Distributed Inference for Quantile Regression Process

Volgushev et al.

July 14, 2020

Quantile Regression

Given data $\{X_i, Y_i\}$,quantile regression for such models is classically formulated through the following minimization problem:

$$\widehat{\boldsymbol{\beta}}_{or}(\tau) := \arg\min_{\mathrm{b}\in\mathbb{R}^m} \sum_{i=1}^{N} \rho_\tau \left\{ Y_i - \mathrm{b}^\top \mathrm{Z}\left(X_i\right) \right\}, \tag{1.1}$$

where $\rho_\tau(u) :=:= \{\tau - 1(u \leq 0)\}u$.

Divide and Conquer

- $N = Sn$: N observations, S machines with n observations each.
- Apply statistical procedures on each machine(worker) and transmit the result to the centre machine(master).

Notation

- Let $\mathcal{X} := supp(X)$.
- Let $Z = Z(X)$ and $Z_i = Z(X_i)$ and assume $\mathcal{T} = [\tau_L, \tau_U]$ for some $0 < \tau_L < \tau_U < 1$.
- $\mathcal{S}^{m-1} \subset \mathbb{R}^m$ is the unit sphere.
- $a_n \asymp b_n$ means that $(|a_n/b_n|)_{n\in\mathbb{N}}$ and $(|b_n/a_n|)_{n\in\mathbb{N}}$ are bounded.
- Define the class of functions

$$\Lambda_c^\eta(\mathcal{T}) := \Big\{ f \in \mathcal{C}^{\lfloor \eta \rfloor}(\mathcal{T}) : \sup_{j\le\lfloor\eta\rfloor} \sup_{\tau\in\mathcal{T}} \big|D^j f(\tau)\big| \le c \\ \sup_{j=\lfloor\eta\rfloor} \sup_{\tau\neq\tau'} \frac{\big|D^j f(\tau) - D^j f(\tau')\big|}{\|\tau-\tau'\|^{\eta-\lfloor\eta\rfloor}} \le c \Big\},$$

where η is called the "degree of Hölder continuity", and $\mathcal{C}^\alpha(\mathcal{X})$ denotes the class of α-continuously differentiable functions on a set $\mathcal{X}$.

The model

We consider a general approximately linear model:

$$Q(x;\tau) \approx Z(x)^T \beta_N(\tau).$$

In this paper, we focus on three classes of transformation $Z(x) \in \mathbb{R}^m$ which include many models as special cases:

- fixed and finite m
- diverging m with local support structure(B-splines)
- diverging m without local support structure

The divide and conquer algorithm

- Divide the data $\{(X_i, Y_i)\}_{i=1}^{N}$ into S sub-samples of size n. Denote the s-th sub-sample as $\{(X_{is}, Y_{is})\}_{i=1}^{n}$ where $s = 1, 2, \ldots, S$.
- For each s and τ, estimate the sub-sample based quantile regression coefficient as follows:

$$\widehat{\boldsymbol{\beta}}^{s}(\tau) := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^m} \sum_{i=1}^{n} \rho_\tau \left\{ Y_{is} - \boldsymbol{\beta}^\top Z\left(X_{is}\right) \right\}.$$

- Each local machine sends $\widehat{\boldsymbol{\beta}}^{s}(\tau) \in \mathbb{R}^m$ to the master that outputs a pooled estimator

$$\overline{\boldsymbol{\beta}}(\tau) := S^{-1} \sum_{s=1}^{S} \widehat{\boldsymbol{\beta}}^{s}(\tau).$$

Quantile Projection

- While $\bar{\beta}(\tau)$ gives an estimator at a fixed $\tau \in \mathcal{T}$, a complete picture of the conditional distribution is often desirable.
- To achieve this, we propose a two-step procedure. First compute $\overline{\boldsymbol{\beta}}(\tau_k) \in \mathbb{R}^m$ for each $\tau_k \in \mathcal{T}_K$, where $\mathcal{T}_K \subset \mathcal{T} = [\tau_L, \tau_U]$ is grid of quantile values in $\mathcal{T}$ with $|\mathcal{T}_K| = K \in \mathbb{N}$.
- Second project each component of the vectors $\{\overline{\boldsymbol{\beta}}(\tau_1), \ldots, \overline{\boldsymbol{\beta}}(\tau_K)\}$ on a space of spline functions in τ.

Quantile Projection

Let

$$\widehat{\boldsymbol{\alpha}}_j := \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \sum_{k=1}^{K} \Big(\bar{\beta}_j(\tau_k) - \boldsymbol{\alpha}^\top \mathrm{B}(\tau_k)\Big)^2, \quad j = 1, \ldots, m,$$

where $\mathrm{B} := (B_1, \ldots, B_q)^T$ is a B-spline basis defined on G equidistant knots $\tau_L = t_1 < \ldots < t_G = \tau_U$ with degree $r_\tau \in \mathbb{N}$. Using $\widehat{\boldsymbol{\alpha}}_j$, we define

$$\widehat{\boldsymbol{\beta}}(\tau) := \widehat{\Xi}^\top \mathrm{B}(\tau),$$

where

$$\widehat{\Xi} := [\widehat{\boldsymbol{\alpha}}_1 \widehat{\boldsymbol{\alpha}}_2 \ldots \widehat{\boldsymbol{\alpha}}_m].$$

The jth element $\widehat{\beta}_j(\tau) = \widehat{\alpha}_j^\top \mathrm{B}(\tau)$ can be viewed as projection,with respect to $||f||_K := (\sum_{k=1}^{K} f^2(\tau_k))^{1/2}$ of $\bar{\beta}_j$ onto the polynomial spline space with basis $B_1, \ldots, B_q$. In what follows, this projection is denoted by Π_k.

Quantile Projection

The algorithm for computing the quantile projection matrix $\hat{\Xi}$ is summarized below, here the divide-and-conquer is applied as a subroutine:

- Define a grid of quantile levels $\tau_k = \tau_L + (k/K)(\tau_U - \tau_L)$ for $k = 1, \ldots, K$. For each τ_k, compute $\bar{\beta}(\tau_k)$.
- For each $j = 1, \ldots, m$, compute

$$\widehat{\boldsymbol{\alpha}}_j = \left(\sum_{k=1}^{K} \mathrm{B}(\tau_k) \mathrm{B}(\tau_k)^\top \right)^{-1} \left(\sum_{k=1}^{K} \mathrm{B}(\tau_k) \bar{\beta}_j(\tau_k) \right),$$

  which is a closed form solution.
- Set the matrix

$$\hat{\Xi} := [\widehat{\boldsymbol{\alpha}}_1 \widehat{\boldsymbol{\alpha}}_2 \ldots \widehat{\boldsymbol{\alpha}}_m].$$

Conditional distribution

A direct application of the above algorithm is to estimate the quantile function for any $\tau \in \mathcal{T}$. More precisely, we consider

$$\widehat{F}_{Y|X}(y|x) := \tau_L + \int_{\tau_L}^{\tau_U} 1\left\{\mathrm{Z}(x)^\top \widehat{\beta}(\tau) < y\right\} d\tau,$$

where τ_L and τ_U are chosen close to 0 and 1. The intuition behind this approach is the observation that

$$\tau_L + \int_{\tau_L}^{\tau_U} 1\{Q(x;\tau) < y\} d\tau = \begin{cases} \tau_L & \text{if } F_{Y|X}(y|x) < \tau_L \\ F_{Y|X}(y|x) & \text{if } \tau_L \leq F_{Y|X}(y|x) \leq \tau_U \\ \tau_U & \text{if } F_{Y|X}(y|x) > \tau_U \end{cases}$$

The function $\hat{F}_{Y|X}$ is smooth functional of the map $\tau \mapsto \mathrm{Z}(x)^\top \widehat{\beta}(\tau)$.

Conditions

The following regularity conditions are needed throughout this paper.

(A1) Assume that $\|Z_i\| \leq \xi_m < \infty$ where ξ_m is allowed to diverge, and that $1/M \leq \lambda_{\min}\left(\mathbb{E}\left[ZZ^\top\right]\right) \leq \lambda_{\max}\left(\mathbb{E}\left[ZZ^\top\right]\right) \leq M$ holds uniformly in N for some fixed constant M.

(A2) The conditional distribution $F_{Y|X}(y|x)$ is twice differentiable w.r.t. y, with the corresponding derivatives $f_{Y|X}(y|x)$ and $f'_{Y|X}(y|x)$.Assume $\bar{f} := \sup_{y\in\mathbb{R},x\in\mathcal{X}}\left|f_{Y|X}(y|x)\right| < \infty, \overline{f'} := \sup_{y\in\mathbb{R},x\in\mathcal{X}}\left|f'_{Y|X}(y|x)\right| < \infty$ uniformly in N.

(A3) Assume that uniformly in N, there exists a constant $f_{\min} < \bar{f}$ such that

$$0 < f_{\min} \leq \inf_{\tau\in\mathcal{T}} \inf_{x\in\mathcal{X}} f_{Y|X}(Q(x;\tau)|x).$$

In these assumptions, we explicitly work with triangular array asymptotics for $\{(X_i, Y_i)\}_{i=1}^N$, where $d = dim(X_i)$ is allowed to grow as well.

Fixed dimensional linear models.

In this section, we assume for all $\tau \in \mathcal{T}$ and $x \in \mathcal{X}$,

$$Q(x;\tau) = \mathrm{Z}(x)^\top \boldsymbol{\beta}(\tau),$$

where $Z(X)$ has fixed dimension m. This simple model setup allows us to derive a simple and clean bound on the difference between $\bar{\beta}, \hat{\beta}$ and the oracle estimator $\hat{\beta}_{or}$.

Theorem 3.1

Assume conditions (A1)-(A3) hold and that $K \ll N^2, S = o(N(\log N)^{-1})$. Then

$$\sup_{\tau \in \mathcal{T}_K} \left\| \overline{\beta}(\tau) - \widehat{\beta}_{or}(\tau) \right\| = O_P\left(\frac{S \log N}{N} + \frac{S^{1/4}(\log N)^{7/4}}{N^{3/4}} \right) + o_P\left(N^{-1/2}\right)$$

If additionally $K \gg G \gg 1$ we also have

$$\begin{aligned}\sup_{\tau \in \mathcal{T}} \left| Z(x_0)^\top \left(\widehat{\beta}(\tau) - \widehat{\beta}_{or}(\tau) \right) \right| \leq & O_P\left(\frac{S \log N}{N} + \frac{S^{1/2}(\log N)^2}{N} \right) \\ & + o_P\left(N^{-1/2}\right) \\ & + \sup_{\tau \in \mathcal{T}} \left| (\Pi_K Q(x_0; \cdot))(\tau) - Q(x_0; \tau) \right| \end{aligned} .$$

Notation

Denote by $\mathcal{P}_1\left(\xi_m, M, \bar{f}, \overline{f'}, f_{\min}\right)$ all pairs (P, Z) of distributions P and transformations Z satisfying (3.1) and (A1)-(A3) with constants $0 < \xi_m, M, \bar{f}, \overline{f'} < \infty, f_{\min} > 0$. Since m, ξ_m are constant in this section, we use the shortened notation $\mathcal{P}_1\left(\xi, M, \bar{f}, \overline{f'}, f_{\min}\right)$

Oracle Estimator

Under (A1)-(A3) it was developed in Belloni et al. (2017) and Chao et al. (2017) who show that

$$\sqrt{N}\left(\widehat{\beta}_{or}(\cdot)-\beta(\cdot)\right) \rightsquigarrow \mathbb{G}(\cdot) \text{ in } \left(\ell^{\infty}(\mathcal{T})\right)^{d}$$

where $\mathbb{G}$ is a centered Gaussian process with covariance structure

$$\begin{aligned} H\left(\tau,\tau'\right) :=& \mathbb{E}\left[\mathbb{G}(\tau)\mathbb{G}\left(\tau'\right)^{\top}\right] \\ =& J_m(\tau)^{-1}\mathbb{E}\left[\mathrm{Z}(X)\mathrm{Z}(X)^{\top}\right] J_m\left(\tau'\right)^{-1}\left(\tau\wedge\tau'-\tau\tau'\right) \end{aligned}$$

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

Oracle rules

Oracle rule for $\bar{\beta}$ A sufficient condition for $\sqrt{N}(\bar{\beta}(\tau) - \beta(\tau)) \rightsquigarrow \mathcal{N}(0, H(\tau, \tau))$ for any $(P, \mathrm{Z}) \in \mathcal{P}_1(\xi, M, f, f', f_{\min})$ is that $S = o(N^{1/2}/\log N)$. A necessary condition for the same result is that $S = o(N^{1/2})$.

Oracle rule for $\hat{\beta}$ Assume that $\tau \mapsto \beta_j(\tau) \in \Lambda_c^\eta(\mathcal{T})$ for $j = 1, \dots, d$ and given $c, \eta > 0$, that $N^2 \gg K \gg G$ and $r_\tau \geq \eta$. A sufficient condition for $\sqrt{N}(\hat{\boldsymbol{\beta}}(\cdot) - \boldsymbol{\beta}(\cdot)) \rightsquigarrow \mathbb{G}(\cdot)$ for any $(P, \mathrm{Z}) \in \mathcal{P}_1(\xi, M, f, f', f_{\min})$ is $S = o\left(N^{1/2}(\log N)^{-1}\right)$ and $G \gg N^{1/(2\eta)}$. A necessary condition for the same result is $S = o(N^{1/2})$ and $G \gg N^{1/(2\eta)}$.

Estimation of Conditional distribution

Define

$$\widehat{F}^{or}_{Y|X}(\cdot \mid x_0) := \tau_L + \int_{\tau_L}^{\tau_U} 1\left\{ \mathrm{Z}(x)^\top \widehat{\boldsymbol{\beta}}_{or}(\tau) < y \right\} d\tau$$

.

The asymptotic distribution of For $\widehat{F}^{or}_{Y|X}$ was derived in Chao, Volgushev and Cheng (2017).

COROLLARY 3.5.

Under the same conditions as Corollary 3.4, we have, for any $x_0 \in \mathcal{X}$,

$$\sqrt{N}\left(\widehat{F}_{Y|X}(\cdot \mid x_0) - F_{Y|X}(\cdot \mid x_0)\right) \rightsquigarrow -f_{Y|X}(\cdot \mid x_0)\, \mathrm{Z}(x_0)^\top \mathbb{G}\left(F_{Y|X}(\cdot \mid x_0)\right)$$
$$\text{in } \ell^\infty\left(\left(Q(x_0; \tau_L), Q(x_0; \tau_U)\right)\right)$$

The same process convergence result holds with $\widehat{F}^{or}_{Y|X}$ replacing $\widehat{F}_{Y|X}(\cdot \mid x_0)$.

Local basis structure.

- We consider models with $Q(x;\tau) \approx \mathrm{Z}(x)^\top \boldsymbol{\beta}(\tau)$ with $m = dim(Z) \to \infty$ as $N \to \infty$ where the transformation Z corresponds to a basis expansion.
- The analysis in this section focuses on the transformations Z with a specific local support structure, which will be defined more formally in Condition (L).
- Since the model $Q(x;\tau) \approx \mathrm{Z}(x)^\top \boldsymbol{\beta}(\tau)$ holds only approximately, there is no unique 'true' value for $\beta(\tau)$.Theoretical results for such models are often stated in terms of the following vector:

$$\gamma_N(\tau) := \arg\min_{\gamma \in \mathbb{R}^m} \mathbb{E}\left[\left(\mathrm{Z}^\top \boldsymbol{\gamma} - Q(X;\tau)\right)^2 f(Q(X;\tau) \mid X)\right]$$

Local Basis Structure

- Note that $\mathrm{Z}^\top \gamma$ can be viewed as the (weighted L_2) projection of $Q(X;\tau)$ onto the approximation space.
- The resulting L_∞ approximation error is defined as

$$c_m(\gamma_N) := \sup_{x \in \mathcal{X}, \tau \in \mathcal{T}} \left| Q(x;\tau) - \gamma_N(\tau)^\top \mathrm{Z}(x) \right|$$

- For any $v \in \mathbb{R}^m$, define the matrix $\widetilde{J_m}(\mathrm{v}) := \mathbb{E}\left[\mathrm{Z}\mathrm{Z}^\top f\left(\mathrm{Z}^\top \mathrm{v} \mid X\right)\right]$.
- For any $a \in \mathbb{R}^m, \mathrm{b}(\cdot) : \mathcal{T} \to \mathbb{R}^m$, define

$$\widetilde{\mathcal{E}}(\mathrm{a}, \mathrm{b}) := \sup_{\tau \in \mathcal{T}} \mathbb{E}\left[\left|\mathrm{a}^\top \widetilde{J_m}^{-1}(\mathrm{b}(\tau))\mathrm{Z}\right|\right].$$

Condition

(L) For each $x \in \mathcal{X}$, the vector $Z(x)$ has zeroes in all but at most r consecutive entries, where r is fixed. Furthermore, $\sup_{x\in\mathcal{X}} \widetilde{\mathcal{E}}(\mathrm{Z}(x), y_N) = O(1)$.

Condition (L) ensures that the matrix $\widetilde{J}_m(\mathrm{v})$ has a band structure for any $v \in \mathbb{R}^m$ such that the off-diagonal entries of $\widetilde{J}_m(\mathrm{v})$ decay exponentially fast.

Univariate polynomial spline

Suppose that (A2-A3) hold and that X has a density on $\mathcal{X} = [0,1]$ uniformly bounded away from zero and infinity. Let $\tilde{\mathrm{B}}(x) = \left(\widetilde{B}_1(x), \ldots, \widetilde{B}_{J-p-1}(x)\right)^\top$ be a polynomial spline basis of degree p defined on J uniformly spaced knots $0 = t_1 < \cdots < t_J = 1$ such that the support of each $\tilde{B}_j$ is contained in the interval$[t_j, t_{j+p+1}]$. Let $\mathrm{Z}(x) := m^{1/2}\left(\widetilde{B}_1(x), \ldots, \widetilde{B}_{J-p-1}(x)\right)^\top$, then there exists a constant $M > 1$, such that $M^{-1} < \mathbb{E}\left[\mathrm{Z}\mathrm{Z}^\top\right] < M$. With this scaling, we have $\xi_m \asymp \sqrt{m}$. Moreover, the first part of assumption (L) holds with $r = p + 1$, while the second part, that is $\sup_{x \in \mathcal{X}} \widetilde{\mathcal{E}}\left(\mathrm{Z}(x), \gamma_N\right) = O(1)$ is verified in Lemma S.2.6.

Theorem 3.7

Suppose that assumptions (A1)–(A3) and (L) hold, that $K \ll N^2$ and $S\xi_m^4 \log N = o(N), c_m(\gamma_N) = o\left(\xi_m^{-1} \wedge (\log N)^{-2}\right)$. Then

$$\begin{aligned}
&\sup_{\tau \in \mathcal{T}_K} \left| Z(x_0)^\top \left(\bar{\beta}(\tau) - \widehat{\beta}_{or}(\tau)\right)\right| \\
&\quad = o_P\left(\|Z(x_0)\| N^{-1/2}\right) \\
&\quad + O_P\left(\left(1 + \frac{\log N}{S^{1/2}}\right)\left(c_m^2(\gamma_N) + \frac{S\xi_m^2(\log N)^2}{N}\right)\right) \\
&\quad + O_P\left(\frac{\|Z(x_0)\|\xi_m S \log N}{N} + \frac{\|Z(x_0)\|}{N^{1/2}}\left(\frac{S\xi_m^2(\log N)^{10}}{N}\right)^{1/4}\right)
\end{aligned}$$

Theorem 3.7:Continue

If additionally $K \gg G \gg 1$ and $c_m^2(\gamma_N) = o\left(N^{-1/2}\right)$, we also have

$$
\begin{aligned}
&\sup_{\tau\in\mathcal{T}}\left|\mathrm{Z}\left(x_0\right)^\top\left(\widehat{\beta}(\tau)-\widehat{\beta}_{or}(\tau)\right)\right| \\
&\quad\le \left\|\mathrm{Z}\left(x_0\right)\right\| \sup_{\tau\in\mathcal{T}_K}\left\|\overline{\beta}(\tau)-\widehat{\beta}_{or}(\tau)\right\| + o_P\left(\left\|\mathrm{Z}\left(x_0\right)\right\| N^{-1/2}\right) \\
&\qquad+ \sup_{T\in\mathcal{T}}\left\{\left|\left(\Pi_K Q\left(x_0;\cdot\right)\right)(\tau)-Q\left(x_0;\tau\right)\right| + \left|\mathrm{Z}\left(x_0\right)^\top\gamma_N(\tau)-Q\left(x_0;\tau\right)\right|\right\}
\end{aligned}
$$

Condition

Denote by $\mathcal{P}_L\left(M, \bar{f}, \overline{f'}, f_{\min}, R\right)$ the collection of all sequences P_N of distributions of (X, Y) on $\mathbb{R}^{d+1}$ and fixed Z with the following properties: (A1-A3) hold with constant M, $\bar{f}, \overline{f'} < \infty, f_{\min} > 0$, (L) holds for some $r < R, \xi_m^4(\log N)^6 = o(N), c_m^2\left(\gamma_N\right) = o\left(N^{-1/2}\right)$.
The following condition characterizes the upper bound on S which is sufficient to ensure the oracle property for $\bar{\beta}(\tau)$.

(L1) Assume that

$$S = o\left(\frac{N}{m\xi_m^2 \log N} \wedge \frac{N}{\xi_m^2(\log N)^{10}} \wedge \frac{N^{1/2}}{\xi_m \log N} \wedge \frac{N^{1/2}\left\|Z\left(x_0\right)\right\|}{\xi_m^2(\log N)^2}\right).$$

Oracle rule for $\bar{\beta}_\tau$

Assume (L1).

$$\frac{\sqrt{N}Z\left(x_0\right)^\top\left(\bar{\beta}(\tau)-\gamma_N(\tau)\right)}{\left(Z\left(x_0\right)^\top J_m(\tau)^{-1}\mathbb{E}\left[ZZ^\top\right]J_m(\tau)^{-1}Z\left(x_0\right)\right)^{1/2}} \rightsquigarrow \mathcal{N}(0,\tau(1-\tau)).$$

This matches the limit behavior of oracle estimator.
If $S = o(N^{1/2}\xi_m^{-1}(\log N)^{-2})$, then (L1) holds.(sufficient condition)
If $S \gtrsim N^{1/2}\xi_m^{-1}$, the weak convergence fails. (necessary condition)

Sufficient Condition for the process r oracle rule

Assume (L1) holds and that $\tau \mapsto Q(x_0;\tau) \in \Lambda_c^{\eta}(\mathcal{T}), r_\tau \geq \eta, \sup_{\tau\in\mathcal{T}} |$
$Z(x_0)^\top \gamma_N(\tau) - Q(x_0\ \tau) | = o\left(\|Z(x_0)\| N^{-1/2}\right)$, that
$N^2 \gg K \gg G \gg N^{1/(2\eta)} \|Z(x_0)\|^{-1/\eta}, c_m^2(\gamma_N) = o(N^{-1/2})$ and that the limit

$$
\begin{aligned}
&H_{x_0}(\tau_1,\tau_2)\\
&\quad := \lim_{N\to\infty} \frac{Z(x_0)^\top J_m^{-1}(\tau_1)\,\mathbb{E}\left[ZZ^\top\right] J_m^{-1}(\tau_2)\,Z(x_0)\,(\tau_1\wedge\tau_2-\tau_1\tau_2)}{\|Z(x_0)\|^2}
\end{aligned}
$$

exists and is nonzero.

Continue

Then,

$$\frac{\sqrt{N}}{\|Z(x_0)\|}\left(Z(x_0)^\top \widehat{\beta}(\cdot) - Q(x_0;\cdot)\right) \rightsquigarrow \mathbb{G}_{x_0}(\cdot) \quad \text{in } \ell^\infty(\mathcal{T})$$

where $\mathbb{G}_{x_0}$ is a centered Gaussian process with $\mathbb{E}\left[\mathbb{G}_{x_0}(\tau)\mathbb{G}_{x_0}(\tau')\right] = H_{x_0}(\tau,\tau')$. This is the same as oracle.
and

$$\frac{\sqrt{N}}{\|Z(x_0)\|}\left(\widehat{F}_{Y|X}(\cdot \mid x_0) - F_{Y|X}(\cdot \mid x_0)\right) \rightsquigarrow -f_{Y|X}(\cdot \mid x_0)\,\mathbb{G}_{x_0}\left(F_{Y|X}(\cdot \mid x_0)\right)$$

Inference utilizing results from subsamples

Now, we consider the practical aspects of inference.
A simple asymptotic level α confidence interval for $Q_{x;\tau}$ is given by

$$\left[Z(x_0)^\top \overline{\beta}(\tau) \pm S^{-1/2} \left(Z(x_0)^\top \widehat{\Sigma}^D Z(x_0) \right)^{1/2} \Phi^{-1}(1-\alpha/2) \right],$$

where $\hat{\Sigma}^D$ is the sample covariance matrix of $\hat{\beta}_1(\tau), \ldots, \hat{\beta}_S(\tau)$. A modification for the confidence interval is

$$\left[Z(x_0)^\top \overline{\beta}(\tau) \pm S^{-1/2} \left(Z(x_0)^\top \widehat{\Sigma}^D Z(x_0) \right)^{1/2} t_{S-1,1-\alpha/2} \right].$$

However,these two approaches are not straightforward to generalize yo inference functionals of $\beta(\tau)$ such as $\hat{F}_{Y|X}(y|x)$.

Inference utilizing results from subsamples

Bootstrap based confidence intervals

- Sample i.i.d. random weights $\{\omega_{s,b}\}_{s=1}, \ldots, S, b = 1, \ldots, B$ from taking value $1 - 1/\sqrt{2}$ with probability $2/3$ and $1 + \sqrt{2}$ with probability $1/3$
- For $b = 1, \ldots, B, k = 1, \ldots, K$, compute the bootstrap estimators

$$\overline{\beta}^{(b)}(\tau_k) := \frac{1}{S} \sum_{s=1}^{S} \frac{\omega_{s,b}}{\bar{\omega}_{\cdot,b}} \widehat{\beta}^{s}(\tau_k)$$

- For a functional of interest Φ approximate quantiles of the distribution of $\Phi(\hat{\beta}(\cdot)) - \Phi(\beta(\cdot))$ by the empirical quantiles.

Inference based on estimating the asymptotic covariance matrix

It is well known that the asymptotic variance–covariance matrix of the difference $\sqrt{n}(\hat{\beta}_{or}(\tau) - \beta\tau)$ takes the sandwich form

$$\Sigma(\tau) = \tau(1-\tau)J_m(\tau)^{-1}\mathbb{E}\left[ZZ^\top\right]J_m(\tau)^{-1},$$

$$J_m(\tau) = \mathbb{E}\left[ZZ^\top f_{Y|X}(Q(X;\tau) \mid X)\right].$$

The middle part $\mathbb{E}\left[ZZ^\top\right]$ is easily estimated by $\frac{1}{nS}\sum_i \sum_s Z_{is}Z_{is}^\top$.

Inference based on estimating the asymptotic covariance matrix

The matrix $J_m(\tau)$ is difficult to estimate. A popular approach is based on Powell's estimator

$$\widehat{J}^P_{ms}(\tau) := \frac{1}{2nh_n} \sum_{i=1}^{n} \mathrm{Z}_{is}\mathrm{Z}_{is}^\top 1 \left\{ \left| Y_{is} - \mathrm{Z}_{is}^\top \widehat{\beta}^s(\tau) \right| \le h_n \right\}.$$

Here, h_n denotes a bandwidth parameter that needs to be chosen carefully in order to balance the resulting bias and variance.

following algorithm can be used in parallel computing

- For $s = 1, 2 \ldots, S$, compute $\widehat{J}^P_{ms}(\tau)$ and $\widehat{\Sigma}_{1s} := \frac{1}{n} \sum_i \mathrm{Z}_{is}\mathrm{Z}_{is}^\top$
- Take average. $\bar{J}^P_m(\tau) := \frac{1}{S} \sum_{s=1}^{S} \widehat{J}^P_{ms}(\tau), \bar{\Sigma}_1 := \frac{1}{S} \sum_{s=1}^{S} \widehat{\Sigma}_{1s}$.
- The final variance estimator is given by $\bar{\Sigma}(\tau) = \tau(1-\tau)\bar{J}^P_m(\tau)^{-1}\bar{\Sigma}_1 \times \bar{J}^P_m(\tau)^{-1}$.

Simulation: Settings

we consider data generated from

$$Y_i = 0.21 + \beta_{m-1}^\top X_i + \varepsilon_i, \quad i = 1, \ldots, N$$

where $\varepsilon_i \sim \mathcal{N}(0, 0.01)$ i.i.d. and $m \in \{4, 16, 32\}$. For each m, the covariate X_i follows a multivariate uniform distribution $\mathcal{U}\left([0,1]^{m-1}\right)$ with $Cov(X_{ij}, X_{ik} = 0.1^2 0.7^{|j-k|}$ for $j, k = 1, \ldots, m-1$, and the vector β_{m-1} takes the form

$$\begin{aligned}
\boldsymbol{\beta}_3 &= (0.21, -0.89, 0.38)^\top \\
\boldsymbol{\beta}_{15} &= \Big(\boldsymbol{\beta}_3^\top, 0.63, 0.11, 1.01, -1.79, -1.39, 0.52, -1.62, \\
&\qquad 1.26, -0.72, 0.43, -0.41, -0.02\Big)^\top \\
\boldsymbol{\beta}_{31} &= \left(\boldsymbol{\beta}_{15}^\top, 0.21, \boldsymbol{\beta}_{15}^\top\right)^\top
\end{aligned}$$

Throughout this section, we fix $\mathcal{T} = [0.05, 0.95]$.

Settings

We consider the following three types of confidence intervals:

- The normal confidence interval
- The confidence interval based on quantiles of the t-distribution
- The bootstrap confidence interval based on sample quantiles

To benchmark our results, we use the infeasible asymptotic confidence interval

$$\left[x_0^\top \bar{\beta}(\tau) \pm N^{-1/2}\sigma(\tau)\Phi^{-1}(1-\alpha/2)\right].$$

Coverage Probability

Figure 2: Results from subsamples

Coverage Probability

Table 1: Results based on estimating the asymptotic covariance matrix

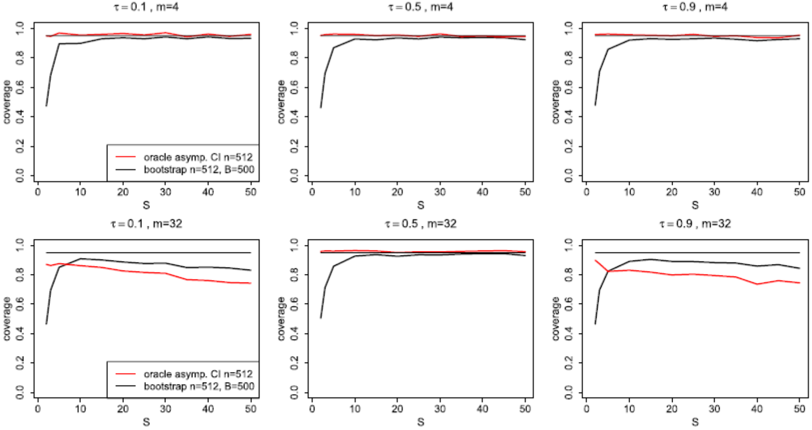

FIG. 3. *Coverage probabilities for oracle confidence intervals (red) and bootstrap confidence intervals (black) for* $F_{Y|X}(y|x_0)$ *for* $x_0 = (1, \dots, 1)/m^{1/2}$ *and* $y = Q(x_0; \tau)$, $\tau = 0.1, 0.5, 0.9$. $n = 512$ *and nominal coverage* 0.95.