Quantile Processes for Semi and Nonparametric Regression

Shih-Kang Chao
Department of Statistics
Purdue University

PURDUE UNIVERSITY

IMS-APRM 2016

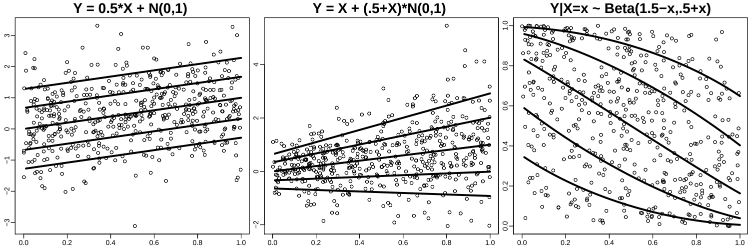
A joint work with Stanislav Volgushev and Guang Cheng

Quantile

Response Y, predictors X. Conditional quantile curve $Q(\cdot;\tau)$ of $Y \in \mathbb{R}$ conditional on X is defined through

$$P(Y \leq Q(X;\tau) | X = x) = \tau \quad \forall x.$$

Examples:

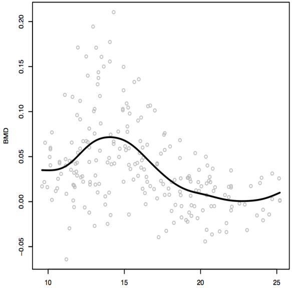


Quantile regression vs mean regression

Mean regression:

$Y_i = m(X_i) + \varepsilon_i, \mathbb{E}[\varepsilon | X = x] = 0$

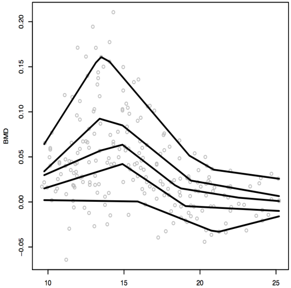
- m: Regression function, object of interest.
- ε_i: 'errors'.



Quantile regression:

$P(Y \leq Q(x; \tau) | X = x) = \tau$

- No strict distinction between 'signal' and 'noise'.
- Object of interest: properties of conditional distribution of $Y | X = x$.
- Contains much richer information than just conditional mean.



Quantile Regression: Estimation

Koenker and Bassett (1978): if $Q(x;\tau) = \boldsymbol{\beta}(\tau)^\top x$, estimate by

$$\widehat{\boldsymbol{\beta}}(\tau) := \arg\min_{\mathbf{b}} \sum_i \rho_\tau(Y_i - \mathbf{b}^\top X_i) \tag{1.1}$$

where $\rho_\tau(u) := \tau u^+ + (1-\tau)u^-$ 'check function'. Well-behaved convex optimization.

$\widehat{Q}(x_0;\tau) := x_0^\top \widehat{\boldsymbol{\beta}}(\tau)$ for any x_0.

▸ Series Model

▸ Partial Linear Model

Quantile Regression Process

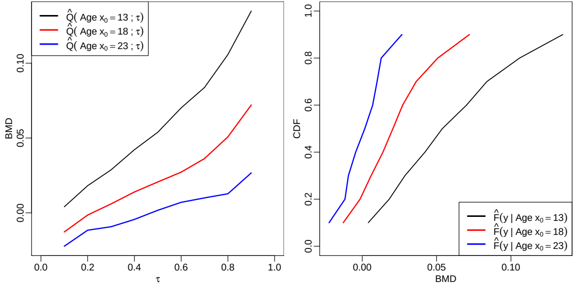
The quantile regression process (QRP) at x_0 is

$$a_n(\widehat{Q}(x_0;\tau) - Q(x_0;\tau)) \in \ell^\infty(\mathcal{T}), \tag{1.2}$$

where $a_n \to \infty$ appropriately chosen, $\mathcal{T} \subset (0,1)$ is compact

For fixed x_0, and τ_L, τ_U "close" to 0 and 1,

$$Q^{-1}(x_0;y) = F(y|x_0) \approx \tau_L + \int_{\tau_L}^{\tau_U} \mathbf{1}\{Q(x_0;\tau) < y\} d\tau =: \Phi_y(Q(x_0;\tau))$$



Convergence of QRP

If for $a_n \to \infty$,

$$a_n(\widehat{Q}(x_0;\cdot) - Q(x_0;\cdot)) \rightsquigarrow \mathbb{G}(\cdot) \text{ in } \ell^\infty(\mathcal{T}), \tag{1.3}$$

where $\mathbb{G}(\cdot)$: Gaussian process, $\ell^\infty(\mathcal{T})$: set of all uniformly bounded, real functions on $\mathcal{T}$, then,

$$a_n\{\widehat{F}_{Y|X}(\cdot|x_0) - F_{Y|X}(\cdot|x_0)\} \rightsquigarrow \\ - f_{Y|X}(\cdot|x_0)\mathbb{G}\big(x_0; F_{Y|X}(\cdot|x_0)\big) \text{ in } \ell^\infty(\mathcal{Y}). \tag{1.4}$$

Proof: $\Phi_y(Q(x;\tau))$ is Hadamard differentiable (Chernozhukov et al., 2010) (tangentially to $\mathcal{C}(0,1)$ at any strictly increasing, differentiable function); functional delta method

Study (1.3):

- Fixed dimension linear model $Q(x;\tau) = x^\top \boldsymbol{\beta}(\tau)$: Koenker and Xiao (2002); Angrist et al. (2006)
- Kernel nonparametric estimation: Qu and Yoon (2015)

A Unified framework (Belloni et al., 2011):

$$Q(x;\tau) \approx \mathbf{Z}(x)^\top \boldsymbol{\beta}(\tau)$$

$\mathbf{Z}(X_i)$ can be a higher dimensional ($\to \infty$) transformation

- $\widehat{Q}(x;\tau) = \mathbf{Z}(x)^\top \widehat{\boldsymbol{\beta}}(\tau)$, where $\widehat{\boldsymbol{\beta}}(\tau)$ is estimated by replacing X_i by $\mathbf{Z}(X_i)$ in ▸ Quantile Regression
- Need to control the bias $Q(x;\tau) - \mathbf{Z}(x)^\top \boldsymbol{\beta}(\tau)$

Overview

We present process convergence results

$$a_n(\widehat{Q}(x_0;\cdot) - Q(x_0;\cdot)) \rightsquigarrow \mathbb{G}(\cdot) \text{ in } \ell^\infty(\mathcal{T})$$

for the models:

- General series estimator
- B-splines: $\mathbf{Z}(x) = \mathbf{B}(x)$
- An application: partial linear models $\mathbf{Z} = (V^\top, \widetilde{\mathbf{Z}}(W)^\top)^\top \in \mathbb{R}^{k+k'}$

Notation:

- $\mathbf{Z}_i := \mathbf{Z}(X_i)$ general basis function (e.g. trigonometric, power, etc.);
- $\mathbf{B}_i := \mathbf{B}(X_i)$ local basis (e.g. B-spline)

Technical Assumptions

Assumption (A): data $(X_i, Y_i)_{i=1,\dots,N}$ form triangular array and are row-wise i.i.d. with

(A1) $m = \mathbf{Z}(x)$. Assume that $\|\mathbf{Z}_i\| \le \xi_m < \infty$, and there exits some fixed constant M so that

$$1/M \le \lambda_{\min}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^T]) \le \lambda_{\max}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^T]) \le M$$

(A2) The conditional distribution $F_{Y|X}(y|x)$ is twice differentiable w.r.t. y. Denote the corresponding derivatives by $f_{Y|X}(y|x)$ and $f'_{Y|X}(y|x)$. Assume that

$$\bar{f} := \sup_{y,x} |f_{Y|X}(y|x)| < \infty, \quad \overline{f'} := \sup_{y,x} |f'_{Y|X}(y|x)| < \infty$$

uniformly in n.

(A3) $0 < f_{\min} \le \inf_{\tau \in \mathcal{T}} \inf_x f_{Y|X}(Q(x;\tau)|x)$ uniformly in n.

A Bahadur Representation

Under Assumption (A), $m\xi_m^2 \log n = o(n)$. For any $\boldsymbol{\beta}_n(\cdot)$ satisfying

$$g_n(\boldsymbol{\beta}_n) := \sup_{\tau\in\mathcal{T}} \left\| \mathbb{E}\left[\mathbf{Z}_i \left\{ F_{Y|X}(\mathbf{Z}_i^\top \boldsymbol{\beta}_n(\tau)|X) - \tau \right\}\right] \right\| = o(\xi_m^{-1})$$
$$c_n(\boldsymbol{\beta}_n) := \sup_{x,\tau\in\mathcal{T}} |Q(x;\tau) - \mathbf{Z}(x)^\top \boldsymbol{\beta}_n(\tau)| = o(1)$$

we have

$$\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_n(\tau) = \underbrace{-\frac{1}{n} J_m(\tau)^{-1} \sum_{i=1}^{n} \mathbf{Z}_i(\mathbf{1}\{Y_i \leq \mathbf{Z}_i^\top \mathbf{b}\} - \tau)}_{(\bigstar)} + \text{Remainder},$$

where $J_m(\tau) := \mathbb{E}[f_{Y|X}(Q(X;\tau))\mathbf{Z}(X)\mathbf{Z}(X)^\top]$.

Assumption $(\diamondsuit)$: $mc_n \log n = o(1)$, $m^3 \xi_m^2 (\log n)^3 = o(n)$, $g_n = o(n^{-1/2})$ and for any $\|\mathbf{u}\| = 1$,

$$\sup_{\tau \in \mathcal{T}} \left| \mathbf{u}^\top J_m(\tau)^{-1} \mathbb{E}\Big[\mathbf{Z}_i \big(\mathbf{1}\{Y_i \le Q(X_i;\tau)\} - \mathbf{1}\{Y_i \le \mathbf{Z}_i^\top \boldsymbol{\beta}_n(\tau)\} \big) \Big] \right| = o(n^{-1/2}).$$

Under Assumption $(\diamondsuit)$, we can replace $(\bigstar)$ by

$$\boldsymbol{U}_n(\tau) := n^{-1} J_m^{-1}(\tau) \sum_{i=1}^{n} \mathbf{Z}_i \big(\mathbf{1}\{Y_i \le Q(X_i;\tau)\} - \tau \big).$$

$\tau \mapsto J_m^{-1}(\tau)\mathbf{Z}_i\big(\mathbf{1}\{Y_i \le Q(X_i;\tau)\} - \tau\big)$:

- A triangular array
- Not Lipschitz in τ

Asymptotic Equicontinuity of Quantile Process:

Under Assumption (A) and $\xi_m^2(\log n)^2 = o(n)$, we have for any $\varepsilon > 0$ and vector $\mathbf{u}_n \in \mathbb{R}^m$ with $\|\mathbf{u}_n\| = 1$,

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\Big(n^{1/2} \sup_{\substack{|\tau_1-\tau_2|\le\delta \\ \tau_1,\tau_2\in\mathcal{T}}} \Big|\mathbf{u}_n^\top \boldsymbol{U}_n(\tau_1) - \mathbf{u}_n^\top \boldsymbol{U}_n(\tau_2)\Big| > \varepsilon\Big) = 0.$$

Proof: Stochastic Equicontinuity + Chaining (Kley et al., 2015; van der Vaart and Wellner, 1996)

Weak Convergence: General Series Estimator

Under Assumption (A) and ($\diamondsuit$). For a sequence $\mathbf{u}_n$, if

$$H(\tau_1, \tau_2; \mathbf{u}_n) := \lim_{n\to\infty} \|\mathbf{u}_n\|^{-2} \mathbf{u}_n^\top J_m^{-1}(\tau_1) \mathbb{E}[\mathbf{Z}\mathbf{Z}^\top] J_m^{-1}(\tau_2) \mathbf{u}_n (\tau_1 \wedge \tau_2 - \tau_1\tau_2)$$

exists for any $\tau_1, \tau_2 \in \mathcal{T}$, then

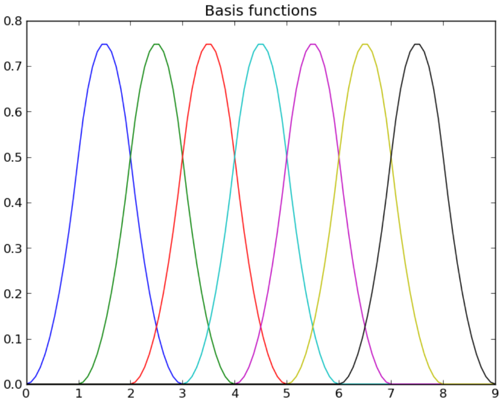
$$\frac{\sqrt{n}}{\|\mathbf{u}_n\|} \Big(\mathbf{u}_n^\top \widehat{\boldsymbol{\beta}}(\cdot) - \mathbf{u}_n^\top \boldsymbol{\beta}_n(\tau) \Big) \rightsquigarrow \mathbb{G}(\cdot) \text{ in } \ell^\infty(\mathcal{T}), \tag{2.1}$$

where $\mathbb{G}(\cdot)$ is a centered Gaussian process with the covariance function H. In particular, there exists a version of $\mathbb{G}$ with almost surely continuous sample paths.

Proof: Asymptotic equicontinuity, and verifying the Lindeberg condition.

Splines

B-Splines $\mathbf{B} = (b_1(x), b_2(x), ..., b_m(x))$ are local basis functions



Notation: For $\mathcal{I} \in \{1, ..., m\}$ and $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{a}^{(\mathcal{I})} = (a'_j)_{j=1}^m \in \mathbb{R}^m$ where $a'_j = 0$ for $j \notin \mathcal{I}$

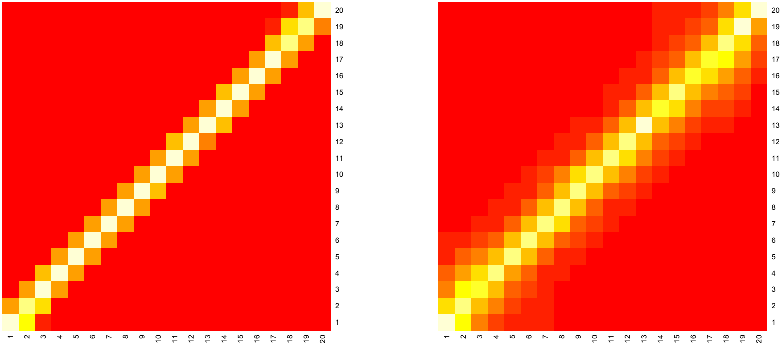
Benefit from Using Splines

$J_m(\tau)$ is a block matrix, where

$$J_m(\tau) = \mathbb{E}[f_{Y|X}(Q(X;\tau))\mathbf{B}(X)\mathbf{B}(X)^\top],$$

entries in $J_m^{-1}(\tau)$ decay geometrically from the main diagonal (Demko et al., 1984)

(a) $J_m(\tau), \dim(X) = 1$ (b) $J_m^{-1}(\tau), \dim(X) = 1$



If $\mathbf{a}$ has at most $\|\mathbf{a}\|_0$ nonzero consecutive entries, we can find set $\mathcal{I}(\mathbf{a}) \subset \{1, ..., m\}$ with $|\mathcal{I}(\mathbf{a})| \asymp \log n$ such that

$$\|\mathbf{a}^\top J_m^{-1}(\tau) - (\mathbf{a}^\top J_m^{-1}(\tau))^{(\mathcal{I}(\mathbf{a}))}\| \lesssim \|\mathbf{a}\|_\infty \|\mathbf{a}\|_0 n^{-c}$$

where $c > 0$ is arbitrary

If $\mathbf{u}_n$ is nonzero at the position in an index set $\mathcal{I}$, which consists of $L < \infty$ consecutive entries

$$\mathbf{u}_n^\top \boldsymbol{U}_n(\tau) \approx \frac{1}{n} \sum_{i=1}^{n} \mathbf{u}_n^\top J_m^{-1}(\tau) \mathbf{B}_i^{(\mathcal{I}(\mathbf{u}_n))} (\mathbf{1}\{Y_i \le \mathbf{B}_i^\top \boldsymbol{\beta}_n(\tau)^{(\mathcal{I}'(\mathbf{u}_n))}\} - \tau)$$

where $\mathcal{I}'(\mathbf{u}_n) = \{1 \le j \le m : \exists i \in \mathcal{I}(\mathbf{u}_n) \text{ such that } |j - i| \le r\}$

A reduction from dimension m to $\log n$!

Define

$$\gamma_n(\tau) := \underset{\mathbf{b}\in\mathbb{R}^m}{\operatorname{argmin}} \mathbb{E}\left[(\mathbf{B}^\top \mathbf{b} - Q(X;\tau))^2 f_{Y|X}(Q(X;\tau)|X)\right],$$

Assumption ($\diamondsuit'$): $\widetilde{c}_n^2 = o(n^{-1/2})$, $\xi_m^4(\log n)^6 = o(n)$, where

$$\widetilde{c}_n(\gamma_n) := \sup_{x,\tau\in\mathcal{T}} |Q(x;\tau) - \mathbf{Z}(x)^\top \boldsymbol{\beta}_n(\tau)|$$

Compare to Assumption ($\diamondsuit$):

- $\xi_m = O(m^{1/2})$: $\xi_m^4(\log n)^6 = o(n)$ is much weaker than $m^3\xi_m^2(\log n)^3 = o(n)$ in Assumption ($\diamondsuit$)
- If $Q(x;\tau)$ is smooth in x for all τ, then $\widetilde{c}_n = o(n^{-2/5})$ when $x \in \mathbb{R}$

Weak Convergence: Spline Series Estimator

Under Assumption (A) and ($\diamondsuit'$). For a sequence $\mathbf{u}_n$, if

$$\widetilde{H}(\tau_1, \tau_2; \mathbf{u}_n) := \lim_{n\to\infty} \|\mathbf{u}_n\|^{-2} \mathbf{u}_n^\top J_m^{-1}(\tau_1)\mathbb{E}[\mathbf{B}\mathbf{B}^\top] J_m^{-1}(\tau_2)\mathbf{u}_n(\tau_1 \wedge \tau_2 - \tau_1\tau_2)$$

exists for any $\tau_1, \tau_2 \in \mathcal{T}$, then

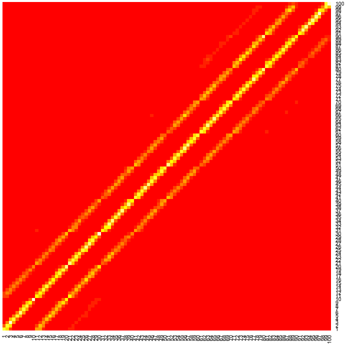
$$\frac{\sqrt{n}}{\|\mathbf{u}_n\|}\Big(\mathbf{u}_n^\top \widehat{\boldsymbol{\beta}}(\cdot) - \mathbf{u}_n^\top \boldsymbol{\gamma}_n(\cdot)\Big) \rightsquigarrow \mathbb{G}(\cdot) \text{ in } \ell^\infty(\mathcal{T}), \tag{3.1}$$

where $\mathbb{G}(\cdot)$ is a centered Gaussian process with the covariance function $\widetilde{H}$. In particular, there exists a version of $\mathbb{G}$ with almost surely continuous sample paths.

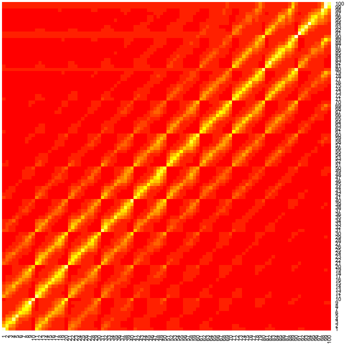
Tensor-Product Spline

Conjecture(Demko et al., 1984, Sec.5): the elements of $J_m^{-1}(\tau)$ from the tensor-product spline **B** are also decreasing geometrically away from the main diagonal

(c) $J_m(\tau), \dim(X) = 2$



(d) $J_m^{-1}(\tau), \dim(X) = 2$



Application: Partial Linear Model

Partial Linear Model:

$$Q(X;\tau) = V^\top \boldsymbol{\alpha}(\tau) + h(W;\tau), \tag{3.2}$$

where $X = (V^\top, W^\top)^\top \in \mathbb{R}^{k+k'}$ and $k, k' \in \mathbb{N}$ are fixed.

Expanding $w \mapsto h(w;\tau)$ in terms of basis vectors $w \mapsto \widetilde{\mathbf{Z}}(w)$, we can approximate (3.2):

$$Q(x;\tau) \approx \mathbf{Z}(x)^\top \boldsymbol{\beta}_n(\tau)$$

where $\mathbf{Z}(x) = (v^\top, \widetilde{\mathbf{Z}}(w)^\top)^\top$ and $\boldsymbol{\beta}_n = (\boldsymbol{\alpha}(\tau)^\top, \boldsymbol{\beta}_n^\dagger(\tau)^\top)^\top$

▸ Quantile Regression gives $\widehat{\boldsymbol{\alpha}}(\tau)$ and $\widehat{h}(w;\tau) = \widetilde{\mathbf{Z}}(w)^\top \widehat{\boldsymbol{\beta}}_n^\dagger(\tau)$

Assumption (B):

(B1) Define $c_n^\dagger := \sup_{\tau,w} |\widetilde{\mathbf{Z}}(w)^\top \boldsymbol{\beta}_n^\dagger(\tau) - h(w;\tau)|$ and assume that

$$\xi_m c_n^\dagger = o(1);$$

$$\sup_{\tau \in \mathcal{T}} \mathbb{E}[f_{Y|X}(Q(X;\tau)|X) \| h_{VW}(W;\tau) - A(\tau)\widetilde{\mathbf{Z}}(W)) \|^2] = O(\lambda_n^2)$$

with $\xi_m \lambda_n^2 = o(1)$

(B2) We have $\max_{j \le k} |V_j| < C$ almost surely for some constant $C > 0$.

(B3)

$$\left(\frac{m \xi_m^{2/3} \log n}{n} \right)^{3/4} + c_n^{\dagger 2} \xi_m = o(n^{-1/2}).$$

Moreover, assume that $c_n^\dagger \lambda_n = o(n^{-1/2})$ and $m c_n^\dagger \log n = o(1)$.

Joint Process Convergence

Under Assumption (A) and (B), suppose at w_0,

$$\Gamma_{22}(\tau_1, \tau_2) = \lim_{n \to \infty} \|\widetilde{\mathbf{Z}}(w_0)\|^{-2} \widetilde{\mathbf{Z}}(w_0)^\top M_2(\tau_1)^{-1} \mathbb{E}[\widetilde{\mathbf{Z}}(W) \widetilde{\mathbf{Z}}(W)^\top] M_2(\tau_2)^{-1} \widetilde{\mathbf{Z}}(w_0)$$

exists, then

$$\begin{pmatrix} \sqrt{n}\{\widehat{\boldsymbol{\alpha}}(\cdot) - \boldsymbol{\alpha}(\cdot)\} \\ \frac{\sqrt{n}}{\|\widetilde{\mathbf{Z}}(w_0)\|}\{\widehat{h}(w_0;\cdot) - h(w_0;\cdot)\} \end{pmatrix} \rightsquigarrow \left(\mathbb{G}_1(\cdot), \ldots, \mathbb{G}_k(\cdot), \mathbb{G}_h(w_0;\cdot)\right)^\top,$$

in $(\ell^\infty(\mathcal{T}))^{k+1}$ where $(\mathbb{G}_1(\cdot), \ldots, \mathbb{G}_k(\cdot), \mathbb{G}_h(w_0;\cdot))$ are centered Gaussian processes with joint covariance function

$$\Gamma(\tau_1, \tau_2; \widetilde{\mathbf{Z}}(w_0)) = (\tau_1 \wedge \tau_2 - \tau_1 \tau_2) \begin{pmatrix} \Gamma_{11}(\tau_1, \tau_2) & \mathbf{0}_k \\ \mathbf{0}_k^\top & \Gamma_{22}(\tau_1, \tau_2) \end{pmatrix}$$

Detail

In particular, there exists a version of $\mathbb{G}_h(w_0;\cdot)$ with almost surely continuous sample paths.

Remarks

- Most asymptotic results for partial linear model separately study the parametric and nonparametric part
- Joint asymptotics phenomenon for mean partial linear model: Cheng and Shang (2015)
- We show such joint asymptotics holds in process sense for quantile model

Thank you for your attention

arXiv 1604.02130

References I

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica*, 74(2):539–563.

Belloni, A., Chernozhukov, V., and Fernández-Val, I. (2011). Conditional quantile processes based on series or many regressors. arXiv preprint arXiv:1105.6154.

Cheng, G. and Shang, Z. (2015). Joint asymptotics for semi-nonparametric regression models with partially linear structure. *Annals of Statistics*, 43(3):1351–1390.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Demko, S., Moss, W. F., and Smith, P. W. (1984). Decay rates for inverses of band matrices. *Mathematics of Computation*, 43(168):491–499.

References II

Kley, T., Volgushev, S., Dette, H., and Hallin, M. (2015+). Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli*.

Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612.

Qu, Z. and Yoon, J. (2015). Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics*, 185:1–19.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Detail for PLM

$$M_2(\tau) := \mathbb{E}\big[\widetilde{\mathbf{Z}}(W)\widetilde{\mathbf{Z}}(W)^\top f_{Y|X}(Q(X;\tau)|X)\big]$$

$$\Gamma_{11}(\tau_1, \tau_2)$$
$$= M_{1,h}(\tau_1)^{-1}\mathbb{E}\big[(V - h_{VW}(W;\tau_1))(V - h_{VW}(W;\tau_2))^\top\big]M_{1,h}(\tau_2)^{-1}$$

where $M_{1,h}(\tau) =$
$\mathbb{E}\big[(V - h_{VW}(W;\tau))(V - h_{VW}(W;\tau))^\top f_{Y|X}(Q(X;\tau)|X)\big]$

Partial Linear Model