a\|V(\theta - \theta^*)\|^2/2$$

▶ return



## References



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.



Y. Golubev and V. Spokoiny

Exponential bounds for minimum contrast estimators

*Electronic Journal of Statistics*, ISSN: 1935-7524.





## References



W. Härdle, P. Janssen and R. Serfling

Strong uniform consistency rates for estimators of conditional functionals

*Ann. Statist.*, 16:1428-1449 1988.



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.





K. Jeong and W. Härdle

A Consistent Nonparametric Test for Causality in Quantile

*SFB 649 Discussion Paper*, 2008.



## References

-  V. Katkovnik and V. Spokoiny  
Spatially adaptive estimation via fitted local likelihood (FLL) techniques
-  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 estimations (with discussion)

*Statistical Sciences*, 12:279-300, 1997.



V. Spokoiny

*Local parametric estimation*

Heidelberg: Springer Verlag, 2008



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.

