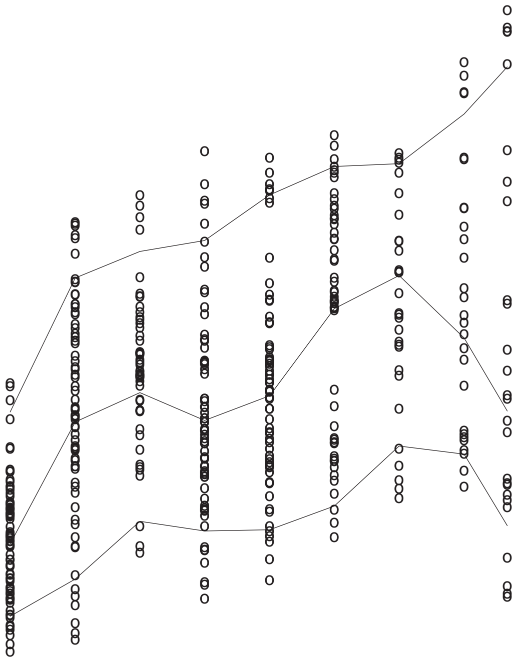
The Stochastic Fluctuation of the Quantile Regression Curve

Wolfgang K. Härdle
Song Song

ISE - Institut für Statistik and Ökonometrie
CASE - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
http://ise.wiwi.hu-berlin.de
http://www.case.hu-berlin.de



- log(Salary) $\sim$ Years
- "The rich got richer and the poor got poorer!"
- Yu et al. (2003)

Conditional Mean Regression

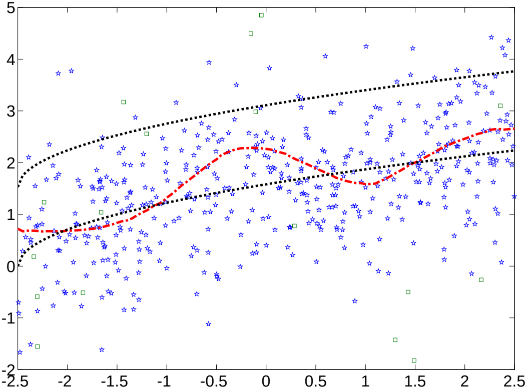


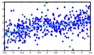
Figure 1: The conditional mean curve, the Nadaraya-Watson estimator and the 0.9-quantile curve. `QR105`

Conditional Quantile Regression

- Median regression = mean regression (symmetric)
- Describes the conditional behavior of a response Y in a broader spectrum
- Stochastic fluctuation effect (open!)
- Parametric model specification (open!)
- ...

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).
  - Causality test, Jeong and Härdle (2008).

Outline

1. Motivation ✓
2. Basic Setup
3. Strong Uniform Consistency Rate
4. Asymptotic Uniform Confidence Band
5. Monte Carlo Simulation
6. Application
7. Further work

Basic Setup

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

Check Function

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

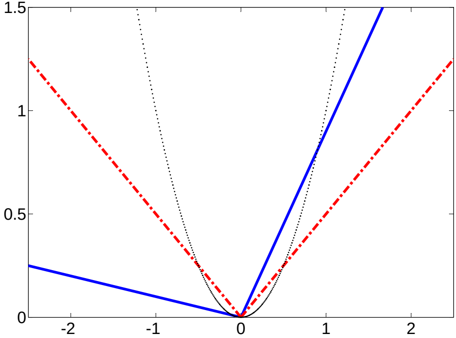
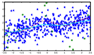


Figure 2: Check function for p=0.9, p=0.5 and weight function in conditional mean regression



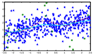
⊡ $l(x)$ minimizes (w.r.t. $\theta \in I = \mathbb{R}$)

$$L(\theta) = \mathsf{E}\{\rho_p(Y - \theta)|X = x\}$$

⊡ $l_n(x)$ minimizes (w.r.t. θ)

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

⊡ $K_h(u) = h^{-1} K(u/h)$ is a kernel with bandwidth h, and has the compact support $[-A, A]$.



- Iteratively reweighted least squares procedure, Lejeune and Sarda (1988), Yu et al. (2003).

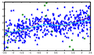
$$L_n(\theta) = n^{-1} \sum_{i=1}^{n} (Y_i - \theta)^2 \left\{ \frac{\rho_p(Y_i - \theta)}{(Y_i - \theta)^2} \right\} K_h(x - X_i)$$

- Define $w_p(Y_i; \theta)$ as $\rho_p(Y_i - \theta)/(Y_i - \theta)^2$

$$L_n(\theta) = n^{-1} \sum_{i=1}^{n} (Y_i - \theta)^2 w_p(Y_i; \theta) K_h(x - X_i)$$

- Integrate w_p into K_h, and rewrite $L_n(\theta)$ as a reweighted sum of squares:

$$L_n(\theta) = n^{-1} \sum_{i=1}^{n} (Y_i - \theta)^2 K_p(x; X_i; Y_i; \theta)$$



Define sequence $l_{n,k}$ (for any initial value $l_{n,1}$) through:

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

Empirically we set $l_{n,1}$ as the global p-quantile.

Theorem

$\exists$ *some* k_0, *s.t.* $l_{n,k}(x) = l_n(x), \forall k \geqslant k_0$.

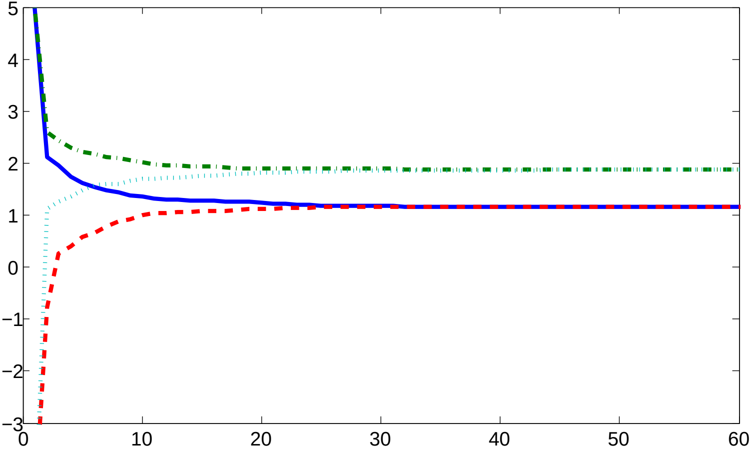


Figure 3: Convergence of $l_{n,k}(x)$ to 0.5-quantile smoother $l_n(x)$ with starting value 5, -5 and x value -1, 1. QR105

Weight Function

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

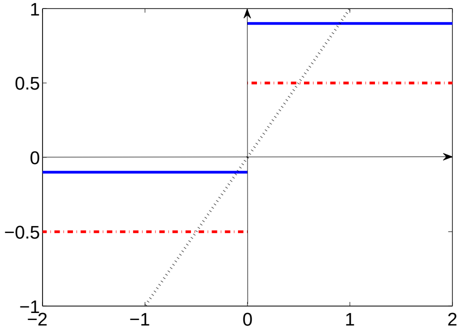
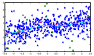


Figure 4: Weight function for p=0.9, p=0.5 and conditional mean regression



The Estimation Idea

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

$$\widetilde{H}_n(\theta, x) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\psi(Y_i - \theta) \tag{3}$$

$$\widetilde{H}(\theta, x) \stackrel{\text{def}}{=} \int_{\mathbb{R}} f(x, y)\psi(y - \theta)dy \tag{4}$$

Outline

1. Motivation $\checkmark$
2. Basic Setup $\checkmark$
3. Strong Uniform Consistency Rate
4. Asymptotic Uniform Confidence Band
5. Monte Carlo Simulation
6. Application
7. Further work

Uniform Strong Consistency Rate

Lemma (Härdle et al. (1988))

For some constant A^* not depending on n, we have a.s. as $n \to \infty$

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

with $f(\cdot|y)$ is ulL-$\tilde{\alpha}$ $(0 < \tilde{\alpha} \leqslant 1)$ on J, uniformly in y.

Require additionally as in Härdle and Luckhaus (1984)

$$\inf_{x \in J} \Big| \int \psi\{y - l(x) + \varepsilon\} dF(y|x) \Big| \geqslant \tilde{q}|\varepsilon|, \quad \text{for } |\varepsilon| \leqslant \delta_1, \qquad (5)$$

with some positive constants $\tilde{q}$ and δ_1.

Theorem

With the additional assumption (5), we have a.s. as $n \to \infty$

$$\sup_{x \in J} |l_n(x) - l(x)| \quad \leq \quad B^* \max\{(nh/\log n)^{-1/2}, h^{\tilde{\alpha}}\} \tag{6}$$

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

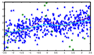
Remark

- Franke and Mwita (2003): without assuming K has compact support, we have a.s. as $n \to \infty$

$$\sup_{x \in J} |l_n(x) - l(x)| \leqslant B^{**}\{(nh/(s_n \log n))^{-1/2} + h^2\}$$

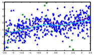
where B^{**} is some constant and $s_n, n \geqslant 1$ is an increasing positive integers sequence, $1 \leqslant s_n \leqslant \frac{n}{2}$ and some other criteria.

- $\{nh/(\log n)\}^{-1/2} \leqslant \{nh/(s_n \log n)\}^{-1/2}$.



Define

$$
\begin{aligned}
H_n(t) &= (nh)^{-1} \sum_{i=1}^{n} K\{(t - X_i)/h\}\psi\{Y_i - l(t)\} \\
D_n(t) &= \partial (nh)^{-1} \sum_{i=1}^{n} K\{(t - X_i)/h\}\psi\{Y_i - \theta\}/\partial\theta \ \{l(t)\} \\
\sigma^2(t) &= \mathsf{E}[\psi^2\{Y - l(t)\}|t] = p(1-p) \\
g'(t) &= \sigma^2(t) f_X(t) = p(1-p) f_X(t) \\
q(t) &= \partial\, \mathsf{E}\{\psi(Y - \theta)|t\}/\partial\theta \ \{l(t)\} \cdot f_X(t) \\
&= f\{l(t)|t\} f_X(t)
\end{aligned}
$$

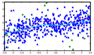


Asymptotic Uniform Confidence Band

Theorem

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

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

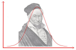


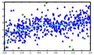
Then

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

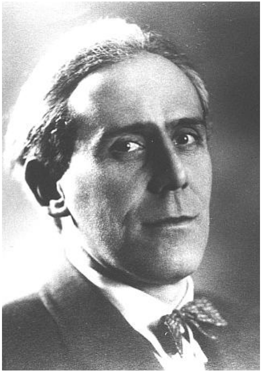
with

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

Emil Julius Gumbel on BBI: 



Emil Julius Gumbel



- Born on 18910718 in München
- Professor of Mathematical Statistics at Heidelberg
- Application of extreme value theory, particularly to climate and hydrology
- Gumbel distribution
- Died on 19660910 in New York

Corollary

An approximate $(1-\alpha)\times 100\%$ *confidence band over* $[0,1]$ *is*

$$\begin{aligned} l_n(t) \quad \pm \quad & (nh)^{-1/2}\{p(1-p)/\hat{f}_X(t)\}^{1/2}\hat{f}^{-1}\{l(t)|t\} \\ & \times\{d_n + c(\alpha)(2\delta\log n)^{-1/2}\}\cdot\{\lambda(K)\}^{1/2}, \end{aligned} \tag{8}$$

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

Remark

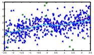
Asymptotic normality at interior points (local constant, local linear, reweighted NW methods, etc.):

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

with $\tau^2(t) = \lambda(K)p(1-p)/[f_X(t)f^2\{l(t)|t\}]$.

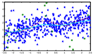
Write (8) as:

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



Optimal Bandwidth Selection

- Minimize the approximation of AMSE (asymptotic mean square error)
- Rule-of-thumb for h_p from Yu and Jones (1998):
  1. Select optimal bandwidth h_{mean} from conditional mean regression
  2. $h_p = [p(1-p)/\varphi^2\{\Phi^{-1}(p)\}]^{1/5} \cdot h_{\text{mean}}$
     with φ, Φ as the pdf and cdf of a standard normal distribution



$$h_p = [p(1-p)/\phi^2\{\Phi^{-1}(p)\}]^{1/5} \cdot h_{\text{mean}}$$

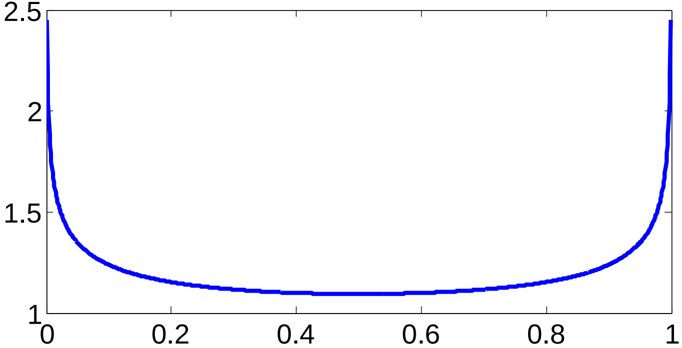


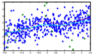
Figure 5: The relationship between h_p and p

Passus Argumentum

1. $l_n(t)$ as a zero (w.r.t. θ) of $\widetilde{H}_n(\theta, t)$, then:

$$\begin{aligned} 0 &= H_n(t) + \{l_n(t) - l(t)\}D_n(t) + \ldots \\ l_n(t) - l(t) &= -\{H_n(t) - \operatorname{E} H_n(t)\}/q(t) - R_n(t) \quad (9) \end{aligned}$$

2. Approximate the leading linear term by a weighted Wiener process similar to Johnston (1982), Bickel and Rosenblatt (1973).



- $\|R_n\| = \sup_{t \in J} |R_n(t)| = \mathcal{O}_p\{(nh \log n)^{-1/2}\}$
- Rosenblatt transformation: $T(x, y) = \{F_{X|y}(x|y), F_Y(y)\}$
- Empirical process: $Z_n(x, y) = n^{1/2}\{F_n(x, y) - F(x, y)\}$
- Brownian bridges: $B_n(x, y) = W_n(x, y) - xyW_n(1, 1)$

Brownian Bridge

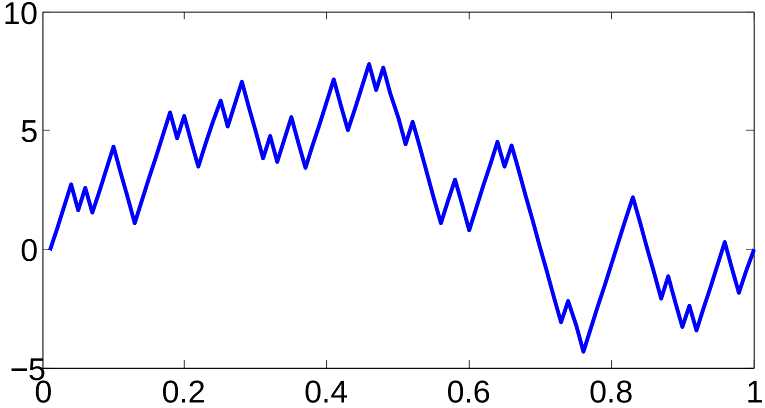
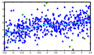


Figure 6: Brownian bridge with $d = 1$ QRbb



Lemma (Tusnady (1977))

On a suitable probability space there exists a sequence of Brownian bridges B_n such that

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

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 $d = 2$, the optimal approximation rate is given in Rio (1996).
For $d > 2$, open problem!

Approximate the Linear Part

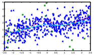
- $Y_n(t) = (nh)^{1/2}\{g'(t)\}^{-1/2}\{H_n(t) - \mathsf{E}\, H_n(t)\}$
- Stochastic integral w.r.t. $Z_n(x, y)$:

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

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

$$\Gamma_n = \{|y| \leqslant a_n\}$$

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



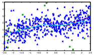
$$Y_{1,n}(t) = \{hg(t)\}^{-1/2} \iint_{\Gamma_n} K\{(t-x)/h\}\psi\{y - l(t)\} dB_n\{T(x,y)\}$$

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

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

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

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



Lemma (Bickel and Rosenblatt (1973))

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.

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

Outline

1. Motivation ✓
2. Basic Setup ✓
3. Strong Uniform Consistency Rate ✓
4. Asymptotic Uniform Confidence Band ✓
5. Monte Carlo Simulation
6. Application
7. Further work

Monte Carlo Simulation

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

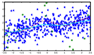
$$\begin{aligned} f(x,y) &= g(y - \sqrt{x+2.5})\mathbf{1}(x \in [-2.5, 2.5]) \qquad (10) \\ g(u) &= \frac{9}{10}\varphi(u) + \frac{1}{90}\varphi(u/9). \end{aligned}$$

$l(x)$ as a zero (w.r.t. θ) of:

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

0.5-quantile curve $l(x) = \sqrt{x+2.5}$, and 0.9-quantile curve $l(x) = 1.5296 + \sqrt{x+2.5}$. We used the quartic kernel

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



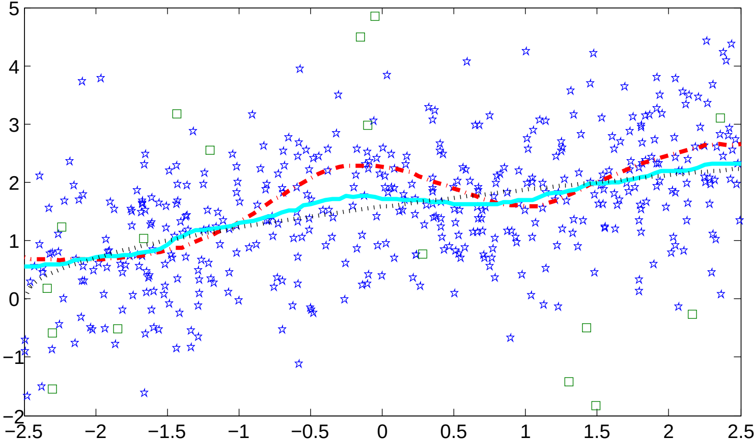


Figure 7: The 0.5-quantile curve, the Nadaraya-Watson estimator $m_n^*(x)$, and the 0.5-quantile smoother $l_n(x)$ with $h_{0.5} = 1.10$. QR105

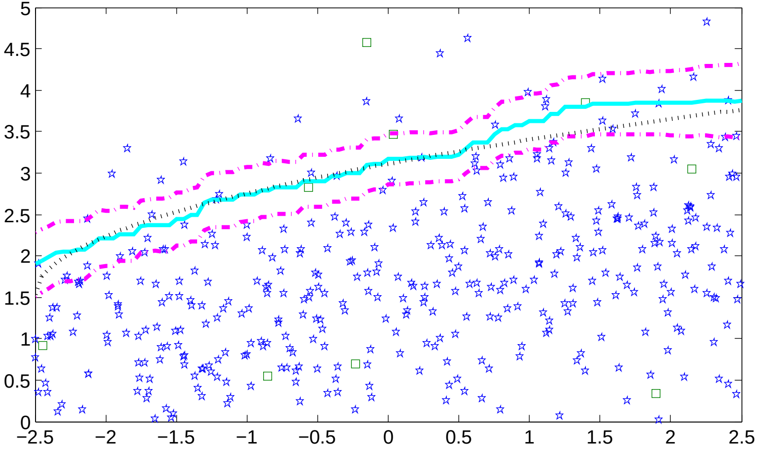
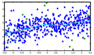


Figure 8: The 0.9-quantile curve, the 0.9-quantile smoother with $h_{0.9} = 1.25$ and 95% confidence bands. `QR1`



Labor Market Application

- For fixed $B(\cdot)$, study nondiscrimination effects in $A(\cdot)$
- Relation: log (Wage) $\sim$ Age
- Data: Current Population Survey (CPS) in 2005
- Male, 25 - 59, full-time, college graduate containing $16,731$ observations

- Conditional mean approach reveals a quadratic relation, Murphy and Welch (1990)
- Conditional quantile approach
- Ages reported as integer values
- Quartic kernel, $h = 2$

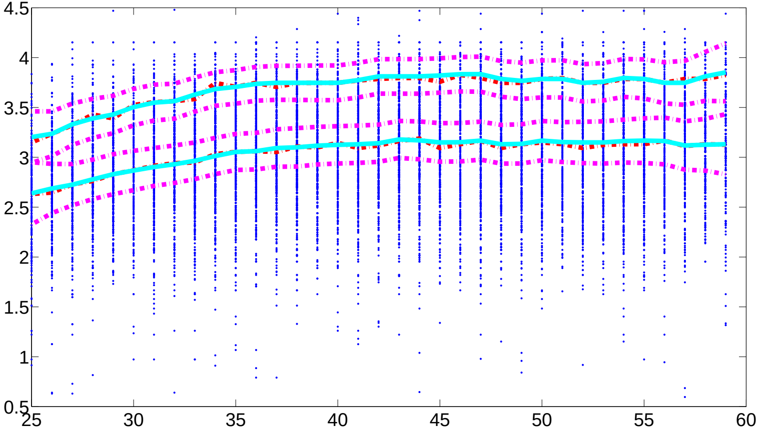


Figure 9: The original observations, local quantiles, 0.5, 0.9-quantile smoothers and corresponding 95% confidence bands. QRCPS

Parametric model specification test

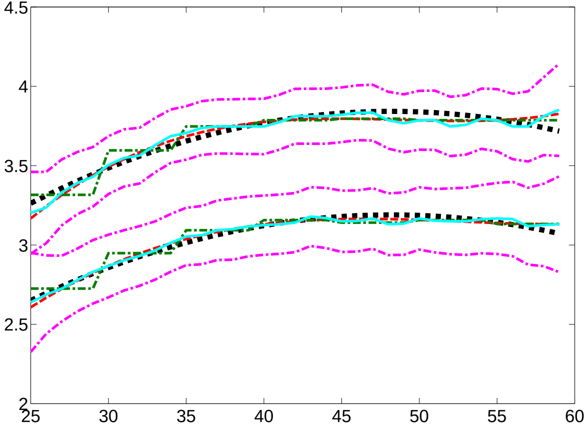
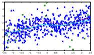


Figure 10: Quadratic, quartic, set of dummies (for age groups) estimates, quantile smoothers and corresponding 95% confidence bands. QRCPS

Conclusion

- At the 5% significance level
  - quadratic $\surd$, quartic $\surd$
  - set of dummies (for age groups) $\surd$
- At the 10% significance level
  - quadratic $\surd$, quartic $\surd$
  - set of dummies (for age groups) $\times$
- Suggest quadratic model measure wage-earning relation for simplicity.



Further work

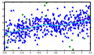
- Panel data/Partial linear quantile regression
- Semiparametric quantile regression adequacy checking $\beta = 0$?
  - Semiparametric quantile regression from Yu
  - Semiparametric mean regression adequacy checking of Zhu, Zhu and Song (2008)
- Nonparametric local adaptive quantile regression
  - Local adaptive mean regression from Spokoiny (2008)
- Strengthen errors-in-variables quantile regression from Fan
  - Linear and partial linear errors-in-variables quantile regression of He and Liang (2000)
- ...

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.

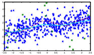


References

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.

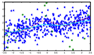


References

- 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.

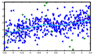
M. G. Lejeune and P. Sarda
Quantile regression: A nonparametric approach
Computational Statistics & Data Analysis, 6:229-239 1988.

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.