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
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.
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 3: Electricity Demand over Time

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


Local Quantile Regression
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.

\[
\log \text{(Income)} = A(\text{year, age, education, etc}) + \beta B(\text{gender, nationality, union status, etc}) + \varepsilon
\]
Parametric e.g. polynomial model

Nonparametric model
- Double-kernel technique, Fan et al. (1996).
- Weighted NW estimation, Hall et al. (1999), Cai (2002).
- 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} \text{ i.i.d., pdf } f(x, y), f(y|x)
- \ell(x) = F_{y|x}^{-1}(\tau), \tau\text{-quantile regression curve}
- \ell_n(x) \text{ 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

\[ Y_i = \theta^* + \varepsilon_i, \varepsilon_i \sim ALD_{\tau}(0, 1) \]

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

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

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

\[ \rho_{\tau}(u) \overset{\text{def}}{=} \tau u 1\{u \in (0, \infty)\} - (1 - \tau) u 1\{u \in (-\infty, 0)\} \]
Check Function

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

Local Quantile Regression
Quasi-likelihood Approach

\[ \theta^* = \arg \min_{\theta} \mathbb{E}_{\lambda_0(.)} \rho_\tau(Y - \theta) \]

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

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 \; \text{as} \; y \rightarrow +/ - \infty \; \text{(heavy tails)}. \]
Conditional Quantile Regression

\[ l(x) = \arg\min_{\theta} \mathbb{E}_{\lambda_0(.)} \{ \rho_{\tau}(Y - \theta) | X = x \} \]

\[ l_n(\theta) = \arg\min_{\theta} n^{-1} \sum_{i=1}^{n} \rho_{\tau}(Y_i - \theta) K_{h(x)}(x - X_i) \]

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

- \( L(\theta, \theta') \overset{\text{def}}{=} L(Y, \theta) - L(Y, \theta') \)
Local Polynomial Quantile Regression

\[ l(u) \approx \theta_0(u) + \theta_1(u)(x - u) + \theta_2(u)(x - u)^2 + \ldots + \theta_p(u)(x - u)^p, \]

\[ \tilde{\theta}(u) = \{\tilde{\theta}_0(u), \tilde{\theta}_1(u), \ldots, \tilde{\theta}_p(u)\}^\top \]

\[ = \arg\max_{\theta \in \Theta} \log \tau(1 - \tau) \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} \rho_\tau(Y_i - \theta^\top \psi_i)w_i \]

\[ = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \rho_\tau(Y_i - \theta^\top \psi_i)w_i, \]

where \( \psi_i = \{(X_i - u), (X_i - u)^2, \ldots, (X_i - u)^p\}^\top \) and \( w_i = K\{(x_i - u)/h\} \), for \( i \) in 1, 2, \ldots, \( n \).
First Step to Happiness
Parametric Exponential Bounds

**Theorem 1** Let $V_0 = V(\theta^*)$. Let $R$ satisfy $\lambda^* \geq q R/\mu$ for $\mu = a/(2\nu_0 q^3)$. We have for any $r < R$, if $R \leq \mu \lambda^*/(\nu_0 q)$, then for any $\zeta > 0$ with $2\zeta/a \leq R^2$, it holds

$$P_{\lambda_0(.)} \{ L(\tilde{\theta}, \theta^*) > \zeta \} \leq \exp\{ -\zeta a/(\nu_0 q^3) - \Omega_0(p, q) \} + \exp\{ -g(R) \},$$

(2)

where $g(R)$ is defined as:

$$g(R) \overset{\text{def}}{=} \inf_{\theta \in \Theta} \sup_{\mu \in M} \{ \mu R + C(\mu, \theta, \theta^*) \}$$

$$= \inf_{\theta \in \Theta} g(R, \theta, \theta^*)$$

Local Quantile Regression
Theorem 2 Under technical assumptions,

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

where $\mathcal{R}_r > 0$ is a constant.
Adaptation Scale

Fix $x$, and a sequence of ordered weights

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

Define

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

LPA: $l(X_i) \approx l_\theta(X_i)$

$$\tilde{\theta}_k = \arg\max_\theta \sum_{i=1}^n \ell\{Y_i, l_\theta(X_i)\} w_i^{(k)}$$

$\equiv \arg\max_\theta L(W^{(k)}, \theta)$

Local Quantile Regression
LMS Procedure

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

- 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

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

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

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

Local Quantile Regression

Calibration by "Propagation" Condition
"Propagation" Condition

The critical value is determined via a false alarm consideration:

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

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

Local Quantile Regression
Calibration by "Propagation" Condition

Critical Values

Consider first $z_1$ letting $z_2 = \ldots = z_{K-1} = \infty$. Leads to the estimates $\hat{\theta}_k(z_1)$ for $k = 2, \ldots, K$.

The value $z_1$ is selected as the minimal one for which

$$ \sup_{\theta^*} E_{\theta^*, \lambda(.)} \left| L\{W^{(k)}, \tilde{\theta}_k, \hat{\theta}_k(z_1)\} \right|^r \leq \frac{\alpha}{K-1}, \quad k = 2, \ldots, K. $$

Set $z_{k+1} = \ldots = z_{K-1} = \infty$ and fix $z_k$ gives the set of parameters $z_1, \ldots, z_k, \infty, \ldots, \infty$ and the estimates $\hat{\theta}_m(z_1, \ldots, z_k)$ for $m = k + 1, \ldots, K$. Select $z_k$ s.t.

$$ \sup_{\theta^*} E_{\theta^*, \lambda(.)} \left| L\{W^{(k)}, \tilde{\theta}_m, \hat{\theta}_m(z_1, z_2, \ldots, z_k)\} \right|^r \leq \frac{k\alpha}{K-1}, \quad m = k + 1, \ldots, K. $$

Local Quantile Regression
Figure 8: Critical value with $\lambda(.)$ corresponding to standard normal distribution, $t(3)$, $\text{pareto}(0.5,1)$ $\rho = 0.2$, $r = 0.5$, $\tau = 0.75$. 
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 $\delta_k = a_0 r \log(\rho^{-1}) + a_1 r \log(h_K/h_k) + a_2 \log(nh_k)$. 
"Small Modeling Bias" and Propagation Properties

**Small Modeling Bias**

\[ \Delta_{\lambda_0(.)}(W^{(k)}, \theta) = \sum_{i=1}^{n} \mathcal{K}\{ P_l(X_i), \lambda_0(.) , P_{l\theta}(X_i), \lambda_0(.) \} 1 \{ w_i^{(k)} > 0 \} \]

"Small Modeling Bias" Condition:

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

\[ \mathbb{E}_{\lambda_0(.)} \log \{ 1 + | L(W^{(k)}, \tilde{\theta}_k, \theta)|^r / R_r(W^{(k)}) \} \leq \Delta + 1, \]

Local Quantile Regression
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

$$E_{\lambda_0(.)} \log \left\{ 1 + \frac{|L(W^{(k^*)}, \tilde{\theta}_{k^*}, \theta)|^r}{\mathcal{R}_r(W^{(k^*)})} \right\} \leq \Delta + 1$$

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

$$+ \log \left\{ 1 + \frac{\delta_{k^*}}{\mathcal{R}_r(W^{(k^*)})} \right\}$$
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)
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.

Local Quantile Regression

Local Quantile Regression
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
Assumptions and Definitions

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

and

\[ \mathcal{Q}(p, q) = -\nu_0 q \epsilon^2 / 2 - \log(2 + 2p) - p \mathcal{C}(q) - \log(1 - a^{-(p+1)}) \]
Appendix

Assumptions and Definitions

\[ \zeta(\theta) \overset{\text{def}}{=} L(\theta) - \mathbb{E} L(\theta) \]  
(3)

\[ \zeta(\theta, \theta^*) \overset{\text{def}}{=} \zeta(\theta) - \zeta(\theta^*) \]

\[ \mathcal{M}(\mu, \theta, \theta^*) \overset{\text{def}}{=} - \log \mathbb{E} \exp\{\mu \zeta(\theta, \theta^*)\} \]

\[ \mathcal{M}(\mu, \theta, \theta^*) \overset{\text{def}}{=} - \log \mathbb{E} \exp\{\mu L(\theta, \theta^*)\} \]
Assumptions and Definitions

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

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

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

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

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

\[
-\mathbb{E} L(\theta, \theta^*) \geq a \|V(\theta - \theta^*)\|^2 / 2
\]
References

Z. W. Cai
Regression quantiles for time series

J. Fan and T. C. Hu and Y. K. Troung
Robust nonparametric function estimation

Y. Golubev and V. Spokoiny
Exponential bounds for minimum contrast estimators
References

W. Härdle, P. Janssen and R. Serfling
Strong uniform consistency rates for estimators of conditional functionals

X. He and H. Liang
Quantile regression estimates for a class of linear and partially linear errors-in-variables model

K. Jeong and W. Härdle
References

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

R. Koenker and G. W. Bassett
Regression quantiles
References

R. Koenker and B. J. Park
An interior point algorithm for nonlinear quantile regression

M. G. Lejeune and P. Sarda
Quantile regression: A nonparametric approach

K. Murphy and F. Welch
Empirical age-earnings profiles
References

S. Portnoy and R. Koenker
The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute error estimations (with discussion)

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
References

- K. Yu and M. C. Jones
  Local linear quantile regression

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