# Confidence Corridors for Multivariate Generalized Quantile Regression

Shih-Kang Chao, Katharina Proksch
Wolfgang Karl Härdle
Holger Dette

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and
Economics
Humboldt-Universität zu Berlin
Chair of Stochastic
Ruhr-Universität Bochum
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://www.ruhr-uni-bochum.de/mathematik3

# Does treatment help?



(a) Before

(b) After

# Stochastic dominance (SD)

Conditional stochastic dominance (CSD)(Delgado and Escanciano, 2013) : $Y_1$ conditionally stochastically dominates $Y_0$:
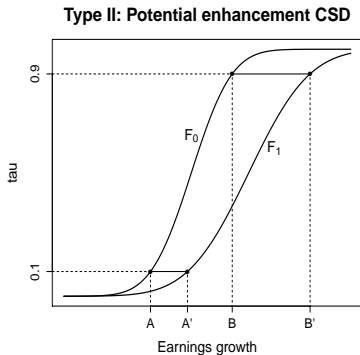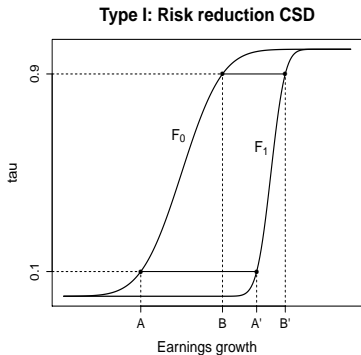
$$F_{1|\boldsymbol{X}}(y|\boldsymbol{x}) \leq F_{0|\boldsymbol{X}}(y|\boldsymbol{x}) \quad \text{a.s. for all } y, \boldsymbol{x}, \tag{1}$$

Take $y = F_{0|\boldsymbol{X}}^{-1}(\tau|\boldsymbol{x})$. Apply $F_{1|\boldsymbol{X}}^{-1}$ to (1):

$$F_{0|\boldsymbol{X}}^{-1}(\tau|\boldsymbol{x}) \leq F_{1|\boldsymbol{X}}^{-1}\big\{F_{0|\boldsymbol{X}}(y|\boldsymbol{x})\big|\boldsymbol{x}\big\} = F_{1|\boldsymbol{X}}^{-1}(\tau|\boldsymbol{x}) \quad \forall \boldsymbol{x}, \tau$$
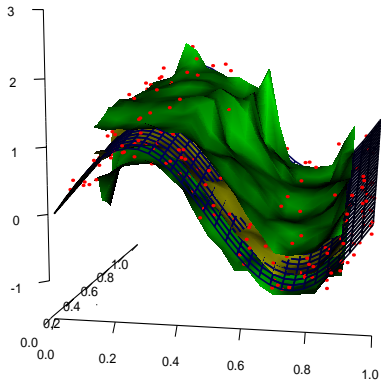
# Which one helps more?

# Confidence corridors (CC)
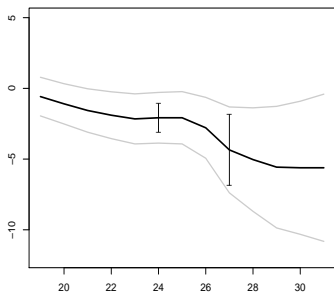


(c) $d = 1$

(d) $d = 2$

# Distribution of the maximal deviation

$$\mathsf{P}\left\{\hat{\theta}_n(\boldsymbol{x}) - \Phi_n(\boldsymbol{x}) \le \theta_0(\boldsymbol{x}) \le \hat{\theta}_n(\boldsymbol{x}) + \Phi_n(\boldsymbol{x}) \,\forall\, \boldsymbol{x} \in \mathcal{D}\right\} = 1 - \alpha$$

$$(2)$$

Find $\Phi_n$ such that the equality (2) holds approximately. Suppose:

$$\mathsf{P}\left\{\sup_{\boldsymbol{x}\in\mathcal{D}} w_n(\boldsymbol{x})\,|\hat{\theta}_n(\boldsymbol{x}) - \theta_0(\boldsymbol{x})| > c(\alpha)\right\} = \alpha$$

Then, with probability $1 - \alpha$,

$$|\hat{\theta}_n(\boldsymbol{x}) - \theta_0(\boldsymbol{x})| \le \frac{c(\alpha)}{w_n(\boldsymbol{x})} \stackrel{\text{def}}{=} \Phi_n(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in \mathcal{D}.$$

Approximation of the distribution of $\sup_{\boldsymbol{x}\in\mathcal{D}} w_n(\boldsymbol{x})\,|\hat{\theta}_n(\boldsymbol{x}) - \theta_0(\boldsymbol{x})|$.

# Confidence corridors: a history



**CPHD**
**GQR estimator: $d \geq 1$**

Smirnov (1950)
Histogram Estimator

Bickel & Rosenblatt (1973b)
density estimation: $d = 1$

Bickel & Rosenblatt (1973a)
Max Dev. Dist. Random Field

Rosenblatt (1976)
density estimation: $d \geq 1$

Guo & Härdle (2011)
ER estimator: $d = 1$

Härdle & Song (2010)
QR estimator: $d = 1$

Hall (1991)
Bootstrap density esti.

Johnston (1982)
function estimation: $d = 1$

Härdle (1989)
$M$-estimator: $d = 1$
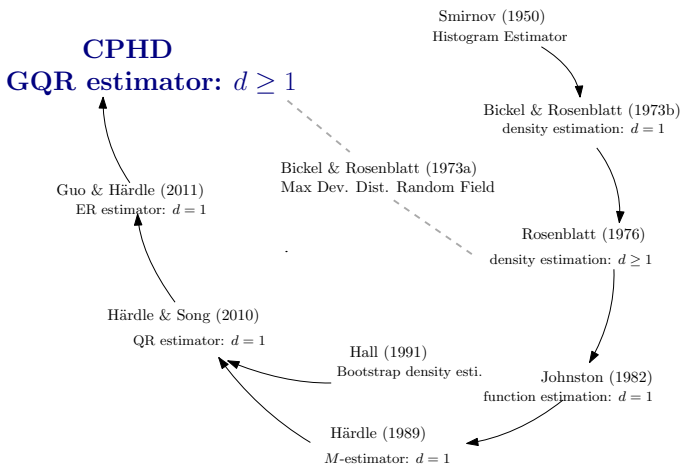
# The story continues

- ⊡ Claeskens and van Keilegom (2003): local polynomial mean estimator
- ⊡ Gené and Nickl (2010): adaptive density estimation with wavelets and kernel
- ⊡ Liu and Wu (2010): long memory, strictly stationary time series density estimation
- ⊡ Fan and Liu (2013): one dimensional, generic (semi)parametric quantile estimator, avoid estimating conditional density

# Outline

# Additive error model

- ⊡ Let $(\boldsymbol{X}_1, Y_1), ..., (\boldsymbol{X}_n, Y_n)$ be a sequence of i.i.d. random vectors in $\mathbb{R}^{d+1}$ and consider the nonparametric regression model

$$Y_i = \theta(\boldsymbol{X}_i) + \varepsilon_i, \quad i = 1, ..., n, \qquad (3)$$

  where $\theta$ is an aspect of $Y$ conditional on $\boldsymbol{X}$ such as the $\tau$-quantile, the $\tau$-expectile regression curve, $\varepsilon_i$ i.i.d. with $\tau$-quantile/expectile 0.

- ⊡ Heterogeneity: $\varepsilon_i$ is allowed to be correlated with $\boldsymbol{X}$

# Estimator and Bahadur representation

⊡ Consider the local constant estimator

$$\hat{\theta}_n(\boldsymbol{x}) \stackrel{\text{def}}{=} \arg\min_{\theta} n^{-1} \sum_{i=1}^{n} K_h(\boldsymbol{x} - \boldsymbol{X}_i) \rho(Y_i - \theta)$$

with a kernel $K$ and loss function $\rho$. ▸ Notations

⊡ Uniform nonparametric Bahadur representation by Kong et al. (2010):

$$\sup_{\boldsymbol{x} \in \mathcal{D}} \left| \hat{\theta}_n(\boldsymbol{x}) - \theta_0(\boldsymbol{x}) - \frac{1}{nS_{n,0,0}(\boldsymbol{x})} \sum_{i=1}^{n} K_h(\boldsymbol{x} - \boldsymbol{X}_i) \psi_\tau \big\{ Y_i - \theta_0(\boldsymbol{x}) \big\} \right|$$

$$= \mathcal{O} \left\{ (\log n / nh^d)^{3/4} \right\}, \quad a.s.[P]$$

# Approximating empirical process

$$V_n^{-1/2} S_{n,0,0}(\boldsymbol{x}) \left\{ \hat{\theta}_n(\boldsymbol{x}) - \theta_0(\boldsymbol{x}) \right\}$$

$$\approx V_n^{-1/2} \frac{1}{n} \sum_{i=1}^{n} K_h(\boldsymbol{x} - \boldsymbol{X}_i) \psi \{ Y_i - \theta_0(\boldsymbol{x}) \}$$

$$\approx \underbrace{\frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x}) \sigma^2(\boldsymbol{x})}} \int \int K\left( \frac{\boldsymbol{x} - \boldsymbol{u}}{h} \right) \psi \{ y - \theta_0(\boldsymbol{x}) \} dZ_n(y, \boldsymbol{u})}_{Y_n(\boldsymbol{x})}$$

⊡ with the centered empirical process
$$Z_n(y, \boldsymbol{u}) \stackrel{\text{def}}{=} n^{1/2} \{ F_n(y, \boldsymbol{u}) - F(y, \boldsymbol{u}) \}.$$

# Maximal deviation for nonparametric QR

### Theorem (1)

*Under regularity conditions,* $vol(\mathcal{D}) = 1$,

▸ Notations    ▸ Assumptions

$$P\left\{(2d\kappa\log n)^{1/2}\left(\sup_{\boldsymbol{x}\in\mathcal{D}}\left[r(\boldsymbol{x})|\hat{\theta}_n(\boldsymbol{x}) - \theta_0(\boldsymbol{x})|\right]/\|K\|_2 - d_n\right) < a\right\}$$

$$\rightarrow \exp\left\{-2\exp(-a)\right\},$$

*as* $n \rightarrow \infty$, *where* $\hat{\theta}_n(\boldsymbol{x})$ *and* $\theta_0(\boldsymbol{x})$ *are the local constant quantile estimator and the true quantile function.*

## Corollary (RQ-CC)

*Under the assumptions of Theorem 1, an approximate $(1 - \alpha) \times 100\%$ confidence corridor over $\alpha \in (0, 1)$ is*

$$\hat{\theta}_n(t) \pm (nh^d)^{-1/2} \big\{ \tau(1 - \tau)\|K\|_2 / \hat{f}_{\boldsymbol{X}}(t) \big\}^{1/2} \hat{f}_{\varepsilon|\boldsymbol{X}} \{0|t\}^{-1}$$
$$\big\{ d_n + c(\alpha)(2d\kappa \log n)^{-1/2} \big\},$$

*where $c(\alpha) = \log 2 - \log \big| \log(1 - \alpha) \big|$ and $\hat{f}_{\boldsymbol{X}}(t)$, $\hat{f}_{\varepsilon|\boldsymbol{X}} \{0|t\}$ are consistent estimates for $f_{\boldsymbol{X}}(t)$, $f_{\varepsilon|\boldsymbol{X}} \{0|t\}$.*

# Maximal deviation for nonparametric ER

## Theorem (2)

*Under regularity conditions, $vol(\mathcal{D}) = 1$,*   ▸ Notations   ▸ Assumptions

$$\mathsf{P}\left\{(2d\kappa\log n)^{1/2}\left(\sup_{\boldsymbol{x}\in\mathcal{D}}\left[r(\boldsymbol{x})|\hat{\theta}_n(\boldsymbol{x}) - \theta_0(\boldsymbol{x})|\right]/\|K\|_2 - d_n\right) < a\right\}$$
$$\rightarrow \exp\left\{-2\exp(-a)\right\},$$

*as $n \rightarrow \infty$, where $\hat{\theta}_n(\boldsymbol{x})$ and $\theta_0(\boldsymbol{x})$ are the local constant expectile estimator and the true expectile function.*

## Corollary (RE-CC)

*Under the assumptions of Theorem 3, an approximate $(1-\alpha) \times 100\%$ confidence corridor over $\alpha \in (0,1)$ is*

$$\hat{\theta}_n(t) \pm (nh^d)^{-1/2} \big\{ \hat{\sigma}^2(x)\|K\|_2/\hat{f}_X(t) \big\}^{1/2} 0.5 \big[\tau - \hat{F}_{\varepsilon|X}(0|x)(2\tau-1)\big]^{-1}$$
$$\big\{ d_n + c(\alpha)(2d\kappa \log n)^{-1/2} \big\},$$

*where $c(\alpha) = \log 2 - \log\big|\log(1-\alpha)\big|$ and $\hat{f}_X(t)$, $\hat{\sigma}^2(x)$ and $\hat{F}_{\varepsilon|X}(0|x)$ are consistent estimates for $f_X(t)$, $\sigma^2(x)$ and $F_{\varepsilon|X}(0|x)$*

# Bootstrap

⊡ Smooth bootstrap:

$$\hat{f}_{\varepsilon,\boldsymbol{X}}(v,\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} g_{h_0}\left(v - \hat{\varepsilon}_i\right) L_{\bar{h}}(\boldsymbol{x} - \boldsymbol{X}_i),, \tag{4}$$

where $g$ and $L$ are kernels and $h_0, \bar{h} \to 0$, $nh_0\bar{h}^d \to \infty$

⊡ Define

$$\hat{\theta}^*(\boldsymbol{x}) - \hat{\theta}_n(\boldsymbol{x}) \tag{5}$$

$$= \frac{1}{n\hat{S}_{n,0,0}(\boldsymbol{x})}\sum_{i=1}^{n} K_h(\boldsymbol{x} - \boldsymbol{X}_i^*)\psi(\varepsilon_i^*) \underbrace{-\mathsf{E}^*\left[K_h(\boldsymbol{x} - \boldsymbol{X}_i^*)\psi(\varepsilon_i^*)\right]}_{\text{Remove the bias}},$$

$$\hat{S}_{n,0,0}(\boldsymbol{x}) = \begin{cases} \hat{f}_{\varepsilon|\boldsymbol{X}}(0|\boldsymbol{x})\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}), & \text{quantile case;} \\ 2\big[\tau - \hat{F}_{\varepsilon|\boldsymbol{X}}(0|\boldsymbol{x})(2\tau - 1)\big]\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}), & \text{expectile case.} \end{cases}$$

(6)

## Theorem (Bootstrap)

*Under regularity conditions, let*    `▸ Assumptions`

$$r^*(\boldsymbol{x}) = \sqrt{\frac{nh^d}{\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})\sigma_*^2(\boldsymbol{x})}}\,\hat{S}_{n,0,0}(\boldsymbol{x}),$$

*Then as $n \to \infty$,*

$$\mathsf{P}^*\left\{(2d\kappa \log n)^{1/2}\left(\sup_{\boldsymbol{x}\in\mathcal{D}}\left[r^*(\boldsymbol{x})|\hat{\theta}^*(\boldsymbol{x}) - \hat{\theta}_n(\boldsymbol{x})|\right]/\|K\|_2 - d_n\right) < a\right\}$$
$$\to \exp\left[-2\exp(-a)\right], \ a.s.$$

## Lemma
*Under regularity conditions, $\|\sigma_*^2(\boldsymbol{x}) - \hat{\sigma}^2(\boldsymbol{x})\| = \mathcal{O}_p^*\left((\log n)^{-1/2}\right)$,*
*a.s.*

## Corollary

*Under the regularity conditions, the bootstrap confidence set is defined by*

$$\left\{ \theta : \sup_{x \in \mathcal{D}} \left| \frac{\hat{S}_{n,0,0}(x)}{\sqrt{\hat{f}_{X}(x)\hat{\sigma}^2(x)}} \left[ \hat{\theta}_n(x) - \theta(x) \right] \right| \leq \xi_\alpha \right\}, \qquad (7)$$

*where $\xi_\alpha$ satisfies*

$$\mathrm{P}^* \left( \sup_{x \in \mathcal{D}} \left| \frac{\hat{S}_{n,0,0}(x)}{\sqrt{\hat{f}_{X}(x)\hat{\sigma}^2(x)}} \left[ \hat{\theta}^*(x) - \hat{\theta}_n(x) \right] \right| \leq \xi_\alpha \right) = 1 - \alpha,$$

*where $\alpha$ is the level of the test and $\hat{S}_{n,0,0}$ is defined as in* (6).

# Implementing Bootstrap CC for QR

Observation:

$$\hat{f}_{\varepsilon|\boldsymbol{X}}(0|\boldsymbol{x}) = n^{-1} \sum_{i=1}^{n} g_{h_0}\left(\hat{\varepsilon}_i\right) L_{\bar{h}}(\boldsymbol{x} - \boldsymbol{X}_i)/\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}) \tag{8}$$

$$\hat{f}_{Y|\boldsymbol{X}}(\hat{\theta}_n(\boldsymbol{x})|\boldsymbol{x}) = n^{-1} \sum_{i=1}^{n} g_{h_1}\left(Y_i - \hat{\theta}_n(\boldsymbol{x})\right) L_{\tilde{h}}(\boldsymbol{x} - \boldsymbol{X}_i)/\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}), \tag{9}$$

are NOT equivalent in finite sample, and $\hat{f}_{Y|\boldsymbol{X}}(\hat{\theta}_n(\boldsymbol{x})|\boldsymbol{x})$ accounts more for the deviation between $\theta_0$ and $\hat{\theta}$

# Bootstrap CC for QR

Hence, we propose to construct CC for QR by

$$\left\{ \theta : \sup_{\boldsymbol{x} \in \mathcal{D}} \left| \sqrt{\hat{f}_{\boldsymbol{X}}(\boldsymbol{x})} \hat{f}_{Y|\boldsymbol{X}} \{ \hat{\theta}_n(\boldsymbol{x}) | \boldsymbol{x} \} \left[ \hat{\theta}_n(\boldsymbol{x}) - \theta(\boldsymbol{x}) \right] \right| \leq \xi_\alpha^\dagger \right\},$$

where $\xi_\alpha^\dagger$ satisfies

$$\mathsf{P}^* \left( \sup_{\boldsymbol{x} \in \mathcal{D}} \left| \hat{f}_{\boldsymbol{X}}(\boldsymbol{x})^{-1/2} \frac{\hat{f}_{Y|\boldsymbol{X}} \{ \hat{\theta}_n(\boldsymbol{x}) | \boldsymbol{x} \}}{\hat{f}_{\varepsilon|\boldsymbol{X}}(0|\boldsymbol{x})} \left[ A_n^*(\boldsymbol{x}) - \mathsf{E}^* A_n^*(\boldsymbol{x}) \right] \right| \leq \xi_\alpha^\dagger \right) \approx 1 - \alpha.$$

# Simulated coverage probabilities

Generating process: $d = 2$

$$Y_i = f(X_{1,i}, X_{2,i}) + \sigma(X_{1,i}, X_{2,i})\varepsilon_i,$$

- $f(x_1, x_2) = \sin(2\pi x_1) + x_2$.
- $(X_1, X_2)$ supported on $[0, 1]^2$ with corr. $= 0.2876$ ▸ Sample Method
- $\varepsilon_i \sim N(0, 1)$ i.i.d.
- Specification for $\sigma(X_1, X_2)$:
  - ▶ Homogeneity: $\sigma(X_1, X_2) = \sigma_0$, for $\sigma_0 = 0.2, 0.5, 0.7$
  - ▶ Heterogeneity:

  $$\sigma(X_1, X_2) = \sigma_0 + 0.8X_1(1 - X_1)X_2(1 - X_2)$$

  for $\sigma_0 = 0.2, 0.5, 0.7$

Table 1: Nonparametric quantile model asymptotic coverage probability. Nominal coverage is 95%. The "()" indicate the volume.

| | Homogeneous | | | Heterogeneous | | |
|---|---|---|---|---|---|---|
| $n$ | $\tau = 0.5$ | $\tau = 0.2$ | $\tau = 0.8$ | $\tau = 0.5$ | $\tau = 0.2$ | $\tau = 0.8$ |
| | | | $\sigma_0 = 0.2$ | | | |
| 100 | .000(0.366) | .109(0.720) | .104(0.718) | .000(0.403) | .120(0.739) | .122(0.744) |
| 300 | .000(0.304) | .130(0.518) | .133(0.519) | .002(0.349) | .136(0.535) | .153(0.537) |
| 500 | .000(0.262) | .117(0.437) | .142(0.437) | .008(0.296) | .156(0.450) | .138(0.450) |
| | | | $\sigma_0 = 0.5$ | | | |
| 100 | .070(0.890) | .269(1.155) | .281(1.155) | .078(0.932) | .300(1.193) | .302(1.192) |
| 300 | .276(0.735) | .369(0.837) | .361(0.835) | .325(0.782) | .380(0.876) | .394(0.877) |
| 500 | .364(0.636) | .392(0.711) | .412(0.712) | .381(0.669) | .418(0.743) | .417(0.742) |
| | | | $\sigma_0 = 0.7$ | | | |
| 100 | .160(1.260) | .381(1.522) | .373(1.519) | .155(1.295) | .364(1.561) | .373(1.566) |
| 300 | .438(1.026) | .450(1.109) | .448(1.110) | .481(1.073) | .457(1.155) | .472(1.152) |
| 500 | .533(0.888) | .470(0.950) | .480(0.949) | .564(0.924) | .490(0.984) | .502(0.986) |

Table 2: Nonparametric quantile model bootstrap coverage probability. Nominal coverage is 95%. The "()" the volume.

| | Homogeneous | | | Heterogeneous | | |
|---|---|---|---|---|---|---|
| $n$ | $\tau = 0.5$ | $\tau = 0.2$ | $\tau = 0.8$ | $\tau = 0.5$ | $\tau = 0.2$ | $\tau = 0.8$ |
| | | | $\sigma_0 = 0.2$ | | | |
| 100 | .325(0.676) | .784(0.954) | .783(0.954) | .409(0.717) | .779(0.983) | .778(0.985) |
| 300 | .442(0.457) | .896(0.609) | .894(0.610) | .580(0.504) | .929(0.650) | .922(0.649) |
| 500 | .743(0.411) | .922(0.502) | .921(0.502) | .839(0.451) | .950(0.535) | .952(0.536) |
| | | | $\sigma_0 = 0.5$ | | | |
| 100 | .929(1.341) | .804(1.591) | .818(1.589) | .938(1.387) | .799(1.645) | .773(1.640) |
| 300 | .950(0.920) | .918(1.093) | .923(1.091) | .958(0.973) | .919(1.155) | .923(1.153) |
| 500 | .988(0.861) | .968(0.943) | .962(0.942) | .990(0.902) | .962(0.986) | .969(0.987) |
| | | | $\sigma_0 = 0.7$ | | | |
| 100 | .976(1.811) | .817(2.112) | .808(2.116) | .981(1.866) | .826(2.178) | .809(2.176) |
| 300 | .986(1.253) | .919(1.478) | .934(1.474) | .983(1.308) | .930(1.537) | .920(1.535) |
| 500 | .996(1.181) | .973(1.280) | .968(1.278) | .997(1.225) | .969(1.325) | .962(1.325) |

Table 3: Nonparametric expectile model asymptotic coverage probability. Nominal coverage is 95%. The "()" indicate the volume.

| | Homogeneous | | | Heterogeneous | | |
|---|---|---|---|---|---|---|
| $n$ | $\tau = 0.5$ | $\tau = 0.2$ | $\tau = 0.8$ | $\tau = 0.5$ | $\tau = 0.2$ | $\tau = 0.8$ |
| | | | $\sigma_0 = 0.2$ | | | |
| 100 | .000(0.428) | .000(0.333) | .000(0.333) | .000(0.463) | .000(0.362) | .000(0.361) |
| 300 | .049(0.341) | .000(0.273) | .000(0.273) | .079(0.389) | .001(0.316) | .002(0.316) |
| 500 | .168(0.297) | .000(0.243) | .000(0.243) | .238(0.336) | .003(0.278) | .002(0.278) |
| | | | $\sigma_0 = 0.5$ | | | |
| 100 | .007(0.953) | .000(0.776) | .000(0.781) | .007(0.997) | .000(0.818) | .000(0.818) |
| 300 | .341(0.814) | .019(0.708) | .017(0.709) | .355(0.862) | .017(0.755) | .018(0.754) |
| 500 | .647(0.721) | .067(0.645) | .065(0.647) | .654(0.759) | .061(0.684) | .068(0.684) |
| | | | $\sigma_0 = 0.7$ | | | |
| 100 | .012(1.324) | .000(1.107) | .000(1.107) | .010(1.367) | .000(1.145) | .000(1.145) |
| 300 | .445(1.134) | .021(1.013) | .013(1.016) | .445(1.182) | .017(1.062) | .016(1.060) |
| 500 | .730(1.006) | .062(0.928) | .078(0.929) | .728(1.045) | .068(0.966) | .066(0.968) |

Table 4: Nonparametric expectile model bootstrap coverage probability. Nominal coverage is 95%. The "()" indicate the volume.

| | Homogeneous | | | Heterogeneous | | |
|---|---|---|---|---|---|---|
| $n$ | $\tau = 0.5$ | $\tau = 0.2$ | $\tau = 0.8$ | $\tau = 0.5$ | $\tau = 0.2$ | $\tau = 0.8$ |
| | | | $\sigma_0 = 0.2$ | | | |
| 100 | .686(2.191) | .781(2.608) | .787(2.546) | .706(2.513) | .810(2.986) | .801(2.943) |
| 300 | .762(0.584) | .860(0.716) | .876(0.722) | .788(0.654) | .877(0.807) | .887(0.805) |
| 500 | .771(0.430) | .870(0.533) | .875(0.531) | .825(0.516) | .907(0.609) | .904(0.615) |
| | | | $\sigma_0 = 0.5$ | | | |
| 100 | .886(5.666) | .906(6.425) | .915(6.722) | .899(5.882) | .927(6.667) | .913(6.571) |
| 300 | .956(1.508) | .958(1.847) | .967(1.913) | .965(1.512) | .962(1.866) | .969(1.877) |
| 500 | .968(1.063) | .972(1.322) | .972(1.332) | .972(1.115) | .971(1.397) | .974(1.391) |
| | | | $\sigma_0 = 0.7$ | | | |
| 100 | .913(7.629) | .922(8.846) | .935(8.643) | .929(8.039) | .935(9.057) | .932(9.152) |
| 300 | .969(2.095) | .969(2.589) | .971(2.612) | .974(2.061) | .972(2.566) | .979(2.604) |
| 500 | .978(1.525) | .976(1.881) | .967(1.937) | .981(1.654) | .978(1.979) | .974(2.089) |

# Application to NSW demonstration data

- ⊡ National Supported Work (NSW): a randomized, temporary employment program carried out in the US in 1970s to help the disadvantages workers

- ⊡ 297 obs. treatment group; 425 obs. control group, all male

- ⊡ Lalonde (1986), Dehejia and Wahba (1999)

- ⊡ Delgado and Escanciano (2013): heterogeneity effect in age; nonnegative treatment effect

- ⊡ $X_1$: Age; $X_2$: schooling in years; $Y$: Earning difference 78-75 (in thousand \$)

- ⊡ Bootstrap: 10,000 repetition

# CSD revisits
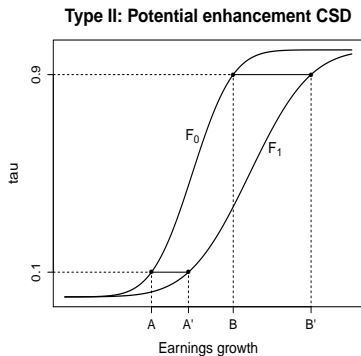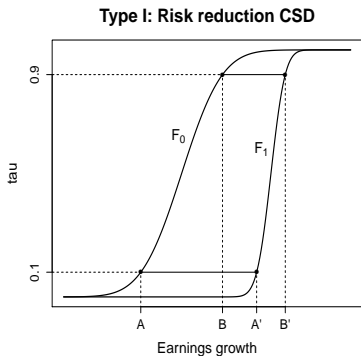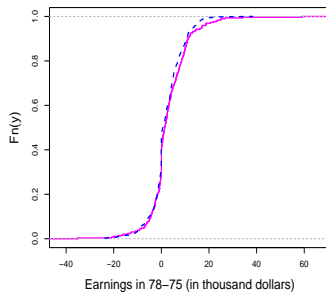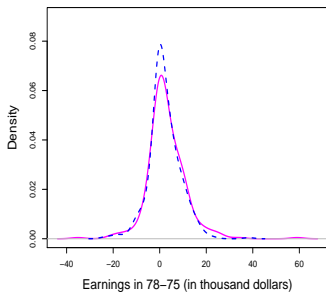
| $\tau$(%) | 10 | 20 | 30 | 50 | 70 | 80 | 90 |
|---|---|---|---|---|---|---|---|
| Treatment | -4.38 | -1.55 | 0.00 | 1.40 | 5.48 | 8.50 | 11.15 |
| Control | -4.91 | -1.73 | -0.17 | 0.74 | 4.44 | 7.16 | 10.56 |

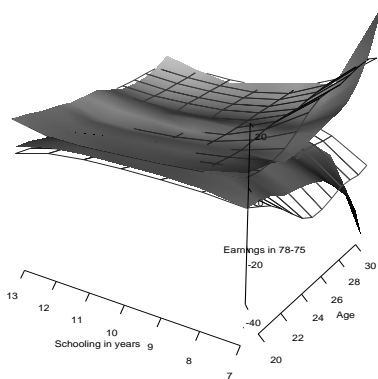Unconditional kernel densities. Magenta: treatment group. Blue: control group. $h_{tr} = 1.652$. $h_{co} = 1.231$.
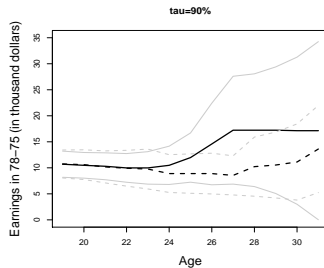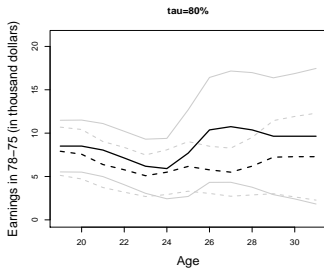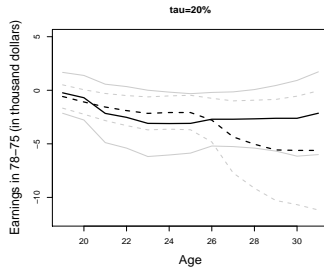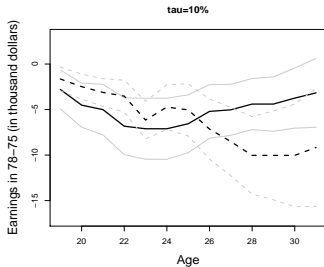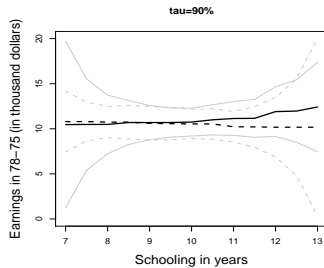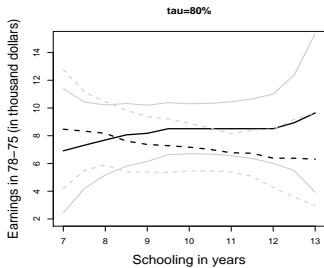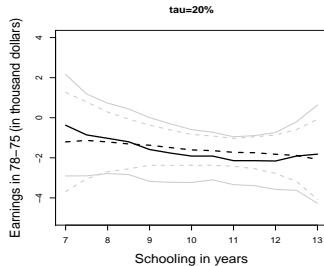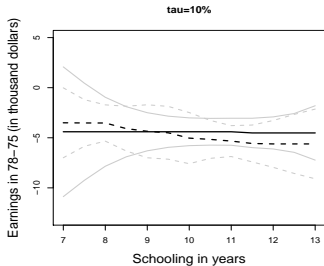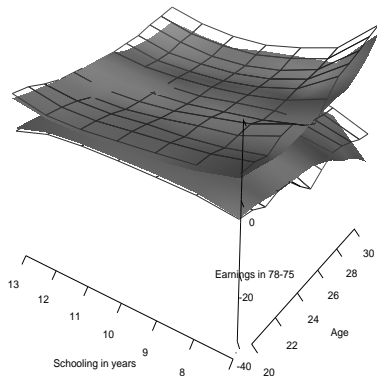
(e) $\tau = 10\%$           (f) $\tau = 20\%$

Figure 1: Confidence corridors for treatment (net surfaces) and control group (solid surfaces).
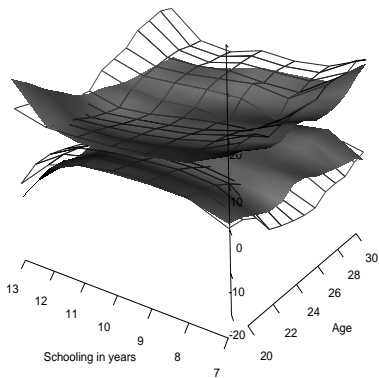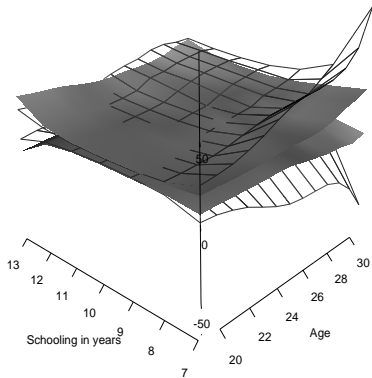
(a) $\tau = 80\%$        (b) $\tau = 90\%$

Figure 2: Confidence corridors for treatment (net surfaces) and control group (solid surfaces).
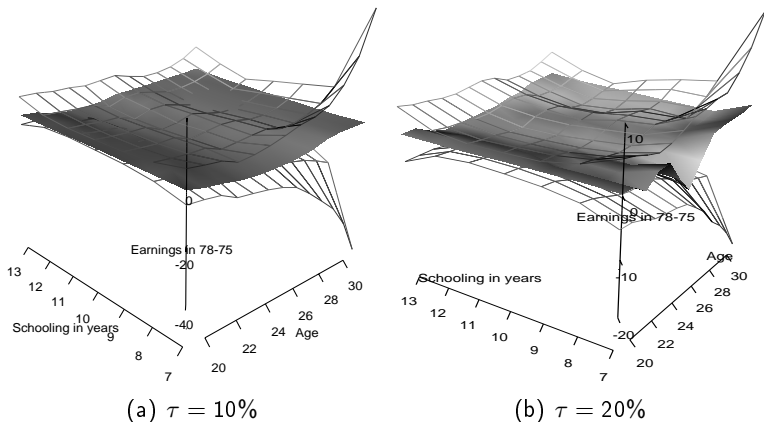
Figure 3: Quantile estimates for treatment (solid surfaces) and confidence corridors control group (net surfaces).
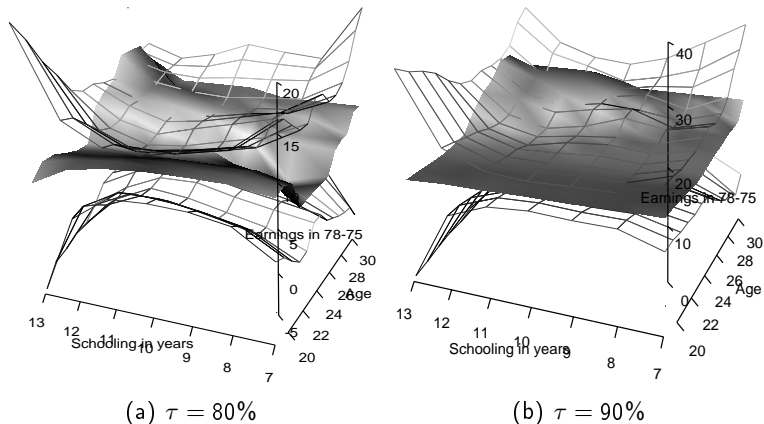
(a) $\tau = 80\%$  (b) $\tau = 90\%$

Figure 4: Quantile estimates for treatment (solid surfaces) and confidence corridors control group (net surfaces).

# Summary

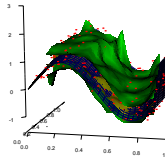- ⊡ The nonnegative CSD is not rejected, confirming the findings of Delgado and Escanciano (2013)

- ⊡ Heterogeneous effect in age and schooling in years: individuals who are older and spend more time in the school benefit more from the treatment

- ⊡ We show: treatment raises the potential for realizing higher earnings growth, but does little in reducing the risk of realizing low earnings growth

# Confidence Corridors for Multivariate Generalized Quantile Regression

Shih-Kang Chao, Katharina Proksch
Wolfgang Karl Härdle
Holger Dette

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
Chair of Stochastic
Ruhr-Universität Bochum
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://www.ruhr-uni-bochum.de/mathematik3

# Assumptions

(A1)  $K$ is of order $s-1$ (see (A3))has bounded support $[-A, A]^d$, continuously differentiable up to order $d$ (and are bounded); i.e. $\partial^\alpha K \in L^1(\mathbb{R}^d)$ exists and is continuous for all multi-indices $\in \{0, 1\}^d$

(A2)  The increasing sequence $\{a_n\}_{n=1}^\infty$ satisfies

$$(\log n)h^{-3d} \int_{|y|>a_n} f_Y(y)dy = \mathcal{O}(1) \tag{10}$$

and

$$(\log n)h^{-d} \int_{|y|>a_n} f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})dy = \mathcal{O}(1), \text{ for all } \boldsymbol{x} \in \mathcal{D}$$

as $n \to \infty$ hold.

(A3)  The true function $\theta_0(\boldsymbol{x})$ is continuously differentiable and is in Hölder class with order $s > d$.

# Assumptions

(A4) $f_{\boldsymbol{X}}(\boldsymbol{x})$ is continuously differentiable and its gradient is uniformly bounded. In particular, $\inf_{\boldsymbol{x} \in \mathcal{D}} f_{\boldsymbol{X}}(\boldsymbol{x}) > 0$.

(A5) The joint probability density function $f(y, \boldsymbol{u})$ is positive and continuously differentiable up to $s$th order (needed for Rosenblatt transform), and the conditional density $f_{Y|\boldsymbol{X}}(y|\boldsymbol{X} = \boldsymbol{x})$ is continuouly differentiable with respect to $\boldsymbol{x}$.

(A6) $h$ satisfies $\sqrt{nh^d}h^s\sqrt{\log n} \to 0$ (undersmoothing), and $nh^{3d} \to \infty$ as $n \to \infty$

▸ Thm RQ-Band

▸ Emp. process QR

# Assumptions

(EA2) $\sup_{\boldsymbol{x} \in \mathcal{D}} \left| \int v^{b_1} f_{\varepsilon|\boldsymbol{X}}(v|\boldsymbol{x}) dv \right| < \infty$, where $b_1$ satisfies

$$n^{-1/6} h^{-d/2-3d/(b_1-2)} = \mathcal{O}(n^{-\nu}), \quad \nu > 0.$$

e.g. when $h = n^{-1/(2s+d)}$, then $b_1 > (4s + 14d)/(2s + d - 3)$.

▸ Thm RE-Band

▸ Emp. process ER

# Assumptions

(B1) $L$ is a Lipschitz, bounded, symmetric kernel. $G$ is Lipschitz continuous cdf, and $g$ is the derivative of $G$ and is also a density, which is Lipschitz continuous, bounded, symmetric and five times continuously differentiable kernel.

(B2) $F_{\varepsilon|\boldsymbol{X}}(v|\boldsymbol{x})$ is in $s'+1$ order Hölder class with respect to $v$ and continuous in $\boldsymbol{x}$, $s' > \max\{2, d\}$. $f_{\boldsymbol{X}}(\boldsymbol{x})$ is in second order Hölder class with respect to $\boldsymbol{x}$ and $v$. $\mathrm{E}[\psi^2(\varepsilon_i)|\boldsymbol{x}]$ is second order continuously differentiable with respect to $\boldsymbol{x} \in \mathcal{D}$.

(B3) $nh_0\bar{h}^d \to \infty$, $h_0, \bar{h} = \mathcal{O}(n^{-\nu})$, where $\nu > 0$.

▶ Scaling factors

# Assumptions

(C1) There exist an increasing sequence $c_n$, $c_n \to \infty$ as $n \to \infty$ such that

$$(\log n)^3 (nh^{6d})^{-1} \int_{|v|>c_n/2} f_\varepsilon(v)dv = \mathcal{O}(1), \qquad (11)$$

as $n \to \infty$.

(EC1) $\sup_{\boldsymbol{x} \in \mathcal{D}} \left| \int v^b f_{\varepsilon|\boldsymbol{X}}(v|\boldsymbol{x})dv \right| < \infty$, where $b$ satisfies

$$n^{-\frac{1}{6} + \frac{4}{b^2} - \frac{1}{b}} h^{-\frac{d}{2} - \frac{6d}{b}} = \mathcal{O}(n^{-\nu}), \quad \nu > 0, \; (\text{Thm. } 18)$$

and

$$b > 2(2s' + d + 1)/(2s' + 3). \; (\text{Lemma } 6)$$

▶ Bootstrap

# Quantile regression notations

$$h = n^{-\kappa}, \quad \rho_\tau(u) = \left| \tau - \mathbf{1}(u < 0) \right| |u|, \quad \psi(u) = \mathbf{1}(u \leq 0) - \tau$$

$$d_n = (2d\kappa \log n)^{1/2} + (2d\kappa(\log n))^{-1/2} \left[ \frac{1}{2}(d-1) \log \log n^\kappa \right.$$

$$\left. + \log \left\{ (2\pi)^{-1/2} H_2(2d)^{(d-1)/2} \right\} \right],$$

$$H_2 = \left( 2\pi \|K\|_2^2 \right)^{-d/2} \det(\Sigma)^{1/2}, \quad \Sigma_{ij} = \int \frac{\partial K}{\partial u_i} \frac{\partial K}{\partial u_j} d\mathbf{u},$$

$$r(\mathbf{x}) = \sqrt{\frac{nh^d f_{\mathbf{X}}(\mathbf{x})}{\tau(1-\tau)}} f_{\varepsilon|\mathbf{x}}(0|\mathbf{x}),$$

▶ Bahadur    ▶ RQ-Band

# Expectile regression notations

$$h = n^{-\kappa}, \quad \rho_\tau(u) = |\tau - \mathbf{1}(u < 0)|u^2, \quad \varphi(u) = -2\{\tau - \mathbf{1}(u < 0)\}|u|$$

$$d_n = (2d\kappa \log n)^{1/2} + \left(2d\kappa(\log n)\right)^{-1/2} \left[\frac{1}{2}(d-1)\log\log n^\kappa\right.$$

$$\left. + \log\left\{(2\pi)^{-1/2} H_2 (2d)^{(d-1)/2}\right\}\right],$$

$$H_2 = \left(2\pi \|K\|_2^2\right)^{-d/2} \det(\Sigma)^{1/2}, \ \Sigma_{ij} = \int \frac{\partial K}{\partial u_i} \frac{\partial K}{\partial u_j} d\boldsymbol{u},$$

$$r(\boldsymbol{x}) = \sqrt{\frac{nh^d f_{\boldsymbol{X}}(\boldsymbol{x})}{\sigma^2(\boldsymbol{x})}} 2\left[\tau - F_{\varepsilon|\boldsymbol{X}}(0|\boldsymbol{x})(2\tau - 1)\right],$$

$$\sigma^2(\boldsymbol{x}) = \mathsf{E}[\varphi^2(Y - \theta_0(\boldsymbol{x}))|\boldsymbol{X} = \boldsymbol{x}].$$

▸ Bahadur   ▸ RE-Band

# Approximations

$$Y_n(\boldsymbol{x}) = \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x})\sigma^2(\boldsymbol{x})}} \int \int K\left(\frac{\boldsymbol{x} - \boldsymbol{u}}{h}\right) \psi\{y - \theta_0(\boldsymbol{x})\} dZ_n(y, \boldsymbol{u})$$

$$Y_{0,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \int \int_{\Gamma_n} K\left(\frac{\boldsymbol{x} - \boldsymbol{u}}{h}\right) \psi\{y - \theta_0(\boldsymbol{x})\} dZ_n(y, \boldsymbol{u})$$

$$Y_{1,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \int \int_{\Gamma_n} K\left(\frac{\boldsymbol{x} - \boldsymbol{u}}{h}\right) \psi\{y - \theta_0(\boldsymbol{x})\} dB_n\big(T(y, \boldsymbol{u})\big)$$

$$Y_{2,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \int \int_{\Gamma_n} K\left(\frac{\boldsymbol{x} - \boldsymbol{u}}{h}\right) \psi\{y - \theta_0(\boldsymbol{x})\} dW_n\big(T(y, \boldsymbol{u})\big)$$

$$Y_{3,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \int \int_{\Gamma_n} K\left(\frac{\boldsymbol{x} - \boldsymbol{u}}{h}\right) \psi\{y - \theta_0(\boldsymbol{u})\} dW_n\big(T(y, \boldsymbol{u})\big)$$

▸ Method

# Approximations

$$Y_{4,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x}) \sigma_n^2(\boldsymbol{x})}} \int \sqrt{\sigma_n^2(\boldsymbol{u}) f_{\boldsymbol{X}}(\boldsymbol{u})} K\left(\frac{\boldsymbol{x} - \boldsymbol{u}}{h}\right) dW(\boldsymbol{u})$$

$$Y_{5,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d}} \int K\left(\frac{\boldsymbol{x} - \boldsymbol{u}}{h}\right) dW(\boldsymbol{u})$$

▸ Method

## Lemma (Bickel and Wichura (1971))

If $\{X_n\}_{n=1}^{\infty}$ is a sequence in $D[0,1]^d$, $P(X \in [0,1]^d) = 1$. For $B, C$ neighboring blocks in $[0,1]^d$, constants $\lambda_1 + \lambda_2 > 1$, $\gamma_1 + \gamma_2 > 0$, $\{X_n\}_{n=1}^{\infty}$ is *tight* if

$$\mathsf{E}[|X_n(B)|^{\gamma_1}|X_n(C)|^{\gamma_2}] \leq \mu(B)^{\lambda_1}\mu(C)^{\lambda_2}, \tag{12}$$

where $\mu(\cdot)$ is a finite nonnegative measure on $[0,1]^d$ (for example, Lebesgue measure), and the increment of $X_n$ on the block $B$, denoted $X_n(B)$, is defined by

$$X_n(B) = \sum_{\boldsymbol{\alpha} \in \{0,1\}^d} (-1)^{d-|\boldsymbol{\alpha}|} X_n\big(\boldsymbol{s} + \odot(\boldsymbol{t} - \boldsymbol{s})\big) \tag{13}$$

▸ Neighboring blocks    ▸ Step 1

# Neighboring Blocks

## Definition

A block $B \subset \mathcal{D}$ is a subset of $\mathcal{D}$ of the form $B = \Pi_i(s_i, t_i]$ with $s$ and $t$ in $\mathcal{D}$; the pth-face of $B$ is $\Pi_{i \neq p}(s_i, t_i]$. Disjoint blocks $B$ and $C$ are p-neigbbors if they abut and have the same pth face; they are neighbors if they are p-neighbors for some $p$ (for example, when $d = 3$, the blocks $(s, t] \times (a, b] \times (c, d]$ and $(t, u] \times (a, b] \times (c, d]$ are 1-neighbors for $s \leq t \leq u$).

# Examples

- $d = 1$: $B = (s, t]$, $X_n(B) = X_n(t) - X_n(s)$;
- $d = 2$: $B = (s_1, t_1] \times (s_2, t_2]$.
  $X_n(B) = X_n(t_1, t_2) - X_n(t_1, s_2) + X_n(s_1, s_2) - X_n(s_1, t_2)$;
- For general $d$, $B = \Pi_{i=1}^d (s_i, t_i]$, let $\boldsymbol{s} = (s_1, ..., s_d)^\top$,
  $\boldsymbol{t} = (t_1, ..., t_d)^\top$, then where $\odot$ denotes the vector of componentwise products.

▸ Bickel & Wichura (1971)

## Lemma (Meerschaert et al. (2013))

*Suppose that $Y = \{Y(t), t \in \mathbb{R}^d\}$ is a centered Gaussian random field with values in $\mathbb{R}$, and denote*
$d(s, t) \stackrel{\text{def}}{=} d_Y(s, t) = (\mathsf{E}\,|Y(t) - Y(s)|^2)^{1/2}, \quad s, t \in \mathbb{R}^d$. *Let $\mathcal{D}$ be a compact set contained in a cube with length $r$ in $\mathbb{R}^d$ and let $\sigma^2 = \sup_{t \in \mathcal{D}} \mathsf{E}[Y(t)^2]$. For any $m > 0$, $\epsilon > 0$, define*

$$\gamma(\epsilon) = \sup_{s, t \in \mathcal{D}, \|s - t\| \leq \epsilon} d(s, t), \quad Q(m) = (2 + \sqrt{2}) \int_1^\infty \gamma(m2^{-y^2}) dy.$$

*Then for all $a > 0$ which staistfy $a \geq (1 + 4d \log 2)^{1/2}(\sigma + a^{-1})$,*

$$\mathsf{P}\left\{\sup_{t \in S} |Y(t)| > a\right\} \leq 2^{2d+2} \left(\frac{r}{Q^{-1}(1/a)} + 1\right)^d \frac{\sigma + a^{-1}}{a} \exp\left\{-\frac{a^2}{2(\sigma + a^{-1})^2}\right\}$$

*where $Q^{-1}(a) = \sup\{m : Q(m) \leq a\}$.*   ▸ Step 4   ▸ Step 6

# Generate Bivariate Uniform Samples

The bivariate samples $(X_1, X_2)$ are generated as follows:

1. Generate $n$ pairs of bivariate normal variables $(Z_1, Z_2)$ with correlation $\rho_N$ and variance 1
2. Transform the normal r.v.: $(X_1, X_2) = (\Phi(Z_1), \Phi(Z_2))$, where $\Phi(\cdot)$ is the standard normal distribution function
3. Let $\rho_U$ be the correlation of $(X_1, X_2)$, the following relation is true:
$$\rho_U = \frac{6}{\pi} \arcsin \frac{\rho_N}{2}.$$

Details: Falk (1999)

▸ Simulation

# Estimating scaling factors

we propose to estimate $F_{\varepsilon|\boldsymbol{X}}$, $f_{\varepsilon|\boldsymbol{X}}$ and $\sigma^2(\boldsymbol{x})$ based on residuals $\hat{\varepsilon}_i = Y_i - \hat{\theta}_n(\boldsymbol{x}_i)$:

$$\hat{F}_{\varepsilon|\boldsymbol{X}}(v|\boldsymbol{x}) = n^{-1} \sum_{i=1}^{n} G\left(\frac{v - \hat{\varepsilon}_i}{h_0}\right) L_{\bar{h}}(\boldsymbol{x} - \boldsymbol{X}_i)/\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}) \qquad (14)$$

$$\hat{f}_{\varepsilon|\boldsymbol{X}}(v|\boldsymbol{x}) = n^{-1} \sum_{i=1}^{n} g_{h_0}(v - \hat{\varepsilon}_i) L_{\bar{h}}(\boldsymbol{x} - \boldsymbol{X}_i)/\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}) \qquad (15)$$

$$\hat{\sigma}^2(\boldsymbol{x}) = n^{-1} \sum_{i=1}^{n} \psi^2(\hat{\varepsilon}_i) L_{\bar{h}}(\boldsymbol{x} - \boldsymbol{X}_i)/\hat{f}_{\boldsymbol{X}}(\boldsymbol{x}) \qquad (16)$$

where $G$ is a CDF, $g$ and $L$ are a kernel functions, $h_0, \bar{h} \to 0$ and $n h_0 \bar{h}^d \to \infty$

## Lemma
*Under regularity conditions, we have*  ▸ Assumptions

1. $\sup_{v \in I} \sup_{\boldsymbol{x} \in \mathcal{D}} \left| \hat{F}_{\varepsilon | \boldsymbol{X}}(v | \boldsymbol{x}) - F_{\varepsilon | \boldsymbol{X}}(v | \boldsymbol{x}) \right| = \mathcal{O}_p(a_n)$

2. $\sup_{v \in I} \sup_{\boldsymbol{x} \in \mathcal{D}} \left| \hat{f}_{\varepsilon | \boldsymbol{X}}(v | \boldsymbol{x}) - f_{\varepsilon | \boldsymbol{X}}(v | \boldsymbol{x}) \right| = \mathcal{O}_p(a_n)$

3. $\sup_{\boldsymbol{x} \in \mathcal{D}} \left| \hat{\sigma}^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{x}) \right| = \mathcal{O}_p(b_n)$

*where* $a_n = h_0^2 + h^s + \bar{h}^2 + (nh_0\bar{h}^d)^{-1/2} \log n + (nh^d)^{-1/2} \log n$, *and*
$b_n = h^s + \bar{h}^2 + (n\bar{h}^d)^{-1/2} \log n + (nh^d)^{-1/2} \log n$.

# References

Bickel, P. J. & Rosenblatt, M.
On Some Global Measures of the Deviations of Density
Function Estimates,
Annals of Statistics, 1973, 1, 1071-1095.

Bickel, P. J. & Rosenblatt, M.
Two-dimensional random fields,
Multivariate Analysis III, 1973, 3-15.

Dedecker, J.; Merlevéde, F. & Rio, E.
Strong Approx. of the empi. distri. function for abs. regular
seq. in $\mathbb{R}^d$
Working Paper, 2013.

# References

📄 Delgado, M. A. & Escanciano, J. C.
Conditional Stochastic Dominance Testing
Journal of Business  Economic Statistics, 2013, 31, 16-28.

📄 Efron, B.
Regression Percentiles using Asymmetric squared error loss,
Statistica Sinica, 1991, 1, 93-125.

📄 Falk, M.
A simple approach to the generation of uniformly distributed random
variables with prescribed correlations.
Comm. Statist. - Simulation Comp. 1999, 28, 785-791.

📄 Guo, M. & Härdle, W.
Simultaneous confidence bands for expectile functions,
AStA Advances in Statistical Analysis, 2012, 96, 517-541.

# References

Härdle, W.

Asymptotic maximal deviation of $M$-smoothers

J. Multivariate Anal, 1989, 29, 163 - 179.

Härdle, W. & Song, S.

Confidence Bands in Quantile Regression,

Econometric Theory, 2010, 26, 1180-1200.

Johnston, G. J.

Prob. of max. dev. for nonp. regression function estimates

J. Multivariate Anal, 1982, 12, 402-414

Kong, E.; Linton, O. Xia, Y. Uniform Bahadur representation for

local polynomial estimates of M-reg. and its application to the additive model

Econometric Theory, 2010, 26, 1529-1564

Li, Q., Lin, J. & Racine, J. S.

Optimal Bandwidth Selection for Nonp. Cond. Dist. and Quantile Functions,

Journal of Business & Economic Statistics, 2013, 31, 57-65.

# References

📄 Mas-Colell, A., Whinston, M. D. Green, J. R.
Microeconomic Theory
Oxford University Press, USA, 1995.

📄 Meerschaert, M. M., Wang, W. and Xiao, Y.
Fernique-Type Inequalities and Moduli of Continuity for Anisotropic Gaussian
Random Fields
Transations of the American Mathematical Society, 2013, 365, 1081-1107

📄 Owen, A. B.
Multidim. Variation for quasi-Monte Carlo
Contemporary Multivariate Analysis and Design of Experiments, World Sci.
Publi., 2005, 2, 49-74

# References

📓 Rosenblatt, M.
On the Max. Dev. of $k$-Dimensional Density Estimates,
The Annals of Probability, 1976, 4, 1009-1015.

📓 Smirnov, N. V.
On the construction of confidence regions for the dens. of distri. of random variables,
Doklady Akad. Nauk SSSR, 1950, 74, 189-191 (Russian)