The Annals of Statistics
2019, Vol. 47, No. 3, 1634–1662
https://doi.org/10.1214/18-AOS1730

DISTRIBUTED INFERENCE FOR QUANTILE REGRESSION PROCESSES

BY STANISLAV VOLGUSHEV[*,1], SHIH-KANG CHAO[†,2] AND GUANG CHENG[†,2,3]

University of Toronto[*] and Purdue University[†]

The increased availability of massive data sets provides a unique opportunity to discover subtle patterns in their distributions, but also imposes overwhelming computational challenges. To fully utilize the information contained in big data, we propose a two-step procedure: (i) estimate conditional quantile functions at different levels in a parallel computing environment; (ii) construct a conditional quantile regression process through projection based on these estimated quantile curves. Our general quantile regression framework covers both linear models with fixed or growing dimension and series approximation models. We prove that the proposed procedure does not sacrifice any statistical inferential accuracy provided that the number of distributed computing units and quantile levels are chosen properly. In particular, a sharp upper bound for the former and a sharp lower bound for the latter are derived to capture the minimal computational cost from a statistical perspective. As an important application, the statistical inference on conditional distribution functions is considered. Moreover, we propose computationally efficient approaches to conducting inference in the distributed estimation setting described above. Those approaches directly utilize the availability of estimators from subsamples and can be carried out at almost no additional computational cost. Simulations confirm our statistical inferential theory.

1. Introduction. The advance of technology in applications such as meteorological and environmental surveillance or e-commerce has led to extremely large data sets which cannot be processed with stand alone computers due to processor, memory or disk bottlenecks. Dividing data and computing among many machines is a common way to alleviate such bottlenecks, and can be implemented by parallel computing systems like Hadoop [White (2012)] with the aid of communication-efficient algorithms.

In statistics, this approach to estimation has recently been adopted under the name divide-and-conquer; see Jordan (2013), Li, Lin and Li (2013). Pioneering

Received February 2017; revised March 2018.

[1]Supported in part by discovery grant from Natural Sciences and Engineering Research Council of Canada.

[2]Supported in part by Office of Naval Research Grant ONR N00014-15-1-2331.

[3]Supported in part by NSF CAREER Award DMS-1151692, DMS-1418042 and DMS-1712907.

MSC2010 subject classifications. 62F12, 62G15, 62G20.

Key words and phrases. B-spline estimation, conditional distribution function, distributed computing, divide-and-conquer, quantile regression process.

contributions on theoretical analysis of divide-and-conquer algorithms focus on mean squared error rates; see Zhang, Duchi and Wainwright (2013, 2015). The analysis therein does not take into account a core question in statistics—inferential procedures. In the last two years, such procedures have been developed for various non and semiparametric estimation approaches that focus on the mean or other notions of central tendency, this line of work includes Banerjee, Durot and Sen (2019), Shang and Cheng (2017), Shi, Lu and Song (2017), Zhao, Cheng and Liu (2016), among others. Focusing on the mean tendency illuminates one important aspect of the dependence between predictors and response, but ignores all other aspects of the conditional distribution of the response which can be equally important. Additionally, most of the work cited above assumes homoskedastic errors or sub-Gaussian tails. Both assumptions are easily violated in modern messy and large-scale data sets, and this limits the applicability of approaches that are available to date.

We propose to use quantile regression in order to extract features of the conditional distribution of the response while avoiding tail conditions and taking heteroskedasticity into account. This approach focuses on estimating the conditional quantile function $x \mapsto Q(x;\tau) := F^{-1}_{Y|X}(\tau|x)$ where $F_{Y|X}$ denotes the conditional distribution of response given predictor. A flexible class of models for conditional quantile functions can be obtained by basis expansions of the form $Q(x;\tau) \approx \mathbf{Z}(x)^\top \boldsymbol{\beta}(\tau)$. This includes linear models of fixed or increasing dimension where the approximation is replaced by an equality, as well as a rich set of non and semiparametric procedures such as tensor product B-splines or additive models.

Given data $\{(X_i, Y_i)\}_{i=1}^N$, quantile regression for such models is classically formulated through the following minimization problem:

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

where $\rho_\tau(u) := \{\tau - \mathbf{1}(u \le 0)\}u$ and $\mathbf{1}(\cdot)$ is the indicator function. However, solving this minimization problem by classical algorithms requires that the complete data set can be processed on a single computer. For large samples, this might not be feasible. This motivates us to utilize the divide-and-conquer approach where the full sample is randomly split across several computers into S smaller subsamples of size n, the minimization in (1.1) is solved on each subsample, and results are averaged in the end to obtain the final estimator $\overline{\boldsymbol{\beta}}(\tau)$ [see (2.3) in Section 2 for a formal definition]. While this approach is easy to implement, the resulting estimator $\overline{\boldsymbol{\beta}}(\tau)$ is typically not a solution to the original minimization problem (1.1). As illustrated in Figure 1, this can be problematic for inference procedures. More precisely, the left panel of Figure 1 depicts the coverage probabilities (on the y-axis) against number of subsamples (on the x-axis) for asymptotic confidence intervals that are based on asymptotic distributions of $\widehat{\boldsymbol{\beta}}_{or}(\tau)$ but computed using $\overline{\boldsymbol{\beta}}(\tau)$

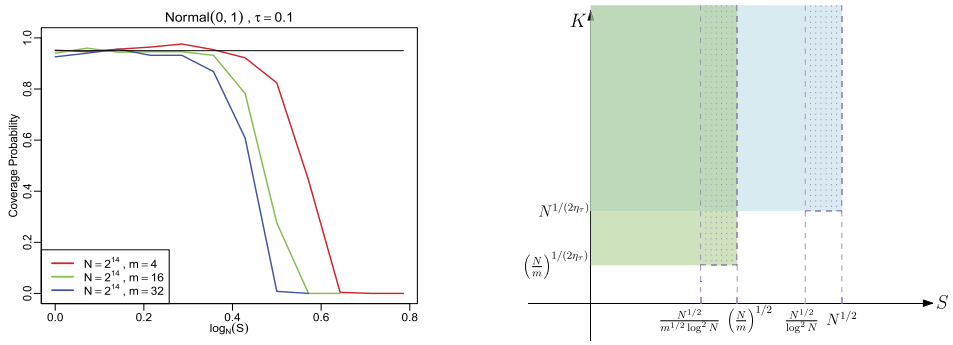

FIG. 1. *Left penal*: *coverage probabilities* (y-*axis*) *of confidence intervals for estimators computed from divide-and-conquer*; x-*axis*: *number of subsamples* S *in log-scale*. *Different colors correspond to linear models with different* $m = \dim(\mathbf{Z}(X))$. *Right panel*: *necessary and sufficient conditions on* (S, K) *for the oracle rule in linear models with fixed dimension* (*blue*) *and B-spline nonparametric models* (*green*). *The dotted region is the discrepancy between the sufficient and necessary conditions*.

for three data generating processes (linear models with different dimension) and a fixed quantile $\tau = 0.1$. This indicates that choosing S too large results in invalid inference, as reflected by a rapid drop in coverage after a certain threshold is reached. For different models, this threshold is different and it intrinsically depends on various properties of the underlying model such as dimension of predictors. These observations indicate that developing *valid* statistical inference procedures based on $\overline{\boldsymbol{\beta}}(\tau)$ requires a careful theoretical analysis of divide-and-conquer procedures. The first major contribution of the present paper is to provide statistical inferential theory for this framework.

The approach described so far provides a way to estimate the conditional quantile function for a single quantile. To obtain a picture of the entire conditional distribution of response given predictors, estimation of the conditional quantile function at several quantile levels is required. Performing such an estimation for a dense grid of quantiles can be computationally costly as estimation at each new quantile requires running the above algorithm anew. To relax this computational burden, we introduce the two-step *quantile projection algorithm*. In the first step, the divide-and-conquer algorithm is used to compute $\overline{\boldsymbol{\beta}}(\tau_1), \ldots, \overline{\boldsymbol{\beta}}(\tau_K)$ for a grid of quantile values $\tau_1, \ldots, \tau_K$. In the second step, a matrix $\widehat{\Xi}$ is computed from $\overline{\boldsymbol{\beta}}(\tau_1), \ldots, \overline{\boldsymbol{\beta}}(\tau_K)$ [see (2.7) for a formal definition]. Given this matrix $\widehat{\Xi}$, the *quantile projection estimator* $\widehat{\boldsymbol{\beta}}(\tau)$ can be computed for any τ by performing a multiplication of this matrix with a known vector (depending on τ) without access to the original dataset. Based on $\widehat{\boldsymbol{\beta}}(\tau)$, we can estimate the conditional distribution function; see (2.8) and (2.9).

The computational cost of our procedure is determined by the number of subsamples S and the number of quantile levels K. To minimize the computational burden, K should be chosen as small as possible. Choosing S large will allow us

to maintain a small sample size n on each local machine, thus enabling us to process larger data sets. At the same time, classical inference procedures for $\widehat{\boldsymbol{\beta}}_{or}$ (e.g., procedures based on asymptotic normality of $\widehat{\boldsymbol{\beta}}_{or}$) should remain valid for $\overline{\boldsymbol{\beta}}(\tau)$, $\widehat{\boldsymbol{\beta}}(\tau)$.

A detailed analysis of conditions which ensure this "oracle rule" is conducted in Section 3. There we derive *sufficient* conditions for both S and K, which are also proved to be necessary in some cases up to $\log N$ terms. Sufficient conditions of S take the form $S = o(N^{1/2} a_{N,m})$ where $a_{N,m} \to 0$. The specific form of $a_{N,m}$ depends on the precise form of the model, with more complex models leading to $a_{N,m}$ converging to zero faster, thus placing more restrictions on S. Necessary and sufficient conditions on K take the form $K \gg N^{1/(2\eta)} \|\mathbf{Z}(x)\|^{-1/\eta}$, where $\eta > 0$ is the degree of Hölder continuity [see (1.2)] of $\tau \mapsto Q(x;\tau)$. In particular, this lower bound is completely independent of S. An interesting insight which will be explained in more detail in Section 3.2 is that for large values of $\|\mathbf{Z}\|$, which typically corresponds to more complex models, the restriction on K becomes weaker; see the paragraph right after Corollary 3.10 for a more detailed explanation.

A graphical summary of the necessary and sufficient conditions on K and S derived in Sections 3.1 and 3.2 is provided in the right panel of Figure 1.

Deriving such necessary conditions on S, K is a crucial step in understanding the limitations of divide-and-conquer procedures, as it shows that the interplay between S, K and model complexity is indeed a feature of the underlying problem and not an artifact of the proof; see also Shang and Cheng (2017).

While the oracle rules described above provide justification for using the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{or}$ for inference, this distribution is typically not pivotal. We discuss several ways to overcome this problem. First, we propose simple and computationally efficient approaches which directly make use of the fact that $\bar{\boldsymbol{\beta}}$ is formed by averaging results from independent subsamples. Those approaches are based on normal and t-approximations as well as a novel bootstrap procedure. Second, we comment on extensions of traditional approaches to inference which rely on estimating the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{or}$ by using kernel smoothing techniques. Simulations demonstrate finite sample properties of the proposed inference strategies.

The rest of this paper is organized as follows. In Section 2, we provide a formal description of all algorithms and estimation procedures discussed in this paper. Section 3 contains the main theoretical results, with additional results presented in the Appendix. In Section 4, we propose several practical approaches to inference. In Section 5, we validate the finite sample relevance of our theoretical results through extensive Monte Carlo experiments. Proofs of all theoretical results are deferred to the Supplementary Material [Volgushev, Chao and Cheng (2019)]. Throughout the manuscript, sections and other results in this Supplementary Material will be referenced by *S.Number*. The following notation will be used throughout the paper.

Notation. Let $\mathcal{X} := \operatorname{supp}(X)$. Let $\mathbf{Z} = \mathbf{Z}(X)$ and $\mathbf{Z}_i = \mathbf{Z}(X_i)$ and assume $\mathcal{T} = [\tau_L, \tau_U]$ for some $0 < \tau_L < \tau_U < 1$. Define the true underlying measure of (Y_i, X_i) by P and denote the corresponding expectation as $\mathbb{E}$. Denote by $\|\mathbf{b}\|$ the L^2-norm of a vector $\mathbf{b}$. $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the smallest and largest eigenvalue of a matrix A. Let $\lfloor \eta \rfloor$ denote the integer part of a real number η, and $|\boldsymbol{j}| = j_1 + \cdots + j_d$ for d-tuple $\boldsymbol{j} = (j_1, \ldots, j_d)$. $\mathcal{S}^{m-1} \subset \mathbb{R}^m$ 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, and $a_n \gg b_n$ means $a_n/b_n \to \infty$. Define the class of functions

$$\Lambda_c^\eta(\mathcal{T}) := \Big\{ f \in \mathcal{C}^{\lfloor \eta \rfloor}(\mathcal{T}) : \sup_{j \le \lfloor \eta \rfloor} \sup_{\tau \in \mathcal{T}} |D^j f(\tau)| \le c, \sup_{j=\lfloor \eta \rfloor} \sup_{\tau \ne \tau'} \frac{|D^j f(\tau) - D^j f(\tau')|}{\|\tau - \tau'\|^{\eta - \lfloor \eta \rfloor}} \le c \Big\}, \tag{1.2}$$

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 $\mathcal{X}$.

2. A two-step procedure for computing quantile process. In this section, we formally introduce two algorithms studied throughout the paper —*divide-and-conquer* and *quantile projection*. The former is used to estimate quantile functions at fixed quantile levels, based on which the latter constructs an estimator for the quantile process. We also note that algorithms presented below can be applied for processing data that are locally stored for any reason and not necessarily large. As an important application, the estimation of conditional distribution functions will be presented.

We consider a general approximately linear model:

$$Q(x; \tau) \approx \mathbf{Z}(x)^\top \boldsymbol{\beta}_N(\tau), \tag{2.1}$$

where $Q(x; \tau)$ is the τth quantile of the distribution of Y conditional on $X = x \in \mathbb{R}^d$, and $\mathbf{Z}(x) \in \mathbb{R}^m$ is a vector of transformation of x. This framework (2.1) incorporates various estimation procedures, for instance, series approximation models; see Belloni et al. (2017) for more discussion. In this paper, we will focus on three classes of transformation $\mathbf{Z}(x) \in \mathbb{R}^m$ which include many models as special cases: fixed and finite m, diverging m with local support structure (e.g., B-splines) and diverging m *without* local support structure.

The *divide-and-conquer* algorithm for estimating $\boldsymbol{\beta}_N(\tau)$ at a fixed $\tau \in (0, 1)$ is described below:

1. Divide the data $\{(X_i, Y_i)\}_{i=1}^N$ into S subsamples of size n.[4] Denote the sth subsample as $\{(X_{is}, Y_{is})\}_{i=1}^n$ where $s = 1, \ldots, S$.

[4]The equal subsample size is assumed for ease of presentation, and can be certainly relaxed.

2. For each s and fixed τ, estimate the subsample based quantile regression coefficient as follows:

$$\widehat{\boldsymbol{\beta}}^s(\tau) := \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^m} \sum_{i=1}^{n} \rho_\tau\{Y_{is} - \boldsymbol{\beta}^\top \mathbf{Z}(X_{is})\}. \tag{2.2}$$

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

$$\overline{\boldsymbol{\beta}}(\tau) := S^{-1} \sum_{s=1}^{S} \widehat{\boldsymbol{\beta}}^s(\tau). \tag{2.3}$$

REMARK 2.1. The minimization problem in (2.2) is in general not strictly convex. Thus, several minimizers could exist. In the rest of this paper, we will only focus on the minimizer with the smallest ℓ_2 norm. This is assumed for presentational convenience, and a close look at the proofs reveals that all statements remain unchanged if any other minimizer is chosen instead.

While $\overline{\boldsymbol{\beta}}(\tau)$ defined in (2.3) 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. More precisely, let

$$\widehat{\boldsymbol{\alpha}}_j := \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q} \sum_{k=1}^{K} \big(\overline{\beta}_j(\tau_k) - \boldsymbol{\alpha}^\top \mathbf{B}(\tau_k)\big)^2, \qquad j = 1, \ldots, m, \tag{2.4}$$

where $\mathbf{B} := (B_1, \ldots, B_q)^\top$ is a B-spline basis defined on G equidistant knots $\tau_L = t_1 < \cdots < t_G = \tau_U$ with degree $r_\tau \in \mathbb{N}$ [we use the normalized version of B-splines as given in Definition 4.19 on page 124 in Schumaker (1981)]. Here, uniformity of knots is assumed for simplicity, all theoretical results can be extended to knots with quasi uniform partitions. Using $\widehat{\boldsymbol{\alpha}}_j$, we define the quantile projection estimator $\widehat{\boldsymbol{\beta}}(\cdot) : \mathcal{T} \to \mathbb{R}^m$:

$$\widehat{\boldsymbol{\beta}}(\tau) := \widehat{\Xi}^\top \mathbf{B}(\tau), \tag{2.5}$$

where $\widehat{\Xi}$ is defined below. The jth element $\widehat{\beta}_j(\tau) = \widehat{\boldsymbol{\alpha}}_j^\top \mathbf{B}(\tau)$ can be viewed as projection, with respect to $\|f\|_K := (\sum_{k=1}^{K} f^2(\tau_k))^{1/2}$, of $\overline{\beta}_j$ onto the polynomial spline space with basis $B_1, \ldots, B_q$. In what follows, this projection is denoted by Π_K. Although we focus on B-splines in this paper, other basis functions are certainly possible.

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

1. 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 $\overline{\boldsymbol{\beta}}(\tau_k)$ as (2.3).

2. For each $j = 1, \ldots, m$, compute

$$\widehat{\boldsymbol{\alpha}}_j = \left(\sum_{k=1}^{K} \mathbf{B}(\tau_k) \mathbf{B}(\tau_k)^\top \right)^{-1} \left(\sum_{k=1}^{K} \mathbf{B}(\tau_k) \overline{\beta}_j(\tau_k) \right), \tag{2.6}$$

which is a closed form solution of (2.4).

3. Set the matrix

$$\widehat{\Xi} := [\widehat{\boldsymbol{\alpha}}_1 \quad \widehat{\boldsymbol{\alpha}}_2 \quad \cdots \quad \widehat{\boldsymbol{\alpha}}_m]. \tag{2.7}$$

A direct application of the above algorithm is to estimate the quantile function for *any* $\tau \in \mathcal{T}$. A more important application of $\widehat{\boldsymbol{\beta}}(\tau)$ is the estimation of the conditional distribution function $F_{Y|X}(y|x)$ for any fixed x. More precisely, we consider

$$\widehat{F}_{Y|X}(y|x) := \tau_L + \int_{\tau_L}^{\tau_U} \mathbf{1}\{\mathbf{Z}(x)^\top \widehat{\boldsymbol{\beta}}(\tau) < y\} \, d\tau, \tag{2.8}$$

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

$$\begin{aligned} &\tau_L + \int_{\tau_L}^{\tau_U} \mathbf{1}\{Q(x;\tau) < y\} \, d\tau \\ &\quad = \begin{cases} \tau_L & \text{if } F_{Y|X}(y|x) < \tau_L; \\ F_{Y|X}(y|x) & \text{if } \tau_L \le F_{Y|X}(y|x) \le \tau_U; \\ \tau_U & \text{if } F_{Y|X}(y|x) > \tau_U; \end{cases} \end{aligned} \tag{2.9}$$

see Chernozhukov, Fernández-Val and Galichon (2010), Volgushev (2013). The estimator $\widehat{F}_{Y|X}$ is a smooth functional of the map $\tau \mapsto \mathbf{Z}(x)^\top \widehat{\boldsymbol{\beta}}(\tau)$ [Chernozhukov, Fernández-Val and Galichon (2010)]. Hence, properties of $\widehat{F}_{Y|X}$ depend on the behavior of $\mathbf{Z}(x)^\top \widehat{\boldsymbol{\beta}}(\tau)$ as a *process* in τ, which we will study in detail in Section 3.

3. Theoretical analysis. The following regularity conditions are needed throughout this paper:

(A1) Assume that $\|\mathbf{Z}_i\| \le \xi_m < \infty$, where ξ_m is allowed to diverge, and that $1/M \le \lambda_{\min}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]) \le \lambda_{\max}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]) \le M$ 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 $\bar{f} := \sup_{y \in \mathbb{R}, x \in \mathcal{X}} |f_{Y|X}(y|x)| < \infty$, $\overline{f'} := \sup_{y \in \mathbb{R}, x \in \mathcal{X}} |f'_{Y|X}(y|x)| < \infty$ uniformly in N.

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

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

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.

Assumptions (A1)–(A3) are fairly standard in the quantile regression literature; see Belloni et al. (2017). The eigenvalue condition in (A1) is imposed for the purpose of normalization, which in turn determines the value of ξ_m. For basis expansions with local support such as B-splines, the right scaling to ensure the eigenvalue condition results is $\xi_m \asymp \sqrt{m}$.

3.1. *Fixed dimensional linear models.* In this section, we assume for all $\tau \in \mathcal{T}$ and $x \in \mathcal{X}$

$$Q(x;\tau) = \mathbf{Z}(x)^\top \boldsymbol{\beta}(\tau), \tag{3.1}$$

where $\mathbf{Z}(x)$ has fixed dimension m. This simple model setup allows us to derive a simple and clean bound on the difference between $\overline{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\beta}}$ and the oracle estimator $\widehat{\boldsymbol{\beta}}_{or}$ and leading to a better understanding of resulting oracle rules. Our first main result is as follows.

THEOREM 3.1. *Suppose that* (3.1) *and assumptions* (A1)–(A3) *hold and that* $K \ll N^2$, $S = o(N(\log N)^{-1})$. *Then*

$$\sup_{\tau\in\mathcal{T}_K} \|\overline{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)\| = O_P\Big(\frac{S\log N}{N} + \frac{S^{1/4}(\log N)^{7/4}}{N^{3/4}}\Big) + o_P\big(N^{-1/2}\big). \tag{3.2}$$

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

$$\begin{aligned}\sup_{\tau\in\mathcal{T}} |\mathbf{Z}(x_0)^\top(\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau))| \le{}& O_P\Big(\frac{S\log N}{N} + \frac{S^{1/2}(\log N)^2}{N}\Big) + o_P\big(N^{-1/2}\big)\\ &+ \sup_{\tau\in\mathcal{T}} |(\Pi_K Q(x_0;\cdot))(\tau) - Q(x_0;\tau)|,\end{aligned}$$

where the projection operator Π_K *was defined right after* (2.5).

The proof of Theorem 3.1 is given in Section S.3.1. To obtain this result, we develop a new Bahadur representation for each local estimator $\widehat{\boldsymbol{\beta}}^s(\tau)$; see Section S.2.1. The main novelty compared to existing results is a sharp bound on the *expectation* of the remainder term. This is crucial to obtain the bound in (3.2). In contrast, relying on previously available versions of the Bahadur representation would result in a bound of the form $(S/N)^{3/4}(\log N)^{3/4}$, which is somewhat more restrictive. See Remark S.2.3 for additional details. Theorem 3.1 can be generalized to potentially miss-specified linear models with dimension that depends on the sample size. Technical details are provided in the Appendix.

The bound in (3.2) quantifies the difference between $\overline{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{or}$ in terms of the number of subsamples S and can be equivalently formulated in terms of the subsample size $n = N/S$ provided that $\log N \sim \log n$. When considering the projection estimator, an additional bias term $\Pi_K Q(x_0;\cdot) - Q(x_0;\cdot)$ is introduced.

Provided that for a given $x_0 \in \mathcal{X}$ one has $\tau \mapsto Q(x_0;\tau) \in \Lambda_c^\eta(\mathcal{T})$, this bias term can be further bounded by $O(G^{-\eta})$. Note that in the setting of Theorem 3.1 the oracle estimator converges to $\boldsymbol{\beta}(\tau)$ at rate $O_P(N^{-1/2})$, uniformly in $\tau \in \mathcal{T}$. By combining the results in Theorem 3.1 with this bound, upper bounds on convergence rates for $\overline{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)$ and $\sup_{\tau\in\mathcal{T}} |\mathbf{Z}(x_0)^\top(\widehat{\boldsymbol{\beta}}(\tau) - Q(x_0;\tau)|$ follow as direct corollaries.

Theorem 3.1 only provides upper bounds on the differences between $\overline{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{or}$. While a more detailed expression for this difference is derived in the proof of Theorem S.2.1, this expression is very unwieldy and does not lead to useful explicit formulas. However, it is possible to prove that the bounds given in Theorem 3.1 are sharp up to $\log N$ factors, which is the main result of the next theorem. Before stating this result, we need to introduce some additional notation. Denote by $\mathscr{P}_1(\xi_m, M, \bar{f}, \overline{f'}, f_{\min})$ all pairs $(P, \mathbf{Z})$ of distributions P and transformations $\mathbf{Z}$ 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 $\mathscr{P}_1(\xi, M, \bar{f}, \overline{f'}, f_{\min})$.

THEOREM 3.2. *For any τ in $\mathcal{T}$ there exists $(P, \mathbf{Z}) \in \mathscr{P}_1(\xi, M, \bar{f}, \overline{f'}, f_{\min})$ and a $C > 0$ such that*

$$\limsup_{N\to\infty} P\left(\|\overline{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)\| \geq \frac{CS}{N}\right) > 0. \tag{3.3}$$

Moreover, for any $c, \eta > 0$ there exist $(P, \mathbf{Z}) \in \mathscr{P}_1(\xi, M, \bar{f}, \overline{f'}, f_{\min})$ such that $\tau \mapsto \beta_j(\tau) \in \Lambda_c^\eta(\mathcal{T})$, $j = 1, \ldots, d$ and

$$\limsup_{N\to\infty} P\left(\sup_{\tau\in\mathcal{T}} \|\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)\| \geq \frac{CS}{N} + CG^{-\eta}\right) > 0. \tag{3.4}$$

The proof of Theorem 3.2 is given in Section S.3.1. The result provided in (3.3) has an interesting implication: the best possible precision of estimating Q in a divide-and-conquer framework is restricted by $n = N/S$, the sample size that can be processed on a single machine, regardless of the total sample size N. A related observation was made in Example 1 of Zhang, Duchi and Wainwright (2013) who construct a data-generating process where the MSE rate of a divide-and-conquer estimator is limited by the sample size that can be processed on a single computer.

As corollaries to the above results, we derive sufficient and necessary conditions on S under which $\overline{\boldsymbol{\beta}}$ and $\mathbf{Z}(x)^\top\widehat{\boldsymbol{\beta}}$ satisfy the oracle rule. Note that the asymptotic distribution of the oracle estimator $\widehat{\boldsymbol{\beta}}_{or}(\tau)$ under various conditions has been known for a long time; see for instance Theorem 4.1 of Koenker (2005). Under (A1)–(A3), it was developed in Belloni et al. (2017) and Chao, Volgushev and Cheng (2017) who show that

$$\sqrt{N}\big(\widehat{\boldsymbol{\beta}}_{or}(\cdot) - \boldsymbol{\beta}(\cdot)\big) \rightsquigarrow \mathbb{G}(\cdot) \qquad \text{in } \big(\ell^\infty(\mathcal{T})\big)^d, \tag{3.5}$$

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

$$H(\tau,\tau') := \mathbb{E}[\mathbb{G}(\tau)\mathbb{G}(\tau')^\top] = J_m(\tau)^{-1}\mathbb{E}[\mathbf{Z}(X)\mathbf{Z}(X)^\top]J_m(\tau')^{-1}(\tau\wedge\tau'-\tau\tau'), \tag{3.6}$$

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

COROLLARY 3.3 (Oracle rules for $\overline{\boldsymbol{\beta}}$). *A sufficient condition for* $\sqrt{N}(\bar{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}(\tau)) \rightsquigarrow \mathcal{N}(0,H(\tau,\tau))$ *for any* $(P,\mathbf{Z})\in\mathscr{P}_1(\xi,M,\bar{f},\overline{f'},f_{\min})$ *is that* $S = o(N^{1/2}/\log N)$. *A necessary condition for the same result is that* $S = o(N^{1/2})$.

The necessary and sufficient conditions above match up to a factor of $\log N$. In order to guarantee the oracle rule for the process, we need to impose additional conditions on the smoothness of $\tau\mapsto\boldsymbol{\beta}(\tau)$ and on the number of grid points G.

COROLLARY 3.4 (Oracle rules for $\widehat{\boldsymbol{\beta}}$). *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* $\sqrt{N}(\widehat{\boldsymbol{\beta}}(\cdot)-\boldsymbol{\beta}(\cdot))\rightsquigarrow\mathbb{G}(\cdot)$ *for any* $(P,\mathbf{Z})\in\mathscr{P}_1(\xi,M,\bar{f},\overline{f'},f_{\min})$ *satisfying above conditions is* $S = o(N^{1/2}(\log N)^{-1})$ *and* $G\gg N^{1/(2\eta)}$. *A necessary condition for the same result is* $S = o(N^{1/2})$ *and* $G\gg N^{1/(2\eta)}$.

Corollary 3.4 characterizes the choice for parameters (S,K) which determine computational cost. The conditions on S remain the same as in Corollary 3.3. The process oracle rule requires restrictions on K based on the smoothness of $\tau\mapsto Q(x_0;\tau)$, denoted as η. Note that, compared to the results in Belloni et al. (2017), smoothness of $\tau\mapsto Q(x_0;\tau)$ is the only additional condition of the data that is needed to ensure process convergence of $\sqrt{N}(\widehat{\boldsymbol{\beta}}(\cdot)-\boldsymbol{\beta}(\cdot))$. Specifically, the lower bound on K in terms of N becomes smaller as η increases, which implies that smoother $\tau\mapsto Q(x_0;\tau)$ allow for larger computational savings. Corollary 3.4 implies that the condition on K is necessary for the oracle rule, no matter how S is chosen.

Next, we apply Corollary 3.4 to the estimation of conditional distribution functions. Define

$$\widehat{F}^{or}_{Y|X}(\cdot|x_0) := \tau_L + \int_{\tau_L}^{\tau_U}\mathbf{1}\{\mathbf{Z}(x)^\top\widehat{\boldsymbol{\beta}}_{or}(\tau)<y\}\,d\tau. \tag{3.7}$$

The asymptotic distribution of $\widehat{F}^{or}_{Y|X}(\cdot|x_0)$ was derived in Chao, Volgushev and Cheng (2017).

COROLLARY 3.5. *Under the same conditions as Corollary 3.4, we have, for any* $x_0\in\mathcal{X}$,

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

where $\widehat{F}_{Y|X}(\cdot|x_0)$ *is defined in* (2.8) *and* $\mathbb{G}$ *is a centered Gaussian process with covariance given in* (3.6). *The same process convergence result holds with* $\widehat{F}^{or}_{Y|X}(\cdot|x_0)$ *replacing* $\widehat{F}_{Y|X}(\cdot|x_0)$.

The proof of Corollary 3.5 uses the fact that $y \mapsto \widehat{F}_{Y|X}(y|x)$ is a Hadamard differentiable functional of $\tau \mapsto \mathbf{Z}(x)^\top \widehat{\boldsymbol{\beta}}(\tau)$ for any fixed x. The result of this corollary can be extended to other functionals with this property.

3.2. *Approximate linear models with local basis structure.* In this section, we consider models with $Q(x;\tau) \approx \mathbf{Z}(x)^\top \boldsymbol{\beta}(\tau)$ with $m = \dim(\mathbf{Z}) \to \infty$ as $N \to \infty$ where the transformation $\mathbf{Z}$ corresponds to a basis expansion. The analysis in this section focuses on transformations $\mathbf{Z}$ with a specific local support structure, which will be defined more formally in Condition (L). Examples of such transformations include polynomial splines or wavelets.

Since the model $Q(x;\tau) \approx \mathbf{Z}(x)^\top \boldsymbol{\beta}(\tau)$ holds only approximately, there is no unique "true" value for $\boldsymbol{\beta}(\tau)$. Theoretical results for such models are often stated in terms of the following vector:

$$\boldsymbol{\gamma}_N(\tau) := \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^m} \mathbb{E}\big[\big(\mathbf{Z}^\top\boldsymbol{\gamma} - Q(X;\tau)\big)^2 f\big(Q(X;\tau)|X\big)\big]; \tag{3.8}$$

see Chao, Volgushev and Cheng (2017) and Remark A.6. Note that $\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau)$ 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

$$c_m(\boldsymbol{\gamma}_N) := \sup_{x\in\mathcal{X},\tau\in\mathcal{T}} \big|Q(x;\tau) - \boldsymbol{\gamma}_N(\tau)^\top \mathbf{Z}(x)\big|. \tag{3.9}$$

For any $\mathbf{v} \in \mathbb{R}^m$, define the matrix $\widetilde{J}_m(\mathbf{v}) := \mathbb{E}[\mathbf{Z}\mathbf{Z}^\top f(\mathbf{Z}^\top\mathbf{v}|X)]$. For any $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{b}(\cdot) : \mathcal{T} \to \mathbb{R}^m$, define $\widetilde{\mathcal{E}}(\mathbf{a},\mathbf{b}) := \sup_{\tau\in\mathcal{T}} \mathbb{E}[|\mathbf{a}^\top \widetilde{J}_m^{-1}(\mathbf{b}(\tau))\mathbf{Z}|]$. Throughout the rest of this subsection, we assume the following condition:

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

Condition (L) ensures that the matrix $\widetilde{J}_m(\mathbf{v})$ has a band structure for any $\mathbf{v} \in \mathbb{R}^m$ such that the off-diagonal entries of $\widetilde{J}_m^{-1}(\mathbf{v})$ decay exponentially fast [Lemma 6.3 in Zhou, Shen and Wolfe (1998)]. Next, we discuss an example of $\mathbf{Z}$ which satisfies (L).

EXAMPLE 3.6 (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 $\widetilde{\mathbf{B}}(x) = (\widetilde{B}_1(x), \ldots, \widetilde{B}_{J-p-1}(x))^\top$ 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 $\widetilde{B}_j$ is contained in the interval $[t_j, t_{j+p+1})$ and normalization is as given in Definition 4.19 on page 124 in Schumaker (1981). Let

$\mathbf{Z}(x) := m^{1/2}(\widetilde{B}_1(x), \ldots, \widetilde{B}_{J-p-1}(x))^\top$, then there exists a constant $M > 1$ such that $M^{-1} < \mathbb{E}[\mathbf{Z}\mathbf{Z}^\top] < M$ [by Lemma 6.2 of Zhou, Shen and Wolfe (1998)]. 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, $\sup_{x\in\mathcal{X}} \widetilde{\mathcal{E}}(\mathbf{Z}(x), \boldsymbol{\gamma}_N) = O(1)$, is verified in Lemma S.2.6.

THEOREM 3.7. *Suppose that assumptions (A1)–(A3) and (L) hold, that* $K \ll N^2$ *and* $S\xi_m^4 \log N = o(N)$, $c_m(\boldsymbol{\gamma}_N) = o(\xi_m^{-1} \wedge (\log N)^{-2})$. *Then*

$$
\begin{aligned}
&\sup_{\tau\in\mathcal{T}_K} |\mathbf{Z}(x_0)^\top(\overline{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau))| \\
&\quad = o_P(\|\mathbf{Z}(x_0)\| N^{-1/2}) \\
&\qquad + O_P\left(\left(1 + \frac{\log N}{S^{1/2}}\right)\left(c_m^2(\boldsymbol{\gamma}_N) + \frac{S\xi_m^2(\log N)^2}{N}\right)\right) \\
&\qquad + O_P\left(\frac{\|\mathbf{Z}(x_0)\|\xi_m S \log N}{N} + \frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\left(\frac{S\xi_m^2(\log N)^{10}}{N}\right)^{1/4}\right).
\end{aligned} \tag{3.10}
$$

If additionally $K \gg G \gg 1$ *and* $c_m^2(\boldsymbol{\gamma}_N) = o(N^{-1/2})$, *we also have*

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

where the projection operator Π_K *was defined right after* (2.5).

The strategy for proving Theorem 3.7 is similar to that for Theorem 3.1, the difference is that we now apply an aggregated Bahadur representation which makes explicit use of the local basis structure of $\mathbf{Z}$ (Section S.2.2).

Similar to Theorem 3.1, Theorem 3.7 only provides upper bounds on the differences between $\overline{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{or}$. As in the setting of fixed-dimensional linear models, this result can be complemented by a corresponding lower bound which we state next.

THEOREM 3.8. *For any* τ *in* $\mathcal{T}$, *there exists a sequence of distributions of* (Y, X) *and sequence of transformations* $\mathbf{Z}$ *such that assumptions (A1)–(A3) and (L) hold, that* $S\xi_m^4(\log N)^{10} = o(N)$, $c_m^2(\boldsymbol{\gamma}_N) = o(N^{-1/2})$ *and there exists a* $C > 0$ *with*

$$
\limsup_{N\to\infty} P\left(|\mathbf{Z}(x_0)^\top\overline{\boldsymbol{\beta}}(\tau) - \mathbf{Z}(x_0)^\top\widehat{\boldsymbol{\beta}}_{or}(\tau)| \ge \frac{CS\xi_m}{N}\right) > 0. \tag{3.11}
$$

Moreover, for any $c, \eta > 0$ there exists a sequence of distributions of (Y, X) and sequence of transformations $\mathbf{Z}$ satisfying the above conditions and a $x_0 \in \mathcal{X}$ with $\tau \mapsto Q(x_0; \tau) \in \Lambda_c^\eta(\mathcal{T})$ such that

$$\limsup_{N\to\infty} P\Big(\sup_{\tau\in\mathcal{T}} |\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}(\tau) - \mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\tau)| \geq \frac{CS\xi_m}{N} + CG^{-\eta}\Big) > 0. \tag{3.12}$$

Compared to Section 3.1, we obtain an interesting insight. The sufficient and necessary conditions on S explicitly depend on m (note that under (A1)–(A3) we have $\xi_m \gtrsim m$) and become more restrictive as m increases. This shows that there is a fundamental limitation on possible computational savings that depends on the model complexity. In other words, more complex models (as measured by m) allow for less splitting and require larger subsamples, resulting in a larger computational burden.

We conclude this section by providing sufficient and necessary conditions for oracle rules in the local basis expansion setting. To state those results, denote by $\mathscr{P}_L(M, \bar{f}, \overline{f'}, f_{\min}, R)$ the collection of all sequences P_N of distributions of (X, Y) on $\mathbb{R}^{d+1}$ and fixed $\mathbf{Z}$ with the following properties: (A1)–(A3) hold with constants $M, \bar{f}, \overline{f'} < \infty$, $f_{\min} > 0$, (L) holds for some $r < R$, $\xi_m^4 (\log N)^6 = o(N)$, $c_m^2(\boldsymbol{\gamma}_N) = o(N^{-1/2})$. The conditions in $\mathscr{P}_L(M, \bar{f}, \overline{f'}, f_{\min}, R)$ ensure the weak convergence of the oracle estimator $\widehat{\boldsymbol{\beta}}_{or}(\tau)$; see Chao, Volgushev and Cheng (2017).

The following condition characterizes the upper bound on S which is sufficient to ensure the oracle rule for $\overline{\boldsymbol{\beta}}(\tau)$.

(L1) Assume that

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

For specific examples, Condition (L1) can be simplified. For instance, in the setting of Example 3.6, we can reduce (L1) to the form

$$S = o\big(N^{1/2}(\log N)^{-2} m^{-1/2} \wedge N(\log N)^{-10} m^{-2}\big).$$

We now present the sufficient and necessary conditions for the oracle rule of $\overline{\boldsymbol{\beta}}(\tau)$ under the Condition (L) for $\mathbf{Z}$.

COROLLARY 3.9 [Oracle rule for $\overline{\boldsymbol{\beta}}(\tau)$]. *Assume that (L1) holds and data are generated from P_N with $(P_N, \mathbf{Z}) \in \mathscr{P}_L(M, \bar{f}, \overline{f'}, f_{\min}, R)$. Then the pooled estimator $\overline{\boldsymbol{\beta}}(\tau)$ defined in (2.3) satisfies for any fixed $\tau \in \mathcal{T}$, $x_0 \in \mathcal{X}$,*

$$\frac{\sqrt{N}\mathbf{Z}(x_0)^\top (\overline{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau))}{(\mathbf{Z}(x_0)^\top J_m(\tau)^{-1} \mathbb{E}[\mathbf{Z}\mathbf{Z}^\top] J_m(\tau)^{-1} \mathbf{Z}(x_0))^{1/2}} \rightsquigarrow \mathcal{N}\big(0, \tau(1-\tau)\big), \tag{3.13}$$

where $J_m(\tau)$ is defined in Corollary 3.3. This matches the limiting behavior of $\widehat{\boldsymbol{\beta}}_{or}$. If $S \gtrsim N^{1/2}\xi_m^{-1}$, the weak convergence result (3.13) fails for some $(P_N, \mathbf{Z}) \in \mathscr{P}_L(1, \bar{f}, \overline{f'}, f_{\min}, R)$, $x_0 \in \mathcal{X}$.

To compare the necessary and sufficient conditions in Corollary 3.9, assume for simplicity that $m \gg N^{\alpha}$ for some $\alpha > 0$ and $\|\mathbf{Z}(x_0)\| \asymp \xi_m$ (this is, for instance, the case for univariate splines as discussed in Example 3.6). Since under (A1)–(A3) we have $m^{1/2} \lesssim \xi_m$, it follows that (L1) holds provided that $S = o(N^{1/2}\xi_m^{-1}(\log N)^{-2})$ and the necessary and sufficient conditions match up to a factor of $(\log N)^2$.

Next, we discuss sufficient conditions for the process oracle rule.

COROLLARY 3.10. *Consider an arbitrary vector $x_0 \in \mathcal{X}$. Assume that data are generated from P_N with $(P_N, \mathbf{Z}) \in \mathscr{P}_L(M, \bar{f}, \overline{f'}, f_{\min}, R)$, that (L1) holds, that $\tau \mapsto Q(x_0;\tau) \in \Lambda_c^{\eta}(\mathcal{T})$, $r_\tau \geq \eta$, $\sup_{\tau\in\mathcal{T}} |\mathbf{Z}(x_0)^\top \boldsymbol{\gamma}_N(\tau) - Q(x_0;\tau)| = o(\|\mathbf{Z}(x_0)\|N^{-1/2})$, that $N^2 \gg K \gg G \gg N^{1/(2\eta)}\|\mathbf{Z}(x_0)\|^{-1/\eta}$, $c_m^2(\boldsymbol{\gamma}_N) = o(N^{-1/2})$ and that the limit*

$$H_{x_0}(\tau_1,\tau_2) := \lim_{N\to\infty} \frac{\mathbf{Z}(x_0)^\top J_m^{-1}(\tau_1)\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]J_m^{-1}(\tau_2)\mathbf{Z}(x_0)(\tau_1\wedge\tau_2 - \tau_1\tau_2)}{\|\mathbf{Z}(x_0)\|^2} \tag{3.14}$$

exists and is nonzero for any $\tau_1, \tau_2 \in \mathcal{T}$, where $J_m(\tau)$ is defined in the statement of Corollary 3.3.

1. *The projection estimator $\widehat{\boldsymbol{\beta}}(\tau)$ defined in (2.5) satisfies*

$$\frac{\sqrt{N}}{\|\mathbf{Z}(x_0)\|}\big(\mathbf{Z}(x_0)^\top\widehat{\boldsymbol{\beta}}(\cdot) - Q(x_0;\cdot)\big) \rightsquigarrow \mathbb{G}_{x_0}(\cdot) \qquad \text{in } \ell^\infty(\mathcal{T}), \tag{3.15}$$

where $\mathbb{G}_{x_0}$ is a centered Gaussian process with $\mathbb{E}[\mathbb{G}_{x_0}(\tau)\mathbb{G}_{x_0}(\tau')] = H_{x_0}(\tau,\tau')$. The same holds for the oracle estimator $\widehat{\boldsymbol{\beta}}_{or}(\tau)$.

If $G \lesssim N^{1/(2\eta)}\|\mathbf{Z}(x_0)\|^{-1/\eta}$ or $S \gtrsim N^{1/2}\xi_m^{-1}$, the weak convergence in (3.15) fails for some $(P_N, \mathbf{Z})$ which satisfies the above conditions.

2. *For $\widehat{F}_{Y|X}(\cdot|x_0)$ defined in (2.8) and $\mathbb{G}$ defined above, we have*

$$\frac{\sqrt{N}}{\|\mathbf{Z}(x_0)\|}\big(\widehat{F}_{Y|X}(\cdot|x_0) - F_{Y|X}(\cdot|x_0)\big) \rightsquigarrow -f_{Y|X}(\cdot|x_0)\mathbb{G}_{x_0}\big(F_{Y|X}(\cdot|x_0)\big) \qquad \text{in } \ell^\infty\big(\big(Q(x_0;\tau_L), Q(x_0;\tau_U)\big)\big).$$

This matches the process convergence for $\widehat{F}^{or}_{Y|X}(\cdot|x_0)$.

The proof of the sufficient conditions in Corollary 3.10 is presented in Section S.3.2.3, and the collapse of the weak convergence (3.15) is shown in Section S.3.3. Similar to the discussion after Corollary 3.4, the process oracle rule does not place additional restrictions on the number of subsamples S besides (L1). However, the process oracle rule requires additional assumptions on the quantile grid $\mathcal{T}_K$. An interesting observation is that $G \gg (N/\|\mathbf{Z}(x_0)\|^2)^{1/(2\eta)}$ in Theorem 3.14 can be weaker than $G \gg N^{1/(2\eta)}$ from Corollary 3.4. For instance, this is true in the

setting of Example 3.6 where $\|\mathbf{Z}(x_0)\| \asymp m^{1/2}$. Consequently, the computational burden is reduced since K can be chosen smaller. The intuition behind this surprising phenomenon is that the convergence rate for the estimator $\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\tau)$ in nonparametric models is typically slower. Thus, less stringent assumptions are needed to ensure that the bias induced by quantile projection is negligible compared to the convergence rate of $\widehat{\boldsymbol{\beta}}_{or}(\tau)$.

The sufficient conditions in this section can be extended to cover approximately linear models with increasing dimension that do not satisfy condition (L). Technical details are provided in the Appendix.

REMARK 3.11. To the best of our knowledge, the only paper that also studies the sharpness of upper bounds for S that guarantee valid inference in a divide-and-conquer setting with nonparametric regression is Shang and Cheng (2017). However, Shang and Cheng (2017) only consider nonparametric *mean* regression in the smoothing spline setup.

4. Practical aspects of inference. In the previous section, we derived conditions which guarantee that the divide-and-conquer estimator $\overline{\boldsymbol{\beta}}$ and the quantile projection estimator $\widehat{\boldsymbol{\beta}}$ share the same asymptotic distribution as the "oracle" estimator $\widehat{\boldsymbol{\beta}}_{or}$, so that inference based on the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{or}$ remains valid. In practice, this result can not be directly utilized since the asymptotic variance of the oracle estimator $\widehat{\boldsymbol{\beta}}_{or}$ is in general not pivotal. Classical approaches to inference for this estimator are typically based on estimating its asymptotic variance from the data directly, or conducting bootstrap to approximate the asymptotic distribution.

Estimating the limiting variance requires the choice of a bandwidth parameter, and existing research indicates that classical rules for bandwidth selection need to be adjusted in a divide-and-conquer setting [see, e.g., Banerjee, Durot and Sen (2019), Xu, Shang and Cheng (2016)]. We discuss related issues for variance estimation in Section 4.2.

Conducting bootstrap substantially increases the computational burden of any procedure, which is problematic in a massive data setting we are considering here. While recent proposals by Kleiner et al. (2014), Sengupta, Volgushev and Shao (2016) provide a way to reduce the computational cost of bootstrap in a big data setting, the approaches described in those papers are not easily combined with the divide-and-conquer setting which we consider here.

As an alternative to variance estimation or classical bootstrap approaches, we propose several simple inference procedures which directly utilize the fact that in a divide-and-conquer setting estimators from subsamples are available. Those procedures are very easy to implement, and require only a very small amount of computation on the central computer without additional communication costs. Details are provided in Section 4.1.

4.1. *Inference utilizing results from subsamples.* We begin by discussing inference at a fixed quantile level τ. The key idea in this section is to make direct use of the fact that S estimators $\widehat{\boldsymbol{\beta}}_s(\tau)$ from subsamples are available. Observe that the estimator $\overline{\boldsymbol{\beta}}(\tau)$ is simply an average of $\widehat{\boldsymbol{\beta}}_1(\tau), \ldots, \widehat{\boldsymbol{\beta}}_S(\tau)$ which can be seen as i.i.d. realizations (provided groups are homogeneous) of a random variable with approximately normal distribution. This suggests two simple options:

1. Use the sample covariance matrix, say $\widehat{\Sigma}^D$, of $\widehat{\boldsymbol{\beta}}_1(\tau), \ldots, \widehat{\boldsymbol{\beta}}_S(\tau)$ in order to conduct inference on $\overline{\boldsymbol{\beta}}(\tau)$ or linear functionals thereof such as $\overline{Q}(x;\tau) := \mathbf{Z}(x)^\top \overline{\boldsymbol{\beta}}(\tau)$. For example, a simple asymptotic level α confidence interval for $Q(x;\tau)$ is given by

$$\big[\mathbf{Z}(x_0)^\top \overline{\boldsymbol{\beta}}(\tau) \pm S^{-1/2}\big(\mathbf{Z}(x_0)^\top \widehat{\Sigma}^D \mathbf{Z}(x_0)\big)^{1/2} \Phi^{-1}(1-\alpha/2)\big]. \tag{4.1}$$

2. A refined version of the previous approach is to additionally exploit the fact that a suitably scaled version of $\mathbf{u}_N^\top \widehat{\boldsymbol{\beta}}_s(\tau)$ should be approximately normal since each $\widehat{\boldsymbol{\beta}}_s(\tau)$ is itself an estimator based on sample of i.i.d. data. Hence for small S (say $S \le 30$) more precise confidence intervals can be obtained by using quantiles of the student t distribution (if $\mathbf{u}_N$ is a vector) or F distribution (if $\mathbf{u}_N$ is a fixed-dimensional matrix). For example, a modification of the confidence interval in (4.1) would take the form

$$\big[\mathbf{Z}(x_0)^\top \overline{\boldsymbol{\beta}}(\tau) \pm S^{-1/2}\big(\mathbf{Z}(x_0)^\top \widehat{\Sigma}^D \mathbf{Z}(x_0)\big)^{1/2} t_{S-1,1-\alpha/2}\big], \tag{4.2}$$

where $t_{S-1,1-\alpha/2}$ denotes the $1-\alpha/2$-quantile of the t-distribution with $S-1$ degrees of freedom. The asymptotic validity of both intervals discussed above is provided in the following theorem.

THEOREM 4.1. *Assume that the conditions of either Corollary 3.3, Corollary 3.9 or Corollary A.2 hold, that* $c_m(\boldsymbol{\gamma}_N) = o(\|\mathbf{Z}(x_0)\| N^{-1/2})$ *and that* $S \ge 2$ (S *can be fixed*). *Then the confidence interval* (4.2) *has asymptotic* (*for* $N \to \infty$) *coverage probability* $1-\alpha$.

If additionally $S \to \infty$, *the confidence interval given in* (4.1) *also has asymptotic coverage probability* $1-\alpha$.

See Section S.4.1 for a proof of Theorem 4.1. The main advantage of the two approaches discussed above lies in their simplicity as they do not require any costly computation or communication between machines. There are two main limitations.

First, for small values of S (say $S \le 30$) the confidence intervals in (4.1) will not have the correct coverage while the interval in (4.2) can be substantially wider than the one based on the exact asymptotic distribution since quantiles of the t-distribution with few degrees of freedom can be substantially larger than corresponding normal quantiles. Moreover, the approach is not applicable if S is smaller than the dimension of the parameter of interest. Second, the approaches are not

straightforward to generalize to inference on nonlinear functionals of $\boldsymbol{\beta}(\tau)$ such as $\widehat{F}_{Y|X}(y|x)$ or inference which is uniform over a continuum of quantiles τ.

The first limitation is not likely to become relevant in a big data setting since here we typically expect that S is large due to computational bottlenecks and memory constraints. For the sake of completeness, we discuss this case in the next section. To deal with the second limitation, we propose to use a bootstrap procedure that can be conducted by solely using the subsample estimators $\{\widehat{\boldsymbol{\beta}}_s(\tau_k)\}_{s=1,\ldots,S,k=1,\ldots,K}$ which are stored on the central machine. Details are provided below:

1. 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 (i.e., weights are chosen such that $\omega_{s,b}\geq 0$, $\mathbb{E}[\omega_{s,b}]=Var(\omega_{s,b})=1$) and let $\bar{\omega}_{\cdot,b}:=S^{-1}\sum_{s=1}^{S}\omega_{s,b}$.
2. For $b=1,\ldots,B$, $k=1,\ldots,K$, compute the bootstrap estimators

$$\overline{\boldsymbol{\beta}}^{(b)}(\tau_k):=\frac{1}{S}\sum_{s=1}^{S}\frac{\omega_{s,b}}{\bar{\omega}_{\cdot,b}}\widehat{\boldsymbol{\beta}}^{s}(\tau_k) \tag{4.3}$$

and define the matrix $\widehat{\Xi}^{(b)}$ from (2.7) with $\overline{\boldsymbol{\beta}}^{(b)}(\tau_k)$ replacing $\overline{\boldsymbol{\beta}}(\tau_k)$.

3. Similarly, as in (2.5) compute $\widehat{\boldsymbol{\beta}}^{(b)}(\cdot)$ from the matrix $\widehat{\Xi}^{(b)}$ defined above. For a functional of interest Φ approximate quantiles of the distribution of $\Phi(\widehat{\boldsymbol{\beta}}(\cdot))-\Phi(\boldsymbol{\beta}(\cdot))$ by the empirical quantiles of $\{\Phi(\widehat{\boldsymbol{\beta}}^{(b)}(\cdot))-\Phi(\widehat{\boldsymbol{\beta}}(\cdot))\}_{b=1,\ldots,B}$.

A formal justification for this bootstrap approach is provided by the following result.

THEOREM 4.2. *Let the assumptions of either Corollary 3.4, Corollary 3.10 or Corollary A.5 hold and assume that additionally $S\to\infty$. Then we have conditionally on the data $(X_i,Y_i)_{i=1,\ldots,N}$,*

$$\frac{\sqrt{N}}{\|\mathbf{Z}(x_0)\|}\big(\mathbf{Z}(x_0)^{\top}\widehat{\boldsymbol{\beta}}^{(1)}(\cdot)-\mathbf{Z}(x_0)^{\top}\widehat{\boldsymbol{\beta}}(\cdot)\big)\rightsquigarrow\mathbb{G}_{x_0}(\cdot)\qquad \text{in } \ell^{\infty}(\mathcal{T}),$$

where the limit $\mathbb{G}_{x_0}$ denotes the centered Gaussian process from Corollary 3.10 or Corollary A.5 under their respective assumptions and $\mathbb{G}_{x_0}=\mathbf{Z}(x_0)^{\top}\mathbb{G}$ under the assumptions of Corollary 3.4.

The proof of Theorem 4.2 is given in Section S.4.2. We conclude this section by remarking that the bootstrap proposed differ from the cluster-robust bootstrap of Hagemann (2017). The main difference is that we propose to directly utilize the subsample estimators $\widehat{\boldsymbol{\beta}}^{s}$ while the approach of Hagemann (2017) requires repeated estimation on the complete sample.

4.2. *Inference based on estimating the asymptotic covariance matrix.* As suggested by a referee, an alternative way to conduct inference is to compute, for

each subsample, not only the estimator $\widehat{\boldsymbol{\beta}}_s$ but also a variance estimator and pool those estimators. Here, we provide some details on this approach. For the sake of simplicity, we only discuss the case where m is fixed and the model is specified correctly, that is, the setting of Section 3.1.

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

$$\Sigma(\tau) = \tau(1-\tau) J_m(\tau)^{-1} \mathbb{E}\big[\mathbf{Z}\mathbf{Z}^\top\big] J_m(\tau)^{-1},$$

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

The middle part $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]$ is easily estimated by $\frac{1}{nS}\sum_i \sum_s \mathbf{Z}_{is}\mathbf{Z}_{is}^\top$. Since this is a simple mean of the subsample-based quantities $\frac{1}{n}\sum_i \mathbf{Z}_{is}\mathbf{Z}_{is}^\top$, implementing this in a distributed computing setting is straightforward.

The matrix $J_m(\tau)$ involves the conditional density $f_{Y|X}(Q(X;\tau)|X)$ and is more difficult to estimate. A popular approach is based on Powell's estimator [Powell (1986)]

$$\widehat{J}_{ms}^P(\tau) := \frac{1}{2nh_n} \sum_{i=1}^n \mathbf{Z}_{is}\mathbf{Z}_{is}^\top \mathbf{1}\big\{\big|Y_{is} - \mathbf{Z}_{is}^\top \widehat{\boldsymbol{\beta}}^s(\tau)\big| \le h_n\big\}. \tag{4.4}$$

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

There are several possible approaches to estimate $J_m^P(\tau)$ in a parallel computing environment. If an additional round of communication is acceptable, it is possible to construct estimators with the same convergence rate and asymptotic distribution as the estimator based on the full sample. Details are provided in Section S.1 in the Supplementary Material [Volgushev, Chao and Cheng (2019)]. If only a single round of communication is allowed, the following algorithm can be used instead:

1. For $s = 1, \dots, S$, in the same round as computing $\widehat{\boldsymbol{\beta}}^s(\tau)$, compute $\widehat{J}_{ms}^P(\tau)$ from (4.4) and $\widehat{\Sigma}_{1s} := \frac{1}{n}\sum_i \mathbf{Z}_{is}\mathbf{Z}_{is}^\top$.
2. Along with $(\widehat{\boldsymbol{\beta}}^s(\tau))_{s=1,\dots,S}$, send $(\widehat{J}_{ms}^P(\tau), \widehat{\Sigma}_{1s})_{s=1,\dots,S}$ to the master machine and compute $\overline{J}_m^P(\tau) := \frac{1}{S}\sum_{s=1}^S \widehat{J}_{ms}^P(\tau)$, $\overline{\Sigma}_1 := \frac{1}{S}\sum_{s=1}^S \widehat{\Sigma}_{1s}$.
3. The final variance estimator is given by $\bar{\Sigma}(\tau) = \tau(1-\tau)\overline{J}_m^P(\tau)^{-1}\overline{\Sigma}_1 \times \overline{J}_m^P(\tau)^{-1}$.

REMARK 4.3. Note that in the above algorithm we first take averages over the subsampled estimators $\widehat{J}_{ms}^P(\tau)$ and only invert the aggregated matrix $\overline{J}_m^P(\tau)$. An alternative approach would have been to compute the estimator $\widehat{J}_{ms}^P(\tau)^{-1}\widehat{\Sigma}_s\widehat{J}_{ms}^P(\tau)^{-1}$ for each subsample and average in the end. However, given that $A \mapsto A^{-1}$ is nonlinear, this might result in additional bias since in general for random matrices $(\mathbb{E}[A])^{-1} \neq \mathbb{E}[A^{-1}]$.

An important question for implementing the above algorithm is the choice of the bandwidth parameter h_n. To gain some intuition about the optimal choice of h_n,

we will formally discuss the case of a linear model fixed dimension. First, observe that by a Taylor expansion we have almost surely

$$\begin{aligned}
&\overline{J}_m^P(\tau)^{-1}\overline{\Sigma}_1\overline{J}_m^P(\tau)^{-1} - J_m(\tau)^{-1}\Sigma_1 J_m(\tau)^{-1} \\
&\quad = -J_m(\tau)^{-1}\big(\overline{J}_m^P(\tau) - J_m(\tau)\big)J_m(\tau)^{-1}\Sigma_1 J_m(\tau)^{-1} \\
&\qquad - J_m(\tau)^{-1}\Sigma_1 J_m(\tau)^{-1}\big(\overline{J}_m^P(\tau) - J_m(\tau)\big)J_m(\tau)^{-1} \\
&\qquad + J_m(\tau)^{-1}(\Sigma_1 - \overline{\Sigma}_1)J_m(\tau)^{-1} \\
&\qquad + O\big(\|\overline{J}_m^P(\tau) - J_m(\tau)\|^2 + \|\overline{\Sigma}_1 - \Sigma_1\|\|\overline{J}_m^P(\tau) - J_m(\tau)\|\big).
\end{aligned}$$

Ignoring higher-order terms and terms that do not depend on h_n it suffices to analyze the properties of $\overline{J}_m^P(\tau) - J_m(\tau)$.

THEOREM 4.4. *Under assumptions (A1)–(A3), assume that additionally $y \mapsto f'_{Y|X}(y|x)$ is continuously differentiable with first derivative being jointly continuous and uniformly bounded as a function of x, y and that $nh_n(\log n)^{-1} \to \infty$. Then*

$$\overline{J}_m^P(\tau) - J_m(\tau) = A_n(\tau) + O_p\bigg(\frac{\log n}{nh_n}\bigg),$$

where the exact form of $A_n(\tau)$ is given in the proof. Moreover,

$$\mathbb{E}\big[A_n(\tau)\big] = \frac{h_n^2}{6}\mathbb{E}\big[\mathbf{Z}\mathbf{Z}^\top f''_{Y|X}\big(\mathbf{Z}^\top\boldsymbol{\beta}(\tau)|X\big)\big] + O\bigg(\frac{\log n}{n}\bigg) + o\big(h_n^2\big)$$

and for $A_{N,(j,k)}$ denoting the entry of A_N in row j and column k

$$\begin{aligned}
\operatorname{Cov}\big(A_{N,(j,k)}(\tau), A_{N,(u,v)}(\tau)\big) = \frac{1}{Nh_n}\mathbb{E}\big[f_{Y|X}\big(\mathbf{Z}^\top\boldsymbol{\beta}(\tau)|X\big)Z_jZ_kZ_uZ_v\big] \\
+ o\bigg(\frac{1}{Nh_n}\bigg).
\end{aligned}$$

The proof of Theorem 4.4 is given in Section S.4.3. Theorem 4.4 has several interesting implications. First, note that the asymptotic MSE of $A_N(\tau)$ is of the order $h_n^4 + (Nh_n)^{-1}$, which is minimized for $h_N \sim N^{-1/5}$ (note that the term $\log n/n$ is negligible). Under the additional condition $S = o(N^{2/5}(\log N)^{-1})$, we have $\log n/(nh_n) = o(N^{-2/5})$, and in this setting the MSE of the first-order expansion of $\overline{J}_m^P(\tau)$ matches that of the Powell "oracle" estimator as derived in Kato (2012). This shows that, despite using estimators from subsamples, the same rate for estimating $J_m^P(\tau)$ as from the full sample can be achieved. That requires a stronger condition on S than the oracle rate for the estimator $\widehat{\boldsymbol{\beta}}$. It is not clear if the latter condition is sharp, and we leave an investigation of this issue to future research.

Second, the choice $h_n \sim n^{-1/5}$ which would have been optimal for estimating $J_m(\tau)$ based on a subsample of size n does not lead to the optimal error rate for the averaged estimator $\overline{J}_m^P(\tau)$. In fact, the optimal bandwidth for $\overline{J}_m^P(\tau)$ is always smaller, which corresponds to undersmoothing. Similar effects were observed in various settings by Zhang, Duchi and Wainwright (2015), Banerjee, Durot and Sen (2019) (see their Appendix A.13) and Shang and Cheng (2017).

5. Monte Carlo experiments. In this section, we demonstrate our theory with simulation experiments. Due to space limitations, we restrict our attention to correctly specified linear models with different dimensions of predictors. More precisely, we consider data generated from

$$Y_i = 0.21 + \boldsymbol{\beta}_{m-1}^\top X_i + \varepsilon_i, \qquad i = 1, \dots, N, \tag{5.1}$$

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 $\mathcal{U}([0,1]^{m-1})$ with $\operatorname{Cov}(X_{ij}, X_{ik}) = 0.1^2 0.7^{|j-k|}$ for $j, k = 1, \dots, m-1$, and the vector $\boldsymbol{\beta}_{m-1}$ takes the form

$$\begin{aligned}
\boldsymbol{\beta}_3 &= (0.21, -0.89, 0.38)^\top; \\
\boldsymbol{\beta}_{15} &= (\boldsymbol{\beta}_3^\top, 0.63, 0.11, 1.01, -1.79, -1.39, 0.52, -1.62, \\
&\qquad 1.26, -0.72, 0.43, -0.41, -0.02)^\top; \\
\boldsymbol{\beta}_{31} &= (\boldsymbol{\beta}_{15}^\top, 0.21, \boldsymbol{\beta}_{15}^\top)^\top.
\end{aligned} \tag{5.2}$$

Throughout this section, we fix $\mathcal{T} = [0.05, 0.95]$. Section 5.1 contains results for the estimator $\overline{\boldsymbol{\beta}}(\tau)$, while results for $\widehat{F}_{Y|X}(y|x)$ are collected in Section 5.2. Additional simulations (including models with heteroskedastic errors) are presented in Section S.9 of the Supplementary Material [Volgushev, Chao and Cheng (2019)].

5.1. *Results for the divide-and-conquer estimator $\overline{\boldsymbol{\beta}}(\tau)$.* We fix the subsample size n and consider the impact of the number of subsamples S on the coverage probabilities of different 95% confidence intervals. To benchmark our results, we use the infeasible asymptotic confidence interval

$$\big[x_0^\top \overline{\boldsymbol{\beta}}(\tau) \pm N^{-1/2} \sigma(\tau) \Phi^{-1}(1-\alpha/2)\big], \tag{5.3}$$

where $\sigma^2(\tau)$ denotes the theoretical asymptotic variance of $\boldsymbol{\beta}^{or}(\tau)$; this CI is valid by the oracle rule but contains unknown quantities. The coverage properties of this interval also indicate whether we are in a regime where the oracle rule holds.

In a first step, we consider the properties of confidence intervals discussed in Section 4.1 which directly utilize the availability of results from subsamples. We consider the following three types of confidence intervals:

1. The normal confidence interval (4.1).
2. The confidence interval (4.2) based on quantiles of the t-distribution.

3. The bootstrap confidence interval based on sample quantiles of $(\bar{\boldsymbol{\beta}}^{(1)}(\tau) - \bar{\boldsymbol{\beta}}(\tau)), \ldots, (\bar{\boldsymbol{\beta}}^{(B)}(\tau) - \bar{\boldsymbol{\beta}}(\tau))$, where $\bar{\boldsymbol{\beta}}^{(1)}(\tau), \ldots, \bar{\boldsymbol{\beta}}^{(B)}(\tau)$ are computed as in (4.3); in this section, we set $B = 500$.

The coverage probabilities of corresponding confidence intervals are summarized in Figure 2 where we fix two subsample sizes $n = 512, 2048$ and present two types of plots: coverage probabilities for "small" values of S ranging from $S = 2$ to $S = 50$ and coverage probabilities for "large" values of $S = 2^k$, $k = 1, \ldots, 10$. Only $m = 4, 32$ is considered here; $m = 16$ can be found in the Supplementary Material [Volgushev, Chao and Cheng (2019)].

Plots for small S help to access how different procedures behave if the number of subsamples is small. As expected from the theory, oracle confidence intervals and simple confidence intervals based on the t-distribution maintain nominal coverage for small values of S. Confidence intervals based on normal approximation and the bootstrap undercover for smaller values of S. Starting from about $S = 20$, coverage probabilities of all four types of intervals are very close. We also see that for $m = 32$, the oracle rule does not apply for a subsample size of $n = 512$ with any number of subsamples $S > 10$; the situation improves for $n = 2048$. This is in line with our asymptotic oracle theory. For "larger" values of S, there is no difference in the coverage probabilities of different intervals. As predicted by the oracle theory, coverage starts to drop earlier for models with larger m.

Next, we analyze the properties of asymptotic confidence intervals which are based on estimating the asymptotic variance of $x_0^\top \boldsymbol{\beta}(\tau)$ from data. We compare three different ways of estimating the asymptotic variance:

1. A simple pooled estimator which uses the default settings in the package `quantreg` to obtain estimated variances in each subsample and takes the average over all subsamples (additional details are provided in the Supplementary Material [Volgushev, Chao and Cheng (2019)], Section S.9.1).
2. The estimator $\bar{\Sigma}$ based on the bandwidth $c^*(\tau)n^{-1/5}$ which minimizes the asymptotic MSE of the estimator J_m within subsamples of size n (additional details on the choice of optimal constant $c^*(\tau)$ are provided in the Supplementary Material [Volgushev, Chao and Cheng (2019)], Section S.9.1).
3. The estimator $\bar{\Sigma}$ based on the bandwidth $c^*(\tau)N^{-1/5}$ which is motivated by the theory developed in Theorem 4.4.

The results are reported in Table 1. Since there is no notable difference between all approaches when S is large, only results for $S \leq 50$ are displayed for the sake of brevity. Interestingly, we do not observe a big difference between the naive bandwidth choice $h_n \sim n^{-1/5}$ and the optimal undersmoothing choice $h_n \sim N^{-1/5}$. This finding is quite intuitive since, once the asymptotic variance is estimated with at most 5–10% relative error, a further increase in estimation accuracy does not lead to substantial improvements in coverage probabilities. The completely automatic choice implemented in the `quantreg` package also performs reasonably well.

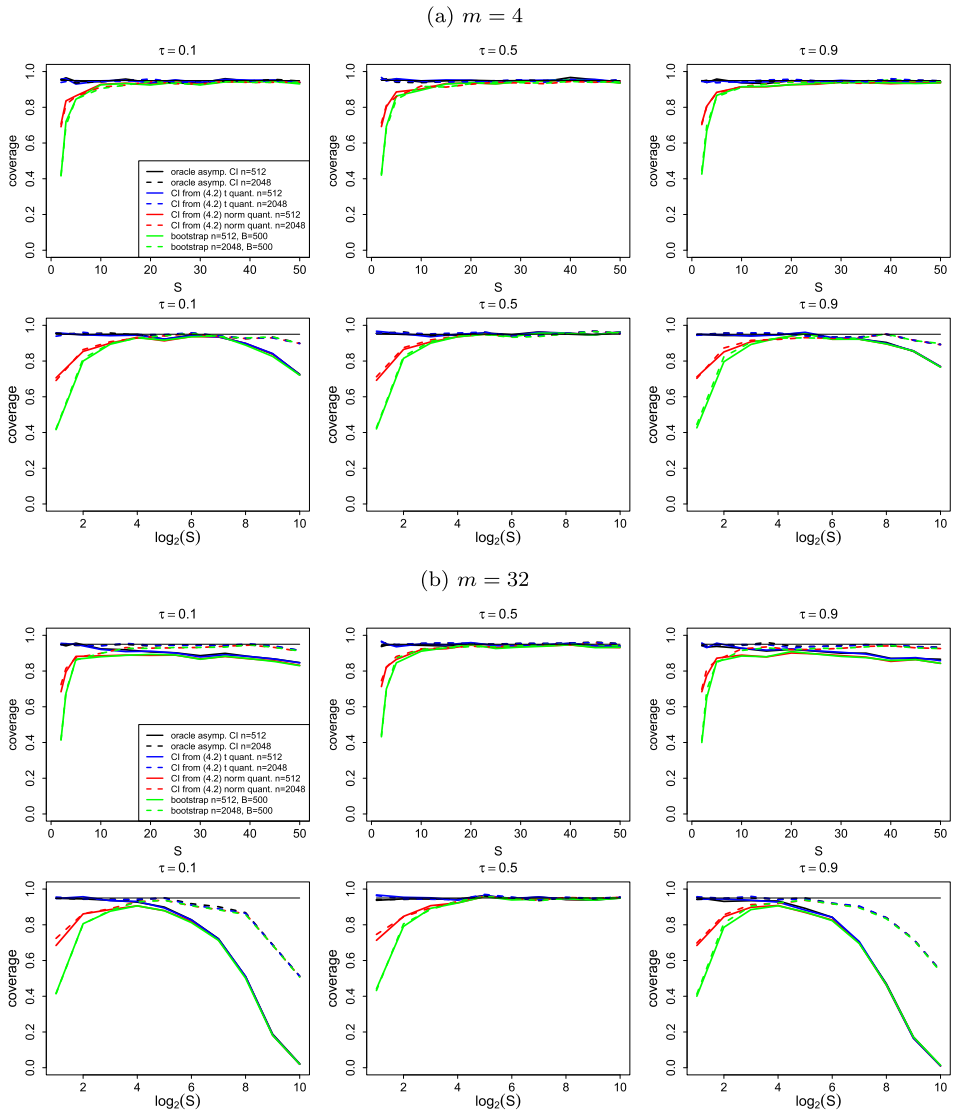

FIG. 2. *Coverage probabilities for* $x_0^\top \boldsymbol{\beta}(\tau)$ *for different values of* S *and* $\tau = 0.1, 0.5, 0.9$ (*left*, *middle*, *right row*). *Solid lines*: $n = 512$, *dashed lines*: $n = 2048$. *Black*: *asymptotic oracle CI*, *blue*: *CI from* (4.2) *based on* t *distribution*, *red*: *CI from* (4.1) *based on normal distribution*, *green*: *bootstrap CI*. *Throughout* $x_0 = (1, \ldots, 1)/m^{1/2}$, *nominal coverage* 0.95.

Finally, note that the pattern of coverage probabilities varies at different τ. For example, in the linear models with normal errors, the coverage probabilities at tail quantiles ($\tau = 0.1, 0.9$) drop to zero much faster than those at $\tau = 0.5$. These empirical observations are not inconsistent with our theory where only the orders of the upper bound for S are shown to be the same irrespective of the value of τ.

TABLE 1

Coverage probabilities based on estimating the asymptotic variance. Different rows correspond to different methods for obtaining covariance matrix or: using true asymptotic variance matrix, def: default choice implemented in quantreg package, nai: asymptotically optimal constant with scaling $h_n \sim n^{-1/5}$, adj: asymptotically optimal constant with scaling $h_n \sim N^{-1/5}$ as suggested by Theorem 4.4

	S											
	1	**10**	**30**	**50**	**1**	**10**	**30**	**50**	**1**	**10**	**30**	**50**
	$n = 512, m = 4, \tau = 0.1$				$n = 512, m = 16, \tau = 0.1$				$n = 512, m = 32, \tau = 0.1$			
or	94.9	94.8	94.9	94.3	94.7	94.3	92.2	90.7	94.4	92.5	88.6	85.5
def	92.6	93.7	93.6	92.8	92	92.3	90.1	88.4	92.5	90.6	86.2	82.9
nai	94.2	93.9	93.6	93.2	96.6	93.3	91	89.4	98.9	92.7	88.4	85
adj	94.2	94.5	94.3	94.2	96.6	94.3	92.1	90.8	98.9	94.4	89.5	87.2
	$n = 512, m = 4, \tau = 0.5$				$n = 512, m = 16, \tau = 0.5$				$n = 512, m = 32, \tau = 0.5$			
or	94.7	96	95.9	95	94.2	95.7	94.8	95.4	95.4	95.5	94.8	95
def	97.8	98.2	98.2	98.2	96.7	98.2	98	98	97.6	97.5	97.3	97.7
nai	95.9	96.8	96.4	96.3	96.7	97	96.4	97	99	97	96.2	96.7
adj	95.9	96.4	96	95.2	96.7	96.4	95.7	96.2	99	96.6	95.7	95.9
	$n = 512, m = 4, \tau = 0.9$				$n = 512, m = 16, \tau = 0.9$				$n = 512, m = 32, \tau = 0.9$			
or	95.4	94.6	94.2	93.6	94.6	94	91.7	90.1	95.2	92.1	90.6	87.1
def	94	93.6	92.8	92.4	92.6	92.2	90	88.1	92.2	90.2	88.2	84.3
nai	94.8	93.8	93.1	92.6	96.6	93.2	90.8	88.6	99	92.5	90.1	86.6
adj	94.8	94.3	93.8	93.7	96.6	94	91.9	90.4	99	93.8	91.4	88.4
	$n = 2048, m = 4, \tau = 0.1$				$n = 2048, m = 16, \tau = 0.1$				$n = 2048, m = 32, \tau = 0.1$			
or	95.6	94.7	94.4	94.3	95	94.2	94.4	94.4	95	95	94	94
def	94.7	94.1	93.8	94	93.8	93.2	93.5	93.7	93.6	94.2	92.7	93.1
nai	95.1	94.1	93.7	94	95.3	93.5	93.5	93.7	95.7	94.6	93.1	93.2
adj	95.1	94.6	94.4	94.5	95.3	94	94.1	94.4	95.7	94.9	93.8	93.8
	$n = 2048, m = 4, \tau = 0.5$				$n = 2048, m = 16, \tau = 0.5$				$n = 2048, m = 32, \tau = 0.5$			
or	94.6	94.9	95.2	95.3	95.3	94.2	94.2	95.5	95.3	95.5	95.1	95.5
def	96	96.2	96.7	96.2	96.7	96.2	95.4	96.7	96	96.4	96.3	96.6
nai	95.1	95.4	95.5	95.6	96.8	95.2	95	96	96.5	95.8	95.7	95.9
adj	95.1	95.1	95.1	95.3	96.8	95	94.7	95.6	96.5	95.6	95.2	95.5
	$n = 2048, m = 4, \tau = 0.9$				$n = 2048, m = 16, \tau = 0.9$				$n = 2048, m = 32, \tau = 0.9$			
or	94.6	95.6	95.8	94.6	95.1	95	94.5	94.7	95.2	94.4	92.9	92.9
def	94.2	94.9	94.9	94.1	94.3	94.5	94	94.3	94.4	93.8	92.6	92.2
nai	94.4	94.9	94.8	94.1	95.5	94.7	94.2	94.4	96.7	94.4	92.7	92.5
adj	94.4	95.3	95.6	94.6	95.5	95.1	94.6	95	96.7	94.9	93.2	93.3

Rather, this phenomenon might be explained by different derivatives of the error density that appear in the estimation bias, and is left for future study.

5.2. *Results for the estimator* $\widehat{F}_{Y|X}(y|x)$. In this section, we consider inference on $F_{Y|X}(y|x)$. We compare the coverage probability of the oracle asymptotic confidence interval

$$\big[\widehat{F}_{Y|X}\big(Q(x_0;\tau)|x_0\big) \pm N^{-1/2}\sigma_F^2(\tau)\Phi^{-1}(1-\alpha/2)\big] \tag{5.4}$$

[here $\sigma_F^2(\tau)$ is the asymptotic variance of the oracle estimator] and the bootstrap confidence interval described above Theorem 4.2. Note that the other approaches described in Section 4 are not directly applicable here since $\widehat{F}_{Y|X}(y|x)$ is a functional of the whole process $\widehat{\boldsymbol{\beta}}(\cdot)$. Since we focus on bootstrap reliability, the number of quantile levels $K = 65$ and knots for spline interpolation $G = 32$ are chosen sufficiently large to ensure nominal coverage of oracle intervals. A detailed study of the impact of K, G on coverage of oracle intervals is provided in Section S.9.3.2 of the Supplementary Material [Volgushev, Chao and Cheng (2019)]. Due to space limitations, we only show the results for small values of S; results for large values of S do not give crucial additional insights and are deferred to the Supplementary Material [Volgushev, Chao and Cheng (2019)]. Coverage probabilities for $m = 4, 32$ are reported in Figure 3. For $m = 4$, the bootstrap and oracle confidence interval show a very similar performance as soon as $S \geq 20$; this is in line with the coverage properties for the bootstrap for $\bar{\boldsymbol{\beta}}$. For $m = 32$, coverage probabilities of the oracle confidence interval indicate that the subsample size n is too small and

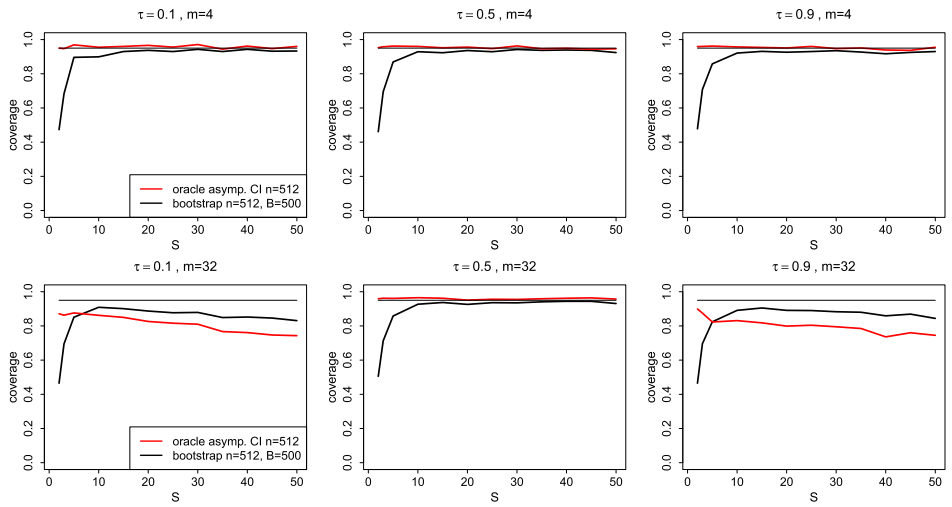

FIG. 3. *Coverage probabilities for oracle confidence intervals* (*red*) *and bootstrap confidence intervals* (*black*) *for* $F_{Y|X}(y|x_0)$ *for* $x_0 = (1,\dots,1)/m^{1/2}$ *and* $y = Q(x_0;\tau)$, $\tau = 0.1, 0.5, 0.9$. $n = 512$ *and nominal coverage* 0.95.

the oracle rule does not apply, even for $S = 2$. Interestingly, coverage probabilities of bootstrap and asymptotic confidence intervals differ in this setting. This does not contradict our theory for the bootstrap since that was only developed under the assumption that we are in a regime where the oracle rule holds.

Summarizing all results obtained so far, we can conclude that inference based on results from subsamples is reliable for $S \geq 20$. Since this does not require additional computation, we recommend using the normal approximation for $S > 20$ for pointwise inference and the bootstrap if process level results are required. For $S < 20$, estimating the asymptotic variance within subsamples and aggregating is recommendable. The simplest approach which is based on averaging variance estimators from the `quanteg` package works well and does not require additional implementation, so we recommend to use this for $S < 20$.

APPENDIX: APPROXIMATE LINEAR MODELS WITHOUT LOCAL BASIS STRUCTURE

In this section, we consider models with transformation $\mathbf{Z}$ of increasing dimension that do not have the special local structure considered in Section 3.2. The price for this generality is that we need to assume a more stringent upper bound for S and put additional growth restrictions on m in order to prove the oracle rule. The conditions on K remain the same.

THEOREM A.1. *Assume that conditions (A1)–(A3) hold and that additionally* $m\xi_m^2 \log N = o(N/S)$, $c_m(\boldsymbol{\gamma}_N) = o(\xi_m^{-1})$, $K \ll N^2$. *Then*

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}_K} \|\overline{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)\| \\
&\quad = O_P\biggl(\biggl(\frac{mc_m^2 \log N}{N}\biggr)^{1/2} + c_m^2\xi_m \\
&\qquad + \biggl(\frac{Sm\xi_m^2 \log N}{N} + c_m^4\xi_m^4\biggr)\biggl(1 + \frac{\log N}{S^{1/2}}\biggr)\biggr) \\
&\qquad + \frac{1}{N^{1/2}} O_P\biggl(\bigl(m\xi_m^2 c_m^2 (\log N)^3\bigr)^{1/2} + \biggl(\frac{Sm^3\xi_m^2 (\log N)^7}{N}\biggr)^{1/4}\biggr) \\
&\qquad + o_P\bigl(N^{-1/2}\bigr).
\end{aligned} \tag{A.1}$$

If additionally $K \gg G \gg 1$, $m^3\xi_m^2 (\log N)^3 = o(N)$, $c_m^2(\boldsymbol{\gamma}_N)\xi_m = o(N^{-1/2})$ *we also have for any* $x_0 \in \mathcal{X}$

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}} \bigl|\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}(\tau) - \mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\tau)\bigr| \\
&\quad \leq \|\mathbf{Z}(x_0)\| \sup_{\tau\in\mathcal{T}_K} \|\overline{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)\| + \sup_{\tau\in\mathcal{T}} \bigl|\bigl(\Pi_K Q(x_0;\cdot)\bigr)(\tau) - Q(x_0;\tau)\bigr| \\
&\qquad + \sup_{\tau\in\mathcal{T}} \bigl|\mathbf{Z}(x_0)^\top \boldsymbol{\beta}_N(\tau) - Q(x_0;\tau)\bigr| + o_P\bigl(\|\mathbf{Z}(x_0)\| N^{-1/2}\bigr),
\end{aligned} \tag{A.2}$$

where the projection operator Π_K *was defined right after* (2.5).

The proof of Theorem A.1 is given in Section S.3.4.2. The general upper bound provided in (A.1) takes a rather complicated form. Under additional assumptions on c_m, major simplifications are possible. Due to space considerations, we do not provide additional details here, rather we implicitly carry out the simplifications when proving the following result.

COROLLARY A.2 (Oracle rule for $\overline{\boldsymbol{\beta}}(\tau)$). *Assume that conditions (A1)–(A3) hold and that additionally* $m^4(\log N)^{10} = o(N)$, $c_m^2(\boldsymbol{\gamma}_N)\xi_m = o(N^{-1/2})$. *Provided that additionally* $S = o(N^{1/2}/(m\xi_m^2(\log N)^2))$ *the estimator* $\overline{\boldsymbol{\beta}}(\tau)$ *defined in* (2.3) *satisfies*

$$\frac{\sqrt{N}\mathbf{u}_N^\top(\overline{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau))}{(\mathbf{u}_N^\top J_m^{-1}(\tau)\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]J_m^{-1}(\tau)\mathbf{u}_N)^{1/2}} \rightsquigarrow \mathcal{N}\big(0, \tau(1-\tau)\big), \tag{A.3}$$

for any $\tau \in \mathcal{T}$, $\mathbf{u}_N \in \mathbb{R}^m$, *where* $J_m(\tau)$ *is defined in the statement of Corollary* 3.3. *The same holds for the oracle estimator* $\widehat{\boldsymbol{\beta}}_{or}(\tau)$.

We note that $m^4(\log N)^{10} = o(N)$ imposes restriction on model complexity. In particular, this requires $m = o(N^{1/4}(\log N)^{-5/2})$. An immediate consequence of Corollary A.2 is the oracle rule for correctly specified models, including linear quantile regression with increasing dimension as a special case. In this case, $c_m(\boldsymbol{\gamma}_N) = 0$.

COROLLARY A.3 (Oracle rule for correctly specified models). *Assume that conditions (A1)–(A3) hold and that the quantile function satisfies* $Q(x;\tau) = \mathbf{Z}(x)^\top\boldsymbol{\gamma}_N(\tau)$ *with a transformation* $\mathbf{Z}(x)$ *of possibly increasing dimension m with each entry bounded almost surely. Then* $\mathbf{u}_N^\top\overline{\boldsymbol{\beta}}(\tau)$ *satisfies the oracle rule provided that* $m^4(\log N)^{10} = o(N)$ *and* $S = o(N^{1/2}m^{-2}(\log N)^{-2})$.

This corollary reduces to Corollary 3.3 in Section 3.1 when m is fixed. It describes the effect of allowing m to increase on the sufficient upper bound for S. We note that $c_m(\boldsymbol{\gamma}_N) = 0$ and $\xi_m \asymp m^{1/2}$ under the settings of Corollary A.3, whose proof follows directly from Corollary A.2.

Both Corollary 3.9 and Corollary A.2 can be applied to local polynomial spline models, but Corollary 3.9 puts less assumptions on S and m than Corollary A.2, because Corollary 3.9 exploits the local support property of splines. This is illustrated in the following Remark A.4 for the specific setting of Example 3.6.

REMARK A.4 (Comparing Corollary 3.9 and Corollary A.2 with univariate splines). Let $\mathbf{Z}$ denote the univariate splines from Example 3.6 and let $\mathbf{u}_N := \mathbf{Z}(x_0)$ for a fixed $x_0 \in \mathcal{X}$. We assume that (A2) and (A3) hold and that X has a density on $\mathcal{X} = [0,1]$ that is uniformly bounded away from zero and infinity.

We will verify in Section S.3.4.4 that (A1) holds with $\xi_m \asymp m^{1/2}$. For simplicity, assume that the bias $c_m(\boldsymbol{\gamma}_N)$ satisfies $\xi_m c_m(\boldsymbol{\gamma}_N)^2 = o(N^{-1/2})$. Corollary 3.9 shows that sufficient conditions for the oracle rule are $m^2(\log N)^6 = o(N)$ and $S = o((Nm^{-1}(\log N)^{-4})^{1/2} \wedge (Nm^{-2}(\log N)^{-10}))$. On the other hand, Corollary A.2 requires the more restrictive conditions $m^4(\log N)^{10} = o(N)$ and $S = o(N^{1/2}m^{-2}(\log N)^{-2})$.

Remark A.4 indicates that Corollary 3.9 gives sharper bounds than Corollary A.2 when both results are applicable. Note however that Corollary A.2 applies to more general settings since, in contrast to Corollary 3.9; it does not require Condition (L). For instance, linear models as discussed in Corollary A.3 can in general not be handled by Corollary 3.9. Finally, we discuss sufficient conditions for the process oracle rule.

COROLLARY A.5 [Oracle rule for $\widehat{\boldsymbol{\beta}}(\tau)$]. *Let $x_0 \in \mathcal{X}$, let the conditions of Corollary A.2 hold. Suppose that $\tau \mapsto Q(x_0;\tau) \in \Lambda_c^\eta(\mathcal{T})$, $r_\tau \geq \eta$, $N^2 \gg K \gg G \gg N^{1/(2\eta)}\|\mathbf{Z}(x_0)\|^{-1/\eta}$, and $\sup_{\tau\in\mathcal{T}} |\mathbf{Z}(x_0)^\top \boldsymbol{\gamma}_N(\tau) - Q(x_0;\tau)| = o(\|\mathbf{Z}(x_0)\| \times N^{-1/2})$. Let the limit $H_{x_0}(\tau_1,\tau_2)$ defined in (3.14) exist and be nonzero for any $\tau_1,\tau_2 \in \mathcal{T}$:*

1. *The projection estimator $\widehat{\boldsymbol{\beta}}(\tau)$ defined in (2.5) satisfies*

$$\frac{\sqrt{N}}{\|\mathbf{Z}(x_0)\|}\big(\mathbf{Z}(x_0)^\top\widehat{\boldsymbol{\beta}}(\cdot) - Q(x_0;\cdot)\big) \rightsquigarrow \mathbb{G}_{x_0}(\cdot) \qquad \text{in } \ell^\infty(\mathcal{T}), \tag{A.4}$$

where $\mathbb{G}_{x_0}$ is a centered Gaussian process with $\mathbb{E}[\mathbb{G}_{x_0}(\tau)\mathbb{G}_{x_0}(\tau')] = H_{x_0}(\tau,\tau')$. This matches the process convergence of $\mathbf{Z}(x_0)^\top\widehat{\boldsymbol{\beta}}_{or}(\tau)$.

2. *The estimator $\widehat{F}_{Y|X}(\cdot|x_0)$ defined in (2.8) satisfies*

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

where and $\mathbb{G}_{x_0}$ is the centered Gaussian process from (A.4). The same is true for $\widehat{F}^{or}_{Y|X}$.

The proof of Corollary A.5 is given in Section S.3.4.5. Note that the condition on K is the same as in Corollary 3.10. Results along the lines of Corollary A.5 can be obtained for any estimator of the form $\mathbf{u}_N^\top\widehat{\boldsymbol{\beta}}(\cdot)$, as long as $\mathbf{u}_N$ satisfies certain technical conditions. For example, the partial derivative, the average derivative and the conditional average derivative of $Q(x;\tau)$ in x fall into this framework. For brevity, we refer the interested reader to Section 2.3 of Belloni et al. (2017) for examples of vectors $\mathbf{u}_N$ and omit the technical details.

REMARK A.6. The weak convergence in (A.3) was derived with the centering $\boldsymbol{\gamma}_N$ for notational convenience. As pointed out in Chao, Volgushev and Cheng (2017), other choices of centering sequences can be advantageous (for instance, this is the case in the setting of partial linear models; see also Belloni et al. (2017) for an alternative centering sequence). The conclusions of Theorem A.2 can be generalized by replacing $\boldsymbol{\gamma}_N$ in there by other vectors $\boldsymbol{\beta}_N(\tau)$ as long as certain technical conditions are satisfied. However, this requires additional notation and technicalities. To keep the presentation simple, we provide those details in Section S.3.4.1 in the Supplementary Material [Volgushev, Chao and Cheng (2019)]. Similarly, the results in Section 3.2 could be stated in terms of more general sequences $\boldsymbol{\beta}_N$ instead of the one considered in (3.8) at the cost of additional notation and conditions.

Acknowledgments. The authors would like to thank the Editors and two anonymous referees for helpful comments that helped to considerably improve an earlier version of this manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "Distributed inference for quantile regression processes" (DOI: 10.1214/18-AOS1730SUPP; .pdf). The supplement contains additional technical remarks, simulation results and all proofs.

REFERENCES

BANERJEE, M., DUROT, C. and SEN, B. (2019). Divide and conquer in non-standard problems and the super-efficiency phenomenon. *Ann. Statist.* To appear.

BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and FERNÁNDEZ-VAL, I. (2017). Conditional quantile processes based on series or many regressors. Preprint. Available at arXiv:1105.6154.

CHAO, S.-K., VOLGUSHEV, S. and CHENG, G. (2017). Quantile processes for semi and nonparametric regression. *Electron. J. Stat.* **11** 3272–3331. MR3708539

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and GALICHON, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78** 1093–1125. MR2667913

HAGEMANN, A. (2017). Cluster-robust bootstrap inference in quantile regression models. *J. Amer. Statist. Assoc.* **112** 446–456. MR3646584

JORDAN, M. I. (2013). On statistics, computation and scalability. *Bernoulli* **19** 1378–1390. MR3102908

KATO, K. (2012). Asymptotic normality of Powell's kernel estimator. *Ann. Inst. Statist. Math.* **64** 255–273. MR2878905

KLEINER, A., TALWALKAR, A., SARKAR, P. and JORDAN, M. I. (2014). A scalable bootstrap for massive data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 795–816. MR3248677

KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657

LI, R., LIN, D. K. J. and LI, B. (2013). Statistical inference in massive data sets. *Appl. Stoch. Models Bus. Ind.* **29** 399–409. MR3117826

POWELL, J. L. (1986). Censored regression quantiles. *J. Econometrics* **32** 143–155. MR0853049

SCHUMAKER, L. L. (1981). *Spline Functions*: *Basic Theory*. Wiley, New York. MR0606200

SENGUPTA, S., VOLGUSHEV, S. and SHAO, X. (2016). A subsampled double bootstrap for massive data. *J. Amer. Statist. Assoc.* **111** 1222–1232. MR3561944

SHANG, Z. and CHENG, G. (2017). Computational limits of a distributed algorithm for smoothing spline. *J. Mach. Learn. Res.* **18** Paper No. 108. MR3725447

SHI, C., LU, W. and SONG, R. (2017). A massive data framework for M-estimators with cubic-rate. *J. Amer. Statist. Assoc.* To appear. DOI:10.1080/01621459.2017.1360779.

VOLGUSHEV, S. (2013). Smoothed quantile regression processes for binary response models. Preprint. Available at arXiv:1302.5644.

VOLGUSHEV, S., CHAO, S.-K. and CHENG, G. (2019). Supplement to "Distributed inference for quantile regression processes." DOI:10.1214/18-AOS1730SUPP.

WHITE, T. (2012). *Hadoop*: *The Definitive Guide*. O'Reilly Media/Yahoo Press.

XU, G., SHANG, Z. and CHENG, G. (2016). Optimal tuning for divide-and-conquer kernel ridge regression with massive data. Preprint. Available at arXiv:1612.05907.

ZHANG, Y., DUCHI, J. C. and WAINWRIGHT, M. J. (2013). Communication-efficient algorithms for statistical optimization. *J. Mach. Learn. Res.* **14** 3321–3363. MR3144464

ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *J. Mach. Learn. Res.* **16** 3299–3340. MR3450540

ZHAO, T., CHENG, G. and LIU, H. (2016). A partially linear framework for massive heterogeneous data. *Ann. Statist.* **44** 1400–1437. MR3519928

ZHOU, S., SHEN, X. and WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26** 1760–1782. MR1673277

S. VOLGUSHEV
DEPARTMENT OF STATISTICAL SCIENCES
UNIVERSITY OF TORONTO
100 ST. GEORGE STREET
TORONTO, ONTARIO M5S 3G3
CANADA
E-MAIL: stanislav.volgushev@utoronto.ca

S.-K. CHAO
G. CHENG
DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
250 N. UNIVERSITY STREET
WEST LAFAYETTE, INDIANA 47906
USA
E-MAIL: skchao74@gmail.com
chengg@purdue.edu
URL: http://www.stat.purdue.edu/~chengg/