Submitted to the Annals of Statistics
arXiv: arXiv:1701.06088

DISTRIBUTED INFERENCE FOR QUANTILE REGRESSION PROCESSES

By Stanislav Volgushev[¶,§], Shih-Kang Chao[‖,†] and Guang Cheng[‖,‡,†]

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 sub-samples 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 standalone computers due to processor, memory, or disk bottlenecks. Dividing data and

*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.

†Partially supported by Office of Naval Research (ONR N00014-15-1-2331).

‡Partially supported by NSF CAREER Award DMS-1151692, DMS-1418042.

§Partially supported by discovery grant from Natural Sciences and Engineering Research Council of Canada.

MSC 2010 subject classifications: Primary 62F12, 62G15, 62G20

Keywords and phrases: B-spline estimation, conditional distribution function, distributed computing, divide-and-conquer, quantile regression process

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 Li et al. (2012); Jordan (2013). Pioneering contributions on theoretical analysis of divide and conquer algorithms focus on mean squared error rates Zhang et al. (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 semi-parametric estimation approaches that focus on the mean or other notions of central tendency, this line of work includes Shang and Cheng (2017); Zhao et al. (2016); Shi et al. (2017); Banerjee et al. (2017), 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 semi-parametric 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 processsed 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 sub-samples of size n, the minimization in (1.1) is solved on

each sub-sample, 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 1.1, this can be problematic for inference procedures. More precisely, the left panel of Figure 1.1 depicts the coverage probabilities (on the y-axis) against number of sub-samples (on the x-axis) for asymptotic confidence intervals that are based on asymptotic distributions of $\widehat{\boldsymbol{\beta}}_{or}(\tau)$ but computed using $\overline{\boldsymbol{\beta}}(\tau)$ 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.

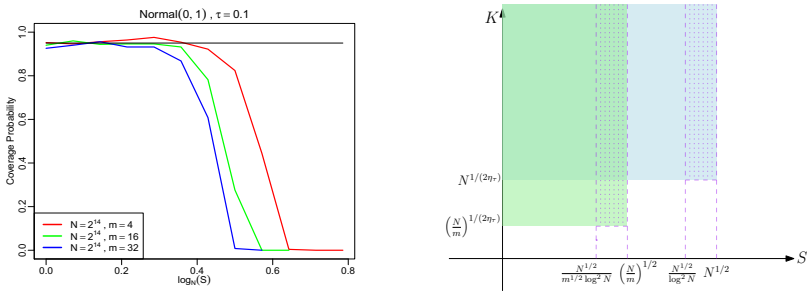


Fig 1.1: Left penal: coverage probabilities (y-axis) of confidence intervals for estimators computed from divide and conquer; x-axis: number of sub-samples S in log-scale. Different colors correspond to linear models with different $m = \dim(\mathbf{Z}(X))$. Right panel: necessary and sufficient conditions 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.

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 $\overline{\boldsymbol{\beta}}(\tau_1), ..., \overline{\boldsymbol{\beta}}(\tau_K)$ for a grid of quantile values $\tau_1, ..., \tau_K$. In the second step, a matrix $\widehat{\Xi}$ is computed from $\overline{\boldsymbol{\beta}}(\tau_1), ..., \overline{\boldsymbol{\beta}}(\tau_K)$ (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 sub-samples 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 $S = o\big(N^{1/2} a_{N,m}\big)$ 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 1.1.

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 sub-samples. 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 Appendix A. 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 online supplement. The following notation will be used throughout the paper.

Notation. Let $\mathcal{X} := \text{supp}(X)$. 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 $|\boldsymbol{j}| = j_1 + ... + j_d$ for d-tuple $\boldsymbol{j} = (j_1, ..., j_d)$. $\mathcal{S}^{m-1} \subset \mathbb{R}^m$ 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

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

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 τ-th 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 sub-samples of size n[1]. Denote the s-th sub-sample as $\{(X_{is}, Y_{is})\}_{i=1}^n$ where $s = 1, ..., S$.
2. For each s and fixed τ, estimate the sub-sample based quantile regression coefficient as follows

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

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 $|\mathcal{T}_K| = K \in \mathbb{N}$.

[1]The equal sub-sample size is assumed for ease of presentation, and can be certainly relaxed.

Second project each component of the vectors $\{\overline{\boldsymbol{\beta}}(\tau_1), ..., \overline{\boldsymbol{\beta}}(\tau_K)\}$ on a space of spline functions in τ. More precisely, let

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

where $\mathbf{B} := (B_1, ..., B_q)^\top$ is a B-spline basis defined on G equidistant knots $\tau_L = t_1 < ... < t_G = \tau_U$ 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 $B_1, ..., B_q$. 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 $k = 1, ..., K$. For each τ_k, compute $\overline{\boldsymbol{\beta}}(\tau_k)$ as (2.3).
2. For each $j = 1, ..., m$, 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

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

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

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

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

(A3) Assume that uniformly in N, there exists a constant $f_{\min} < \overline{f}$ such that
$$0 < f_{\min} \leq \inf_{\tau\in\mathcal{T}} \inf_{x\in\mathcal{X}} f_{Y|X}(Q(x;\tau)|x).$$

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.

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*

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

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

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

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 misspecified linear models with dimension that depends on the sample size. Technical details are provided in the Appendix, Section A.

The bound in (3.2) quantifies the difference between $\overline{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{or}$ in terms of the number of sub-samples S and can be equivalently formulated in terms of the sub-sample 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 $\mathcal{P}_1(\xi_m, M, \bar{f}, \overline{f'}, f_{min})$ 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 $\mathcal{P}_1(\xi, M, \bar{f}, \overline{f'}, f_{min})$.

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

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

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

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

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 et al. (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 et al. (2017) who show that

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

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

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

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* $(P,\mathbf{Z}) \in \mathcal{P}_1(\xi, M, \bar{f}, \overline{f'}, f_{min})$ *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* $j = 1, ..., d$ *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* $(P,\mathbf{Z}) \in \mathcal{P}_1(\xi, M, \bar{f}, \overline{f'}, f_{min})$ *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 et al. (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

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

see Chao et al. (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 et al. (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 $\widetilde{\mathbf{B}}(x) = (\widetilde{B}_1(x), ..., \widetilde{B}_{J-p-1}(x))^\top$ be a polynomial spline basis of degree p defined on J uniformly spaced knots $0 = t_1 < ... < t_J = 1$, 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 $\mathbf{Z}(x) := m^{1/2}(\widetilde{B}_1(x), ..., \widetilde{B}_{J-p-1}(x))^\top$, 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 et al. (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, i.e., $\sup_{x\in\mathcal{X}} \widetilde{\mathcal{E}}(\mathbf{Z}(x), \boldsymbol{\gamma}_N) = O(1)$, 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*

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

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

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

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).

Similarly 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*

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

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 an* $x_0 \in \mathcal{X}$ *with* $\tau \mapsto Q(x_0;\tau) \in \Lambda_c^\eta(\mathcal{T})$ *such that*

(3.12)

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

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 $\mathcal{P}_L(M, \bar{f}, \overline{f'}, f_{min}, R)$ 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 $\mathcal{P}_L(M, \bar{f}, \overline{f'}, f_{min}, R)$ ensure the weak convergence of the oracle estimator $\widehat{\boldsymbol{\beta}}_{or}(\tau)$, see Chao et al. (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

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

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

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

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 $(P_N, \mathbf{Z}) \in \mathcal{P}_L(M, \bar{f}, \overline{f'}, f_{min}, R)$. 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 behaviour of $\widehat{\boldsymbol{\beta}}_{or}$. If $S \gtrsim N^{1/2} \xi_m^{-1}$ the weak convergence result (3.13) fails for some $(P_N, \mathbf{Z}) \in \mathcal{P}_L(1, \bar{f}, \overline{f'}, f_{min}, R), x_0 \in \mathcal{X}$.

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 $(P_N, \mathbf{Z}) \in \mathcal{P}_L(M, \bar{f}, \overline{f'}, f_{min}, R)$, 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*

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

exists and is non-zero 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

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

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

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

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, Section A.

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 et al. (2017); Xu et al. (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 et al. (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 sub-samples 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 $\widehat{\boldsymbol{\beta}}_1(\tau), ..., \widehat{\boldsymbol{\beta}}_S(\tau)$ which can be seen as iid 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 $\widehat{\boldsymbol{\beta}}_1(\tau), ..., \widehat{\boldsymbol{\beta}}_S(\tau)$ 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

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

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 iid data. Hence for small S (say $S \leq 30$) 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

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

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* $S \geq 2$ *(*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 $S \leq 30$) 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 non-linear 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 $\{\widehat{\boldsymbol{\beta}}_s(\tau_k)\}_{s=1,..,S,k=1,...,K}$ which are stored on the central machine. Details are provided below.

1. Sample iid random weights $\{\omega_{s,b}\}_{s=1,...,S,b=1,...,B}$ 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 $b = 1, ..., B, k = 1, ..., K$ 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 $\{\Phi(\widehat{\boldsymbol{\beta}}^{(b)}(\cdot))-\Phi(\widehat{\boldsymbol{\beta}}(\cdot))\}_{b=1,...,B}$.

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 $(X_i, Y_i)_{i=1,...,N}$*

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

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 sub-sample 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 sub-sample, 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, i.e. 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

$$\Sigma(\tau) = \tau(1-\tau) J_m(\tau)^{-1} \mathbb{E}[\mathbf{Z}\mathbf{Z}^\top] J_m(\tau)^{-1}, \quad J_m(\tau) = \mathbb{E}[\mathbf{Z}\mathbf{Z}^\top f_{Y|X}(Q(X;\tau)|X)].$$

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))

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

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 $J^P_m(\tau)$ 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 appendix. If only a single round of communication is allowed, the following algorithm can be used instead.

1. For $s = 1, ..., S$, in the same round as computing $\widehat{\boldsymbol{\beta}}^s(\tau)$, compute $\widehat{J}^P_{ms}(\tau)$ 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 $(\widehat{J}^P_{ms}(\tau), \widehat{\Sigma}_{1s})_{s=1,\dots,S}$ to the master machine and compute $\overline{J}^P_m(\tau) := \frac{1}{S}\sum_{s=1}^S \widehat{J}^P_{ms}(\tau)$, $\overline{\Sigma}_1 := \frac{1}{S}\sum_{s=1}^S \widehat{\Sigma}_{1s}$.
3. The final variance estimator is given by $\bar{\Sigma}(\tau) = \tau(1-\tau)\overline{J}^P_m(\tau)^{-1}\overline{\Sigma}_1\overline{J}^P_m(\tau)^{-1}$.

REMARK 4.3. Note that in the above algorithm we first take averages over the subsampled estimators $\widehat{J}^P_{ms}(\tau)$ and only invert the aggregated matrix $\overline{J}^P_m(\tau)$. An alternative approach would have been to compute the estimator $\widehat{J}^P_{ms}(\tau)^{-1}\widehat{\Sigma}_s\widehat{J}^P_{ms}(\tau)^{-1}$ for each subsample and average in the end. However, given that $A \mapsto A^{-1}$ is non-linear, 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

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

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

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*

$$\overline{J}^P_m(\tau) - J_m(\tau) = A_n(\tau) + O_p\Big(\frac{\log n}{nh_n}\Big)$$

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

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

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

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

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 et al. (2015), Banerjee et al. (2017) (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

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

where $\varepsilon_i \sim \mathcal{N}(0, 0.01)$ iid 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 $\mathrm{Cov}(X_{ij}, X_{ik}) = 0.1^2 0.7^{|j-k|}$ for $j, k = 1, ..., m-1$, 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 online supplement.

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 sub-samples S on the

coverage probabilities of different 95% confidence intervals. To benchmark our results, we use the infeasible asymptotic confidence interval

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

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 sub-samples. 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 $(\bar{\boldsymbol{\beta}}^{(1)}(\tau) - \bar{\boldsymbol{\beta}}(\tau)), ..., (\bar{\boldsymbol{\beta}}^{(B)}(\tau) - \bar{\boldsymbol{\beta}}(\tau))$, where $\bar{\boldsymbol{\beta}}^{(1)}(\tau), ..., \bar{\boldsymbol{\beta}}^{(B)}(\tau)$ are computed as in (4.3); in this section we set $B = 500$.

The coverage probabilities of corresponding confidence intervals are summarized in Figure 5.1 where we fix two sub-sample 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 $S = 2^k, k = 1, ..., 10$. Only $m = 4, 32$ is considered here, $m = 16$ can be found in the online supplement.

Plots for small S help to access how different procedures behave if the number of sub-samples 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 under-cover 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 sub-sample size of $n = 512$ with any number of sub-samples $S > 10$; 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 sub-sample and

takes the average over all sub-samples (additional details are provided in the online Supplement, 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 sub-samples of size n (additional details on the choice of optimal constant $c^*(\tau)$ are provided in the online Supplement, 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 5.1. 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.

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. 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

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

(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 online supplement. Due to space limitations, we only show the results for small values of S, results for

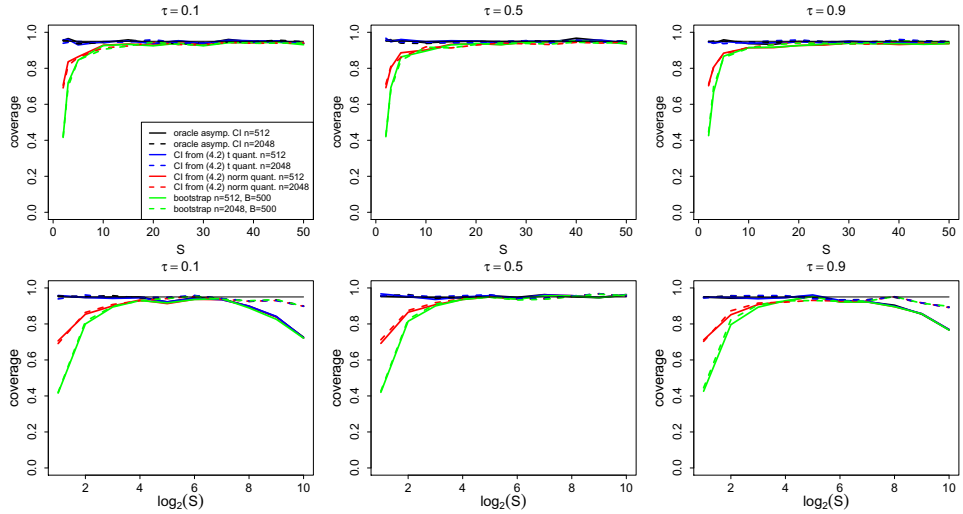
S	1	10	30	50	1	10	30	50	1	10	30	50
	n = 512, m = 4, $\tau = 0.1$				n = 512, m = 16, $\tau = 0.1$				n = 512, m = 32, $\tau = 0.1$			
or	94.9	94.8	94.9	94.3	94.7	94.3	92.2	90.7	94.4	92.5	88.6	85.5
def	92.6	93.7	93.6	92.8	92	92.3	90.1	88.4	92.5	90.6	86.2	82.9
nai	94.2	93.9	93.6	93.2	96.6	93.3	91	89.4	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

TABLE 5.1

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.

large values of S do not give crucial additional insights and are deferred to the online supplement. Coverage probabilities for $m = 4, 32$ are reported in Figure 5.2. 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 sub-sample size n is too small and 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.

(a) $m = 4$



(b) $m = 32$

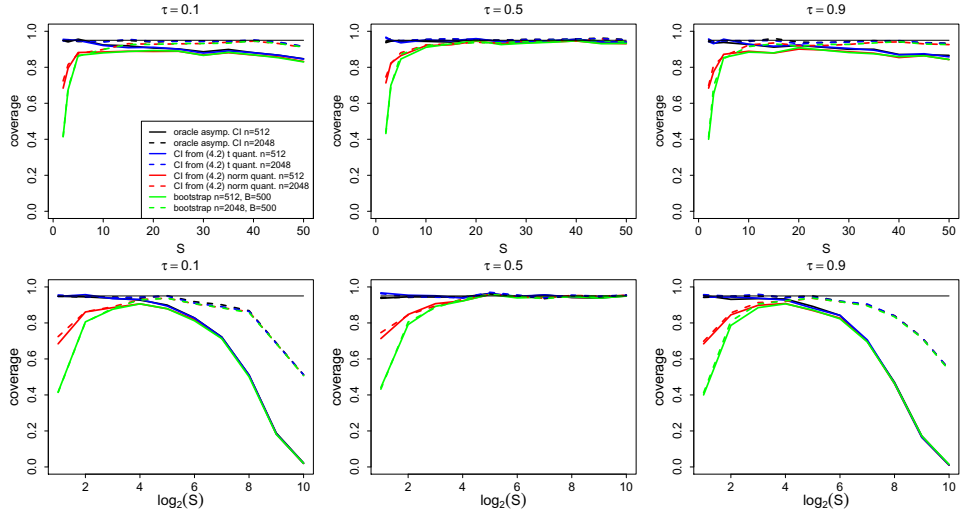


Fig 5.1: Coverage probabilities for $x_0^\top \boldsymbol{\beta}(\tau)$ for different values of S and $\tau = 0.1, 0.5, 0.9$ (left, middle, right row). Solid lines: $n = 512$, dashed lines: $n = 2048$. Black: asymptotic oracle CI, blue: CI from (4.2) based on t distribution, red: CI from (4.1) based on normal distribution, green: bootstrap CI. Throughout $x_0 = (1, ..., 1)/m^{1/2}$, nominal coverage 0.95.

Summarizing all results obtained so far, we can conclude that inference based on results from sub-samples is reliable for $S \geq 20$. Since this does not require additional computation, we recommend using the normal approximation for $S > 20$ for point-wise inference and the bootstrap if process level results are required. For $S < 20$, estimating the asymptotic variance within sub-samples 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 $S < 20$.

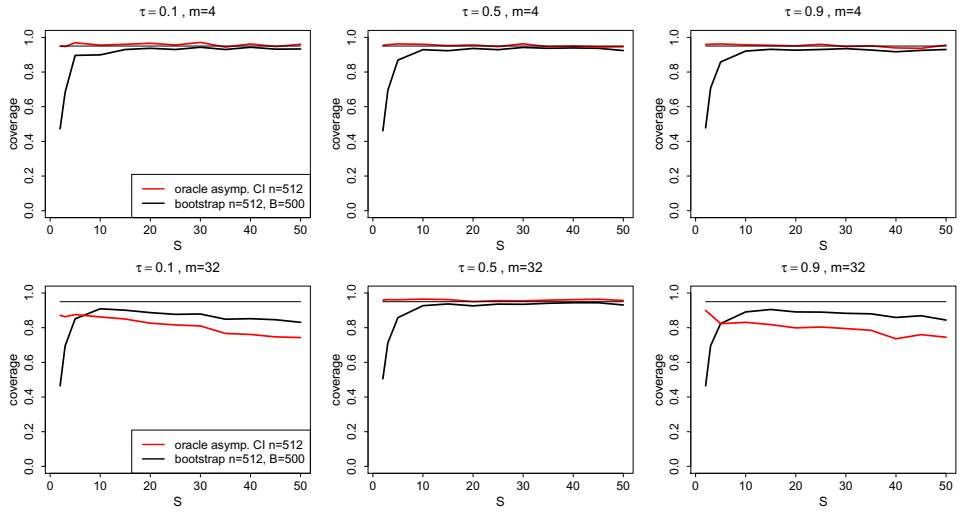


Fig 5.2: Coverage probabilities for oracle confidence intervals (red) and bootstrap confidence intervals (black) for $F_{Y|X}(y|x_0)$ for $x_0 = (1, ..., 1)/m^{1/2}$ and $y = Q(x_0; \tau)$, $\tau = 0.1, 0.5, 0.9$. $n = 512$ and nominal coverage 0.95.

References.

Banerjee, M., Durot, C., and Sen, B. (2017). Divide and conquer in non-standard problems and the super-efficiency phenomenon. *The Annals of Statistics*, page to appear.

Belloni, A., Chernozhukov, V., Chetverikov, D., and Fernández-Val, I. (2017). Conditional quantile processes based on series or many regressors. arXiv preprint arXiv:1105.6154.

Billingsley, P. (1995). *Probability and Measure*. Wiley.

Bücher, A. and Kojadinovic, I. (2017). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *arXiv preprint arXiv:1706.01031*.

Chao, S., Volgushev, S., and Cheng, G. (2017). Quantile process for semi and nonparametric regression. *Electronic Journal of Statistics*, 11(2):3272–3331.

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

Hagemann, A. (2017). Cluster-robust bootstrap inference in quantile regression models. *Journal of the American Statistical Association*, 112(517):446–456.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics*, 31(5):1600–1635.

Jordan, M. I. (2013). On statistics, computation and scalability. *Bernoulli*, 19(4):1378–1390.

Kato, K. (2012). Asymptotic normality of Powell's kernel estimator. *Annals of the Institute of Statistical Mathematics*, 64(2):255–273.

Kleiner, A., Talwalkar, A., Sarkar, P., and Jordan, M. I. (2014). A scalable bootstrap for massive data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4):795–816.

Kley, T., Volgushev, S., Dette, H., and Hallin, M. (2016). Quantile spectral processes: Asymptotic analysis and inference. *Bernoulli*, 22(3):1770–1807.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.

Koenker, R. (2016). *quantreg: Quantile Regression*. R package version 5.29.

Koltchinskii, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Annals of Statistics*, 34(6):2593–2656.

Ledoux, M. and Talagrand, M. (1991). Probability in banach spaces, volume 23 of ergebnisse der mathematik und ihrer grenzgebiete (3)[results in mathematics and related areas (3)].

Li, R., Lin, D. K. J., and Li, B. (2012). Statistical inference in massive data sets. *Applied Stochastic Models in Business and Industry*, 29:399–409.

Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Annals of Probability*, pages 863–884.

McDiarmid, C. (1989). On the method of bounded differences. In Siemons, J., editor, *Surveys in Combinatorics*, London Mathematical Society Lecture Note Series 141, pages 148–188. Cambridge University Press.

Powell, J. L. (1986). Censored regression quantiles. *Journal of econometrics*, 32(1):143–155.

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

Sengupta, S., Volgushev, S., and Shao, X. (2016). A subsampled double bootstrap for massive data. *Journal of the American Statistical Association*, 111(515):1222–1232.

Shang, Z. and Cheng, G. (2017). Computational limits of a distributed algorithm for smoothing spline. *Journal of Machine Learning Research*, 18(108):1–37.

Shi, C., Lu, W., and Song, R. (2017). A massive data framework for M-estimators with cubic-rate. *Journal of the American Statistical Association*, page to appear.

Volgushev, S. (2013). Smoothed quantile regression processes for binary response models. *arXiv preprint arXiv:1302.5644*.

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. *arXiv preprint arXiv:1612.05907*.

Zhang, Y., Duchi, J., and Wainwright, M. J. (2013). Communication-efficient algorithms for statistical optimization. *Journal of Machine Learning Research*, 14:3321–3363.

Zhang, Y., Duchi, J., and Wainwright, M. J. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *Journal of Machine Learning Research*, 16:3299–3340.

Zhao, T., Cheng, G., and Liu, H. (2016). A partially linear framework for massive heterogeneous data. *Annals of Statistics*, 44(4):1400–1437.

Zhou, S., Shen, X., and Wolfe, D. (1998). Local asymptotics for regression splines and confidence regions. *Annals of Statistics*, 26(5):1760–1782.

APPENDIX A: 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*

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

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* $x_0 \in \mathcal{X}$

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

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*

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

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 $\sup_{\tau\in\mathcal{T}} |\mathbf{Z}(x_0)^\top \boldsymbol{\gamma}_N(\tau) - Q(x_0;\tau)| = o(\|\mathbf{Z}(x_0)\|N^{-1/2})$. Let the limit $H_{x_0}(\tau_1,\tau_2)$ defined in (3.14) exist and be non-zero for any $\tau_1,\tau_2 \in \mathcal{T}$*

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

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

   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,*

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

   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 et al. (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 appendix. 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.

SUPPLEMENTARY MATERIAL

ONLINE SUPPLEMENTARY MATERIAL: Distributed inference for quantile regression processes

(doi: COMPLETED BY THE TYPESETTER; .pdf). The supplementary materials contain all the proofs, additional technical details, and additional simulation results.

Additional Notation. This notation will be used throughout the supplement. Let $\mathcal{S}^{m-1} := \{\mathbf{u} \in \mathbb{R}^m : \|\mathbf{u}\| = 1\}$ denote the unit sphere in $\mathbb{R}^m$. For a set $\mathcal{I} \subset \{1, ..., m\}$, define

$$\mathbb{R}^m_{\mathcal{I}} := \{\mathbf{u} = (u_1, ..., u_m)^\top \in \mathbb{R}^m : u_j \neq 0 \text{ if and only if } j \in \mathcal{I}\} \tag{S.0.5}$$

$$\mathcal{S}^{m-1}_{\mathcal{I}} := \{\mathbf{u} = (u_1, ..., u_m)^\top \in \mathcal{S}^{m-1} : u_j \neq 0 \text{ if and only if } j \in \mathcal{I}\} \tag{S.0.6}$$

For simplicity, we sometimes write $\sup_\tau(\inf_\tau)$ and $\sup_x(\inf_x)$ instead of $\sup_{\tau\in\mathcal{T}}(\inf_{\tau\in\mathcal{T}})$ and $\sup_{x\in\mathcal{X}}(\inf_{x\in\mathcal{X}})$ throughout the Appendix. I_k is the k-dimensional identity matrix for $k \in \mathbb{N}$.

APPENDIX S.1: AN ALTERNATIVE ESTIMATOR FOR $J^P_M(\tau)$

Here, we provide a simple way to estimate the matrix $J_m(\tau)$ in a divide and conquer setup provided that an additional round of communication is acceptable.

1. Compute $\overline{\boldsymbol{\beta}}(\tau)$ on the master computer and send to each distributed unit.
2. On each machine, compute
$$\widetilde{J}^P_{ms}(\tau) := \frac{1}{2nh_n}\sum_{i=1}^{n} \mathbf{Z}_{is}\mathbf{Z}_{is}^\top \mathbf{1}\Big\{|Y_{is} - \mathbf{Z}_{is}^\top\overline{\boldsymbol{\beta}}(\tau)| \leq h_n\Big\}$$
and send results back to master machine.
3. On the master, compute $\widetilde{J}^P_m(\tau) := \frac{1}{S}\sum_{s=1}^{S}\widetilde{J}^P_{ms}(\tau)$. The final variance estimator is given by $\tau(1-\tau)\widetilde{J}^P_m(\tau)^{-1}\overline{\Sigma}_1\widetilde{J}^P_m(\tau)^{-1}$.

Provided that $\overline{\boldsymbol{\beta}}(\tau)$ satisfies the oracle rule, the theory developed in Kato (2012) for Powell's estimator applies (note that the results developed in Kato (2012) only require $\sqrt{N}$ consistency of the estimator $\overline{\boldsymbol{\beta}}$, which is guaranteed by the oracle rule). In particular, following Kato (2012), the optimal rate for the bandwidth h_n is given by $h_N \sim N^{-1/5}$, and an explicit formula for the optimal bandwidth can be found on page 263 of Kato (2012). In practice, the 'rule of thumb' provided on page 264 in Kato (2012) can be used to determine the bandwidth h_n.

The algorithm described above provides a simple way to estimate the asymptotic covariance matrix of $\overline{\boldsymbol{\beta}}(\tau)$ at oracle rate $N^{-2/5}$ under the same assumptions that guarantee $\overline{\boldsymbol{\beta}}$ to satisfy the oracle rule. The additional communication and computational costs are also fairly low since only the vectors $\overline{\boldsymbol{\beta}}$ and the matrices $\widetilde{J}^P_{ms}(\tau)$ need to be communicated and computation of $\widetilde{J}^P_{ms}(\tau)$ on each sub-sample does not involve any optimization.

APPENDIX S.2: AGGREGATED BAHADUR REPRESENTATIONS

S.2.1. Aggregated Bahadur representation for general series approximation model. Note that Assumptions (A1)-(A3) imply that for any sequence of $\mathbb{R}^m$-valued functions $\boldsymbol{\beta}_N(\tau)$ satisfying $\sup_{\tau\in\mathcal{T}}\sup_x|\boldsymbol{\beta}_N(\tau)^\top\mathbf{Z}(x)-Q(x;\tau)|=o(1)$, the smallest eigenvalues of the matrices

(S.2.1)
$$\widetilde{J}_m(\boldsymbol{\beta}_N(\tau)):=\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top f_{Y|X}(\boldsymbol{\beta}_N(\tau)^\top\mathbf{Z}|X)],\quad J_m(\tau):=\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top f_{Y|X}(Q(X;\tau)|X)]$$

are bounded away from zero uniformly in τ for sufficiently large n. Define for any vector of functions $\mathbf{b}_N:\mathcal{T}\to\mathbb{R}^m$,

$$g_N(\mathbf{b}_N):=\sup_{\tau\in\mathcal{T}}\big\|\mathbb{E}[\mathbf{Z}_i(\mathbf{1}\{Y_i\le\mathbf{Z}_i^\top\mathbf{b}_N(\tau)\}-\tau)]\big\|. \tag{S.2.2}$$

THEOREM S.2.1 (ABR for General Series Model). *Assume that the underlying distribution of (X_i,Y_i) satisfies conditions (A1)-(A3) and that $m\xi_m^2\log n=o(n)$. Consider any sequence $\boldsymbol{\beta}_N(\tau)$ such that $g_N:=g_N(\boldsymbol{\beta}_N)=o(\xi_m^{-1}), c_m:=c_m(\boldsymbol{\beta}_N)=o(1)$. Then we have*

$$\overline{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_N(\tau)=-\frac{1}{N}J_m(\tau)^{-1}\sum_{i=1}^N\mathbf{Z}_i(\mathbf{1}\{Y_i\le\mathbf{Z}_i^\top\boldsymbol{\beta}_N(\tau)\}-\tau)+r_N^{(1)}(\tau)+r_N^{(2)}(\tau)$$

where, for any $\kappa_n\ll n/\xi_m^2$, any sufficiently large n, and a constant C independent of n

$$P\Big(\sup_{\tau\in\mathcal{T}}\|r_N^{(1)}(\tau)\|\ge CR_{1,n}(\kappa_n)\Big)\le Se^{-\kappa_n} \tag{S.2.3}$$

with

(S.2.4)
$$R_{1,n}(\kappa_n):=\xi_m\Big(\Big(\frac{m}{n}\log n\Big)^{1/2}+\Big(\frac{\kappa_n}{n}\Big)^{1/2}\Big)^2+c_m\Big(\Big(\frac{m}{n}\log n\Big)^{1/2}+\Big(\frac{\kappa_n}{n}\Big)^{1/2}\Big)+g_N.$$

Moreover there exists a constant C independent of n, S such that for n sufficiently large and $A \geq 1$

$$\sup_{\tau\in\mathcal{T}} P\Big(\big\|r_N^{(2)}(\tau)\big\| > CR_{2,n}(A,\kappa_n)\Big) \leq 2Sn^{-A} + 2e^{-\kappa_n^2} \tag{S.2.5}$$

where

$$R_{2,n}(A,\kappa_n) := \frac{A\kappa_n}{S^{1/2}}\Big(\frac{m}{n}\Big)^{1/2}\Big((\xi_m g_N \log n)^{1/2} + \Big(\frac{m\xi_m^2(\log n)^3}{n}\Big)^{1/4}\Big) + \Big(1+\frac{\kappa_n}{S^{1/2}}\Big)\Big(g_N^2\xi_m^2 + \frac{m\xi_m^2\log n}{n}\Big). \tag{S.2.6}$$

See Section S.5.1 of the online Appendix for a proof for Theorem S.2.1. We give a simpler expression for the remainder terms in the following remark.

REMARK S.2.2. *For subsequent applications of the Bahadur representation Theorem S.2.1, we will often set $\kappa_n = A\log n$ for a sufficiently large constant $A > 0$, and assume the remainder terms $r_N^{(1)}(\tau) + r_N^{(2)}(\tau)$, defined in Theorem S.2.1, are bounded with certain order.*

Under the conditions $\kappa_n = A\log n$, $(\log n)^2 S^{-1} = o(1)$ and $m \geq 1$, $R_{1,n}(\kappa_n) + R_{2,n}(\kappa_n)$ can be bounded by (neglecting a positive constant)

$$c_m\Big(\frac{m}{n}\log n\Big)^{1/2} + g_N + \frac{\log n}{S^{1/2}}\Big(\frac{m}{n}\Big)^{1/2}\Big((\xi_m g_N\log n)^{1/2} + \Big(\frac{m\xi_m^2(\log n)^3}{n}\Big)^{1/4}\Big) + g_N^2\xi_m^2 + \frac{m\xi_m^2\log n}{n}. \tag{S.2.7}$$

We note that without the condition $(\log n)^2 S^{-1} = o(1)$, we have an additional $\log n$ factor to the last two terms in (S.2.7).

REMARK S.2.3. *A result along the lines of Theorem S.2.1 can also be obtained by a direct application of the classical Bahadur representation (e.g. Theorem 5.1 in Chao et al. (2017)), but that leads to bounds on remainder terms that are less sharp compared to the ones provided in (S.2.3)-(S.2.6). To illustrate this point, assume that $Q(x;\tau) = x^\top\boldsymbol{\beta}(\tau)$ with x denoting a covariate of fixed dimension. In this case $g_N(\boldsymbol{\beta}_N) = c_m(\boldsymbol{\beta}_N) = 0$ and ξ_m, m are constant. The classical Bahadur representation in Theorem 5.1 of Chao et al. (2017) yields an expansion of the form*

$$\widehat{\boldsymbol{\beta}}^s(\tau) - \boldsymbol{\beta}_N(\tau) = -\frac{1}{n}J_m(\tau)^{-1}\sum_{i=1}^n X_{is}(\mathbf{1}\{Y_{is} \leq Q(X_{is};\tau)\} - \tau) + r_n^s$$

where $J_m(\tau) := \mathbb{E}[XX^\top f_{Y|X}(Q(X;\tau)|X)]$, $r_n^s = O_P(n^{-3/4}\log n)$ *and the exponent of* n *in the bound on* r_n^s *can in general not be improved. Applying this expansion to the estimator in each group we obtain*

(S.2.8)

$$\overline{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_N(\tau) = -\frac{1}{N} J_m(\tau)^{-1} \sum_{i=1}^{N} X_i(\mathbf{1}\{Y_i \le Q(X_i;\tau)\} - \tau) + \frac{1}{S}\sum_{s=1}^{S} r_n^s.$$

Without additional information on the moments of r_n^s*, the bound* $r_n^s = O_P(n^{-3/4}\log n)$ *would at best provide a bound of the form* $\frac{1}{S}\sum_{s=1}^{S} r_n^s = O_P(n^{-3/4}\log n)$*. Thus by the representation* (S.2.8) *the oracle rule holds provided that* $n^{-3/4}\log n = o(N^{-1/2})$*, which is equivalent to* $S = o(N^{1/3}(\log N)^{-4/3})$ *for an arbitrarily small* $\varepsilon > 0$*. In contrast, for* $c_m(\boldsymbol{\beta}_N) = g_N(\boldsymbol{\beta}_N) = 0$ *the bound in* (S.2.7) *reduces to* $n^{-1}\log n + n^{-3/4}S^{-1/2}(\log n)^{7/4}$*. Since* $N = Sn$ *this is of order* $o(N^{-1/2})$ *if* $S = o(N^{1/2}(\log N)^{-2})$*. This is a considerably less restrictive condition than* $S = o(N^{1/3}(\log N)^{-4/3})$ *which was derived by the 'direct' approach above. Similar arguments apply in a more general setting of a linear model with increasing dimension or series expansions.*

S.2.2. Aggregated Bahadur representation for local basis functions.

THEOREM S.2.4 (ABR for local basis functions). *Assume that the underlying distribution satisfies conditions (A1)-(A3) and let Condition (L) hold. Assume that* $m\xi_m^2 \log n = o(n)$*,* $c_m(\boldsymbol{\gamma}_N) = o(1)$ *and that* $c_m(\boldsymbol{\gamma}_N)^2 = o(\xi_m^{-1})$*. Then we have for* $\mathbf{u}_N \in \mathcal{S}_{\mathcal{I}}^{m-1}$ *(defined in* (S.0.6)*) with* $\mathcal{I}$ *consisting of* $L = O(1)$ *consecutive integers (*$\mathcal{I}$ *is allowed to depend on* n*)*

(S.2.9)

$$\mathbf{u}_N^\top \overline{\boldsymbol{\beta}}(\tau) - \mathbf{u}_N^\top \boldsymbol{\gamma}_N(\tau) = -\frac{1}{N}\mathbf{u}_N^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1} \sum_{i=1}^{N} \mathbf{Z}_i(\mathbf{1}\{Y_i \le \mathbf{Z}_i^\top \boldsymbol{\gamma}_N(\tau)\} - \tau) + r_N^{(1)}(\tau, \mathbf{u}_N) + r_N^{(2)}(\tau, \mathbf{u}_N)$$

where, for any $\kappa_n \ll n/\xi_m^2$*, any sufficiently large* n*, and a constant* C *independent of* n

$$P\Big(\sup_{\tau\in\mathcal{T}} \sup_{\mathbf{u}_N \in \mathcal{S}_{\mathcal{I}}^{m-1}} \|r_N^{(1)}(\tau, \mathbf{u}_N)\| \ge C R_{1,n}^{(L)}(\kappa_n)\Big) \le S e^{-\kappa_n}. \tag{S.2.10}$$

where

(S.2.11)

$$R_{1,n}^{(L)}(\kappa_n) := \frac{\xi_m \log n}{n} + \sup_{\mathbf{u}_N \in \mathcal{S}_{\mathcal{I}}^{m-1}} \widetilde{\mathcal{E}}(\mathbf{u}_N, \boldsymbol{\gamma}_N)\Big(c_m(\boldsymbol{\gamma}_N)^2 + \frac{\xi_m^2(\log n)^2 + \xi_m^2\kappa_n}{n}\Big).$$

Moreover there exists a constant C independent of n, S such that for n sufficiently large and $A \geq 1$

$$\sup_{\tau} P\Big(\sup_{\mathbf{u}_N \in \mathcal{S}_{\mathcal{I}}^{m-1}} \Big| r_N^{(2)}(\tau, \mathbf{u}_N) \Big| > C R_{n,2}^{(L)}(A, \kappa_n) \Big) \leq 2Sn^{-A} + 2e^{-\kappa_n^2} \tag{S.2.12}$$

where

$$R_{n,2}^{(L)}(A, \kappa_n) := A \sup_{\mathbf{u}_N \in \mathcal{S}_{\mathcal{I}}^{m-1}} \widetilde{\mathcal{E}}(\mathbf{u}_N, \boldsymbol{\gamma}_N) \Big(1 + \frac{\kappa_n}{S^{1/2}}\Big) \Big(c_m(\boldsymbol{\gamma}_N)^4 + \frac{\xi_m^2 (\log n)^2}{n} \Big)$$

$$+ A \frac{\kappa_n}{n^{1/2} S^{1/2}} \Big(c_m(\boldsymbol{\gamma}_N)(\kappa_n^{1/2} \vee \log n) + \frac{\xi_m^{1/2} (\kappa_n^{1/2} \vee \log n)^{3/2}}{n^{1/4}} \Big). \tag{S.2.13}$$

See Section S.5.2 of the online Appendix for a proof of Theorem S.2.4.

The following corollary gives an ABR representation which centered at the true function.

COROLLARY S.2.5. *If the conditions in Theorem S.2.4 hold, and additionally for a fixed $x_0 \in \mathcal{X}$, $c_m(\boldsymbol{\gamma}_N) = o(\|\mathbf{Z}(x_0)\| N^{-1/2})$ then for any $\tau \in \mathcal{T}$, (S.2.9) in Theorem S.2.4 holds with $\mathbf{u}_N$ and $\mathbf{u}_N^\top \boldsymbol{\gamma}_N(\tau)$ being replaced by $\mathbf{Z}(x_0)$ and $Q(x_0; \tau)$.*

LEMMA S.2.6. *In setting of Example 3.6 we have*

$$\sup_{\tau \in \mathcal{T}, x \in [0,1]} \mathbb{E} |\mathbf{Z}(x) \widetilde{J}_m^{-1}(\boldsymbol{\gamma}_N(\tau)) \mathbf{Z}| = O(1)$$

$$\sup_{\tau \in \mathcal{T}, x \in [0,1]} \mathbb{E} |\mathbf{Z}(x) J_m^{-1}(\tau) \mathbf{Z}| = O(1)$$

The proof of Lemma S.2.6 follows by similar arguments as the proof of Example 2.3 on page 3279 in Chao et al. (2017) and is omitted for brevity.

APPENDIX S.3: PROOFS FOR SECTION 3 AND SECTION A

S.3.1. Proofs for Section 3.1. Here we provide proofs for Theorem 3.1 and the sufficient conditions for the oracle rule in Corollary 3.3 and Corollary 3.4. Theorem 3.2 and the necessary conditions of Corollary 3.3 and Corollary 3.4 will be established in Section S.3.3.

Theorem 3.1 follows from Theorem A.1 by noting that for any $P \in \mathcal{P}_1$, the quantities m, ξ_m are constants that do not depend on N. Moreover, since $J_m(\tau)$ is invertible by assumptions (A1)-(A3), $\boldsymbol{\gamma}_N \equiv \boldsymbol{\beta}$, and additionally we have $c_m(\boldsymbol{\gamma}_N) = g_N(\boldsymbol{\gamma}_N) = 0$.

The sufficiency of $S = o(N^{1/2}(\log N)^{-1})$ for weak convergence of $\bar{\boldsymbol{\beta}}$ in Corollary 3.3 follows from the corresponding weak convergence of $\widehat{\boldsymbol{\beta}}_{or}$ and (3.2) which implies that under $S = o(N^{1/2}(\log N)^{-1})$ we have $\bar{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{or} = o_P(N^{-1/2})$.

It remains to prove the 'sufficient' part of Corollary 3.4. As a direct consequence of the second part of Theorem 3.1 we obtain

$$\sup_{\tau\in\mathcal{T}} |\widehat{\beta}_j(\tau) - \widehat{\beta}_{or,j}(\tau))| \leq \|\Pi_K \beta_j(\cdot) - \beta_j(\cdot)\|_\infty + o_P(N^{-1/2}).$$

Next, note that by assumption, $\beta_j \in \Lambda_c^\eta(\mathcal{T})$. Applying Theorem 2.62 and Theorem 6.25 of Schumaker (1981), we find that

$$\inf_{g\in\Theta_G} \|\beta_j(\cdot) - g\|_\infty \lesssim G^{-\eta}, \tag{S.3.1}$$

where Θ_G denotes the space of splines of degree r_τ over the grid $t_1, ..., t_G$ on $\mathcal{T} = [\tau_L, \tau_U] \subset (0, 1)$. This together with Lemma 5.1 of Huang (2003) shows that

$$\|\Pi_K \beta_j(\cdot) - \beta_j(\cdot)\|_\infty = O(G^{-\eta}). \tag{S.3.2}$$

Thus the 'sufficient' part of the statement of Corollary 3.4 follows. □

S.3.2. Proofs for Section 3.2. This section contains the proofs of Theorem 3.7 and the sufficient conditions for weak convergence in Corollary 3.9 and Corollary 3.10. The proof of Theorem 3.8 and the necessary conditions in Corollary 3.9 and Corollary 3.10 are given in section S.3.3.

S.3.2.1. *Proof of Theorem 3.7.* We begin with a proof of (3.10). Define

$$\begin{aligned} R_n(S) := \Big(1 + \frac{\log N}{S^{1/2}}\Big)\Big(c_m^2(\boldsymbol{\gamma}_N) + \frac{S\xi_m^2(\log N)^2}{N}\Big) \\ + \frac{\|\mathbf{Z}(x_0)\|\xi_m S \log N}{N} + \frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big(\frac{S\xi_m^2(\log N)^{10}}{N}\Big)^{1/4}. \end{aligned} \tag{S.3.3}$$

From Theorem S.6.2 in the supplementary material we obtain the following representation for the oracle estimator

$$\sup_{\tau\in\mathcal{T}} \Big| \frac{\mathbf{Z}(x_0)^\top}{\|\mathbf{Z}(x_0)\|}(\widehat{\boldsymbol{\beta}}_{or}(\tau) - \boldsymbol{\gamma}_N(\tau)) - \widetilde{\boldsymbol{U}}_{1,N}(\tau)\Big| = O_P(R_N(1)) \tag{S.3.4}$$

where

(S.3.5)

$$\widetilde{\boldsymbol{U}}_{1,N}(\tau) := -N^{-1}\frac{\mathbf{Z}(x_0)^\top}{\|\mathbf{Z}(x_0)\|}\widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1}\sum_{i=1}^N \mathbf{Z}_i(\mathbf{1}\{Y_i \leq \mathbf{Z}_i^\top \boldsymbol{\gamma}_N(\tau)\} - \tau).$$

Let $\mathbf{u}_N := \frac{\mathbf{Z}(x_0)}{\|\mathbf{Z}(x_0)\|}$ and apply Theorem S.2.4 to obtain

$$\mathbf{u}_N^\top(\overline{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau)) = \widetilde{\boldsymbol{U}}_{1,N}(\tau) + r_N^{(1)}(\tau, \mathbf{u}_N) + r_N^{(2)}(\tau, \mathbf{u}_N) \tag{S.3.6}$$

with

$$P\Big(\sup_{\tau\in\mathcal{T}_K} |r_N^{(1)}(\tau, \mathbf{u}_N) + r_N^{(2)}(\tau, \mathbf{u}_N)| \geq \frac{C}{2} R_{1,n}^{(L)}(A\log n) + \frac{C}{2} R_{2,n}^{(L)}(A, A\log n)\Big)$$
$$\leq 2N^2(3S+2)N^{-A}$$

and $R_{1,n}^{(L)}(A\log n), R_{2,n}^{(L)}(A, A\log n)$ defined in Theorem S.2.4. Note that under assumption (L) we have $\widetilde{\mathcal{E}}(\mathbf{u}_N, \boldsymbol{\gamma}_N) = O(\xi_m^{-1})$. Now a straightforward but tedious calculation shows that under the assumptions of Theorem 3.7 we have, for any fixed constant $A > 0$,

$$R_{1,n}^{(L)}(A\log n) + R_{2,n}^{(L)}(A, A\log n) = O(R_n(S)).$$

Choosing $A = 3$, a combination of (S.3.4) and the bound above yields (3.10).

To prove the second claim we first shall prove that

$$\sup_{\tau\in\mathcal{T}} \big\|\widetilde{\boldsymbol{U}}_{1,N}(\tau) - \boldsymbol{U}_N(\tau)\big\| = o_P\big(N^{-1/2}\big), \tag{S.3.7}$$

where

$$\boldsymbol{U}_N(\tau) := -N^{-1}\frac{\mathbf{Z}(x_0)^\top}{\|\mathbf{Z}(x_0)\|} J_m(\tau)^{-1} \sum_{i=1}^N \mathbf{Z}_i(\mathbf{1}\{Y_i \leq Q(X_i;\tau)\} - \tau). \tag{S.3.8}$$

This follows by similar arguments as Step 1 in the proof of Theorem 2.4 in Section A.2 on pages 3294-3296 of Chao et al. (2017), by noting that the conditions required there follow exactly by our (A1)-(A3), (L), (L1) where we use that $\xi_m^4(\log N)^6 = o(N)$, $c_m(\boldsymbol{\gamma}_N)^2 = o(N^{-1/2})$ by (L1), and $\mathcal{I}$ and L in the statement of Theorem 2.4 of Chao et al. (2017) being the coordinates of the non-zero components of $\mathbf{Z}(x_0)$ and r, respectively.

The proof of the second claim is very similar to the proof of (A.2). Use (S.3.7) instead of (S.3.34) and (S.3.4) instead of (S.3.30) and similar arguments as in the proof of (A.2) to obtain

$$\sup_{\tau\in\mathcal{T}} \Big|(\Pi_K \mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\cdot))(\tau) - \mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\tau)\Big| \leq o_P(\|\mathbf{Z}(x_0)\| N^{-1/2})$$
$$+ \sup_{\tau\in\mathcal{T}} \Big|[\Pi_K Q(x_0;\cdot)](\tau) - Q(x_0;\tau)\Big| + \sup_{\tau\in\mathcal{T}} |\mathbf{Z}(x_0)^\top \boldsymbol{\gamma}_N(\tau) - Q(x_0;\tau)|$$
$$+ \sup_{\tau\in\mathcal{T}} \Big|[\Pi_K \mathbf{Z}(x_0)^\top \boldsymbol{U}_N](\tau) - \mathbf{Z}(x_0)^\top \boldsymbol{U}_N(\tau)\Big|,$$

and bound the last term exactly as in the proof of (A.2).

S.3.2.2. *Proof of* (3.13) *and the weak convergence of* $\widehat{\boldsymbol{\beta}}_{or}(\tau)$ *in Corollary 3.9.* Combining (S.3.4) and (S.3.7) we obtain the representation

$$\begin{aligned}&\mathbf{Z}(x_0)^\top(\widehat{\boldsymbol{\beta}}_{or}(\tau)-\boldsymbol{\gamma}_N(\tau))\\&=-N^{-1}\mathbf{Z}(x_0)^\top J_m(\tau)^{-1}\sum_{i=1}^N\mathbf{Z}_i(\mathbf{1}\{Y_i\le Q(X_i;\tau)\}-\tau)+o_P\Big(\frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big).\end{aligned}$$

Moreover, by straightforward calculations bounding the terms in (3.10) we obtain under assumption (L1) the representation $\mathbf{Z}(x_0)^\top(\widehat{\boldsymbol{\beta}}_{or}(\tau)-\widehat{\boldsymbol{\beta}}(\tau))=o_P(\|\mathbf{Z}(x_0)\|N^{-1/2})$.

To prove the weak convergence result for both $\widehat{\boldsymbol{\beta}}_{or}$ and $\overline{\boldsymbol{\beta}}$, it thus suffices to prove that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\frac{\mathbf{Z}(x_0)^\top J_m(\tau)^{-1}\mathbf{Z}_i(\mathbf{1}\{Y_i\le Q(X_i;\tau)\}-\tau)}{(\mathbf{Z}(x_0)^\top J_m(\tau)^{-1}\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]J_m(\tau)^{-1}\mathbf{Z}(x_0))^{1/2}}\rightsquigarrow\mathcal{N}(0,\tau(1-\tau)).$$

This follows by an application of the Lindeberg CLT. The verification of the Lindeberg condition here is a simple modification from finite dimensional joint convergence in the proof of (2.4) in Section A.1.2 in Chao et al. (2017) to pointwise convergence. □

S.3.2.3. *Proof of* (3.15) *and part 2 of Corollary 3.10.* The proof of the process convergence of $\widehat{\boldsymbol{\beta}}_{or}(\tau)$ follows directly by Theorem 2.4 of Chao et al. (2017). Moreover, the bound

$$\sup_{\tau\in\mathcal{T}}|\mathbf{Z}(x_0)(\widehat{\boldsymbol{\beta}}(\tau)-\widehat{\boldsymbol{\beta}}_{or}(\tau))|=o_P(N^{-1/2}\|\mathbf{Z}(x_0)\|)$$

follows from Theorem 3.7 after some simple computations to bound the term $\sup_{\tau\in\mathcal{T}_K}|\mathbf{Z}(x_0)(\overline{\boldsymbol{\beta}}(\tau)-\widehat{\boldsymbol{\beta}}_{or}(\tau))|$ and using similar arguments as in the proof of (S.3.2) to obtain the bound

$$\sup_{\tau\in\mathcal{T}}\Big|[\Pi_K Q(x_0;\cdot)](\tau)-Q(x_0;\tau)\Big|=O(G^{-\eta}).$$

This implies (3.15). The proof of the second part follows by similar arguments as the proof of Corollary 4.1 in Chao et al. (2017) and is omitted for the sake of brevity. □

S.3.3. Proof of Theorem 3.2, Theorem 3.8 and necessary conditions for S, G in Section 3.1 and Section 3.2.

First, we will prove in Section S.3.3.1 that the oracle rule in Corollary 3.4 fails whenever $G \lesssim N^{1/(2\eta)}$ and that the oracle rule in Corollary 3.10 fail whenever $G \lesssim N^{1/(2\eta)}\|\mathbf{Z}(x_0)\|^{-1/\eta}$, no matter how S is chosen.

Second, in Section S.3.3.2 we will construct a special class of data generation processes. This construction will be utilized in Section S.3.3.3 to prove that the oracle rules in Corollary 3.3 and Corollary 3.9 fail if $S \gtrsim N^{1/2}$ or $S \gtrsim N^{1/2}\xi_m^{-1}$, respectively.

Third, in Section S.3.3.4 we will prove that, provided $G \gg N^{1/(2\eta_\tau)}$ and $S \gtrsim N^{1/2}$, the oracle rule in Corollary 3.4 fails and that, for $G \gg N^{1/(2\eta)}\|\mathbf{Z}(x_0)\|^{-1/\eta_\tau}$ and $S \gtrsim N^{1/2}\xi_m^{-1}$, the oracle rule in Corollary 3.10 fails.

Fourth, in Section S.3.3.5 we will derive the lower bounds in Theorem 3.2 and Theorem 3.8 from the necessity of the conditions in Corollary 3.9 and 3.10.

S.3.3.1. *Sharpness of the conditions on G in Corollary 3.4 and 3.10.* Fix an integer $\eta_\tau \geq 2$ and assume that $G = G_N \lesssim N^{1/(2\eta_\tau)}$. We will now prove that weak convergence in (3.5) fails for some $(P, \mathbf{Z}) \in \mathcal{P}_1(d^{1/2}, 1, \overline{f}, \overline{f'}, f_{min})$, independently of the choice of S.

We begin by noting that it is possible to construct a function g on $\mathbb{R}$ with the following properties: g has support $[-1, 1]$, g is η_τ times continuously differentiable on $\mathbb{R}$ with absolute values of all the function and derivatives bounded uniformly by 1, g does not vanish on $\mathbb{R}$. From the properties listed here it is clear that g is not a polynomial and hence, $d_{r_\tau} := \inf_{p\in\mathcal{P}_{r_\tau}} \sup_{x\in[-1,1]} |g(x) - p(x)| > 0$ for any fixed $r_\tau \geq \eta_\tau$ and $r_\tau \in \mathbb{N}$, where $\mathcal{P}_{r_\tau}$ is the space of polynomials of degree r_τ.

We now construct a function Q on $\mathcal{T} = [\tau_L, \tau_U]$ that is a proper quantile function with the following properties: the corresponding distribution function F is twice continuously differentiable, its density f satisfies (A2)-(A3) and there exists a subsequence N_k and $\varepsilon > 0$ such that for all k

$$G_{N_k}^{\eta_\tau} \inf_{p\in\mathcal{PP}_{r_\tau}(G_{N_k})} \sup_{\tau\in\mathcal{T}} |Q(\tau) - p(\tau)| \geq \varepsilon, \tag{S.3.9}$$

where $\mathcal{PP}_{r_\tau}(G_N)$ is the set of piecewise polynomials of degree r_τ defined on a the equally spaced grid $\tau_L = t_{1,N} < ... < t_{G_N,N} = \tau_U$. Begin by considering the case $\limsup_{N\to\infty} G_N = \infty$. Pick a subsequence N_k of N and a sequence

of integers $\nu_k \leq G_{N_k}$ such that

$$I_k := \Big[\tau_L + \frac{\nu_k(\tau_U - \tau_L)}{G_{N_k}}, \tau_L + \frac{(\nu_k+1)(\tau_U-\tau_L)}{G_{N_k}}\Big] \subset \Big(\tau_L + (\tau_U-\tau_L)\sum_{j=1}^{k-1} 2^{-j}, \tau_L + (\tau_U-\tau_L)\sum_{j=1}^{k} 2^{-j}\Big).$$

By construction, the intervals I_k are mutually disjoint and each interval is contained between two adjacent knots in the grid $t_{1,N_k}, ..., t_{G_{N_k},N_k}$.

Define for some $a, c > 0$ (specified later):

$$Q(\tau) := a\tau + c\sum_{k=2}^{\infty}(2G_{N_k})^{-\eta_\tau} g(2G_{N_k}(\tau - s_k)/(\tau_U - \tau_L)) \tag{S.3.10}$$

where s_k denotes the midpoint of I_k. Note that by construction the supports of $\tau \mapsto g(2G_{N_k}(\tau - s_k)/(\tau_U - \tau_L))$ do not overlap for different k, so for each fixed τ only one summand in the sum above does not vanish. Moreover, again by construction, the support of $\tau \mapsto g(2G_{N_k}(\tau - s_k)/(\tau_U - \tau_L))$ is contained between two knots of the form $(\tau_U - \tau_L)\nu_k/G_{N_k}, (\tau_U - \tau_L)(\nu_k + 1)/G_{N_k}$ for some $\nu_k \in \mathbb{N}$. Hence, (S.3.10) implies that

$$\inf_{p \in \mathcal{PP}_{r_\tau}(G_{N_k})} \sup_{\tau \in \mathcal{T}} |Q(\tau) - p(\tau)| \geq c(2G_{N_k})^{-\eta_\tau} d_{r_\tau}. \tag{S.3.11}$$

Suppose now that $\limsup_{N\to\infty} G_N < \infty$. Since in this case G_N is a bounded sequence of natural numbers, it must take at least one value, say N_0, infinitely often. Let $s := \tau_L + (\tau_U - \tau_L)/(2G_{N_0})$. Define for $a, c > 0$,

$$Q(\tau) := a\tau + c(2G_{N_0})^{-\eta_\tau} g(2G_{N_0}(\tau - s)/(\tau_U - \tau_L)). \tag{S.3.12}$$

Similar to (S.3.11), we have for $N_k \equiv N_0$

$$\inf_{p \in \mathcal{PP}_{r_\tau}(G_{N_k})} \sup_{\tau \in \mathcal{T}} |Q(\tau) - p(\tau)| \geq c(2G_{N_k})^{-\eta_\tau} d_{r_\tau}. \tag{S.3.13}$$

Simple computations show that for the Q defined in (S.3.10) or (S.3.12) $\sup_\tau |Q'(\tau)| \leq a + c, \sup_\tau |Q''(\tau)| \leq c, \inf_\tau Q'(\tau) \geq a - c$. Hence $a, c > 0$ can be chosen such that (A2)-(A3) hold for $F = Q^{-1}$. Thus we have established (S.3.9).

Now let $P(X_1 = m^{1/2}\mathbf{e}_j) = 1/m$ for $j = 1, ..., m$ and let $\mathbf{Z}(x) \equiv x$ and assume $X_1, ..., X_N$ are i.i.d. Let $Y_1, ..., Y_N$ be independent of $\{X_i\}_{1=1,...,N}$ and have quantile function Q for which (S.3.9) holds and (A2)-(A3) hold

for $F = Q^{-1}$. To simplify notation, assume that (S.3.9) holds with $N \equiv N_k$. Observe that by definition $\tau \mapsto \widehat{\beta}_j(\tau) \in \mathcal{PP}_{r_\tau}(G_N)$ for any $j = 1, ..., d$, where we recall that $\mathbf{B}(\tau)$ is a vector of B-spline basis with degree r_τ. Thus we have almost surely for all N

(S.3.14)
$$G_N^{\eta_\tau} \sup_{\tau\in\mathcal{T}} |\widehat{\beta}_j(\tau) - \beta_j(\tau)| \geq G_N^{\eta_\tau} \inf_{p\in\mathcal{PP}_{r_\tau}(G_{N_k})} \sup_{\tau\in\mathcal{T}} |p(\tau) - Q(\tau)| \geq \varepsilon > 0.$$

Note that (S.3.14) holds regardless of the number of groups S. Now, since $G_N \lesssim N^{1/(2\eta_\tau)}$, we have $C := \sup_{N\geq 1} G_N^{\eta_\tau}/N^{1/2} < \infty$ and thus (S.3.14) implies that for all N

$$N^{1/2} \sup_{\tau\in\mathcal{T}} |\widehat{\beta}_j(\tau) - \beta_j(\tau)| \geq \varepsilon/C > 0 \quad a.s. \tag{S.3.15}$$

Now assume that the weak convergence (3.5) holds. We have by the continuous mapping theorem

$$\sqrt{N} \sup_{\tau\in\mathcal{T}} |\widehat{\beta}_j(\tau) - \beta_j(\tau)| \rightsquigarrow \sup_{\tau\in\mathcal{T}} |\mathbb{G}(\tau)|, \tag{S.3.16}$$

where $\mathbb{G}$ is a tight, centered Gaussian process with continuous sample paths. By the arguments given on pages 60-61 of Ledoux and Talagrand (1991), it follows that $\sup_{\tau\in\mathcal{T}} |\mathbb{G}(\tau)|$ has a continuous distribution with left support point zero, and thus $P(\sup_{\tau\in\mathcal{T}} |\mathbb{G}(\tau)| < \delta) > 0$ for all $\delta > 0$. This also implies that $[\varepsilon/(2C), \infty)$ is a continuity set of the distribution of $\sup_{\tau\in\mathcal{T}} |\mathbb{G}(\tau)|$, and thus the weak convergence together with the Portmanteau Theorem implies that

$$\lim_{N\to\infty} P\Big(\sqrt{N} \sup_{\tau\in\mathcal{T}} |\widehat{\beta}_j(\tau) - \beta_j(\tau)| \geq \varepsilon/(2C)\Big) = P\Big(\sup_{\tau\in\mathcal{T}} |\mathbb{G}(\tau)| \geq \varepsilon/(2C)\Big) < 1.$$

This contradicts (S.3.15), and thus the weak convergence (3.5) can not hold. Arguments for the case in Corollary 3.10 with increasing m are similar as above, except that the rate of weak convergence in (S.3.16) is $\sqrt{N}\|\mathbf{Z}(x_0)\|^{-1}$, and $G_N \lesssim N^{1/(2\eta_\tau)}\|\mathbf{Z}(x_0)\|^{-1/\eta_\tau}$. □

S.3.3.2. *A special class of data generation processes.* Consider a family of quantile functions $Q_{a,b}(\tau) = a\tau^2 + b\tau$, $\tau \in (0, 1)$ indexed by $a \geq 0, b > 0$. The corresponding distribution functions $F_{a,b}$ have support $[0, a + b]$ and take the form

$$F_{a,b}(y) = \begin{cases} \frac{1}{2a}(-b + (b^2 + 4ay)^{1/2}), & a > 0, b > 0; \\ \frac{y}{b}, & a = 0, b > 0. \end{cases} \tag{S.3.17}$$

Note that the first two derivatives of the function $F_{a,b}$ defined in (S.3.17) are

$$F'_{a,b}(y) = (b^2 + 4ay)^{-1/2}, \quad F''_{a,b}(y) = -2a(b^2 + 4ay)^{-3/2}$$

and in particular

$$\inf_{y\in[0,a+b]} F'_{a,b}(y) \geq (b+2a)^{-1} \qquad \sup_{y\in[0,a+b]} |F'_{a,b}(y)| \leq b^{-1}, \qquad \sup_{y\in[0,a+b]} |F''_{a,b}(y)| \leq 2ab^{-3}.$$

Choosing $b = 1/\overline{f}$ and a such that

$$0 \leq 2a < \left(1/f_{min} - 1/\overline{f}\right) \wedge \left(\overline{f'}(\overline{f})^{-3}\right) \tag{S.3.18}$$

ensures that

$$\inf_{y\in[0,a+b]} F'_{a,b}(y) \geq f_{min}, \qquad \sup_{y\in[0,a+b]} |F'_{a,b}(y)| \leq \overline{f}, \qquad \sup_{y\in[0,a+b]} |F''_{a,b}(y)| \leq \overline{f'}. \tag{S.3.19}$$

Let $Y_1, ..., Y_n$ be i.i.d. with distribution function $F_{a,b}$ and $X_1, ..., X_n$ i.i.d., independent of $Y_1, ..., Y_n$, and such that $P(X_1 = \mathbf{e}_j\sqrt{m}) = 1/m, j = 1, ..., m$ where $\mathbf{e}_j \in \mathbb{R}^m$ denotes the jth unit vector. In this model, we have

$$\sum_{i=1}^{n} \rho_\tau(Y_i - \boldsymbol{\beta}^\top X_i) = \sum_{j=1}^{m} \sum_{i:X_i=\mathbf{e}_j} \rho_\tau(Y_i - \sqrt{m}\beta_j).$$

Define $\mathcal{A}_j := \{i : X_i = \mathbf{e}_j\}$, $n_j := \#\mathcal{A}_j$. Letting

$$\widetilde{\boldsymbol{\beta}}(\tau) = \operatorname*{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^m} \sum_{i=1}^{n} \rho_\tau(Y_i - \boldsymbol{\beta}^\top X_i) \tag{S.3.20}$$

we find that $\widetilde{\beta}_j(\tau) = \operatorname{argmin}_{b\in\mathbb{R}} \sum_{i:X_i=\mathbf{e}_j} \rho_\tau(Y_i - \sqrt{m}b)$, and by p.9 in Koenker (2005) the solution of this minimization problem with smallest absolute value is $\widetilde{\beta}_j(\tau) = m^{-1/2} Y^j_{(\lceil n_j\tau \rceil)} \mathbf{1}\{n_j > 0\}$ where $Y^j_{(k)}$ is the k-th order statistic of the sample $\{Y_i\}_{i\in\mathcal{A}_j}$ and $\lceil u \rceil$ denotes the smallest integer greater or equal u.

Now assume that m, n, K, G are sequences of positive integers indexed by $N \to \infty$ such that $n \geq 2, m \leq n, n = N/S$ for an integer S.

In Section S.7 of the online Appendix we shall prove the following facts about $\widetilde{\beta}_j(\tau)$ defined above

1. For any fixed $\tau \in \mathcal{T}$, $b > 0$ and any $a_{max} > 0$ there exists $a \in [0, a_{max}]$ such that

$$\limsup_{N\to\infty} \frac{n}{m} \left| \mathbb{E}[m^{1/2}\widetilde{\beta}_j(\tau)] - Q_{a,b}(\tau) \right| > 0 \tag{S.3.21}$$

$$\frac{n}{m} \mathrm{Var}(m^{1/2}\widetilde{\beta}_j(\tau)) \lesssim 1 \tag{S.3.22}$$

2. Assume that $G \gg (N/m)^{\alpha}$ for some $\alpha > 0$ and $N/m \geq 1$. Let $\check{\beta}_j(\tau) := \sum_{k=1}^{K} A_k(\tau_k)\widetilde{\beta}_j(\tau_k)$ where

$$A_k(\tau) := \mathbf{B}(\tau)^{\top}\left(\sum_{k=1}^{K}\mathbf{B}(\tau_k)\mathbf{B}(\tau_k)^{\top}\right)^{-1}\mathbf{B}(\tau_k), \tag{S.3.23}$$

and we recall that $\mathbf{B} = (B_1, ..., B_q)^{\top}$ where $\{B_1, ..., B_q\}$ B-spline basis with equidistant knots $\tau_L = t_1 < ... < t_G = \tau_U$ and degree $r_\tau \geq \eta_\tau$. For any closed interval $\mathcal{T}_0 \subset \mathcal{T}$ with non-empty interior and any $a_{max}, b > 0$ there exists $a \in [0, a_{max}]$ and a sequence τ_N in $\mathcal{T}_0$ such that

$$\limsup_{N\to\infty}\frac{n}{m}\left|\mathbb{E}[m^{1/2}\check{\beta}_j(\tau_N)] - Q_{a,b}(\tau_N)\right| > 0 \tag{S.3.24}$$

$$\frac{n}{m}\mathrm{Var}(m^{1/2}\check{\beta}_j(\tau_N)) \lesssim 1. \tag{S.3.25}$$

S.3.3.3. *Necessity of conditions of oracle rules in Corollary 3.3 and Corollary 3.9.* Fix an arbitrary $\tau \in \mathcal{T}$. Let $Y_1, ..., Y_N$ be i.i.d. with distribution function $F_{a,b}$ defined in (S.3.17) and independent of $X_1, ..., X_N$ where X_i i.i.d. with $P(X_1 = \mathbf{e}_j\sqrt{m}) = 1/m, j = 1, ..., m$. Let $\mathbf{Z}(x) := x$. Define $b = 1/\overline{f}$ and pick a satisfying (S.3.18) such that (S.3.21)-(S.3.22) hold.

We begin by considering the case where $m = d$ is fixed. Denote by P the measure corresponding to X_1, Y_1. Due to the choice of a, b and the distribution of X the pair $(\mathbf{Z}, P)$ is an element of $\mathcal{P}_1(d^{1/2}, 1, \overline{f}, \overline{f'}, f_{min})$. Moreover, by construction, $\{\widehat{\beta}_j^s(\tau)\}_{s=1,...,S}$ are i.i.d. and $\widehat{\beta}_j^1(\tau)$ has the same distribution as $\widetilde{\beta}_j(\tau)$ defined in (S.3.20).

Next, we shall prove that weak convergence of $\overline{\boldsymbol{\beta}}$ in Corollary 3.3 fails for $\mathbf{u} = \mathbf{e}_1$. Given (S.3.22), a simple computation shows that

$$\mathbb{E}[(\sqrt{N}(\overline{\beta}_1(\tau) - \mathbb{E}[\overline{\beta}_1(\tau)]))^2] = \frac{N}{S}\mathrm{Var}(\widetilde{\beta}_1(\tau)) = n\mathrm{Var}(\widetilde{\beta}_1(\tau)) \lesssim 1 \tag{S.3.26}$$

for any fixed $\tau \in \mathcal{T}$. Thus the sequence $\sqrt{N}(\overline{\beta}_1(\tau) - \mathbb{E}[\overline{\beta}_1(\tau)])$ is uniformly integrable.

Now assume that the that weak convergence of $\overline{\boldsymbol{\beta}}$ in Corollary 3.3 holds. By (S.3.21) we know that for $S \gtrsim N^{1/2}$ and any fixed $\tau \in \mathcal{T}$

$$\left|\mathbb{E}[\overline{\beta}_1(\tau)] - \beta_1(\tau)\right| \gtrsim 1/n = S/N \gtrsim N^{-1/2}. \tag{S.3.27}$$

If $\left|\mathbb{E}[\overline{\beta}_1(\tau)] - \beta_1(\tau)\right| \asymp N^{-1/2}$, uniform integrability of $\sqrt{N}(\overline{\beta}_1(\tau) - \mathbb{E}[\overline{\beta}_1(\tau)])$ implies uniform integrability of $\sqrt{N}(\overline{\beta}_1(\tau) - \beta_1(\tau))$. In that case weak convergence of $\overline{\boldsymbol{\beta}}$ implies $\mathbb{E}[\sqrt{N}(\overline{\beta}_1(\tau) - \beta_1(\tau))] \to 0$ by Theorem 25.12 on p.338 of Billingsley (1995), which contradicts (S.3.27).

If $\left|\mathbb{E}[\overline{\beta}_1(\tau)]-\beta_1(\tau)\right| \gg N^{-1/2}$, (S.3.26) implies that $N\mathbb{E}[(\overline{\beta}_1(\tau)-\beta_1(\tau))^2] \to \infty$, and thus $\sqrt{N}\left|\overline{\beta}_1(\tau)-\beta_1(\tau)\right|$ diverges to infinity in probability, which contradicts weak convergence of $\overline{\boldsymbol{\beta}}$.

Now consider the setting of Section 3.2 where $m \to \infty$. With the choice of X described in Section S.3.3.2 we have $\boldsymbol{\gamma}_N(\tau) = m^{-1/2}Q_{a,b}(\tau)(1,...,1)^\top \in \mathbb{R}^m$, $\xi_m = m^{1/2}$, $c_m(\boldsymbol{\gamma}_N) = 0$. The matrix $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]$ is the $m \times m$ identity matrix while $J_m(\tau), \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))$ are diagonal with entries bounded away from zero and infinity. A simple computation shows that (L) holds with $r = 1$. Denote by P_N the sequence of measures corresponding to X_i, Y_i. Picking m such that $m^2(\log N)^6 = o(N)$ ensures that $(\mathbf{Z}, P_N)$ lies in $\mathcal{P}_L(1, \overline{f}, \overline{f'}, f_{min}, R)$ for any $R \geq 1$. Letting $x_0 = \mathbf{Z}(x_0) = \mathbf{e}_1 \in \mathbb{R}^m$ the weak convergence in (3.13) takes the form

$$\sqrt{\frac{N}{m}}(m^{1/2}\overline{\beta}_1(\tau) - Q_{a,b}(\tau)) \rightsquigarrow \mathcal{N}(0, \sigma(\tau))$$

for some constant $\sigma(\tau)$. Moreover, (S.3.22) yields

$$\begin{aligned}\mathbb{E}[(\sqrt{N}m^{-1/2}(m^{1/2}\overline{\beta}_1(\tau) - m^{1/2}\mathbb{E}[\overline{\beta}_1(\tau)]))^2] \\ = \frac{N}{Sm}\mathrm{Var}(m^{1/2}\widetilde{\beta}_1(\tau)) = \frac{n}{m}\mathrm{Var}(m^{1/2}\widetilde{\beta}_1(\tau)) \lesssim 1\end{aligned}$$

while (S.3.21) implies that for $S \gtrsim N^{1/2}/m^{1/2} = N^{1/2}\xi_m^{-1}$

$$\left|\mathbb{E}[m^{1/2}\overline{\beta}_1(\tau)] - Q_{a,b}(\tau)\right| \gtrsim m/n = mS/N \gtrsim m^{1/2}N^{-1/2}.$$

Now failure of weak convergence in (3.13) can be established for $x_0 = \mathbf{Z}(x_0) = \mathbf{e}_1 \in \mathbb{R}^m$ by exactly the same arguments as given in the first part of the proof for the fixed m case, details are omitted for the sake of brevity.

S.3.3.4. *Necessity of conditions for oracle rules in Corollary 3.4 and Corollary 3.10.* Fix a subset $\mathcal{T}_0 \subset \mathcal{T}$ which is closed and has non-empty interior. Let $Y_1, ..., Y_N$ be i.i.d. with distribution function $F_{a,b}$ defined in (S.3.17) and independent of $X_1, ..., X_N$ where X_i i.i.d. with $P(X_1 = \mathbf{e}_j\sqrt{m}) = 1/m, j = 1, ..., m$. Let $b = 1/\overline{f}$ and pick a satisfying (S.3.18) and a sequence τ_N in $\mathcal{T}_0$ such that (S.3.24)-(S.3.25) hold. Since τ_N is a sequence in a compact set, it must have a convergent subsequence $\tau_{N_k} \to \tau_0 \in \mathcal{T}_0$. From now on, we will without loss of generality assume that $N = N_k$.

Assume that (3.5) holds. Process convergence in (3.5), along with the continuity of the sample paths of the limiting processes, implies that the sequence of random variables $\sqrt{N}(m^{1/2}\widehat{\beta}_1(\tau_N) - Q_{a,b}(\tau_N))$ converges to a centered normal random variable. This can be led to a contradiction with

(S.3.24)-(S.3.25) by similar arguments as in Section S.3.3.3 after observing that by definition $\widehat{\beta}_1(\tau) = S^{-1}\sum_{s=1}^{S}\check{\beta}_1(\tau)$.

The proof that (3.15) fails can be done similarly, and details are omitted for the sake of brevity. □

S.3.3.5. *Proof of Theorem 3.8 and Theorem 3.2.* The proofs of both results follow by the same type of arguments, and thus we only give details for the proof of Theorem 3.2. This proof will be based on the necessary conditions in Corollary 3.3 and Corollary 3.4.

First we derive (3.3) by contradiction. Assume that for some $\tau \in \mathcal{T}$ we have for any $(P, \mathbf{Z}) \in \mathcal{P}_1$ and $C > 0$ that $P(\|\bar{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)\| > CS/N) \to 0$. Then $\bar{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau) = o_P(S/N)$ and thus $\sqrt{N}(\bar{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)) \rightsquigarrow \mathcal{N}(0, H(\tau,\tau))$ even if $S = N^{1/2}$. However, this contradicts the necessity statement in Corollary 3.3 (in the proof of Corollary 3.3 we show that for any $\tau \in \mathcal{T}$ if $S/N^{1/2} \gtrsim 1$ then there exists $(P, \mathbf{Z}) \in \mathcal{P}_1$ such that this weak convergence fails).

The claim in (3.4) will be derived from the 'necessary' part of Corollary 3.4 by similar arguments. Assume that (3.4) does not hold, i.e. there exist $c, \eta > 0$ such that for any $(P, \mathbf{Z}) \in \mathcal{P}_1$ and any $x_0 \in \mathcal{X}$ with $\tau \mapsto Q(x_0;\tau) \in \Lambda_c^\eta$ we have for any $C > 0$

$$P\Big(\sup_{\tau\in\mathcal{T}} \|\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)\| \geq \frac{CS}{N} + CG^{-\eta}\Big) \to 0.$$

This implies

$$\sup_{\tau\in\mathcal{T}} \|\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)\| = o_P\Big(\frac{S}{N} + G^{-\eta}\Big),$$

and the process convergence $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{or} - \boldsymbol{\beta}) \rightsquigarrow \mathbb{G}$ implies $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightsquigarrow \mathbb{G}$. This contradicts the necessity of $S = o(N^{1/2})$ *and* $G \gg N^{1/(2\eta)}$ in Corollary 3.4. Hence (3.4) follows. □

S.3.4. Proofs for Section A. Recall the definition of $g_N(\mathbf{b}_N)$ in (S.2.2) and $c_m(\mathbf{b}_N)$ in (3.9). As mentioned in Remark A.6, the results in Section A hold for general 'centerings' $\boldsymbol{\beta}_N$ (which includes $\boldsymbol{\gamma}_N$ as a special case) provided that certain technical conditions are satisfied. The precise form of those conditions is discussed in section S.3.4.1. There, we show that $\boldsymbol{\gamma}_N$ satisfies those general conditions. In sections S.3.4.2 and S.3.4.5, all theoretical results will be proved for $\boldsymbol{\beta}_N$ instead of $\boldsymbol{\gamma}_N$ under the general conditions described in section S.3.4.1. The results with $\boldsymbol{\gamma}_N$ in Section A follow as a special case.

S.3.4.1. *Details of Remark A.6.* Assume that (A1)-(A3) hold. Corollary A.2 holds for any sequence of centering coefficients $\boldsymbol{\beta}_N$ under the following conditions: $S = o(N^{1/2}/(m\xi_m^2(\log N)^2))$, $Sc_m^2(\boldsymbol{\beta}_N)m\log N = o(1)$, and

(B1) $g_N(\boldsymbol{\beta}_N) = o(N^{-1/2})$ and $m^4(\log N)^{10} + m^2\xi_m^4 = o(N)$.
(B2) $mc_m(\boldsymbol{\beta}_N)\log N = o(1)$ and the sequence $\mathbf{u}_N$ (in Corollary A.2) satisfies

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

Furthermore, Corollary A.5 holds under the additional restriction $c_m(\boldsymbol{\beta}_N) = o(\|\mathbf{Z}(x_0)\|N^{-1/2})$ and the other conditions on Q, G, K stated in Corollary A.5.

In the sequel, we verify that, assuming the conditions stated in Corollary A.2 and Corollary A.5 respectively, $\boldsymbol{\gamma}_N$ satisfies the conditions given above. Note that some conditions in these theorems overlap with the conditions stated above, so we only need to verify that the conditions $c_m^2(\boldsymbol{\gamma}_N)\xi_m = o(N^{-1/2})$ and $S = o(N^{1/2}/(m\xi_m^2(\log N)^2))$ in both theorems imply $g_N(\boldsymbol{\gamma}_N) = o(N^{-1/2})$, $\chi_N(\mathbf{u}_N, \boldsymbol{\gamma}_N) = o(\|\mathbf{u}_N\|N^{-1/2})$, $Sc_m^2(\boldsymbol{\gamma}_N)m\log N = o(1)$ and moreover $mc_m(\boldsymbol{\gamma}_N)\log N = o(1)$.

Under the definition of $\boldsymbol{\gamma}_N$ in (3.8) and (A1)-(A3), we have $g_N(\boldsymbol{\gamma}_N) = O(\xi_m c_m^2(\boldsymbol{\gamma}_N))$. To see this, note that under (A2),

$$\begin{aligned}\mathbf{Z}(F_{Y|X}(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau)|X) - \tau) &= \mathbf{Z}f_{Y|X}(Q(X;\tau)|X)(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau) - Q(X;\tau)) \\ &\quad + \frac{1}{2}\mathbf{Z}f'_{Y|X}(\zeta_n(X,\tau)|X)(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau) - Q(X;\tau))^2, \end{aligned} \tag{S.3.28}$$

where $\zeta_n(X,\tau)$ is a value between $\mathbf{Z}(X)^\top\boldsymbol{\gamma}_N(\tau)$ and $Q(X;\tau)$. Thus

$$\begin{aligned} g_N(\boldsymbol{\gamma}_N) &= \sup_{\tau\in\mathcal{T}}\left\|\mathbb{E}[\mathbf{Z}(F_{Y|X}(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau)|X) - \tau)]\right\| \\ &\le \frac{\overline{f'}c_m(\boldsymbol{\gamma}_N)^2}{2}\|\mathbf{Z}\| = O(\xi_m c_m^2(\boldsymbol{\gamma}_N)), \end{aligned} \tag{S.3.29}$$

where the second inequality follows from the first order condition

$$\mathbb{E}[\mathbf{Z}f_{Y|X}(Q(X;\tau)|X)(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau) - Q(X;\tau))] = 0$$

by the definition of $\boldsymbol{\gamma}_N$ in (3.8) and (S.3.28). Moreover,

$$\chi_N(\mathbf{u}_N, \boldsymbol{\gamma}_N) = \sup_{\tau\in\mathcal{T}}\left|\mathbf{u}_N^\top J_m(\tau)^{-1}\mu(\boldsymbol{\gamma}_N(\tau),\tau)\right| \lesssim \|\mathbf{u}_N\|_2 g_N(\boldsymbol{\gamma}_N).$$

Hence, $c_m^2(\boldsymbol{\gamma}_N)\xi_m = o(N^{-1/2})$ implies $g_N(\boldsymbol{\gamma}_N) = o(N^{-1/2})$ and $\chi_N(\mathbf{u}_N, \boldsymbol{\gamma}_N) = o(\|\mathbf{u}_N\|N^{-1/2})$. In addition,

$$Sc_m^2(\boldsymbol{\gamma}_N)m\log N = o(SN^{-1/2}\xi_m^{-1}m\log N) = o(1),$$

where the first equality is from $c_m^2(\boldsymbol{\gamma}_N)\xi_m = o(N^{-1/2})$ and the second from $S = o(N^{1/2}/(m\xi_m^2(\log N)^2))$. Finally

$$mc_m(\boldsymbol{\gamma}_N)\log N = o(mN^{-1/4}\log N) = o((m^4N^{-1}(\log N)^4)^{1/4}) = o(1)$$

where the first bound follows from $c_m^2(\boldsymbol{\gamma}_N)\xi_m = o(N^{-1/2}), \xi_m \gtrsim 1$ and the third bound from $m^4(\log N)^{10} = o(N)$.

□

S.3.4.2. *Proof of Theorem A.1.* We begin by proving of (A.1). From Theorem S.6.1 in the supplementary material, we obtain the following representation for the oracle estimator

(S.3.30)
$$\widehat{\boldsymbol{\beta}}_{or}(\tau) - \boldsymbol{\beta}_N(\tau) = -J_m(\tau)^{-1}\frac{1}{N}\sum_{i=1}^N \mathbf{Z}_i(\mathbf{1}\{Y_i \le \mathbf{Z}_i^\top\boldsymbol{\beta}_N(\tau)\} - \tau) + r_n^{or}(\tau)$$

where $\sup_{\tau\in\mathcal{T}}\|r_n^{or}(\tau)\| = o_P(R_N^{or})$ and

$$R_N^{or} := c_m\Big(\frac{m\log N}{N}\Big)^{1/2} + c_m^2\xi_m + \Big(\frac{m\xi_m^2\log N}{N} + c_m^4\xi_m^4\Big) + \frac{1}{N^{1/2}}\Big(\Big(m\xi_m^2c_m^2(\log N)^3\Big)^{1/2} + \Big(\frac{m^3\xi_m^2(\log N)^7}{N}\Big)^{1/4}\Big).$$

Next we will prove that under the assumptions of Theorem A.1 we also have

(S.3.31)
$$\sup_{\tau\in\mathcal{T}_K}\Big\|\overline{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_N(\tau) + J_m(\tau)^{-1}\frac{1}{N}\sum_{i=1}^N \mathbf{Z}_i(\mathbf{1}\{Y_i \le \mathbf{Z}_i^\top\boldsymbol{\beta}_N(\tau)\} - \tau)\Big\| = O_P(R_N)$$

where

$$R_N := c_m\Big(\frac{m\log N}{N}\Big)^{1/2} + c_m^2\xi_m + \Big(\frac{Sm\xi_m^2\log N}{N} + c_m^4\xi_m^4\Big)\Big(1 + \frac{\log N}{S^{1/2}}\Big) + \frac{1}{N^{1/2}}\Big(\Big(m\xi_m^2c_m^2(\log N)^3\Big)^{1/2} + \Big(\frac{Sm^3\xi_m^2(\log N)^7}{N}\Big)^{1/4}\Big).$$

To this end, apply Theorem S.2.1. A simple computation shows that the quantities $R_{1,n}, R_{2,n}$ defined in (S.2.4) and (S.2.6) of Theorem S.2.1 satisfy

$$R_{1,n}(A\log n) + R_{2,n}(A, A\log n) \leq CR_N$$

for any fixed $A > 0$. Thus we obtain from (S.2.3) and (S.2.5) that there exists N_0 such that for all $N \geq N_0$ and any $\tau \in \mathcal{T}$

$$P\big(\big\|r_N^{(1)}(\tau) + r_N^{(2)}(\tau)\big\| > \widetilde{C}R_N\big) \leq 2n^{-A}(3S+1), \tag{S.3.32}$$

where

$$\overline{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_N(\tau) + J_m(\tau)^{-1}\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i(\mathbf{1}\{Y_i \leq \mathbf{Z}_i^\top \boldsymbol{\beta}_N(\tau)\} - \tau) = r_N^{(1)}(\tau) + r_N^{(2)}(\tau).$$

Choosing A large enough ensures that

$$\sup_{\tau\in\mathcal{T}_K}\big\|r_N^{(1)}(\tau) + r_N^{(2)}(\tau)\big\| = o_P(R_N), \tag{S.3.33}$$

and we obtain (S.3.31).

Next we prove (A.2). Recall the definition of the projection operator Π_K right after (2.5).

Step 1: $\|\Pi_K\|_\infty := \sup_{\|f\|_\infty=1}\|\Pi_K f\|_\infty < C$ for a constant $C > 0$ independent of K.
We will apply Theorem A.1 of Huang (2003), and for that we need to validate conditions A.1-A.3 of Huang (2003). Note that *none* of these conditions is associated with the functions being projected on the spline space – all conditions are associated with the design and the selection of basis functions.

We first show that Condition A.1 of Huang (2003) holds in our context. By the definition of grids $t_1, ..., t_G$ and $\tau_1, ..., \tau_K$, each interval $[t_k, t_{k+1}]$ contains $c_{k,K}$ spaced grid points of the form $\tau_U + l(\tau_L - \tau_U)/K$. Hence, applying Lemma S.8.2, we obtain for $j = 0, ..., G-1$ and any $f \in \Theta_G$

$$K^{-1}\sum_{k:\tau_k\in(t_j,t_{j+1}]} f^2(\tau_k) \leq \Big(1 + \frac{C_{r_\tau}}{c_{j,K}-1}\Big)\int_{t_j}^{t_{j+1}} f^2(x)dx,$$

and

$$K^{-1}\sum_{k:\tau_k\in(t_j,t_{j+1}]} f^2(\tau_k) \geq \Big(1 - \frac{C_{r_\tau}}{c_{j,K}-1}\Big)\int_{t_j}^{t_{j+1}} f^2(x)dx,$$

for a constant C_{r_τ} that depends only on r_τ. Since $\min_{k=1,\dots,K} c_{j,K} \to \infty$ it follows that for sufficiently large K Condition A.1 of Huang (2003) holds with $\gamma_1 = 1/2$ and $\gamma_2 = 2$.

Conditions A.2 (i) and (ii) of Huang (2003) easily follow from the fact that $\mathcal{T}$ is bounded. Finally, the definition of the knots ensures that A.2 (iii) in Huang (2003) holds. Condition A.3 holds since B-spline basis is used in the projection procedure, cf. the discussion following Condition A.3 on page 1630 of Huang (2003).

Thus conditions A.1-A.3 in Huang (2003) are satisfied, and Theorem A.1 of Huang (2003) shows the existence of a constant C such that $\|\Pi_K f\|_\infty \leq C\|f\|_\infty$ for all functions $f \in \ell^\infty(\mathcal{T})$. This completes the proof of step 1.

From the definition of $\widehat{\boldsymbol{\beta}}$ and linearity of Π_K we obtain

$$\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}(\tau) = (\Pi_K \mathbf{Z}(x_0)^\top \overline{\boldsymbol{\beta}}(\cdot))(\tau)$$

and thus

$$\begin{aligned}\mathbf{Z}(x_0)^\top(\widehat{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)) = \ & (\Pi_K\{\mathbf{Z}(x_0)^\top[\overline{\boldsymbol{\beta}}(\cdot) - \widehat{\boldsymbol{\beta}}_{or}(\cdot)]\})(\tau) \\ & + (\Pi_K\{\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\cdot)\})(\tau) - \mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\tau).\end{aligned}$$

Observe that

$$\sup_{\tau\in\mathcal{T}} \left|(\Pi_K\{\mathbf{Z}(x_0)^\top[\overline{\boldsymbol{\beta}}(\cdot) - \widehat{\boldsymbol{\beta}}_{or}(\cdot)]\})(\tau)\right| \lesssim \sup_{\tau\in\mathcal{T}_K} \left|\mathbf{Z}(x_0)^\top[\overline{\boldsymbol{\beta}}(\tau) - \widehat{\boldsymbol{\beta}}_{or}(\tau)]\right|.$$

Hence it remains to prove that

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

From Theorem 2.1 of Chao et al. (2017) we obtain

$$\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\tau) = Q(x_0;\tau) + \mathbf{Z}(x_0)^\top \boldsymbol{U}_N(\tau) + r_N(\tau) \tag{S.3.34}$$

where

$$\boldsymbol{U}_N(\tau) := -N^{-1} J_m(\tau)^{-1} \sum_{i=1}^N \mathbf{Z}_i(\mathbf{1}\{Y_i \leq Q(X_i;\tau)\} - \tau) \tag{S.3.35}$$

and $\sup_{\tau\in\mathcal{T}}|r_N(\tau)| \le \sup_{\tau\in\mathcal{T}}|\mathbf{Z}(x_0)^\top\boldsymbol{\gamma}_N(\tau)-Q(x_0;\tau)|+o_P(\|\mathbf{Z}(x_0)\|N^{-1/2})$. This implies

$$\begin{aligned}\sup_{\tau\in\mathcal{T}}\Big|(\Pi_K\mathbf{Z}(x_0)^\top\widehat{\boldsymbol{\beta}}_{or}(\cdot))(\tau)-\mathbf{Z}(x_0)^\top\widehat{\boldsymbol{\beta}}_{or}(\tau)\Big| &\le o_P\Big(\frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big)\\ +\sup_{\tau\in\mathcal{T}}\Big|[\Pi_K Q(x_0;\cdot)](\tau)-Q(x_0;\tau)\Big| &+\sup_{\tau\in\mathcal{T}}|\mathbf{Z}(x_0)^\top\boldsymbol{\gamma}_N(\tau)-Q(x_0;\tau)|\\ &+\sup_{\tau\in\mathcal{T}}\Big|[\Pi_K\mathbf{Z}(x_0)^\top\boldsymbol{U}_N](\tau)-\mathbf{Z}(x_0)^\top\boldsymbol{U}_N(\tau)\Big|.\end{aligned}$$

and it remains to bound the last term. From Lemma 5.1 of Huang (2003) we obtain that

(S.3.36)

$$\sup_{\tau\in\mathcal{T}}\big|(\Pi_K\mathbf{Z}(x_0)^\top\boldsymbol{U}_N)(\tau)-\mathbf{Z}(x_0)^\top\boldsymbol{U}_N(\tau)|\le C\inf_{\widetilde{f}\in\Theta_G}\sup_{\tau\in\mathcal{T}}|\mathbf{Z}(x_0)^\top\boldsymbol{U}_N(\tau)-\widetilde{f}(\tau)|,$$

where C is an absolute constant and Θ_G is the space of splines of degree r_τ with knots $t_1,...,t_G$ on $\mathcal{T}=[\tau_L,\tau_U]\subset(0,1)$. The right hand side of (S.3.36) can be related to the *modulus of continuity*. Indeed, using Theorem 6.27 of Schumaker (1981) we have for an absolute constant $C>0$,

$$\begin{aligned}&\inf_{\widetilde{f}\in\Theta_G}\sup_{\tau\in\mathcal{T}}\big|\mathbf{Z}(x_0)^\top\boldsymbol{U}_N(\tau)-\widetilde{f}(\tau)\big|\\ &\le C\sup_{\tau,\tau'\in\mathcal{T},|\tau-\tau'|\le\delta}|\mathbf{Z}(x_0)^\top\boldsymbol{U}_N(\tau)-\mathbf{Z}(x_0)^\top\boldsymbol{U}_N(\tau')|=o_P\Big(\frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big),\end{aligned}\tag{S.3.37}$$

where $\delta=\max_{0\le j\le K-1}(t_{j+1}-t_j)=o(1)$, and the last equality follows from Lemma S.8.1. Thus the proof of (A.2) is complete. □

S.3.4.3. *Proof of Corollary A.2.* From Theorem 2.1 of Chao et al. (2017) we obtain the representation

$$\begin{aligned}&\mathbf{Z}(x_0)^\top(\widehat{\boldsymbol{\beta}}_{or}(\tau)-\boldsymbol{\gamma}_N(\tau))\\ &=-\mathbf{Z}(x_0)^\top J_m(\tau)^{-1}\frac{1}{N}\sum_{i=1}^N\mathbf{Z}_i(\mathbf{1}\{Y_i\le Q(X_i;\tau)\}-\tau)+o_P\Big(\frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big).\end{aligned}$$

Moreover, by straightforward calculations bounding the terms in (A.1) we obtain the representation $\mathbf{Z}(x_0)^\top(\widehat{\boldsymbol{\beta}}_{or}(\tau)-\overline{\boldsymbol{\beta}}(\tau))=o_P(\|\mathbf{Z}(x_0)\|N^{-1/2})$. Hence to prove the weak convergence result for both $\widehat{\boldsymbol{\beta}}_{or}$ and $\overline{\boldsymbol{\beta}}$, it suffices to prove that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\frac{\mathbf{Z}(x_0)^\top J_m(\tau)^{-1}\mathbf{Z}_i(\mathbf{1}\{Y_i\le Q(X_i;\tau)\}-\tau)}{(\mathbf{Z}(x_0)^\top J_m(\tau)^{-1}\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]J_m(\tau)^{-1}\mathbf{Z}(x_0))^{1/2}}\rightsquigarrow\mathcal{N}(0,\tau(1-\tau)).$$

This follows by an application of the Lindeberg CLT. The verification of the Lindeberg condition here is a simple modification from finite dimensional joint convergence in the proof of (2.4) on page 3292 in Chao et al. (2017) to pointwise convergence. □

S.3.4.4. *Details for Remark A.4.* In the proof below, we will use the notations in Example 3.6: polynomial spline $\widetilde{B}_j$, j'th knot t_j and degree q.

We first verify (A1). The bounded spectrum $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]$ follows by the argument in Example 3.6. We note that every univariate polynomial spline function satisfies $|\widetilde{B}_j| \le C$ almost surely for some constant $C > 0$ for all $j \le m$. Moreover, the support of $\widetilde{B}_j$ and $\widetilde{B}_k$ is disjoint unless $|j-k| \le q+1$. Hence, we have almost surely

$$\|\mathbf{Z}_i\| = m^{1/2}\Bigg(\sum_{j=1}^{k-q-1} \widetilde{B}_j(X_i)^2\Bigg)^{1/2} \lesssim m^{1/2}\Bigg(\sum_{j=1}^{k-q-1} \mathbf{1}\{t_j \le X_i \le t_{j+q+1}\}\Bigg)^{1/2} \lesssim m^{1/2}. \tag{S.3.38}$$

On the other hand, since each $\widetilde{B}_j(X_i)$ is nonzero almost surely, $\|\mathbf{Z}_i\| \gtrsim m^{1/2}$ and $\xi_m \asymp m^{1/2}$.

We now verify the sufficient conditions of Corollary 3.9 and Corollary A.2 under the conditions made in this remark. Condition (L) holds given the discussion in Example 3.6. (L1) and the conditions $\xi_m^4(\log N)^6 = o(N), c_m^2(\boldsymbol{\gamma}_N) = o(N^{-1/2})$ hold under the assumptions of this remark. On the other hand, the sufficient conditions of Corollary A.2 can be easily verified by the assumptions of this remark since $\xi_m \asymp m^{1/2}$.

S.3.4.5. *Proof of Corollary A.5.* Similar to Section S.3.4.2, we will prove the weak convergence of $\widehat{\boldsymbol{\beta}}(\tau)$ in models with centering coefficient $\boldsymbol{\beta}_N(\tau)$ from a class specified in Section S.3.4.1.

The process convergence of $\frac{\sqrt{N}}{\|\mathbf{Z}(x_0)\|}\big(\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\cdot) - Q(x_0;\cdot)\big)$ is exactly Theorem 2.1 of Chao et al. (2017). A straightforward computation shows that under the assumptions of Corollary A.5 the results in (A.1) and (A.2) yield

$$\sup_{\tau\in\mathcal{T}} \frac{\sqrt{N}}{\|\mathbf{Z}(x_0)\|}\Big|\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}_{or}(\tau) - \mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}(\tau)\Big| = o_P(1).$$

This yields the claim in the first part. The proof of the second part follows by similar arguments as the proof of Corollary 4.1 in Chao et al. (2017), and is therefore omitted for brevity. □

APPENDIX S.4: PROOFS FOR SECTION 4

S.4.1. Proof of Theorem 4.1. Since the proofs under the assumptions of Corollary 3.4, Corollary 3.10 and Corollary A.5 are similar, we will only give a proof under the conditions of Corollary A.5. Define

$$W_N := \sqrt{n}(\mathbf{Z}(x_0)^\top \bar{\boldsymbol{\beta}}(\tau) - Q(x_0;\tau)), \quad \widehat{\sigma}^2 := n\mathbf{Z}(x_0)^\top \widehat{\Sigma}^D \mathbf{Z}(x_0)$$

and

$$p_N := P\Big(S^{1/2}|W_N| \leq \widehat{\sigma} t_{S-1,1-\alpha/2}\Big).$$

Throughout this proof, we will write S_N instead of S in order to emphasize that S depends on N. First, observe that along any sub-sequence N_k such that S_{N_k} only takes the constant value $S_0 > 2$ we have

$$\frac{\sqrt{n}}{\sigma_n}(\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}^1(\tau) - Q(x_0;\tau), ..., \mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}^{S_0}(\tau) - Q(x_0;\tau)) \rightsquigarrow \mathcal{N}(0, I_{S_0})$$

where I_{S_0} denotes the $S_0 \times S_0$ identity matrix and

$$\sigma_n^2 = \tau(1-\tau)\frac{\mathbf{Z}(x_0)^\top J_m(\tau)^{-1}\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]J_m(\tau)^{-1}\mathbf{Z}(x_0)}{\|\mathbf{Z}(x_0)\|^2}.$$

Observe that $\mathbf{Z}(x_0)^\top \widehat{\Sigma}^D \mathbf{Z}(x_0)$ is the sample variance of $\{\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}^s\}_{s=1,...,S}$. As a direct consequence of the continuous mapping theorem, we obtain $p_{N_k} \to (1-\alpha)$. By the definition of convergence of a sequence of real numbers to a limit, this also implies that for any fixed $S_0 > 2$

$$\lim_{N_0 \to \infty} \sup_{N \geq N_0 : S_N = S_0} |p_N - (1-\alpha)| = 0,$$

which in turn yields that for any fixed $\bar{S}$

$$\lim_{N_0 \to \infty} \sup_{N \geq N_0 : 2 < S_N \leq \bar{S}} |p_N - (1-\alpha)| = 0.$$

Next observe that for any fixed $\varepsilon > 0$

$$\begin{aligned}\Big|p_N - (1-\alpha)\Big| \leq\ & P\Big(\Big|\frac{\widehat{\sigma} t_{S-1,1-\alpha/2}}{\sigma_N \Phi^{-1}(1-\alpha/2)} - 1\Big| \geq \varepsilon\Big) \\ & + P\Big(\Big|S^{1/2}\sigma_N^{-1}|W_N| - \Phi^{-1}(1-\alpha/2)\Big| \leq \varepsilon\Phi^{-1}(1-\alpha/2)\Big) \\ & + \Big|P\Big(S^{1/2}\sigma_N^{-1}|W_N| \leq \Phi^{-1}(1-\alpha/2)\Big) - (1-\alpha)\Big|.\end{aligned}$$

Under the assumptions of Corollary A.5, we have the decomposition

$$\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}^s(\tau) = Q(x_0;\tau) + L_n^s(\tau) + r_n^s(\tau)$$

where $\sup_{s=1,\dots,S} |r_n^s(\tau)| = o_P(\|\mathbf{Z}(x_0)\| n^{-1/2})$ and

$$L_n^s(\tau) := \frac{1}{n}\sum_{i=1}^n \mathbf{Z}(x_0)^\top J_m(\tau)^{-1}\mathbf{Z}(X_{is})(\mathbf{1}\{Y_{is} \le Q(X_{is};\tau)\} - \tau).$$

This implies

$$\begin{aligned}
\widehat{\sigma}^2 &= \frac{n}{S_N - 1}\sum_{s=1}^{S_N}\Big(\mathbf{Z}(x_0)^\top \widehat{\boldsymbol{\beta}}^s(\tau) - \mathbf{Z}(x_0)^\top \bar{\boldsymbol{\beta}}(\tau)\Big)^2 \\
&= \frac{n}{S_N - 1}\sum_{s=1}^{S_N}\Big(L_n^s(\tau) - S_N^{-1}\sum_{s=1}^{S} L_n^s(\tau)\Big)^2 + o_P(1) \\
&= \frac{1}{S_N - 1}\sum_{s=1}^{S_N}(\sqrt{n}L_n^s(\tau))^2 - \frac{1}{S_N(S_N-1)}\Big(\sum_{s=1}^{S_N}\sqrt{n}L_n^s(\tau)\Big)^2 + o_P(1) \\
&=: \widehat{\sigma}_{L,N}^2 + o_P(\|\mathbf{Z}(x_0)\|^2)
\end{aligned}$$

where the $o_P(1)$ term is for $N \to \infty$ for arbitrary sequences S_N. Noting that $\mathbb{E}[(\sqrt{n}L_n^s(\tau))^4/\sigma_N^4]$ is bounded uniformly in n, s and that $\mathbb{E}[L_n^s(\tau)] = 0$, $\mathbb{E}[(\sqrt{n}L_n^s(\tau))^2] = \sigma_N^2$, an application of the Markov inequality shows that there exists a constant C such that for any fixed $\varepsilon > 0$ and any $S_N > 2$

$$P\Big(\Big|\frac{\widehat{\sigma}_{L,N}^2}{\sigma_N^2} - 1\Big| \ge \varepsilon\Big) \le C S_N^{-1}\varepsilon^{-2}.$$

Next, observe that $|\widehat{\sigma}_{L,N}^2/\sigma_N^2 - 1| \ge |\widehat{\sigma}_{L,N}/\sigma_N - 1|$ and thus

$$\begin{aligned}
\Big|\frac{\widehat{\sigma} t_{S-1,1-\alpha/2}}{\sigma_N \Phi^{-1}(1-\alpha/2)} - 1\Big| &\le \Big|\frac{\widehat{\sigma}}{\sigma_N} - 1\Big|\frac{t_{1,1-\alpha/2}}{\Phi^{-1}(1-\alpha/2)} + \Big|\frac{t_{S-1,1-\alpha/2}}{\Phi^{-1}(1-\alpha/2)} - 1\Big| \\
&\le \Big|\frac{\widehat{\sigma}^2}{\sigma_N^2} - 1\Big|\frac{t_{1,1-\alpha/2}}{\Phi^{-1}(1-\alpha/2)} + \Big|\frac{t_{S-1,1-\alpha/2}}{\Phi^{-1}(1-\alpha/2)} - 1\Big|.
\end{aligned}$$

It follows that for any fixed $\varepsilon > 0$ and $\bar{S}$ with $|t_{\bar{S}-1,1-\alpha/2}/\Phi^{-1}(1-\alpha/2) - 1| < \varepsilon/2$ we have for some constant $\widetilde{C}$

$$\limsup_{N\to\infty: S_N \ge \bar{S}} P\Big(\Big|\frac{\widehat{\sigma} t_{S-1,1-\alpha/2}}{\sigma_N \Phi^{-1}(1-\alpha/2)} - 1\Big| \ge \varepsilon\Big) \le \widetilde{C}\bar{S}^{-1}\Big(\frac{t_{1,1-\alpha/2}}{\Phi^{-1}(1-\alpha/2)}\Big)^2 \varepsilon^{-2}$$

Finally, noting that $S^{1/2}\sigma_N^{-1}|W_N| \rightsquigarrow |\mathcal{N}(0,1)|$, we have

$$\limsup_{N\to\infty} P\Big(\Big|S^{1/2}\sigma_N^{-1}|W_N| - \Phi^{-1}(1-\alpha/2)\Big| \leq \varepsilon\Phi^{-1}(1-\alpha/2)\Big) \\ = P\Big(\Big||\mathcal{N}(0,1)| - \Phi^{-1}(1-\alpha/2)\Big| \leq \varepsilon\Phi^{-1}(1-\alpha/2)\Big)$$

and

$$\Big|P\Big(S^{1/2}\sigma_N^{-1}|W_N| \leq \Phi^{-1}(1-\alpha/2)\Big) - (1-\alpha)\Big| \to 0.$$

Combining all results so far, we obtain for any fixed $\varepsilon > 0$ and $\bar{S}$ with $|t_{\bar{S}-1,1-\alpha/2}/\Phi^{-1}(1-\alpha/2) - 1| < \varepsilon/2$

$$\limsup_{N\to\infty} \Big|p_N - (1-\alpha)\Big| \leq \widetilde{C}\bar{S}^{-1}\Big(\frac{t_{1,1-\alpha/2}}{\Phi^{-1}(1-\alpha/2)}\Big)^2\varepsilon^{-2} \\ + P\Big(\Big||\mathcal{N}(0,1)| - \Phi^{-1}(1-\alpha/2)\Big| \leq \varepsilon\Phi^{-1}(1-\alpha/2)\Big).$$

The right-hand side can be made arbitrarily small by the choice of fixed values for $\varepsilon, \bar{S}$, and thus $p_N \to (1-\alpha)$. This completes the proof of the first part of Theorem 4.1. The second part follows easily since $t_{S_N-1,1-\alpha/2} \to \Phi^{-1}(1-\alpha/2)$ if $S \to \infty$. □

S.4.2. Proof of Theorem 4.2. Since the arguments under the assumptions of Corollary 3.10 are very similar to the arguments under the assumptions of Corollary A.5, we will only give a proof in the latter case. Note also that Corollary 3.4 follows from Corollary A.5. Define

$$\mathbb{G}_N^{(1)}(\tau) := -\frac{1}{N}\sum_{s=1}^{S}(\omega_{s,1}-1)\mathbf{Z}(x_0)^\top J_m^{-1}(\tau)\sum_{i=1}^{n}\mathbf{Z}(X_{is})(\mathbf{1}\{Y_{is} \leq Q(X_{is};\tau)\}-\tau)$$

and

$$\mathbb{G}_N(\tau) := -\frac{1}{N}\sum_{s=1}^{S}\mathbf{Z}(x_0)^\top J_m^{-1}(\tau)\sum_{i=1}^{n}\mathbf{Z}(X_{is})(\mathbf{1}\{Y_{is} \leq Q(X_{is};\tau)\} - \tau).$$

Throughout the proof, let $C_\omega := 1+\sqrt{2}$ and note that $\omega_{s,b} \leq C_\omega$ almost surely. The proof will rely on the following two key steps:

Step 1: prove that for any fixed $B \in \mathbb{N}$

$$\text{(S.4.1)}\quad \frac{\sqrt{N}}{\|\mathbf{Z}(x_0)\|}(\mathbb{G}_N(\cdot), \mathbb{G}_N^{(1)}(\cdot), \ldots, \mathbb{G}_N^{(B)}(\cdot)) \rightsquigarrow (\mathbb{G}_{x_0}(\cdot), \mathbb{G}_{x_0}^{(1)}(\cdot), ..., \mathbb{G}_{x_0}^{(B)}(\cdot))$$

as elements in $(\ell^\infty(\mathcal{T}))^{B+1}$ where the processes $\mathbb{G}_{x_0}^{(1)}, ..., \mathbb{G}_{x_0}^{(B)}$ iid copies of the process $\mathbb{G}_{x_0}$ defined in (A.4) and that $\mathbb{G}_N^{(1)}(\cdot)$ is asymptotically uniformly equicontinuous in probability.

Step 2: prove that uniformly in $\tau \in \mathcal{T}_k$ we have for $j = 1, ..., B$

$$\mathbf{Z}(x_0)^\top(\overline{\boldsymbol{\beta}}^{(j)}(\tau) - \overline{\boldsymbol{\beta}}(\tau)) = \mathbb{G}_N^{(j)}(\tau) + o_P\Big(\frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big). \tag{S.4.2}$$

Given the results in step 1 and step 2, exactly the same arguments as those in the proof of Corollary A.5 (replacing (S.3.34) by (S.4.2) and (S.3.37) by uniform asymptotic equicontinuity of $\mathbb{G}_{x_0}^{(1)}$; note that $\mathbb{G}_{x_0}^{(1)}$ in the present setting corresponds to $\mathbf{Z}(x_0)^\top \boldsymbol{U}_N(\tau)$ in the proof of Corollary A.5) show that for $j = 1, ..., B$

$$\sup_{\tau\in\mathcal{T}} \Big|\mathbf{Z}(x_0)^\top(\widehat{\boldsymbol{\beta}}^{(j)}(\tau) - \widehat{\boldsymbol{\beta}}(\tau)) - \mathbb{G}_N^{(j)}(\tau)\Big| = o_P\Big(\frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big),$$

and together with Step 2, (S.4.1) and Lemma 3.1 from Bücher and Kojadinovic (2017) the assertion follows.

It remains to establish the claims in step 1 and step 2.

Proof of step 1 Finite-dimensional convergence follows by a simple application of the triangular array Lyapunov CLT with $\delta = 2$ after observing that $\frac{\sqrt{N}}{\|\mathbf{Z}(x_0)\|}\mathbb{G}_N^{(1)}(\tau) = \sum_{s=1}^S V_{N,s}^{(1)}(\tau)$ where

$$V_{N,s}^{(1)}(\tau) := \frac{(\omega_{s,1} - 1)}{\sqrt{N}\|\mathbf{Z}(x_0)\|}\mathbf{Z}(x_0)^\top J_m^{-1}(\tau)\sum_{i=1}^n \mathbf{Z}(X_{is})(\mathbf{1}\{Y_{is} \le Q(X_{is};\tau)\} - \tau)$$

are independent summands. Hence it suffices to establish asymptotic uniform equicontinuity, which entails tightness. The proof is based on similar ideas as the proof of Lemma A.3 in Chao et al. (2017), and we will only stress the main differences. Following the latter paper, define

$$U_N(\tau) := \frac{1}{S}\sum_{s=1}^S(\omega_{s,1} - 1)J_m(\tau)^{-1}\frac{1}{n}\sum_{s=1}^S \mathbf{Z}(X_{is})(\mathbf{1}\{Y_{is} \le Q(X_{is};\tau)\} - \tau)$$
$$\Delta_{is}(\tau,\eta) := \mathbf{1}\{Y_{is} \le Q(X_{is};\tau)\} - \mathbf{1}\{Y_{is} \le Q(X_{is};\eta)\} - (\tau - \eta).$$

Similarly to the proof of equations (A.34) and (A.35) in the latter paper we find that for any vector $\mathbf{u}_N \in \mathbb{R}^m$ with $\|\mathbf{u}_N\| = 1$

$$\mathbb{E}\Big[|(\omega_{s,1} - 1)\mathbf{u}_N^\top J_m(\tau)^{-1}\mathbf{Z}(X_{is})\Delta_{is}(\tau,\eta)|^4\Big] \lesssim \xi_m^2|\tau - \eta|$$

and

$$\mathbb{E}\Big[|(\omega_{s,1}-1)\mathbf{u}_N^\top(J_m^{-1}(\tau)-J_m^{-1}(\eta))\mathbf{Z}(X_{is})(\mathbf{1}\{Y_{is}\le Q(X_{is};\tau)\}-\tau)|^4\Big] \lesssim \xi_m^2|\tau-\eta|^2$$

where the constants on the right-hand side are independent of τ, η, n. Next, similar computations as on page 3307 (above the derivation of equation (A.37)) in Chao et al. (2017) yield for any $\tau, \eta \in \mathcal{T}$

$$\mathbb{E}\Big[|N^{1/2}\mathbf{u}_N^\top(U_N(\tau)-U_N(\eta))|^4\Big] \lesssim |\tau-\eta|^2 + \frac{\xi_m^2}{N}|\tau-\eta|. \tag{S.4.3}$$

In particular we have for $|\tau-\eta| \ge \xi_m^6/N^3$

$$\mathbb{E}\Big[|N^{1/2}\mathbf{u}_N^\top(U_N(\tau)-U_N(\eta))|^4\Big] \lesssim |\tau-\eta|^{4/3}. \tag{S.4.4}$$

Apply Lemma A.1 of Kley et al. (2016) with $T=\mathcal{T}$, $d(s,t)=|t-s|^{1/3}$, $\bar\eta_n = \xi_m^2/N$, $\Psi(x)=x^4$. Note that with the notation introduced in the latter reference $D(\varepsilon,d)\lesssim\varepsilon^{-3}$ and thus for any $v \ge \xi_m^2/N$

$$\sup_{|\tau-\eta|\le\delta}|N^{1/2}\mathbf{u}_N^\top(U_N(\tau)-U_N(\eta))| \le S_1(v) + 2\sup_{|\tau-\eta|^{1/3}\le\xi_m^2/N,\tau\in\widetilde{T}}|N^{1/2}\mathbf{u}_N^\top(U_N(\tau)-U_N(\eta))|$$

where the set $\widetilde{T}$ contains at most $O(\xi_m^{-6}N^3)$ points and the random variable $S_1(v)$ satisfies

$$\|S_1(v)\|_4 \lesssim \int_0^v \varepsilon^{-3/4}d\varepsilon + (\delta^{1/3}+2\xi_m^2/N)v^{-6/4}.$$

Later we will prove that

$$\sup_{|\tau-\eta|^{1/3}\le\xi_m^2/N,\tau\in\widetilde{T}}|N^{1/2}\mathbf{u}_N^\top(U_N(\tau)-U_N(\eta))| = o_P(1). \tag{S.4.5}$$

Given (S.4.5), it follows that for any $c, v > 0$

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\limsup_{N\to\infty} P\Big(\sup_{|\tau-\eta|\le\delta} |N^{1/2}\mathbf{u}_N^\top(U_N(\tau)-U_N(\eta))| \ge cN^{-1/2}\Big)\\
\le&\limsup_{N\to\infty} P\Big(\sup_{|\tau-\eta|^{1/3}\le\xi_m^2/N,\tau\in\widetilde{T}} |N^{1/2}\mathbf{u}_N^\top(U_N(\tau)-U_N(\eta))| \ge cN^{-1/2}/2\Big)\\
&+\lim_{\delta\downarrow 0}\limsup_{N\to\infty} P(S_1(v)\ge c/2)\\
\lesssim&\lim_{\delta\downarrow 0}\limsup_{N\to\infty}\frac{16}{c^4}\Big[\int_0^v \varepsilon^{-3/4}d\varepsilon + (\delta^{1/3}+2\xi_m^2/N)v^{-6/4}\Big]^4\\
=&\frac{16}{c^4}\Big[\int_0^v \varepsilon^{-3/4}d\varepsilon\Big]^4.
\end{aligned}$$

Since the right-hand side can be made arbitrarily small by choosing v small, it follows that

$$\lim_{\delta\downarrow 0}\limsup_{N\to\infty} P\Big(\sup_{|\tau-\eta|\le\delta} |N^{1/2}\mathbf{u}_N^\top(U_N(\tau)-U_N(\eta))| \ge c\Big) = 0,$$

and hence it remains to prove (S.4.5). To do so, consider the decomposition

$$\begin{aligned}
&\mathbf{u}_N^\top(U_N(\tau)-U_N(\eta))\\
=&\frac{1}{N}\mathbf{u}_N^\top(J_m^{-1}(\tau)-J_m^{-1}(\eta))\sum_{s,i}(\omega_{s,1}-1)\mathbf{Z}(X_{is})(\mathbf{1}\{Y_{is}\le Q(X_{is};\tau)\}-\tau)\\
&+\frac{1}{N}\mathbf{u}_N^\top J_m^{-1}(\eta)\sum_{s,i}(\omega_{s,1}-1)\mathbf{Z}(X_{is})\Delta_{is}(\tau,\eta).
\end{aligned}$$

Similarly as in the proof of Lemma A.3 in Chao et al. (2017) (in the calculation below (A.34) on page 3306) we obtain

$$\|J_m^{-1}(\tau)-J_m^{-1}(\eta)\|_{op} \lesssim |\tau-\eta|$$

where $\|\cdot\|_{op}$ denotes the operator norm, and thus for $|\tau-\eta|\le\xi_m^6 N^{-3}$

$$\begin{aligned}
&\Big|\frac{1}{N}\mathbf{u}_N^\top(J_m^{-1}(\tau)-J_m^{-1}(\eta))\sum_{s,i}(\omega_{s,1}-1)\mathbf{Z}(X_{is})(\mathbf{1}\{Y_{is}\le Q(X_{is};\tau)\}-\tau)\Big|\\
&\lesssim|\tau-\eta|\xi_m = O(\xi_m^7 N^{-3}) = o(N^{-1}) \quad a.s.
\end{aligned} \tag{S.4.6}$$

Next observe that for $\tau\le\eta$

$$|\Delta_{is}(\tau,\eta)| \le \mathbf{1}\{Q(X_{is};\tau)\le Y_{is}\le Q(X_{is};\eta)\}+|\tau-\eta|.$$

This implies that for any fixed τ we have a.s. for a constant C independent of τ, η, N

$$\begin{aligned}&\sup_{|\tau-\eta|\le\varepsilon_N}\Big|\mathbf{u}_N^\top J_m^{-1}(\eta)N^{-1}\sum_{s,i}(\omega_{s,1}-1)\mathbf{Z}(X_{is})\Delta_{is}(\tau,\eta)\Big|\\ \le& C\frac{\xi_m}{N}\sum_{i,s}\Big[\varepsilon_N+\mathbf{1}\{Q(X_{is};\tau-\varepsilon_N)\le Y_{is}\le Q(X_{is};\tau+\varepsilon_N)\}\Big].\end{aligned}$$

Thus

$$\begin{aligned}&\sup_{\tau\in\widetilde{T}}\sup_{|\tau-\eta|\le\xi_m^6N^{-3}}\Big|\mathbf{u}_N^\top J_m^{-1}(\eta)N^{-1/2}\sum_{s,i}(\omega_{s,1}-1)\mathbf{Z}(X_{is})\Delta_{is}(\tau,\eta)\Big|\\ \lesssim&\frac{\varepsilon_N\xi_m}{N}+\xi_m\sup_{\tau\in\widetilde{T}}B_N(\tau,\varepsilon_N)\end{aligned}$$

where

$$B_N(\tau,\varepsilon_N):=\sum_{i,s}\mathbf{1}\{Q(X_{is};\tau-\varepsilon_N)\le Y_{is}\le Q(X_{is};\tau+\varepsilon_N)\}.$$

Since $B_N(\tau,\varepsilon_N)\sim Bin(N,2\varepsilon_N)$ distribution and by applying the multiplicative Chernoff bound for Binomial random variables we find that for any $t\ge 1$

$$P\Big(B_N(\tau,\varepsilon_N)>N(1+t)\varepsilon_N\Big)\le\exp(-2tN\varepsilon_N/3).$$

Plugging in $\varepsilon_N=\xi_m^6N^{-3}$ and $t=3AN^2\log N/(2\xi_m^6)\gg 1$ we obtain

$$P\Big(B_N(\tau,2\xi_m^6N^{-3})>2\xi_m^6N^{-2}+3A\log N\Big)\le N^{-A}.$$

By choosing $A=3$ and applying the union bound for probabilities we finally obtain for N sufficiently large and a constant $0<c<\infty$

$$\begin{aligned}&P\Big(\sup_{\tau\in\widetilde{T}}\sup_{|\tau-\eta|\le\xi_m^6N^{-3}}\Big|\mathbf{u}_N^\top J_m^{-1}(\eta)\sum_{s,i}(\omega_{s,1}-1)\mathbf{Z}(X_{is})\Delta_{is}(\tau,\eta)\Big|\ge c\xi_m\log N\Big)\\ \le&|\widetilde{T}|N^{-3}=o(1).\end{aligned}$$

Combining this with (S.4.6), (S.4.5) follows.

Proof of step 2 Observe that

$$\begin{aligned}
&\mathbf{Z}(x_0)^\top(\overline{\boldsymbol{\beta}}^{(1)}(\tau)-\overline{\boldsymbol{\beta}}(\tau))\\
=\;&\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}\Big(\frac{\omega_{s,1}}{\bar\omega_{\cdot,1}}-1\Big)\widehat{\boldsymbol{\beta}}^s(\tau)\\
=\;&\frac{1}{\bar\omega_{\cdot,1}}\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}\Big(\omega_{s,1}-\bar\omega_{\cdot,1}\Big)\Big(\widehat{\boldsymbol{\beta}}^s(\tau)-\boldsymbol{\beta}_N(\tau)\Big)\\
=\;&\frac{1}{\bar\omega_{\cdot,1}}\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}\Big(\omega_{s,1}-1\Big)\Big(\widehat{\boldsymbol{\beta}}^s(\tau)-\boldsymbol{\beta}_N(\tau)\Big)\\
&-\frac{\bar\omega_{\cdot,1}-1}{\bar\omega_{\cdot,1}}\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}\Big(\widehat{\boldsymbol{\beta}}^s(\tau)-\boldsymbol{\beta}_N(\tau)\Big)\\
=\;&\frac{1}{\bar\omega_{\cdot,1}}\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}\Big(\omega_{s,1}-1\Big)\Big(\widehat{\boldsymbol{\beta}}^s(\tau)-\boldsymbol{\beta}_N(\tau)\Big)+o_P\Big(\frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big).
\end{aligned}$$

Note that under the assumptions on the weights $\bar\omega_{\cdot,1}=1+o_P(1)$. Thus (S.4.2) will follow from step 1 once we prove that, uniformly in $\tau\in\mathcal{T}_K$,

$$\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}\Big(\omega_{s,1}-1\Big)\Big(\widehat{\boldsymbol{\beta}}^s(\tau)-\boldsymbol{\beta}_N(\tau)\Big)=\mathbb{G}_N^{(1)}(\tau)+o_P\Big(\frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big). \tag{S.4.7}$$

To this end apply Theorem S.6.1, which implies that

$$\begin{aligned}
&\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}(\omega_{s,1}-1)\Big(\widehat{\boldsymbol{\beta}}^s(\tau)-\boldsymbol{\beta}_N(\tau)\Big)-\widetilde{\mathbb{G}}_N^{(1)}(\tau)\\
&\qquad=\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}(\omega_{s,1}-1)\Big(r^s_{n,1}(\tau)+r^s_{n,2}(\tau)+r^s_{n,3}(\tau)+r^s_{n,4}(\tau)\Big)
\end{aligned}$$

where

$$\widetilde{\mathbb{G}}_N^{(1)}(\tau):=-\frac{1}{N}\sum_{s=1}^{S}(\omega_{s,1}-1)\mathbf{Z}(x_0)^\top J_m^{-1}(\tau)\sum_{i=1}^{n}\psi(Y_{is},\mathbf{Z}(X_{is});\boldsymbol{\beta}_N(\tau),\tau)$$

and the remainder terms $r^s_{n,j}(\tau)$ correspond to $r_{n,j}(\tau)$ defined in Theo-

rem S.6.1. We will prove that

(S.4.8)
$$\sup_{\tau\in\mathcal{T}_K}\left|\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}(\omega_{s,1}-1)\Big(r^s_{n,1}(\tau)+r^s_{n,2}(\tau)+r^s_{n,3}(\tau)+r^s_{n,4}(\tau)\Big)\right| = o_P(\|\mathbf{Z}(x_0)\|N^{-1/2}),$$

(S.4.9)
$$\sup_{\tau\in\mathcal{T}_K}\left|\widetilde{\mathbb{G}}^{(1)}_N(\tau)-\mathbb{G}^{(1)}_N(\tau)\right| = o_P(\|\mathbf{Z}(x_0)\|N^{-1/2}).$$

We begin by a proof of (S.4.8). Applying Theorem S.6.1 with $\kappa_n = A\log n$ and A sufficiently large we find that

$$\begin{aligned}&\left|\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}(\omega_{s,1}-1)\Big(r^s_{n,1}(\tau)+r^s_{n,2}(\tau)+r^s_{n,4}(\tau)\Big)\right|\\ \le&\sup_s\sup_{\tau\in\mathcal{T}}\left|\mathbf{Z}(x_0)^\top(r^s_{n,1}(\tau)+r^s_{n,2}(\tau)+r^s_{n,4}(\tau))\right|\frac{1}{S}\sum_{j=1}^{S}|\omega_{j,1}-1|\\ =&o_P\Big(\frac{\|\mathbf{Z}(x_0)\|}{N^{1/2}}\Big).\end{aligned}$$

Next, note that by independence between $\omega_{s,1}$ and the sample we have $\mathbb{E}[(\omega_{s,1}-1)r^s_{n,3}(\tau)]=0$ for all s,τ. Let

$$D_n(A) := C\Re_3(A\log n)$$

where $\Re_3(\cdot)$ is defined in (S.6.5), and C is the constant appearing in (S.6.3). Consider the decomposition

$$\begin{aligned}(\omega_{s,1}-1)\mathbf{Z}(x_0)^\top r^s_{n,3}(\tau) =&\mathbf{Z}(x_0)^\top(\omega_{s,1}-1)r^s_{n,3}(\tau)\mathbf{1}\{|r^s_{n,3}(\tau)|\le D_n(A)C_\omega\}\\ &+\mathbf{Z}(x_0)^\top(\omega_{s,1}-1)r^s_{n,3}(\tau)\mathbf{1}\{|r^s_{n,3}(\tau)|> D_n(A)C_\omega\}\\ =:&R^{(s)}_{n,1}(\tau)+R^{(s)}_{n,2}(\tau)\end{aligned}$$

By (S.6.3), (S.6.8) in Theorem S.6.1, we see that

$$P\Big(\sup_{s=1,\dots,S}\sup_\tau\|R^{(s)}_{n,2}(\tau)\|\neq 0\Big)\le 2Sn^{-A}. \tag{S.4.10}$$

Moreover, by independence of the weights $\omega_{s,1}$ from the sample $\mathbb{E}[R^{(s)}_{n,1}(\tau)]=0$ and by construction $|R^{(s)}_{n,1}(\tau)|\le\|\mathbf{Z}(x_0)\|C_\omega D_N(A)$ a.s. Apply Hoeffding's

inequality to obtain

$$P\Big(\Big|\frac{1}{S}\sum_{s=1}^{S} R_{n,1}^{(s)}(\tau)\Big| \geq t\|\mathbf{Z}(x_0)\|\Big) \leq 2\exp\Big(-\frac{2St^2}{4C_\omega^2 D_N^2(A)}\Big)$$

Letting $t = 4\sqrt{\log n}S^{-1/2}D_N(6)C_\omega$ and combining this with (S.4.10) shows that

$$\begin{aligned}&P\Big(\sup_{\tau\in\mathcal{T}_K}\Big|\mathbf{Z}(x_0)^\top\frac{1}{S}\sum_{s=1}^{S}(\omega_{s,1}-1)r_{n,3}^s(\tau)\Big| \geq 4\|\mathbf{Z}(x_0)\|\sqrt{\log n}S^{-1/2}D_N(6)C_\omega\Big)\\ \leq& N^2\Big(2n^{-8}+2Sn^{-6}\Big) = o(1).\end{aligned}$$

Straightforward but tedious computations show that

$$\sqrt{\log n}S^{-1/2}D_N(6) = o(N^{-1/2}),$$

and thus the proof of (S.4.8) is complete. Next we prove (S.4.9). Note that

$$\begin{aligned}&\Big|\widetilde{\mathbb{G}}_N^{(1)}(\tau) - \mathbb{G}_N^{(1)}(\tau)\Big|\\ \leq\ & C_\omega\frac{1}{N}\sum_{i,s}|\mathbf{Z}(x_0)^\top J_m^{-1}(\tau)\mathbf{Z}(X_{is})|\mathbf{1}\{|Y_{is}-Q(X_{is};\tau)| \leq c_m(\boldsymbol{\gamma}_N)\}\\ =:\ & C_\omega\|\mathbf{Z}(x_0)\|\frac{1}{N}\sum_{i,s}W_{is}(\tau).\end{aligned}$$

We have

$$\sup_{\tau\in\mathcal{T}}\mathbb{E}[W_{is}(\tau)] = O(c_m(\boldsymbol{\gamma}_N)) = o(N^{-1/2}\|\mathbf{Z}(x_0)\|)$$

and $\sup_{\tau\in\mathcal{T}}|W_{is}(\tau)| \leq C\xi_m$ almost surely for some constant $C < \infty$. Moreover, the $W_{is}(\tau)$ are iid across i, s and

$$\sup_{\tau\in\mathcal{T}}\mathrm{Var}(W_{11}(\tau)) = O(c_N(\boldsymbol{\gamma}_N)) = O(\|\mathbf{Z}(x_0)\|N^{-1/2}) = o(N^{-1/4}).$$

Hence Bernstein's inequality and the union bound for probabilities yields

$$\begin{aligned}&P\Big(\sup_{\tau\in\mathcal{T}_K}\Big|\sum_{i,s}W_{is}(\tau) - \mathbb{E}[W_{is}(\tau)]\Big| \geq N^{1/2}(\log N)^{-1}\Big)\\ \leq& 2|\mathcal{T}_K|\exp\Big(-\frac{N(\log N)^{-2}/2}{N\sup_{\tau\in\mathcal{T}}\mathrm{Var}(W_{11}(\tau)) + C\xi_m N^{1/2}(\log N)^{-1}/3}\Big) = o(1).\end{aligned}$$

Thus the proof of (S.4.9) is complete.

□

S.4.3. Proof of Theorem 4.4. Define the class of functions

$$\mathcal{G}_3(\delta) := \Big\{(z,y) \mapsto \|zz^\top\| \mathbf{1}\Big\{|y - z^\top \mathbf{b} + t| \le \delta\Big\} \mathbf{1}\{\|z\| \le \xi\} \Big| t \in \mathbb{R}, \mathbf{b} \in \mathbb{R}^m\Big\}.$$

Let

$$\widetilde{\boldsymbol{\beta}}^{s,-i}(\tau) := \boldsymbol{\beta}(\tau) - (n-1)^{-1/2} J_m(\tau)^{-1} \mathbb{G}_n^{s,-i} \psi(\cdot; \widehat{\boldsymbol{\beta}}^{s,-i}(\tau), \tau)$$

where $\mathbb{G}_n^{s,-i}$ is the empirical process corresponding to sub-sample s with i'th observation removed and

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

By Lemma S.6.3 we obtain for $A \ge 1$ and some constant C

$$P\Big(\sup_{s,\tau,j} \|\widehat{\boldsymbol{\beta}}^{s,-j}(\tau) - \widehat{\boldsymbol{\beta}}^{s}(\tau)\| \ge CA \frac{\log n}{n}\Big) \le 2Sn^{1-A}.$$

Following the arguments in the proof of Lemma S.6.3 we also find that for some constant C_1 we have almost surely

$$\Big\|\widetilde{\boldsymbol{\beta}}^{s,-j}(\tau) - \widehat{\boldsymbol{\beta}}^{s,-j}(\tau)\Big\| \le C_1\Big(\frac{1}{n} + \|\widehat{\boldsymbol{\beta}}^{s,-j}(\tau) - \boldsymbol{\beta}(\tau)\|^2\Big),$$

so that by (S.6.18) there exists a constant C_2 such that

$$P\Big(\sup_{s,\tau,j} \|\widetilde{\boldsymbol{\beta}}^{s,-j}(\tau) - \widehat{\boldsymbol{\beta}}^{s}(\tau)\| \ge C_2 A \frac{\log n}{n}\Big) \le Sn^{1-A}. \tag{S.4.11}$$

Moreover, the arguments in the proof of Theorem S.6.1 show that

$$\begin{aligned} &\sup_{s,i,\tau} \Big\|\mathbb{E}[\widetilde{\boldsymbol{\beta}}^{s,-i}(\tau)] - \boldsymbol{\beta}(\tau)\Big\| \\ &= \sup_{s,i,\tau} \Big\|\mathbb{E}\Big[n^{-1/2} J_m(\tau)^{-1} \mathbb{G}_n^{s,-i} \psi(\cdot; \widehat{\boldsymbol{\beta}}^{s,-i}(\tau), \tau)\Big]\Big\| = O\Big(\frac{\log n}{n}\Big). \end{aligned} \tag{S.4.12}$$

Define the events

$$\Omega_{1,n}^{s,j} := \Big\{\|\widetilde{\boldsymbol{\beta}}^{s,-j}(\tau) - \widehat{\boldsymbol{\beta}}^{s}(\tau)\| \le C_2 4 \frac{\log n}{n}\Big\}, \quad \Omega_{1,n} := \cap_{s,j} \Omega_{1,n}^{s,j}.$$

and observe that for $\delta_n := 4C_2\frac{\log n}{n}$

$$\begin{aligned}
&\frac{1}{2nh_n}\Big\|\sum_{i=1}^n \mathbf{Z}_{is}\mathbf{Z}_{is}^\top\Big[\mathbf{1}\Big\{|Y_{is}-\mathbf{Z}_{is}^\top\widehat{\boldsymbol{\beta}}^s(\tau)|\le h_n\Big\}-\mathbf{1}\Big\{|Y_{is}-\mathbf{Z}_{is}^\top\widetilde{\boldsymbol{\beta}}^{s,-i}(\tau)|\le h_n\Big\}\Big]\Big\| \\
\le\ & \frac{1}{h_n}\sup_{\mathbf{b}\in\mathbb{R}^m,t\in\mathbb{R}}\frac{1}{n}\sum_{i=1}^n\|\mathbf{Z}_{is}\mathbf{Z}_{is}^\top\|\mathbf{1}\Big\{|Y_{is}-\mathbf{Z}_{is}^\top\mathbf{b}-t|\le\xi\delta_n\Big\}+\frac{\xi^2}{h_n}\mathbf{1}_{\Omega_{1,n}^C} \\
\le\ & \frac{1}{h_n}\sup_{\mathbf{b}\in\mathbb{R}^m,t\in\mathbb{R}}\mathbb{E}\Big[\|\mathbf{Z}_{1s}\mathbf{Z}_{1s}^\top\|\mathbf{1}\Big\{|Y_{1s}-\mathbf{Z}_{1s}^\top\mathbf{b}-t|\le\xi\delta_n\Big\}\Big] \\
&+\frac{1}{h_n}\|\mathbb{P}_{n,s}-P\|_{\mathcal{G}_3(\delta_n)}+\frac{\xi^2}{h_n}\mathbf{1}_{\Omega_{1,n}^C} \\
=\ & \frac{1}{h_n}\|\mathbb{P}_{n,s}-P\|_{\mathcal{G}_3(\xi\delta_n)}+O(h_n^{-1}\delta_n)+\frac{\xi^2}{h_n}\mathbf{1}_{\Omega_{1,n}^C}.
\end{aligned}$$

Standard arguments show that $\mathcal{G}_3(\xi\delta_n)$ satisfies (S.8.1) with fixed constant V and thus by (S.8.2) we have

$$\sup_s\mathbb{E}\|\mathbb{P}_{n,s}-P\|_{\mathcal{G}_3(\xi\delta_n)}=O\Big(\frac{\log n}{n}\Big).$$

This implies

$$\sup_\tau\Big\|\overline{J}_m^P(\tau)-\frac{1}{S}\sum_{s=1}^S\frac{1}{2nh_n}\sum_{i=1}^n\mathbf{Z}_{is}\mathbf{Z}_{is}^\top\mathbf{1}\Big\{|Y_{is}-\mathbf{Z}_{is}^\top\widetilde{\boldsymbol{\beta}}^{s,-i}(\tau)|\le h_n\Big\}\Big\|=O_P\Big(\frac{\log n}{nh_n}\Big).$$

Next note that by a Taylor expansion we have

$$F_{Y|X}(a+h_n|x)-F_{Y|X}(a-h_n|x)=2h_nf_{Y|X}(a|x)+\frac{h_n^3}{3}f''_{Y|X}(a|x)+o(h_n^3)$$

where the remainder term is uniform in $a \in \mathbb{R}$. Thus we have by (S.4.12)

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{2h_n}\mathbf{Z}_{1s}\mathbf{Z}_{1s}^\top \mathbf{1}\Big\{|Y_{1s} - \mathbf{Z}_{1s}^\top \widetilde{\boldsymbol{\beta}}^{s,-1}(\tau)| \le h_n\Big\}\Big] \\
= &\int \frac{1}{2h_n}\mathbf{Z}(x)\mathbf{Z}(x)^\top \mathbb{E}\Big[\Big(F_{Y|X}(\mathbf{Z}(x)^\top \widetilde{\boldsymbol{\beta}}^{s,-1}(\tau) + h_n|x) \\
&\qquad - F_{Y|X}(\mathbf{Z}(x)^\top \widetilde{\boldsymbol{\beta}}^{s,-1}(\tau) - h_n|x)\Big)\Big] dP^X(x) \\
= &\int \mathbf{Z}(x)\mathbf{Z}(x)^\top \mathbb{E}\Big[\Big(f_{Y|X}(\mathbf{Z}(x)^\top \widetilde{\boldsymbol{\beta}}^{s,-1}(\tau)|x) \\
&\qquad + \frac{h_n^2}{6} f''_{Y|X}(\mathbf{Z}(x)^\top \widetilde{\boldsymbol{\beta}}^{s,-1}(\tau)|x)\Big)\Big] dP^X(x) + o(h_n^2) \\
= &\int \mathbf{Z}(x)\mathbf{Z}(x)^\top \Big[\Big(f_{Y|X}(\mathbf{Z}(x)^\top \boldsymbol{\beta}(\tau)|x) + \frac{h_n^2}{6} f''_{Y|X}(\mathbf{Z}(x)^\top \boldsymbol{\beta}(\tau)|x)\Big)\Big] dP^X(x) \\
&+ O\Big(\frac{\log n}{n}\Big) + o(h_n^2).
\end{aligned}$$

Similar but simpler computations show that for

$$\mathbf{W}_s(\mathbf{b}) := \mathrm{vec}\Big(\frac{1}{2h_n}\mathbf{Z}_{1s}\mathbf{Z}_{1s}^\top \mathbf{1}\Big\{|Y_{1s} - \mathbf{Z}_{1s}^\top \mathbf{b}| \le h_n\Big\}\Big)$$

we have uniformly in s

$$\mathrm{Var}(\mathbf{W}_s(\widetilde{\boldsymbol{\beta}}^{s,-1}(\tau))) = \mathrm{Var}(\mathbf{W}_s(\boldsymbol{\beta}(\tau))) + o(h_n^{-1}).$$

This completes the proof. □

APPENDIX S.5: PROOFS FOR APPENDIX S.2

S.5.1. Proof of Theorem S.2.1. From Theorem S.6.1 we obtain the representation

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^s(\tau) - \boldsymbol{\beta}_N(\tau) = &-n^{-1/2} J_m(\tau)^{-1} \mathbb{G}_n^s(\psi(\cdot; \boldsymbol{\beta}_N(\tau), \tau)) \\
&+ r_{n,1}^s(\tau) + r_{n,2}^s(\tau) + r_{n,3}^s(\tau) + r_{n,4}^s(\tau).
\end{aligned}$$

where the quantities $r_{n,j}^1, ..., r_{n,j}^S$ are i.i.d. for each $j = 1, ..., 4$ and

$$\mathbb{G}_n^s(\psi(\cdot; \boldsymbol{\beta}_N(\tau), \tau)) = n^{-1/2} \sum_{i=1}^n \psi(Y_{is}, \mathbf{Z}(X_{is}); \boldsymbol{\beta}_N(\tau), \tau).$$

Letting

$$r_N^{(1)}(\tau) := \frac{1}{S}\sum_{s=1}^{S}\Big(r_{n,1}^s(\tau) + r_{n,2}^s(\tau) + r_{n,4}^s(\tau)\Big)$$

we find that the bound in (S.2.3) is a direct consequence of Theorem S.6.1, the union probability bound and the observation that $\xi_m g_N^2 = o(g_N)$ by the assumption $g_N = o(\xi_m^{-1})$.

Next, let

$$r_N^{(2)}(\tau) := \frac{1}{S}\sum_{s=1}^{S} r_{n,3}^s(\tau).$$

Define $\widetilde{r}_{n,3}^s(\tau) := r_{n,3}^s(\tau) - \mathbb{E}[r_{n,3}^s(\tau)]$ and consider the decomposition

$$\begin{aligned} r_{n,3}^s(\tau) &= \widetilde{r}_{n,3}^s \mathbf{1}\big\{\|\widetilde{r}_{n,3}^s(\tau)\| \le D_n(A)\big\} \\ &\quad + \widetilde{r}_{n,3}^s(\tau)\mathbf{1}\big\{\|\widetilde{r}_{n,3}^s(\tau)\| > D_n(A)\big\} + \mathbb{E}[r_{n,3}^s(\tau)] =: \sum_{j=1}^{3} R_{n,j}^{(s)}(\tau) \end{aligned}$$

where, for a sufficiently large constant C,

$$D_n(A) := C\Re_3(A\log n) + C_{3,2}\xi_m\Big(g_N^2\xi_m + \frac{m\xi_m\log n}{n}\Big),$$

$\Re_3(\cdot)$ is defined in (S.6.5) and $C_{3,2}$ denotes the constant appearing in (S.6.8). By (S.6.3), (S.6.8) in Theorem S.6.1, we see that

$$P\Big(\sup_{s=1,\dots,S}\sup_{\tau}\|R_{n,2}^{(s)}(\tau)\| \ne 0\Big) \le 2Sn^{-A}; \tag{S.5.1}$$

and by (S.6.8) in Theorem S.6.1,

$$\sup_{\tau}\|R_{n,3}^{(s)}(\tau)\| \le C_{3,2}\xi_m\Big(g_N^2\xi_m + \frac{m\xi_m\log n}{n}\Big). \tag{S.5.2}$$

Next we deal with $R_{n,1}^{(s)}(\tau)$. Consider the function $g : (\mathbb{R}^m)^S \to \mathbb{R}$,

$$g(x_1,\dots,x_S) := \Big\|\frac{1}{S}\sum_{s=1}^{S} x_s\Big\|.$$

Direct computations show that

$$|g(x_1,\dots,x_S) - g(x_1,\dots,x_i',\dots,x_S)| \le \frac{\|x_i' - x_i\|}{S}.$$

Note that by construction $\|R_{n,1}^{(s)}(\tau)\| \leq D_n(A)$ almost surely. Apply McDiarmid's inequality (Lemma 1.2 on page 149 in McDiarmid (1989)) to $g(R_{n,1}^{(1)}(\tau), ..., R_{n,1}^{(S)}(\tau))$ to obtain for all $t > 0$

$$\sup_\tau P\Big(\Big\|\frac{1}{S}\sum_{s=1}^{S} R_{n,1}^{(s)}(\tau)\Big\| \geq t\Big) \leq 2\exp\Big(\frac{-St^2}{2D_n^2(A)}\Big)$$

Letting $t = \sqrt{2}\kappa_n S^{-1/2} D_n(A)$ and combining this with (S.5.1) and (S.5.2) shows that

$$\sup_\tau P\Big(\big\|r_N^{(2)}(\tau)\big\| > \sqrt{2}\kappa_n S^{-1/2} D_n(A) + C_{3,2}\xi_m\Big(g_N^2\xi_m + \frac{m\xi_m \log n}{n}\Big)\Big) \leq 2Sn^{-A} + 2e^{-\kappa_n^2}.$$

Finally, from an elementary but tedious calculation and the definition of $\Re_3(A\log n)$ in (S.6.5), we obtain

$$\kappa_n S^{-1/2}\Re_3(A\log n) \lesssim A\frac{\kappa_n}{S^{1/2}}\Big(\frac{m}{n}\Big)^{1/2}\Big((\xi_m g_N \log n)^{1/2} + \Big(\frac{m\xi_m^2(\log n)^3}{n}\Big)^{1/4}\Big).$$

Thus the proof of Theorem S.2.1 is complete. □

S.5.2. Proof of Theorem S.2.4. The proof proceeds similarly to the proof of Theorem S.2.1. From Theorem S.6.2 we obtain the representation

$$\begin{aligned}
&\mathbf{u}_N^\top(\widehat{\boldsymbol{\beta}}^s(\tau) - \boldsymbol{\gamma}_N(\tau)) \\
&= -n^{-1}\mathbf{u}_N^\top J_m(\boldsymbol{\gamma}_N(\tau))^{-1}\sum_{i=1}^{n}\mathbf{Z}_i^s(\mathbf{1}\{Y_i^s \leq (\mathbf{Z}_i^s)^\top\boldsymbol{\gamma}_N(\tau)\} - \tau) + \mathbf{u}_N^\top\sum_{k=1}^{4} r_{n,k}^s(\tau) \\
&\quad + \mathbf{u}_N^\top J_m(\boldsymbol{\gamma}_N(\tau))^{-1}\mathbb{E}[\mathbf{Z}(F_{Y|X}(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau)|X) - \tau)].
\end{aligned}$$

where the quantities $r_{n,j}^1(\tau), ..., r_{n,j}^S(\tau)$ are i.i.d. for each $j = 1, ..., 4$ and

$$\mathbb{G}_n^s(\psi(\cdot;\boldsymbol{\gamma}_N(\tau),\tau)) = n^{-1/2}\sum_{i=1}^{n}\psi(Y_{is}, \mathbf{Z}(X_{is}); \boldsymbol{\gamma}_N(\tau), \tau).$$

Next, note that

$$\begin{aligned}
\mathbf{Z}(F_{Y|X}(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau)|X) - \tau) &= \mathbf{Z}f_{Y|X}(Q(X;\tau)|X)(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau) - Q(X;\tau)) \\
&\quad + \frac{1}{2}\mathbf{Z}f'_{Y|X}(\zeta_n(X,\tau)|X)(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau) - Q(X;\tau))^2
\end{aligned}$$

where $\zeta_n(X,\tau)$ is a value between $\mathbf{Z}(X)^\top\boldsymbol{\gamma}_N(\tau)$ and $Q(X;\tau)$. From the definition of $\boldsymbol{\gamma}_N$ as (3.8), we obtain $\mathbb{E}[\mathbf{Z}f_{Y|X}(Q(X;\tau)|X)(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau)-Q(X;\tau))] = 0$ so that

$$\sup_{\tau\in\mathcal{T}}\left|\mathbf{u}_N^\top\widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1}\mathbb{E}[\mathbf{Z}(F_{Y|X}(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau)|X)-\tau)]\right| \lesssim \widetilde{\mathcal{E}}(\mathbf{u}_N,\boldsymbol{\gamma}_N)c_n^2(\boldsymbol{\gamma}_N).$$

Letting

$$r_N^{(1)}(\tau,\mathbf{u}_N) := \Big|\frac{1}{S}\mathbf{u}_N^\top\sum_{s=1}^{S}\Big(r_{n,1}^s(\tau)+r_{n,2}^s(\tau)+r_{n,4}^s(\tau)\Big) + \mathbf{u}_N^\top J_m(\boldsymbol{\gamma}_N(\tau))^{-1}\mathbb{E}[\mathbf{Z}(F_{Y|X}(\mathbf{Z}^\top\boldsymbol{\gamma}_N(\tau)|X)-\tau)]\Big|$$

we find that the bound on $r_N^{(1)}(\tau,\mathbf{u}_N)$ is a direct consequence of Theorem S.6.2. Next, let

$$r_N^{(2)}(\tau,\mathbf{u}_N) := \frac{1}{S}\sum_{s=1}^{S}\mathbf{u}_N^\top r_{n,3}^s(\tau).$$

Define $\widetilde{r}_{n,3}^s(\tau) := r_{n,3}^s(\tau) - \mathbb{E}[r_{n,3}^s(\tau)]$ and consider the decomposition

$$\begin{aligned} r_{n,3}^s(\tau) =\ & \widetilde{r}_{n,3}^s\mathbf{1}\Big\{\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}|\mathbf{u}_N^\top\widetilde{r}_{n,3}^s(\tau)|\le D_n^{(L)}(A)\Big\} \\ & + \widetilde{r}_{n,3}^s(\tau)\mathbf{1}\Big\{\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}|\mathbf{u}_N^\top\widetilde{r}_{n,3}^s(\tau)|> D_n^{(L)}(A)\Big\} + \mathbb{E}[r_{n,3}^1(\tau)] \\ =:\ & \sum_{j=1}^{3}R_{n,j}^{(s)}(\tau) \end{aligned}$$

where for a constant $A>0$, and a sufficiently large constant C

$$D_n^{(L)}(A) := C\Re_3^{(L)}(A\log n)+C_{3,2}\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}\widetilde{\mathcal{E}}(\mathbf{u}_N,\boldsymbol{\gamma}_N)\Big(c_m^4(\boldsymbol{\gamma}_N)+\frac{\xi_m^2(\log n)^2}{n}\Big),$$

$\Re_3^{(L)}$ is defined in (S.6.13), $C_{3,2}$ denotes the constant appearing in (S.6.15).

By (S.6.11) in Theorem S.6.2,

$$P\Big(\sup_{s=1,\dots,S}\sup_{\tau}\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}|\mathbf{u}_N^\top R_{n,2}^{(s)}(\tau)|\neq 0\Big)\le 2Sn^{-A}. \tag{S.5.3}$$

Applying (S.6.15) in Theorem S.6.2 yields

(S.5.4)

$$\sup_{\tau}\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}|\mathbf{u}_N^\top R_{n,3}^{(s)}(\tau)|\leq C_{3,2}\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}\widetilde{\mathcal{E}}(\mathbf{u}_N,\boldsymbol{\gamma}_N)\Big(c_m^4(\boldsymbol{\gamma}_N)+\frac{\xi_m^2(\log n)^2}{n}\Big).$$

Next consider he function $g:(\mathbb{R}^m)^S\to\mathbb{R}$,

$$g(x_1,...,x_S):=\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}\Big|\mathbf{u}_N^\top\frac{1}{S}\sum_{s=1}^S x_s\Big|=\Big\|\Big(\frac{1}{S}\sum_{s=1}^S x_s\Big)^{(\mathcal{I})}\Big\|.$$

The reverse triangle inequality shows that

$$|g(x_1,...,x_S)-g(x_1,...,x_i',...,x_S)|\leq\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}\frac{|\mathbf{u}_N^\top(x_i'-x_i)|}{S}.$$

Note that by construction $\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}|\mathbf{u}_N^\top R_{n,1}^{(s)}(\tau)|\leq D_n^{(L)}(A)$ almost surely. Apply McDiarmid's inequality (Lemma 1.2 on page 149 in McDiarmid (1989)) to $g(R_{n,1}^{(1)}(\tau),...,R_{n,1}^{(S)}(\tau))$ to obtain for all $t>0$

$$\sup_\tau P\Big(\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}\Big|\frac{1}{S}\mathbf{u}_N^\top\sum_{s=1}^S R_{n,1}^{(s)}(\tau)\Big|\geq t\Big)\leq 2\exp\Big(\frac{-St^2}{2(D_n^{(L)}(A))^2}\Big).$$

Letting $t=\sqrt{2}\kappa_n S^{-1/2}D_n^{(L)}(A)$ and combining this with (S.5.3)-(S.5.4) shows that

$$\begin{aligned}\sup_\tau P\Big(\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}\Big|r_N^{(2)}(\tau,\mathbf{u}_N)\Big|>\sqrt{2}\kappa_n S^{-1/2}D_n^{(L)}(A)\\ +C_{3,2}\sup_{\mathbf{u}_N\in\mathcal{S}_{\mathcal{I}}^{m-1}}\widetilde{\mathcal{E}}(\mathbf{u}_N,\boldsymbol{\gamma}_N)\Big(c_m^4(\boldsymbol{\gamma}_N)+\frac{\xi_m^2(\log n)^2}{n}\Big)\Big)\\ \leq 2Sn^{-A}+2e^{-\kappa_n^2}.\end{aligned}$$

Thus the proof of Theorem S.2.4 is complete. □

APPENDIX S.6: REFINED BAHADUR REPRESENTATIONS FOR SUB-SAMPLES AND THEIR PROOFS

In this section, we consider a triangular array $\{(X_{is},Y_{is})\}_{i=1}^n$ from the s-th group. In order to keep the notation simple, the sample will be denoted by $\{(X_i,Y_i)\}_{i=1}^n$. In addition to the notation introduced in the main

part of the manuscript, we will make use of the following notation throughout the appendix. Denote the empirical measure of n samples $(Y_i, \mathbf{Z}_i)$ by $\mathbb{P}_n$. For a function $x \mapsto f(x)$ define $\mathbb{G}_n(f) := n^{1/2} \int f(x)(d\mathbb{P}_n(x) - dP(x))$ and $\|f\|_{L_p(P)} = (\int |f(x)|^p dP(x))^{1/p}$ for $0 < p < \infty$. For a class of functions $\mathcal{G}$, let $\|\mathbb{P}_n - P\|_{\mathcal{G}} := \sup_{f\in\mathcal{G}} |\mathbb{P}_n f - Pf|$. For any $\epsilon > 0$, the covering number $N(\epsilon, \mathcal{G}, L_p)$ is the minimal number of balls of radius ϵ (under L_p-norm) that is needed to cover $\mathcal{G}$. The bracketing number $N_{[\,]}(\epsilon, \mathcal{G}, L_p)$ is the minimal number of ϵ-brackets that is needed to cover $\mathcal{G}$. An ϵ-bracket refers to a pair of functions within an ϵ distance: $\|u - l\|_{L_p} < \epsilon$. Throughout the proofs, C, C_1, C_j etc. will denote constants which do not depend on n but may have different values in different lines. For $\tau \in \mathcal{T}$, define $\psi(Y_i, \mathbf{Z}_i; \mathbf{b}, \tau) := \mathbf{Z}_i(\mathbf{1}\{Y_i \le \mathbf{Z}_i^\top \mathbf{b}\} - \tau)$ and $\mu(\mathbf{b}, \tau) := \mathbb{E}\big[\psi(Y_i, \mathbf{Z}_i; \mathbf{b}, \tau)\big] = \mathbb{E}\big[\mathbf{Z}_i\big\{F_{Y|X}(\mathbf{Z}_i^\top \mathbf{b}|X) - \tau\big\}\big]$.

The aim of this section is to provide precise bounds on the remainder terms in the Bahadur representation of the estimator $\breve{\boldsymbol{\beta}}(\tau)$ which is defined as

$$\breve{\boldsymbol{\beta}}(\tau) := \arg\min_{\mathbf{b}\in\mathbb{R}^m} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{b}^\top \mathbf{Z}(X_i)), \quad \tau \in \mathcal{T}.$$

THEOREM S.6.1. *Suppose Conditions (A1)-(A3) hold. Assume that additionally* $m\xi_m^2 \log n = o(n)$. *Then, for any* $\boldsymbol{\beta}_N(\cdot)$ *satisfying* $g_N(\boldsymbol{\beta}_N) = o(\xi_m^{-1})$ *and* $c_m(\boldsymbol{\beta}_N) = o(1)$, *we have*

$$\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_N(\tau) = -n^{-1/2} J_m(\tau)^{-1} \mathbb{G}_n(\psi(\cdot; \boldsymbol{\beta}_N(\tau), \tau)) + r_{n,1}(\tau) + r_{n,2}(\tau) + r_{n,3}(\tau) + r_{n,4}(\tau). \tag{S.6.1}$$

The remainder terms $r_{n,j}$*'s can be bounded as follows:*

$$\sup_\tau \|r_{n,1}(\tau)\| \le C_1 \frac{m\xi_m}{n} \quad a.s. \tag{S.6.2}$$

for a constant C_1 *independent of* n. *Moreover, we have for any* $\kappa_n \ll n/\xi_m^2$, *all sufficiently large* n, *and a constant* C *independent of* n

$$P\Big(\sup_\tau \|r_{n,j}(\tau)\| \le C\Re_j(\kappa_n)\Big) \ge 1 - 2e^{-\kappa_n}, \quad j = 2, 3, 4, \tag{S.6.3}$$

where

(S.6.4)

$$\Re_2(\kappa_n) := \xi_m\Big(\Big(\frac{m}{n}\log n\Big)^{1/2} + n^{-1/2}\kappa_n^{1/2} + g_N\Big)^2,$$

(S.6.5)

$$\Re_3(\kappa_n) := \Big(\Big(\frac{m\log n}{n}\Big)^{1/2} + \Big(\frac{\kappa_n}{n}\Big)^{1/2} + g_N\Big)^{1/2}\Big(\Big(\frac{m\xi_m\log n}{n}\Big)^{1/2} + \Big(\frac{\xi_m\kappa_n}{n}\Big)^{1/2}\Big),$$

(S.6.6)

$$\Re_4(\kappa_n) := c_m(\boldsymbol{\beta}_N)\Big(\Big(\frac{m}{n}\log n\Big)^{1/2} + \frac{\kappa_n^{1/2}}{n^{1/2}}\Big) + g_N.$$

Additionally,

$$\sup_\tau \|r_{n,3}(\tau)\| \le C_{3,1}\xi_m, \quad a.s., \tag{S.6.7}$$

and it holds that for sufficiently large n,

$$\sup_\tau \|\mathbb{E}[r_{n,3}(\tau)]\| \le C_{3,2}\xi_m^2\Big(g_N^2 + \frac{m\log n}{n}\Big) \tag{S.6.8}$$

where $C_{3,1}, C_{3,2}$ are constants that are independent of n.

The statements in (S.6.2)-(S.6.6) are from Theorem 5.1 of Chao et al. (2017) and reproduced here for the sake of completeness. The bound in (S.6.8) is new and crucial for the ABR in Theorem S.2.1 in the main manuscript. See Section S.6.1 in the supplementary material for a proof.

The next result provides a refined version of the previous theorem for local basis functions that satisfy Condition (L).

THEOREM S.6.2. *Suppose Conditions (A1)-(A3) and (L) hold. Assume that additionally $m\xi_m^2\log n = o(n)$. Assume that the set $\mathcal{I}$ consists of at most $L \ge 1$ consecutive integers. Then, for $\boldsymbol{\gamma}_N$ defined in (3.8) we have*

$$\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau) = -n^{-1/2}\widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1}\mathbb{G}_n(\psi(\cdot;\boldsymbol{\gamma}_N(\tau),\tau)) + \sum_{k=1}^4 r_{n,k}^{(L)}(\tau).$$

where remainder terms $r_{n,j}$'s can be bounded as follows (C_1, C_4 are constants that do not depend on n):

$$\sup_{\mathbf{a}\in\mathcal{S}_{\mathcal{I}}^{m-1}}\sup_\tau |\mathbf{a}^\top r_{n,1}^{(L)}(\tau)| \le C_1\frac{\xi_m\log n}{n} \quad a.s., \tag{S.6.9}$$

$$\sup_{\mathbf{a}\in\mathcal{S}_{\mathcal{I}}^{m-1}}\sup_\tau |\mathbf{a}^\top r_{n,4}^{(L)}(\tau)| \le \frac{1}{n} + C_4c_m^2(\boldsymbol{\gamma}_N)\sup_{\mathbf{a}\in\mathcal{S}_{\mathcal{I}}^{m-1}}\widetilde{\mathcal{E}}(\mathbf{a},\boldsymbol{\gamma}_N) \quad a.s. \tag{S.6.10}$$

Moreover, we have for any $\kappa_n \ll n/\xi_m^2$, all sufficiently large n, and a constant C independent of n

$$P\Big(\sup_{\mathbf{a}\in\mathcal{S}_{\mathcal{I}}^{m-1}}\sup_{\tau}|\mathbf{a}^\top r_{n,j}^{(L)}(\tau)| \le C\Re_j^{(L)}(\kappa_n)\Big) \ge 1 - n^2 e^{-\kappa_n}, \quad j = 2,3 \tag{S.6.11}$$

where

$$\Re_2^{(L)}(\kappa_n) := \sup_{\mathbf{a}\in\mathcal{S}_{\mathcal{I}}^{m-1}} \widetilde{\mathcal{E}}(\mathbf{a},\boldsymbol{\gamma}_N)\Big(\frac{\xi_m(\log n + \kappa_n^{1/2})}{n^{1/2}} + c_m^2(\boldsymbol{\gamma}_N)\Big)^2, \tag{S.6.12}$$

$$\Re_3^{(L)}(\kappa_n) := \Big(c_m(\boldsymbol{\gamma}_N)\frac{\kappa_n^{1/2}\vee\log n}{n^{1/2}} + \frac{\xi_m^{1/2}(\kappa_n^{1/2}\vee\log n)^{3/2}}{n^{3/4}}\Big), \tag{S.6.13}$$

and $\widetilde{\mathcal{E}}(\mathbf{a},\boldsymbol{\gamma}_N) := \sup_{\tau\in\mathcal{T}}\mathbb{E}[\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1}\mathbf{Z}]$. *Additionally,*

$$\sup_{\tau}\sup_{\mathbf{a}\in\mathcal{S}_{\mathcal{I}}^{m-1}}|\mathbf{a}^\top r_{n,3}^{(L)}(\tau)| \le C_{3,1}\xi_m \quad a.s. \tag{S.6.14}$$

and for sufficiently large n we have

$$\sup_{\mathbf{a}\in\mathcal{S}_{\mathcal{I}}^{m-1}}\sup_{\tau}|\mathbf{a}^\top\mathbb{E}[r_{n,3}^{(L)}(\tau)]| \le C_{3,2}\sup_{\mathbf{a}\in\mathcal{S}_{\mathcal{I}}^{m-1}}\widetilde{\mathcal{E}}(\mathbf{a},\boldsymbol{\gamma}_N)\Big(c_m^4(\boldsymbol{\gamma}_N) + \frac{\xi_m^2(\log n)^2}{n}\Big). \tag{S.6.15}$$

where $C_{3,1}, C_{3,2}$ are constants that do not depend on n.

Similarly to the setting in Theorem S.6.1, the statements in (S.6.9)-(S.6.13) are proved in Theorem 5.2 of Chao et al. (2017) and reproduced here for the sake of completeness. The bound in (S.6.15) is new and crucial for the ABR in Theorem S.2.4.

In both proofs, we will repeatedly use the following leave-one-out estimator

$$\breve{\boldsymbol{\beta}}^{(-j)}(\tau) := \arg\min_{\mathbf{b}\in\mathbb{R}^m}\sum_{i\neq j}\rho_\tau(Y_i - \mathbf{b}^\top\mathbf{Z}(X_i)), \quad \tau\in\mathcal{T}, j = 1,...,n \tag{S.6.16}$$

S.6.1. Proof of Theorem S.6.1. Observe the representation

$$\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_N(\tau) = -n^{-1/2}J_m(\tau)^{-1}\mathbb{G}_n(\psi(\cdot;\boldsymbol{\beta}_N(\tau),\tau)) + r_{n,1}(\tau) + r_{n,2}(\tau) + r_{n,3}(\tau) + r_{n,4}(\tau)$$

where

$$
\begin{aligned}
r_{n,1}(\tau) &:= \widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1}\mathbb{P}_n\psi(\cdot;\breve{\boldsymbol{\beta}}(\tau),\tau),\\
r_{n,2}(\tau) &:= -\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1}\Big(\mu(\breve{\boldsymbol{\beta}}(\tau),\tau)-\mu(\boldsymbol{\beta}_N(\tau),\tau)\\
&\qquad -\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))(\breve{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_N(\tau))\Big),\\
r_{n,3}(\tau) &:= -n^{-1/2}\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1}\Big(\mathbb{G}_n(\psi(\cdot;\breve{\boldsymbol{\beta}}(\tau),\tau))-\mathbb{G}_n(\psi(\cdot;\boldsymbol{\beta}_N(\tau),\tau))\Big),\\
r_{n,4}(\tau) &:= -n^{-1/2}(J_m(\tau)^{-1}-\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1})\mathbb{G}_n(\psi(\cdot;\boldsymbol{\beta}_N(\tau),\tau))\\
&\quad -\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1}\mu(\boldsymbol{\beta}_N(\tau),\tau).
\end{aligned}
$$

The bounds in (S.6.2)-(S.6.6) follow similarly as in the proof of Theorem 5.1 in Chao et al. (2017). The bound in (S.6.7) follows from the definition of $r_{n,3}$. It remains to prove (S.6.8). Note that the expectation of $r_{n,3}(\tau)$ is not zero in general. Write

$$
\begin{aligned}
&r_{n,3}(\tau)\\
&=-\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1}\Big[-n^{-1/2}\mathbb{G}_n\psi(\cdot;\boldsymbol{\beta}_N(\tau),\tau)\\
&\qquad +n^{-1}\sum_{j=1}^{n}\Big(\psi(Y_j,\mathbf{Z}_j;\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau)-\mu(\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau)\Big)\\
&\quad -n^{-1}\sum_{j=1}^{n}\Big(\psi(Y_j,\mathbf{Z}_j;\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau)-\psi(Y_j,\mathbf{Z}_j;\breve{\boldsymbol{\beta}}(\tau),\tau)\\
&\qquad\qquad -\mu(\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau)+\mu(\breve{\boldsymbol{\beta}}(\tau),\tau)\Big)\Big].
\end{aligned}
$$

By definition of $\breve{\boldsymbol{\beta}}^{(-j)}(\tau)$ in (S.6.16), the expectation of the first line in the above representation is zero. To bound the expectation of the second line, observe that for

$$
\delta_n := \widetilde{C}\xi_m\Big(g_N^2(\boldsymbol{\beta}_N)+\frac{m\log n}{n}\Big)
$$

with $\widetilde{C}$ sufficiently large we have

$$
\begin{aligned}
&\sup_{\tau\in\mathcal{T}}\Big\|\mathbb{E}\Big[n^{-1}\sum_{j=1}^{n}\psi(Y_j,\mathbf{Z}_j;\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau)-\psi(Y_j,\mathbf{Z}_j;\breve{\boldsymbol{\beta}}(\tau),\tau)\Big]\Big\|\\
=&\sup_{\tau\in\mathcal{T}}\sup_{\|\mathbf{a}\|=1}\Big|\mathbb{E}\Big[n^{-1}\sum_{j=1}^{n}\mathbf{a}^\top\mathbf{Z}_j\big(\mathbf{1}\{Y_j\le\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)\}-\mathbf{1}\{Y_j\le\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}(\tau)\}\big)\Big]\Big|\\
\le&\sup_{\tau\in\mathcal{T}}\sup_{\|\mathbf{a}\|=1}n^{-1}\sum_{j=1}^{n}\mathbb{E}\Big[|\mathbf{a}^\top\mathbf{Z}_j|\mathbf{1}\big\{|Y_j-\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)|\le\xi_m\delta_n\big\}\\
&\qquad\qquad\times\mathbf{1}\big\{\sup_j\big|\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}(\tau)-\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)\big|\le\xi_m\delta_n\big\}\Big]\\
&+\xi_m\sup_{\tau\in\mathcal{T}}P\Big(\sup_j\big|\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}(\tau)-\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)\big|>\xi_m\delta_n\Big)\\
\le&\sup_{\tau\in\mathcal{T}}\sup_{\|\mathbf{a}\|=1}n^{-1}\sum_{j=1}^{n}\mathbb{E}\Big[|\mathbf{a}^\top\mathbf{Z}_j|\big|F_{Y|X}(\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)+\xi_m\delta_n|X_j)\\
&\qquad\qquad-F_{Y|X}(\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)-\xi_m\delta_n|X_j)\big|\Big]\\
&+\xi_mP\Big(\sup_{\tau\in\mathcal{T}}\sup_j\big|\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}(\tau)-\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)\big|>\xi_m\delta_n\Big)\\
\le&\xi_mP\Big(\sup_{\tau\in\mathcal{T}}\sup_j\big|\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}(\tau)-\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)\big|>\xi_m\delta_n\Big)\\
&+2\xi_m\delta_n\bar{f}\sup_{\|\mathbf{a}\|=1}n^{-1}\sum_{j=1}^{n}\mathbb{E}\big[|\mathbf{a}^\top\mathbf{Z}_j|\big]\\
\le&\ 2\xi_m\delta_n\bar{f}\lambda_{\max}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top])^{1/2}+\xi_mP\Big(\sup_{\tau\in\mathcal{T}}\sup_j\big|\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}(\tau)-\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)\big|>\xi_m\delta_n\Big)\\
\le&C_{3,2}\xi_m^2\Big(g_N^2(\boldsymbol{\beta}_N)+\frac{m\log n}{n}\Big)
\end{aligned}
$$

for all sufficiently large n and a constant $C_{3,2}$ independent of τ, n. Here, the last line follows from Lemma S.6.3 and the definition of δ_n provided that $\widetilde{C}$ in the definition of δ_n is sufficiently large. Thus it remains to bound the expectation of $\mu(\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau)-\mu(\breve{\boldsymbol{\beta}}(\tau),\tau)$. A Taylor expansion shows that

$$
\begin{aligned}
\sup_\tau\|\mu(\mathbf{b}_1,\tau)-\mu(\mathbf{b}_2,\tau)\|&\le\sup_{\|\mathbf{a}\|=1}\mathbb{E}\big[|\mathbf{a}^\top\mathbf{Z}|\mathbf{1}\big\{|Y-\mathbf{Z}^\top\mathbf{b}_1|\le|\mathbf{Z}^\top(\mathbf{b}_1-\mathbf{b}_2)|\big\}\big]\\
&\le 2\bar{f}\sup_{\|\mathbf{a}\|=1}\mathbb{E}\big[|\mathbf{a}^\top\mathbf{Z}||\mathbf{Z}^\top(\mathbf{b}_1-\mathbf{b}_2)|\big]\\
&\le 2\bar{f}\lambda_{\max}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top])\|\mathbf{b}_1-\mathbf{b}_2\|,
\end{aligned}
$$

where the last inequality follows by Cauchy-Schwarz inequality. On the other hand $\|\mu(\mathbf{b}_1,\tau)-\mu(\mathbf{b}_2,\tau)\| \le 2\xi_m$. Thus we have for any $\delta_n > 0$,

$$\sup_{\tau,j}\Big\|\mathbb{E}\big[\mu(\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau)-\mu(\breve{\boldsymbol{\beta}}(\tau),\tau)\big]\Big\| \le 2\bar{f}\lambda_{\max}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top])\delta_n + 2\xi_m P\Big(\sup_{\tau,j}\|\breve{\boldsymbol{\beta}}^{(-j)}(\tau)-\breve{\boldsymbol{\beta}}(\tau)\| > \delta_n\Big).$$

Choosing δ_n as above completes the proof. □

LEMMA S.6.3. *Under the assumptions of Theorem S.6.1 we have for any $\kappa_n \ll n/\xi_m^2$, all sufficiently large n, and a constant C independent of n we have for $\breve{\boldsymbol{\beta}}^{(-j)}$ defined in (S.6.16)*

$$P\Big(\max_{j=1,\ldots,n}\sup_{\tau}\|\breve{\boldsymbol{\beta}}^{(-j)}(\tau)-\breve{\boldsymbol{\beta}}(\tau)\| \ge C\Big(g_N^2(\boldsymbol{\beta}_N)\xi_m+\frac{\xi_m\kappa_n}{n}+\frac{m\xi_m\log n}{n}\Big)\Big) \le 2ne^{-\kappa_n}.$$

Proof of Lemma S.6.3
Let $\mathbb{P}_n^{(-j)}(f) := (n-1)^{-1}\sum_{i\neq j} f(X_i,Y_i)$, $\mathbb{G}_n^{(-j)}(f) := \sqrt{n-1}(\mathbb{P}_n^{(-j)}(f) - \mathbb{E}f(X_1,Y_1))$. Similarly to the decomposition considered previously (S.6.1), we have

(S.6.17)

$$\breve{\boldsymbol{\beta}}^{(-j)}(\tau)-\boldsymbol{\beta}_N(\tau) = -(n-1)^{-1/2}J_m(\tau)^{-1}\mathbb{G}_n^{(-j)}(\psi(\cdot;\boldsymbol{\beta}_N(\tau),\tau))+\sum_{k=1}^4 r_{n,k}^{(-j)}(\tau)$$

where

$$\begin{aligned}
r_{n,1}^{(-j)}(\tau) &:= \widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1}\mathbb{P}_n^{(-j)}\psi(\cdot;\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau),\\
r_{n,2}^{(-j)}(\tau) &:= -\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1}\Big(\mu(\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau)-\mu(\boldsymbol{\beta}_N(\tau),\tau)\\
&\qquad -\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))(\breve{\boldsymbol{\beta}}^{(-j)}(\tau)-\boldsymbol{\beta}_N(\tau))\Big),\\
r_{n,3}^{(-j)}(\tau) &:= \frac{-\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1}}{\sqrt{n-1}}\Big(\mathbb{G}_n^{(-j)}(\psi(\cdot;\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau))-\mathbb{G}_n^{(-j)}(\psi(\cdot;\boldsymbol{\beta}_N(\tau),\tau))\Big),\\
r_{n,4}^{(-j)}(\tau) &:= -(n-1)^{-1/2}(J_m(\tau)^{-1}-\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1})\mathbb{G}_n^{(-j)}(\psi(\cdot;\boldsymbol{\beta}_N(\tau),\tau))\\
&\quad -\widetilde{J}_m(\boldsymbol{\beta}_N(\tau))^{-1}\mu(\boldsymbol{\beta}_N(\tau),\tau).
\end{aligned}$$

By similar arguments as in the proof for $r_{n,2}$ in Theorem 5.1 in Chao et al. (2017) there exists C independent of n such that $P(\Omega_{1,n}^{(j)}(C)) \le e^{-\kappa_n}, j = 1,\ldots,n$ where we defined the event

(S.6.18)

$$\Omega_{1,n}^{(j)}(C) := \Big\{\sup_{\tau}\|\breve{\boldsymbol{\beta}}^{(-j)}(\tau)-\boldsymbol{\beta}_N(\tau)\| \ge C\Big(g_N(\boldsymbol{\beta}_N)+\frac{\kappa_n^{1/2}}{n^{1/2}}+\Big(\frac{m\log n}{n}\Big)^{1/2}\Big)\Big\}.$$

Combining (S.6.1) and (S.6.17) we find that, almost surely,

$$\sup_j \sup_\tau \Big\| \breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \breve{\boldsymbol{\beta}}(\tau) - \sum_{k=1}^{4} \Big(r_{n,k}^{(-j)}(\tau) - r_{n,k}(\tau) \Big) \Big\| \lesssim \frac{\xi_m}{n}.$$

As in the proof of Lemma 34 on page 106 in Belloni et al. (2017), arguments from linear optimization lead to the bound

$$\sup_j \sup_\tau \| r_{n,1}^{(-j)}(\tau) - r_{n,1}(\tau) \| \leq \frac{2}{\inf_\tau \lambda_{min}(\widetilde{J}_m(\boldsymbol{\beta}_N(\tau)))} \frac{m\xi_m}{n-1} \quad a.s.$$

Direct calculations show that

$$\sup_j \sup_\tau \| r_{n,4}^{(-j)}(\tau) - r_{n,4}(\tau) \| \leq C_1 \frac{\xi_m}{n},$$

and from Lemma S.6.5 we obtain for a constant C independent of j, τ, n

$$\| r_{n,2}^{(-j)}(\tau) - r_{n,2}(\tau) \| \leq C\xi_m \Big(\| \breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \boldsymbol{\beta}_N(\tau) \|^2 + \| \breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_N(\tau) \|^2 \Big).$$

The probability of the event

$$\Omega_{1,n}(C) := \Big\{ \sup_{\tau,j} \Big(\| \breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \boldsymbol{\beta}_N(\tau) \|^2 + \| \breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_N(\tau) \|^2 \Big) \geq C \Big(g_N^2(\boldsymbol{\beta}_N) + \frac{\kappa_n}{n} + \frac{m \log n}{n} \Big) \Big\}$$

can be bounded by $(n+1)e^{-\kappa_n}$ if C is chosen suitably, this follows from the bound $P(\Omega_{1,n}^{(j)}(C)) \leq e^{-\kappa_n}$ in (S.6.18). Finally, define

(S.6.19)

$$\mathcal{G}_2(\delta) := \Big\{ (z,y) \mapsto \mathbf{a}^\top z (\mathbf{1}\{y \leq z^\top \mathbf{b}_1\} - \mathbf{1}\{y \leq z^\top \mathbf{b}_2\}) \mathbf{1}\{\|z\| \leq \xi_m\} \Big| \mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^m, \|\mathbf{b}_1 - \mathbf{b}_2\| \leq \delta, \mathbf{a} \in \mathcal{S}^{m-1} \Big\}.$$

Letting $\delta_n := \sup_j \sup_\tau \| \breve{\boldsymbol{\beta}}(\tau)^{(-j)} - \breve{\boldsymbol{\beta}}(\tau) \|$, we have

$$\sup_j \sup_\tau \| r_{n,3}^{(-j)}(\tau) - r_{n,3}(\tau) \| \leq C_3 \Big(\frac{\xi_m}{n} + \| \mathbb{P}_n - P \|_{\mathcal{G}_2(\delta_n)} \Big).$$

Summarizing the bounds obtained so far we find for a constant C_4 independent of n

$$\delta_n \leq C_4 \Big(\frac{m\xi_m}{n} + \| \mathbb{P}_n - P \|_{\mathcal{G}_2(\delta_n)} + \sup_{\tau,j} \xi_m \big(\| \breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \boldsymbol{\beta}_N(\tau) \|^2 + \| \breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_N(\tau) \|^2 \big) \Big).$$

Consider the event

$$\Omega_{2,n}(C) := \Big\{ \sup_{0\le t\le 1} \frac{\|\mathbb{P}_n - P\|_{\mathcal{G}_2(t)}}{\chi_n(t,\kappa_n)} \ge C \Big\},$$

where

$$\chi_n(t,\kappa_n) = \xi_m^{1/2} t^{1/2} \Big(\frac{m}{n}\log(\xi_m \vee n)\Big)^{1/2} + \frac{m\xi_m}{n}\log(\xi_m\vee n) + \Big(\frac{\xi_m t\kappa_n}{n}\Big)^{1/2} + \frac{\xi_m\kappa_n}{n}$$

defined as (S.6.22) in Lemma S.6.4. It follows from (S.6.21) that $P(\Omega_{2,n}(C)) \le e^{-\kappa_n}\log_2 n$ if we choose C suitably. Thus, δ_n satisfies the inequality

$$\begin{aligned}
\delta_n &\le C_8\xi_m\Big(g_N^2(\boldsymbol{\beta}_N) + \frac{\kappa_n}{n} + \frac{m\log n}{n}\Big) + C_8\chi_n(\delta_n,\kappa_n) \\
&\le C_8\xi_m\Big(g_N^2(\boldsymbol{\beta}_N) + \frac{\kappa_n}{n} + \frac{m\log n}{n}\Big) + C_8\frac{m\xi_m}{n}\log(\xi_m\vee n) + C_8\frac{\xi_m\kappa_n}{n} \\
&\quad + C_9\delta_n^{1/2}\Big(\frac{\xi_m m}{n}\log(\xi_m\vee n) + \frac{\xi_m\kappa_n}{n}\Big)^{1/2} \\
&:= a_n + \delta_n^{1/2} b_n
\end{aligned}$$

with probability at least $1 - 2ne^{-\kappa_n}$ (note that $(n+1)e^{-\kappa_n} + \log_2 n e^{-\kappa_n} \le 2ne^{-\kappa_n}$). A simple calculation shows that the inequality $0 \le \delta_n \le a_n + b_n\delta_n^{1/2}$ implies $0 \le \delta_n \le 4\max(a_n, b_n^2)$. Straightforward calculations show that under the assumption of the theorem we have

$$a_n + b_n^2 \lesssim g_N^2(\boldsymbol{\beta}_N)\xi_m + \frac{\xi_m\kappa_n}{n} + \frac{m\xi_m\log n}{n}$$

provided that $\kappa_n \to \infty$. Thus we have for some constant C independent of n

$$P\Big(\delta_n \ge C\Big(g_N^2(\boldsymbol{\beta}_N)\xi_m + \frac{\xi_m\kappa_n}{n} + \frac{m\xi_m\log n}{n}\Big)\Big) \le 2ne^{-\kappa_n}.$$

This completes the proof. □

LEMMA S.6.4. *Consider the class of functions $\mathcal{G}_2(\delta_n)$ defined in (S.6.19). Under assumptions (A1)-(A3) and $\xi_m\delta_n \gg n^{-1}, \xi_m = O(n^b)$ for some $b < \infty$, we have for some constant C independent of n, sufficiently large n and arbitrary $\kappa_n > 0$,*

$$P\Big(\|\mathbb{P}_n - P\|_{\mathcal{G}_2(\delta_n)} \ge C\chi_n(\delta_n,\kappa_n)\Big) \le e^{-\kappa_n} \tag{S.6.20}$$

and for $\kappa_n \ge 1$

$$P\Big(\sup_{0\le t\le 1}\frac{\|\mathbb{P}_n - P\|_{\mathcal{G}_2(t)}}{\chi_n(t,\kappa_n)} \ge 2C\Big) \le e^{-\kappa_n}\log_2 n \tag{S.6.21}$$

where

(S.6.22)
$$\chi_n(t,\kappa_n) := \Big(\frac{m\xi_m}{n}\log(\xi_m \vee n)t\Big)^{1/2} + \frac{m\xi_m}{n}\log(\xi_m \vee n) + \Big(\frac{\xi_m\kappa_n t}{n}\Big)^{1/2} + \frac{\xi_m\kappa_n}{n}.$$

Proof of Lemma S.6.4. For a proof of (S.6.20) see Lemma C.3 in the Appendix of Chao et al. (2017). For a proof of (S.6.21), observe that for $\kappa_n \geq 1$ we have $\chi_n(t/2,\kappa_n) \geq \chi_n(t,\kappa_n)/2, \chi_n(0,\kappa_n) \geq \chi_n(n^{-1},\kappa_n)/2$, and thus

$$\begin{aligned}
&\sup_{0\leq t\leq 1}\frac{\|\mathbb{P}_n - P\|_{\mathcal{G}_2(t)}}{\chi_n(t,\kappa_n)} \\
&\leq \Big(\sup_{0\leq t\leq n^{-1}}\frac{\|\mathbb{P}_n - P\|_{\mathcal{G}_2(t)}}{\chi_n(t,\kappa_n)}\Big) \vee \Big(\max_{k:n^{-1}\leq 2^{-k}\leq 1}\sup_{2^{-k-1}\leq t\leq 2^{-k}}\frac{\|\mathbb{P}_n - P\|_{\mathcal{G}_2(t)}}{\chi_n(t,\kappa_n)}\Big) \\
&\leq \Big(\max_{k:n^{-1}\leq 2^{-k}\leq 1}\frac{2\|\mathbb{P}_n - P\|_{\mathcal{G}_2(2^{-k})}}{\chi_n(2^{-k},\kappa_n)}\Big) \vee \Big(\frac{2\|\mathbb{P}_n - P\|_{\mathcal{G}_2(n^{-1})}}{\chi_n(n^{-1},\kappa_n)}\Big)
\end{aligned}$$

Now the set $\{k : n^{-1} \leq 2^{-k} \leq 1\}$ contains not more than $\log_2 n$ elements, and (S.6.21) thus follows from (S.6.20). □

LEMMA S.6.5. *Under assumptions (A1)-(A3) we have*

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}}\sup_{\|\mathbf{b}-\boldsymbol{\beta}_N(\tau)\|\leq\omega}\|\mu(\mathbf{b},\tau) - \mu(\boldsymbol{\beta}_N(\tau),\tau) - \widetilde{J}_m(\boldsymbol{\beta}_N(\tau))(\mathbf{b}-\boldsymbol{\beta}_N(\tau))\| \\
\leq\ &\lambda_{\max}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top])\overline{f'}\omega^2\xi_m,
\end{aligned}$$

Proof of Lemma S.6.5 see the proof of Lemma C.1 in the Appendix of Chao et al. (2017). □

LEMMA S.6.6. *Let assumptions (A1)-(A3) hold. Then, for any $t > 1$*

$$\begin{aligned}
&\Big\{\sup_{\tau\in\mathcal{T}}\|\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_N(\tau)\| \leq \frac{2t(s_{n,1} + g_N(\boldsymbol{\beta}_N))}{\inf_{\tau\in\mathcal{T}}\lambda_{min}(\widetilde{J}_m(\boldsymbol{\beta}_N(\tau)))}\Big\} \\
&\qquad\qquad \supseteq \Big\{(s_{n,1} + g_N(\boldsymbol{\beta}_N)) < \frac{\inf_{\tau\in\mathcal{T}}\lambda^2_{min}(\widetilde{J}_m(\boldsymbol{\beta}_N(\tau)))}{4t\xi_m\overline{f'}\lambda^2_{\max}(\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top])}\Big\}.
\end{aligned}$$

where $s_{n,1} := \|\mathbb{P}_n - P\|_{\mathcal{G}_1}$ and

$$\mathcal{G}_1 := \Big\{(z,y)\mapsto \mathbf{a}^\top z(\mathbf{1}\{y\leq z^\top\mathbf{b}\}-\tau)\mathbf{1}\{\|z\|\leq\xi_m\}\Big|\tau\in\mathcal{T},\mathbf{b}\in\mathbb{R}^m,\mathbf{a}\in\mathcal{S}^m\Big\}$$

Proof of Lemma S.6.6. See the proof of Lemma C.2 in the Appendix of Chao et al. (2017). □

S.6.2. Proof of Theorem S.6.2. We begin by introducing some notation and useful preliminary results which will be used throughout this section. For a vector $\mathbf{v} \in \mathbb{R}^m$ and a set $\mathcal{I} \subset \{1, ..., m\}$ let $\mathbf{v}^{(\mathcal{I})} \in \mathbb{R}^m$ denote the vector that has entries v_i for $i \in \mathcal{I}$ and zero otherwise. For a vector $\mathbf{a} \in \mathbb{R}^m$ let $\|\mathbf{a}\|_0$ denote the number of non-zero entries in this vector, let $k_{\mathbf{a}}$ denote the position of the first non-zero entry and define

$$\mathcal{I}_{\mathbf{a}}(D) := \big\{ i \in \{1, ..., m\} : |i - k_{\mathbf{a}}| \leq \|\mathbf{a}\|_0 + D \big\}. \tag{S.6.23}$$

Under (A1)-(A3) and (L) we obtain by similar arguments as in Lemma A.1 in Chao et al. (2017)

$$\|\mathbf{a}^\top \widetilde{J}_m^{-1}(\boldsymbol{\gamma}_N(\tau)) - (\mathbf{a}^\top \widetilde{J}_m^{-1}(\boldsymbol{\gamma}_N(\tau)))^{(\mathcal{I}_{\mathbf{a}}(D))}\| \leq C\|\mathbf{a}\|_\infty \|\mathbf{a}\|_0 \lambda^D \tag{S.6.24}$$

for constants $\lambda \in (0, 1), C > 0$ independent of n, τ.

In the sequel, we will make use of the following decomposition which holds for any $\mathbf{b} \in \mathbb{R}^m$ such that $\widetilde{J}_m(\mathbf{b})$ is invertible

$$\breve{\boldsymbol{\beta}}(\tau) - \mathbf{b} = -n^{-1/2} \widetilde{J}_m^{-1}(\mathbf{b}) \mathbb{G}_n(\psi(\cdot; \mathbf{b}, \tau)) + \widetilde{J}_m^{-1}(\mathbf{b}) \sum_{k=1}^{4} R_{n,k}(\tau, \mathbf{b}) \tag{S.6.25}$$

where

$$R_{n,1}(\tau, \mathbf{b}) := \mathbb{P}_n \psi(\cdot; \breve{\boldsymbol{\beta}}(\tau), \tau), \tag{S.6.26}$$

$$R_{n,2}(\tau, \mathbf{b}) := -\Big(\mu(\breve{\boldsymbol{\beta}}(\tau), \tau) - \mu(\mathbf{b}, \tau) - \widetilde{J}_m(\mathbf{b})(\breve{\boldsymbol{\beta}}(\tau) - \mathbf{b}) \Big), \tag{S.6.27}$$

$$R_{n,3}(\tau, \mathbf{b}) := -n^{-1/2} \Big(\mathbb{G}_n(\psi(\cdot; \breve{\boldsymbol{\beta}}(\tau), \tau)) - \mathbb{G}_n(\psi(\cdot; \mathbf{b}, \tau)) \Big), \tag{S.6.28}$$

$$R_{n,4}(\tau, \mathbf{b}) := -\mu(\mathbf{b}, \tau). \tag{S.6.29}$$

Let $r_{n,k}^{(L)}(\tau) := \widetilde{J}_m^{-1}(\boldsymbol{\gamma}_N(\tau)) R_{n,k}(\tau, \boldsymbol{\gamma}_N(\tau)), k = 1, ..., 4$. The bounds in (S.6.9)-(S.6.10) follow similarly to the bounds (5.6)-(5.9) in Theorem 5.2 in Chao et al. (2017). The bound in (S.6.14) follows from the definition of $r_{n,3}$. Thus it remains to establish (S.6.15). Write

$$\begin{aligned} R_{n,3}(\tau, \boldsymbol{\gamma}_N(\tau)) = & -\Big[-n^{-1/2} \mathbb{G}_n \psi(\cdot; \boldsymbol{\gamma}_N(\tau), \tau) \\ & + n^{-1} \sum_{j=1}^{n} \Big(\psi(Y_j, \mathbf{Z}_j; \breve{\boldsymbol{\beta}}^{(-j)}(\tau), \tau) - \mu(\breve{\boldsymbol{\beta}}^{(-j)}(\tau), \tau) \Big) \\ & - n^{-1} \sum_{j=1}^{n} \Big(\psi(Y_j, \mathbf{Z}_j; \breve{\boldsymbol{\beta}}^{(-j)}(\tau), \tau) - \psi(Y_j, \mathbf{Z}_j; \breve{\boldsymbol{\beta}}(\tau), \tau) \\ & - \mu(\breve{\boldsymbol{\beta}}^{(-j)}(\tau), \tau) + \mu(\breve{\boldsymbol{\beta}}(\tau), \tau) \Big) \Big]. \end{aligned}$$

Define

$$\widetilde{\delta}_n := \widetilde{C}\Big(c_m(\boldsymbol{\gamma}_N)^4 + \frac{\xi_m^2(\log n)^2}{n}\Big).$$

and observe that for any $\mathbf{a} \in \mathcal{S}_{\mathcal{I}}^{m-1}$, $\tau \in \mathcal{T}$,

$$\begin{aligned}
&\Big|(\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\mathbb{E}\Big[n^{-1}\sum_{j=1}^n \psi(Y_j, \mathbf{Z}_j; \breve{\boldsymbol{\beta}}^{(-j)}(\tau), \tau) - \psi(Y_j, \mathbf{Z}_j; \breve{\boldsymbol{\beta}}(\tau), \tau)\Big]\Big| \\
=&\Big|\mathbb{E}\Big[n^{-1}\sum_{j=1}^n (\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\mathbf{Z}_j(\mathbf{1}\{Y_i \le \mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}^{(-j)}(\tau)\} - \mathbf{1}\{Y_i \le \mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}(\tau)\})\Big]\Big| \\
\lesssim\ & n^{-1}\sum_{j=1}^n \mathbb{E}\Big[|(\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\mathbf{Z}_j|\mathbf{1}\Big\{|Y_j - \mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}^{(-j)}(\tau)| \le \widetilde{\delta}_n\Big\} \\
&\qquad \times \mathbf{1}\Big\{\sup_j |\mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}(\tau) - \mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}^{(-j)}(\tau)| \le \widetilde{\delta}_n\Big\}\Big] \\
&+ \xi_m P\Big(\sup_j |\mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}(\tau) - \mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}^{(-j)}(\tau)| > \widetilde{\delta}_n\Big) \\
\le\ & n^{-1}\sum_{j=1}^n \mathbb{E}\Big[|(\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\mathbf{Z}_j|\Big(F_{Y|X}(\mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}^{(-j)}(\tau) + \widetilde{\delta}_n|X_j) \\
&\qquad - F_{Y|X}(\mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \widetilde{\delta}_n|X_j)\Big)\Big] \\
&+ \xi_m P\Big(\sup_j |\mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}(\tau) - \mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}^{(-j)}(\tau)| > \widetilde{\delta}_n\Big) \\
\le\ & 2\widetilde{\delta}_n \bar{f} n^{-1}\sum_{j=1}^n \mathbb{E}\Big[|(\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\mathbf{Z}_j|\Big] + \xi_m P\Big(\sup_j |\mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}(\tau) - \mathbf{Z}_j^\top \breve{\boldsymbol{\beta}}^{(-j)}(\tau)| > \widetilde{\delta}_n\Big) \\
\le\ & C_{3,2}\widetilde{\mathcal{E}}(\mathbf{a}, \boldsymbol{\gamma}_N)\Big(c_m^4(\boldsymbol{\gamma}_N) + \frac{\xi_m^2(\log n)^2}{n}\Big)
\end{aligned}$$

for all sufficiently large n and a constant $C_{3,2}$ independent of $\tau, n, \mathbf{a}$. Here, the last line follows from Lemma S.6.7 with $\widetilde{C}$ in the definition of $\widetilde{\delta}_n$ chosen sufficiently large. Thus it remains to bound $(\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\mathbb{E}[\mu(\breve{\boldsymbol{\beta}}^{(-j)}(\tau), \tau) - \mu(\breve{\boldsymbol{\beta}}(\tau), \tau)]$. A Taylor expansion shows that for $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^m$

$$\begin{aligned}
&\sup_\tau |(\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\mu(\mathbf{b}_1, \tau) - \mu(\mathbf{b}_2, \tau)| \\
&\le \mathbb{E}\Big[|(\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\mathbf{Z}_j|\mathbf{1}\Big\{|Y_j - \mathbf{Z}_j^\top \mathbf{b}_1| \le |\mathbf{Z}_j^\top(\mathbf{b}_1 - \mathbf{b}_2)|\Big\}\Big] \\
&\le 2\bar{f}\widetilde{\mathcal{E}}(\mathbf{a}, \boldsymbol{\gamma}_N) \sup_{x \in \mathcal{X}} |\mathbf{Z}(x)^\top(\mathbf{b}_1 - \mathbf{b}_2)|.
\end{aligned}$$

On the other hand $\|\mu(\mathbf{b}_1,\tau)-\mu(\mathbf{b}_2,\tau)\| \le 2\xi_m$. Thus we have for any $\widetilde{\delta}_n>0, \mathbf{a}\in\mathcal{S}_{\mathcal{I}}^{m-1}$

$$\sup_{\tau,j}\Big\|(\mathbf{a}^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\mathbb{E}[\mu(\breve{\boldsymbol{\beta}}^{(-j)}(\tau),\tau)-\mu(\breve{\boldsymbol{\beta}}(\tau),\tau)]\Big\| \lesssim \widetilde{\mathcal{E}}(\mathbf{a},\boldsymbol{\gamma}_N)\widetilde{\delta}_n+\xi_m P\Big(\sup_{\tau,j}\|\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}(\tau)-\mathbf{Z}_j^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)\|>\widetilde{\delta}_n\Big).$$

Choosing $\widetilde{\delta}_n$ as above completes the proof. □

LEMMA S.6.7. *Under the assumptions of Theorem S.6.2 we have for any $\kappa_n \ll n/\xi_m^2$, all sufficiently large n, and a constant C independent of n*

$$P\Big(\sup_{j,\tau,x}\|\mathbf{Z}(x)^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau)-\mathbf{Z}(x)^\top\breve{\boldsymbol{\beta}}(\tau)\|\ge C\Big(c_m^4(\boldsymbol{\gamma}_N)+\frac{\xi_m^2(\kappa_n+(\log n)^2)}{n}\Big)\Big) \le 3m(n+1)e^{-\kappa_n}.$$

Proof of Lemma S.6.7

Recall that for a vector $\mathbf{a}\in\mathbb{R}^m$ $\|\mathbf{a}\|_0$ denotes the number of non-zero entries in $\mathbf{a}$ and $k_{\mathbf{a}}$ denotes the position of the first non-zero entry. Consider the following additional notation which is used exclusively in this proof:

$$\mathcal{I}(x) := \{i\in\{1,...,m\} : |i-k_{\mathbf{Z}(x)}|\le r+D\log n\}, \tag{S.6.30}$$

$$\mathcal{I}'(x) := \{i\in\{1,...,m\} : \exists\, j\in\mathcal{I}(x) \text{ such that } |i-j|\le r\}, \tag{S.6.31}$$

where r is defined in (L), and we suppress the dependence of $\mathcal{I}(x)$ and $\mathcal{I}'(x)$ on D for the sake of a simpler notation.

Let $\mathbb{P}_n^{(-j)}f := (n-1)^{-1}\sum_{i\neq j}f(X_i,Y_i)$, $\mathbb{G}_n^{(-j)}f := \sqrt{n-1}(\mathbb{P}_n^{(-j)}f-\mathbb{E}f(X_1,Y_1))$. We have

$$\mathbf{Z}(x)^\top(\breve{\boldsymbol{\beta}}^{(-j)}(\tau)-\boldsymbol{\gamma}_N(\tau)) = -(n-1)^{-1/2}(\mathbf{Z}(x)^\top\widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})^{(\mathcal{I}(x))}\mathbb{G}_n^{(-j)}(\psi(\cdot;\boldsymbol{\gamma}_N(\tau),\tau))+\sum_{k=1}^{5}r_{n,k}^{(-j)}(\tau,x) \tag{S.6.32}$$

where $r_{n,k}^{(-j)}(\tau,x) := (\mathbf{Z}(x)^\top\widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})^{(\mathcal{I}(x))}R_{n,k}^{(-j)}(\tau,\boldsymbol{\gamma}_N(\tau)), k=1,...,4$, $R_{n,k}^{(-j)}$ is defined exactly as $R_{n,k}$ in (S.6.26)-(S.6.29) with $n,\breve{\boldsymbol{\beta}},\mathbb{P}_n,\mathbb{G}_n$ replaced

by $n-1, \breve{\boldsymbol{\beta}}^{(-j)}, \mathbb{P}_n^{(-j)}, \mathbb{G}_n^{(-j)}$, respectively, and

$$\begin{aligned}
&r_{n,5}^{(-j)}(\tau,x)\\
&:= -\Big[(\mathbf{Z}(x)^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})^{(\mathcal{I}(x))} - (\mathbf{Z}(x)^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\Big]\\
&\quad\times -\Big[-(n-1)^{-1/2}\mathbb{G}_n^{(-j)}(\psi(\cdot;\boldsymbol{\gamma}_N(\tau),\tau)) + \sum_{k=1}^{4} R_{n,k}^{(-j)}(\tau,\boldsymbol{\gamma}_N(\tau))\Big]\\
&=\Big[(\mathbf{Z}(x)^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})^{(\mathcal{I}(x))} - (\mathbf{Z}(x)^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\Big](\breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \boldsymbol{\gamma}_N(\tau)),
\end{aligned}$$

where the last expression follows from (S.6.25).

Similarly we have

$$\begin{aligned}
&\mathbf{Z}(x)^\top(\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau))\\
&\quad = -n^{-1/2}(\mathbf{Z}(x)^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})^{(\mathcal{I}(x))}\mathbb{G}_n(\psi(\cdot;\boldsymbol{\gamma}_N(\tau),\tau)) + \sum_{k=1}^{5} r_{n,k}(\tau)
\end{aligned}$$

where

$$\begin{aligned}
&r_{n,5}(\tau,x)\\
&:= -\Big[(\mathbf{Z}(x)^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})^{(\mathcal{I}(x))} - (\mathbf{Z}(x)^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\Big]\\
&\quad\times -\Big[-n^{-1/2}\mathbb{G}_n(\psi(\cdot;\boldsymbol{\gamma}_N(\tau),\tau)) + \sum_{k=1}^{4} R_{n,k}(\tau,\boldsymbol{\gamma}_N(\tau))\Big].\\
&=\Big[(\mathbf{Z}(x)^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})^{(\mathcal{I}(x))} - (\mathbf{Z}(x)^\top \widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1})\Big](\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau)),
\end{aligned}$$

where the last expression follows from (S.6.25).

We obtain the representation

$$\begin{aligned}
\sup_{j,\tau,x}\Big|&\mathbf{Z}(x)^\top\breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \mathbf{Z}(x)^\top\breve{\boldsymbol{\beta}}(\tau) - (r_{n,5}^{(-j)}(\tau,x) - r_{n,5}(\tau,x))\\
&- \sum_{k=1}^{3}\Big(r_{n,k}^{(-j)}(\tau,x) - r_{n,k}(\tau,x)\Big)\Big| \lesssim \frac{\xi_m^2}{n} \quad a.s.
\end{aligned}$$

Observe that $r_{n,4}^{(-j)}(\tau) - r_{n,4}(\tau) = 0$ for all τ because $r_{n,4}$ does not depend on $\breve{\boldsymbol{\beta}}(\tau)$ or $\breve{\boldsymbol{\beta}}^{(-j)}(\tau)$.

The results in Lemma S.6.11 show that

$$\sup_{j,\tau,x}\|r_{n,1}^{(-j)}(\tau,x) - r_{n,1}(\tau,x)\| \lesssim \frac{\xi_m^2\log n}{n} \quad a.s.$$

From Lemma S.6.10 and condition (L) we obtain for a constant C independent of j, τ, n, x

$$\begin{aligned}&\|r_{n,2}^{(-j)}(\tau) - r_{n,2}(\tau)\| \\ &\leq C\Big(|\mathbf{Z}(x)^\top \breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \mathbf{Z}(x)^\top \boldsymbol{\gamma}_N(\tau)|^2 + |\mathbf{Z}(x)^\top \breve{\boldsymbol{\beta}}(\tau) - \mathbf{Z}(x)^\top \boldsymbol{\gamma}_N(\tau)|^2\Big).\end{aligned}$$

Let $w_n := \Big(c_m^4(\boldsymbol{\gamma}_N) + \frac{\xi_m^2 (\log n)^2 + \xi_m^2 \kappa_n}{n}\Big)$ and define the event

$$\widetilde{\Omega}_{1,n}(C) := \Big\{ \sup_{\tau,j,x} \Big(|\mathbf{Z}(x)^\top (\widehat{\boldsymbol{\beta}}^{(-j)}(\tau) - \boldsymbol{\gamma}_N(\tau))|^2 + |\mathbf{Z}(x)^\top (\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau))|^2\Big) \geq C w_n \Big\}.$$

By an application of Lemma S.6.8, $P(\widetilde{\Omega}_{1,n}(C))$ can be bounded by $m(n+1)e^{-\kappa_n}$ if C is chosen suitably.

Next, let $\widetilde{\delta}_n := \sup_{j,\tau,x} \|\mathbf{Z}(x)^\top \breve{\boldsymbol{\beta}}(\tau)^{(-j)} - \mathbf{Z}(x)^\top \breve{\boldsymbol{\beta}}(\tau)\|$. We obtain for the class of functions $\mathcal{G}_2$ defined in (S.6.34) and using (L)

$$\sup_{j,\tau,x} |r_{n,3}^{(-j)}(\tau, x) - r_{n,3}(\tau, x)| \leq C_3\Big(\frac{\xi_m^2}{n} + \xi_m \sup_x \|\mathbb{P}_n - P\|_{\mathcal{G}_2(\widetilde{\delta}_n, \mathcal{I}(x), \mathcal{I}'(x))}\Big).$$

Consider the event [here $\widetilde{\chi}_n(t, \mathcal{I}_1, \mathcal{I}_j, \kappa_n)$ is defined as (S.6.38) in Lemma S.6.9]

$$\widetilde{\Omega}_{2,n}(C) := \Big\{ \sup_x \sup_{0 \leq t \leq 1} \frac{\|\mathbb{P}_n - P\|_{\mathcal{G}_2(t, \mathcal{I}(x), \mathcal{I}'(x))}}{\widetilde{\chi}_n(t, \mathcal{I}(x), \mathcal{I}'(x), \kappa_n)} \geq C \Big\}.$$

Since for $x \in \mathcal{X}$ the pair of sets $(\mathcal{I}(x), \mathcal{I}'(x))$ takes at most m distinct values, it follows from (S.6.37) that $P(\widetilde{\Omega}_{2,n}(C)) \leq m e^{-\kappa_n} \log_2 n$ if we choose C suitably.

Finally, apply (S.6.24) to find that

$$\sup_{j,\tau,x} |r_{n,5}(\tau, x)| + |r_{n,5}^{(-j)}(\tau, x)| \lesssim \xi_m \lambda^{D \log n} \Big(\|\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau)\| + \|\breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \boldsymbol{\gamma}_N(\tau)\|\Big),$$

almost surely. Applying Theorem S.6.1 with $\boldsymbol{\beta}_N = \boldsymbol{\gamma}_N$ allows to bound the probability of the event $\Omega_{3,n}(C) := \{\|\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau)\| + \|\breve{\boldsymbol{\beta}}^{(-j)}(\tau) - \boldsymbol{\gamma}_N(\tau)\| > C\}$ by $e^{-\kappa_n}$, and by choosing D sufficiently large it follows that $\sup_{j,\tau,x} |r_{n,5}(\tau, x)| + |r_{n,5}^{(-j)}(\tau, x)| \leq n^{-1}$ with probability at least $1 - e^{-\kappa_n}$.

Summarizing the bounds obtained so far, there is a constant C_j independent of n such that $P(\widetilde{\Omega}_{1,n}(C_2) \cap \widetilde{\Omega}_{2,n}(C_2) \cap \Omega_{3,n}(C_2)) \leq m(n+1)e^{-\kappa_n} +$

$m \log_2 n e^{-\kappa_n} + e^{-\kappa_n} \le 3m(n+1)e^{-\kappa_n}$. On $\widetilde{\Omega}_{1,n}(C_2) \cap \widetilde{\Omega}_{2,n}(C_2) \cap \Omega_{3,n}(C_2)$,

$$\begin{aligned}
\delta_n &\le C_2\Big(c_m^4(\boldsymbol{\gamma}_N) + \frac{\xi_m^2\kappa_n}{n} + \frac{\xi_m^2(\log n)^2}{n}\Big) + C_2 \sup_x \xi_m \chi_n(\delta_n, \mathcal{I}(x), \mathcal{I}'(x), \kappa_n) \\
&\le C_2\Big(c_m^4(\boldsymbol{\gamma}_N) + \frac{\xi_m^2\kappa_n}{n} + \frac{\xi_m^2(\log n)^2}{n}\Big) + C_2\Big(\frac{\xi_m^2 \log^2(\xi_m \vee n)}{n} + \frac{\xi_m^2\kappa_n}{n}\Big) \\
&\quad + C_2\delta_n^{1/2}\Big(\frac{\xi_m^2 \log^2(\xi_m \vee n)}{n} + \frac{\xi_m^2\kappa_n}{n}\Big)^{1/2} \\
&:= a_n + \delta_n^{1/2} b_n.
\end{aligned}$$

A simple calculation shows that the inequality $0 \le \delta_n \le a_n + b_n\delta_n^{1/2}$ implies $0 \le \delta_n \le 4\max(a_n, b_n^2)$. Straightforward calculations show that under the assumption of the theorem we have

$$a_n + b_n^2 \lesssim c_m^4(\boldsymbol{\gamma}_N) + \frac{\xi_m^2 \log^2(\xi_m \vee n)}{n} + \frac{\xi_m^2\kappa_n}{n}.$$

Thus we have for some constant C_3 independent of n

$$P\Big(\delta_n \ge C_3\Big(c_m^4(\boldsymbol{\gamma}_N) + \frac{\xi_m^2 \log^2(\xi_m \vee n)}{n} + \frac{\xi_m^2\kappa_n}{n}\Big)\Big) \le 3m(n+1)e^{-\kappa_n}.$$

This completes the proof. □

LEMMA S.6.8. *Under the assumptions of Theorem S.6.2 we have for sufficiently large n and any $\kappa_n \ll n/\xi_m^2$*

$$P\Big(\sup_{\tau,x} |\mathbf{Z}(x)^\top(\breve{\boldsymbol{\beta}}(\tau) - \boldsymbol{\gamma}_N(\tau))| \ge C\Big(\frac{\xi_m \log n + \xi_m\kappa_n^{1/2}}{n^{1/2}} + c_m^2(\boldsymbol{\gamma}_N)\Big)\Big) \le (m+1)e^{-\kappa_n}.$$

where the constant C does not depend on n.

Proof of Lemma S.6.8. See the proof for Lemma C.4 in the Appendix of Chao et al. (2017). □

LEMMA S.6.9. *Let $\mathcal{Z} := \{\mathbf{Z}(x) | x \in \mathcal{X}\}$ where $\mathcal{X}$ is the support of X. For arbitrary index sets $\mathcal{I}, \mathcal{I}' \subset \{1, ..., m\}$, define the classes of functions*

$$\begin{aligned}
\text{(S.6.33)}\quad \mathcal{G}_1(\mathcal{I}, \mathcal{I}') := \Big\{(z,y) \mapsto \mathbf{a}^\top z^{(\mathcal{I})}(\mathbf{1}\{y \le z^\top \mathbf{b}^{(\mathcal{I}')}\} - \tau)\mathbf{1}\{\|z\| \le \xi_m\} \\
\Big| \tau \in \mathcal{T}, \mathbf{b} \in \mathbb{R}^m, \mathbf{a} \in \mathcal{S}^{m-1}\Big\},
\end{aligned}$$

(S.6.34)
$$\mathcal{G}_2(\delta,\mathcal{I},\mathcal{I}') := \Big\{(z,y)\mapsto \mathbf{a}^\top z^{(\mathcal{I})}(\mathbf{1}\{y\le z^\top \mathbf{b}_1^{(\mathcal{I}')}\}-\mathbf{1}\{y\le z^\top \mathbf{b}_j^{(\mathcal{I}')}\})\mathbf{1}\{z\in\mathcal{Z}\}\Big| \\ \mathbf{b}_1,\mathbf{b}_j\in\mathbb{R}^m, \sup_{v\in\mathcal{Z}}\|v^\top\mathbf{b}_1 - v^\top\mathbf{b}_j\|\le\delta, \mathbf{a}\in\mathcal{S}^{m-1}\Big\}.$$

Under assumptions (A1)-(A3) we have

(S.6.35)
$$P\Big(\|\mathbb{P}_n - P\|_{\mathcal{G}_1} \ge C\Big[\Big(\frac{\max(|\mathcal{I}|,|\mathcal{I}'|)}{n}\log\xi_m\Big)^{1/2} \\ + \frac{\max(|\mathcal{I}|,|\mathcal{I}'|)\xi_m}{n}\log\xi_m + \frac{\kappa_n^{1/2}}{n^{1/2}} + \frac{\xi_m\kappa_n}{n}\Big]\Big) \le e^{-\kappa_n}$$

and for $\xi_m\delta_n \gg n^{-1}$ we have for sufficiently large n and arbitrary $\kappa_n > 0$

(S.6.36)
$$P\Big(\|\mathbb{P}_n - P\|_{\mathcal{G}_2(\delta,\mathcal{I},\mathcal{I}')} \ge C\widetilde{\chi}_n(\delta_n,\mathcal{I},\mathcal{I}',\kappa_n)\Big) \le e^{-\kappa_n}$$

and for $\kappa_n \ge 1$

(S.6.37)
$$P\Big(\sup_{0\le t\le 1}\frac{\|\mathbb{P}_n - P\|_{\mathcal{G}_2(t,\mathcal{I},\mathcal{I}')}}{\widetilde{\chi}_n(t,\mathcal{I},\mathcal{I}',\kappa_n)} \ge 2C\Big) \le e^{-\kappa_n}\log_2 n$$

where

(S.6.38)
$$\widetilde{\chi}_n(t,\mathcal{I},\mathcal{I}',\kappa_n) := t^{1/2}\Big(\frac{\max(|\mathcal{I}|,|\mathcal{I}'|)}{n}\log(\xi_m\vee n)\Big)^{1/2} \\ + \frac{\max(|\mathcal{I}|,|\mathcal{I}'|)\xi_m}{n}\log(\xi_m\vee n) + \Big(\frac{t\kappa_n}{n}\Big)^{1/2} + \frac{\xi_m\kappa_n}{n}.$$

Proof of Lemma S.6.9. See the proof for Lemma C.5 in the Appendix of Chao et al. (2017). The proof of (S.6.37) is similar to the proof for (S.6.21). □

LEMMA S.6.10. *Under assumptions (A1)-(A3) we have for any $\mathbf{a}\in\mathbb{R}^m$*

$$\Big|\mathbf{a}^\top\widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1}\mu(\mathbf{b},\tau) - \mathbf{a}^\top\widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1}\mu(\boldsymbol{\gamma}_N(\tau),\tau) - \mathbf{a}^\top(\mathbf{b}-\boldsymbol{\gamma}_N(\tau))\Big| \\ \le \overline{f'}\sup_x|\mathbf{Z}(x)^\top\mathbf{b} - \mathbf{Z}(x)^\top\boldsymbol{\gamma}_N(\tau)|^2\mathbb{E}[|\mathbf{a}^\top\widetilde{J}_m(\boldsymbol{\gamma}_N(\tau))^{-1}\mathbf{Z}|].$$

Proof of Lemma S.6.10 See the proof for Lemma C.6 in the Appendix of Chao et al. (2017). □

LEMMA S.6.11. *Under assumptions (A1)-(A3) and (L) we have for any* $\mathbf{a} \in \mathbb{R}^m$ *having zero entries everywhere except at* L *consecutive positions:*

$$|\mathbf{a}^\top \mathbb{P}_n \psi(\cdot; \breve{\boldsymbol{\beta}}(\tau), \tau)| \leq \frac{(L+2r)\|\mathbf{a}\|\xi_m}{n}.$$

Proof of Lemma S.6.11. See the proof for Lemma C.7 in the Appendix of Chao et al. (2017). □

APPENDIX S.7: PROOF OF (S.3.21)-(S.3.22) AND (S.3.24)-(S.3.25)

Let $U_i := F_{a,b}(Y_i)$. By strict monotonicity of $F_{a,b}$ on its support we have $U_i \sim U[0,1]$. Moreover $Y^j_{(k)} = Q_{a,b}(U^j_{(k)})$ where $U^j_{(k)}$ denotes the kth order statistic of the sample $\{U_i\}_{i\in\mathcal{A}_j}$ if $\mathcal{A}_j$ is non-empty and $U^j_{(k)} := 0$ if $\mathcal{A}_j$ is empty. By independence of $\{Y_1, ..., Y_n\}$ and $\{X_1, ..., X_n\}$ we have conditionally on $\{X_1, ..., X_n\}$,

$$U^j_{(k)} \sim \mathbf{1}\{n_j > 0\} Beta(k, n_j + 1 - k). \tag{S.7.1}$$

Throughout this section we use the notation introduced in section S.3.3.2.

S.7.1. Proof of (S.3.21) and (S.3.24) (bias). Proof of (S.3.21): note that

$$m^{1/2}\widetilde{\beta}_j(\tau) = Y^j_{(\lceil n_j\tau\rceil)}\mathbf{1}\{n_j > 0\} = Q_{a,b}(U^j_{(\lceil n_j\tau\rceil)})\mathbf{1}\{n_j > 0\} = Q_{a,b}(U^j_{(\lceil n_j\tau\rceil)})$$

where $n_j \sim Bin(n, 1/m)$ and the last equality uses the definition of $U^j_{(\lceil n_j\tau\rceil)}$ and the fact that $Q_{a,b}(0) = 0$. Now we distinguish two cases.

If $\limsup_{N\to\infty}(n/m)\left|\mathbb{E}[U^j_{(\lceil n_j\tau\rceil)} - \tau]\right| > 0$, using the definition of $Q_{a,b}$:

$$\left|\mathbb{E}[Q_{a,b}(U^j_{(\lceil n_j\tau\rceil)}) - Q_{a,b}(\tau)]\right| = \left|a(\mathbb{E}[(U^j_{(\lceil n_j\tau\rceil)})^2 - \tau^2]) + b(\mathbb{E}[U^j_{(\lceil n_j\tau\rceil)} - \tau])\right|, \tag{S.7.2}$$

so (S.3.21) holds for any $b > 0$ and $a = 0$.

If $\limsup_{N\to\infty}(n/m)\left|\mathbb{E}[U^j_{(\lceil n_j\tau\rceil)} - \tau]\right| = 0$ apply (S.7.1) and (S.7.13) in Lemma S.7.1 to obtain

$$\limsup_{N\to\infty} \frac{n}{m}\left|\mathbb{E}[Q_{a,b}(U^j_{(\lceil n_j\tau\rceil)})] - Q_{a,b}(\tau)\right| \geq \frac{a\tau(1-\tau)}{200} > 0 \tag{S.7.3}$$

since $\mathcal{T} = [\tau_L, \tau_U]$ with $0 < \tau_L < \tau_U < 1$. Thus (S.3.21) holds for any $b > 0$ and $0 < a < a_{max}$.

Proof of (S.3.22): note that $Q_{a,b}$ is a polynomial of degree 2 and

$$\left\| \sum_{k=1}^{K} A_k(\cdot) Q_{a,b}(\tau_k) - Q_{a,b}(\cdot) \right\|_\infty = \left\| \Pi_K Q_{a,b} - Q_{a,b} \right\|_\infty$$

where $\Pi_K f$ denotes the projection, with respect to the inner product $\langle f, g \rangle := \sum_k f(\tau_k) g(\tau_k)$, of a function f onto the space of functions spanned by $B_1, ..., B_q$. Since $B_1, ..., B_q$ are splines of degree $r_\tau \geq \eta_\tau \geq 2$, $Q_{a,b}$ lies in the space spanned by $B_1, ..., B_q$ (see p.111 of Schumaker (1981)). In other words, $Q_{a,b} \in \Theta_G$ where Θ_G is defined in the beginning of Section S.3.4.5. Therefore, by similar arguments as in Section S.3.4.5

$$\left\| \Pi_K Q_{a,b} - Q_{a,b} \right\|_\infty \lesssim \inf_{p \in \Theta_G} \| p - Q_{a,b} \|_\infty = 0. \tag{S.7.4}$$

We have from (S.7.4),

$$\begin{aligned} &\mathbb{E}[m^{1/2} \check{\beta}_j(\tau) - Q_{a,b}(\tau)] \\ &= \sum_{k=1}^{K} A_k(\tau) \mathbb{E}[m^{1/2} \widetilde{\beta}_j(\tau_k) - Q_{a,b}(\tau_k)] \\ &= \sum_{k: |\tau_k - \tau| \leq (\log G)^2 / G} A_k(\tau) \mathbb{E}[Q_{a,b}(U^j_{(\lceil n_j \tau_k \rceil)}) - Q_{a,b}(\tau_k)] + o(m/n), \end{aligned} \tag{S.7.5}$$

where the $o(m/n)$ is uniform in $\tau \in \mathcal{T}$. The first equality above is from (S.7.4), and the second follows from the fact that $\sup_\tau \left| \mathbb{E}[m^{1/2} \widetilde{\beta}_j(\tau) - Q_{a,b}(\tau)] \right| \leq 2(a+b)$ and

$$\begin{aligned} &\left| \sum_{k: |\tau_k - \tau| > (\log G)^2 / G} A_k(\tau) \mathbb{E}\big[m^{1/2} \widetilde{\beta}_j(\tau_k) - Q_{a,b}(\tau_k) \big] \right| \\ &\qquad \leq 2(a+b) \sum_{k: |\tau_k - \tau| > (\log G)^2 / G} |A_k(\tau)| \lesssim \gamma^{(\log G)^2 / 2} = o(m/n), \end{aligned} \tag{S.7.6}$$

where $0 < \gamma < 1$; in the second above inequality we apply the bound (S.7.33) in Lemma S.7.3.

Hence, by (S.7.5), it suffices to prove that for $b > 0$ and any $a_{max} > 0$, there exists $a \in [0, a_{max}]$ and a sequence τ_N such that

$$\limsup_{N \to \infty} \frac{n}{m} \left| \sum_{k: |\tau_k - \tau_N| \leq (\log G)^2 / G} A_k(\tau_N) \mathbb{E}\big[Q_{a,b}(U^j_{(\lceil n_j \tau_k \rceil)}) - Q_{a,b}(\tau_k) \big] \right| > 0. \tag{S.7.7}$$

In the following, we will distinguish the cases $\limsup_{N\to\infty}(n/m) < \infty$ and $\limsup_{N\to\infty}(n/m) = \infty$.

First consider the case $\limsup_{N\to\infty}(n/m) < \infty$. Invoke (S.7.17) in Lemma S.7.2 to see that there exist a sequence τ_N in $\mathcal{T}$ such that

$$\limsup_{N\to\infty} \frac{n}{m}\left| \sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N)\mathbb{E}\big[Q_{a,b}(U^j_{(\lceil n_j\tau_k\rceil)}) - Q_{a,b}(\tau_k)\big]\right|$$
$$\ge \limsup_{N\to\infty} \frac{n}{m}\left|\mathbb{E}\big[Q_{a,b}(U^j_{(\lceil n_j\tau_N\rceil)}) - Q_{a,b}(\tau_N)\big]\right|$$

If $\limsup_{N\to\infty}(n/m)\left|\mathbb{E}[U^j_{(\lceil n_j\tau_N\rceil)} - \tau_N]\right| \neq 0$, use (S.7.2), to see that the right-hand side in the equation above is non-zero for any $b > 0$ and $a = 0$.

If $\limsup_{N\to\infty}(n/m)\left|\mathbb{E}[U^j_{(\lceil n_j\tau_N\rceil)} - \tau_N]\right| = 0$ use (S.7.13) in Lemma S.7.1 to obtain

$$\limsup_{N\to\infty} \frac{n}{m}\left|\mathbb{E}\big[Q_{a,b}(U^j_{(\lceil n_j\tau_N\rceil)}) - Q_{a,b}(\tau_N)\big]\right| \ge \limsup_{N\to\infty} \frac{a\tau_N(1-\tau_N)}{200} > 0$$

where the last bound follows since $0 < \tau_L \le \tau_N \le \tau_U < 1$ for all N. Thus we have established (S.7.7) for the case $\limsup_{N\to\infty}(n/m) > 0$.

Next consider the case $\limsup_{N\to\infty}(n/m) = \infty$. By (S.7.18) in Lemma S.7.2 there exist a sequence $\tau_N \in \mathcal{T}$ such that

$$\limsup_{N\to\infty} \frac{n}{m}\left| \sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N)\mathbb{E}\big[Q_{a,b}(U^j_{(\lceil n_j\tau_k\rceil)}) - Q_{a,b}(\tau_k)\big]\right|$$
$$\ge \limsup_{N\to\infty} \frac{n}{m}\left|a\frac{\tau_N(1-\tau_N)}{(n/m)+1} + (b+2a\tau_N) \sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N)\Big(\mathbb{E}[U^j_{(\lceil n_j\tau_k\rceil)}] - \tau_N\Big)\right|.$$

We distinguish two cases.

If $\limsup_{N\to\infty}(n/m)\left|\sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N)\Big(\mathbb{E}[U^j_{(\lceil n_j\tau_k\rceil)}] - \tau_N\Big)\right| > 0$, (S.7.7) holds for τ_N with any $b > 0$ and $a = 0$.

If $\limsup_{N\to\infty}(n/m)\left|\sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N)\Big(\mathbb{E}[U^j_{(\lceil n_j\tau_k\rceil)}] - \tau_N\Big)\right| = 0$, we can pick an arbitrary $0 < a \le a_{max}$ since $\tau_N \in \mathcal{T} = [\tau_L, \tau_U]$ with $0 < \tau_L < \tau_U < 1$ and thus

$$\limsup_{N\to\infty} \frac{n}{m}\left|a\frac{\tau_N(1-\tau_N)}{(n/m)+1}\right| \ge \frac{1}{2}\limsup_{N\to\infty} a\tau_N(1-\tau_N)$$
$$\ge \frac{1}{2}a\min\big\{\tau_L(1-\tau_L), \tau_U(1-\tau_U)\big\} > 0.$$

Thus, (S.7.7) holds when $\limsup_{N\to\infty}(n/m) = \infty$. Combining the two cases above completes the proof of (S.7.7). □

S.7.2. Proof of (S.3.22) and (S.3.25). We begin by proving a slightly stronger version of (S.3.22) for any j:

$$\sup_{\tau\in\mathcal{T}} \mathrm{Var}(\widetilde{\beta}_j(\tau)) \lesssim 1/n. \tag{S.7.8}$$

Observe that, as discussed in te beginning of Section S.7.1, $m^{1/2}\widetilde{\beta}_j(\tau) = Q_{a,b}(U^j_{(\lceil n_j\tau\rceil)})$, and

$$\begin{aligned}
&\mathrm{Var}\big(m^{1/2}\widetilde{\beta}_j(\tau)\big)\\
&= \mathrm{Var}\big(Q_{a,b}(U^j_{(\lceil n_j\tau\rceil)})\big)\\
&= \mathbb{E}\big[\big\{a\big((U^j_{(\lceil n_j\tau\rceil)})^2 - \mathbb{E}[(U^j_{(\lceil n_j\tau\rceil)})^2]\big) + b\big(U^j_{(\lceil n_j\tau\rceil)} - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}]\big)\big\}^2\big]\\
&= \mathbb{E}\big[\big\{a\big((U^j_{(\lceil n_j\tau\rceil)})^2 - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}]^2 - \mathrm{Var}(U^j_{(\lceil n_j\tau\rceil)})\big) + b\big(U^j_{(\lceil n_j\tau\rceil)} - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}]\big)\big\}^2\big]\\
&\le 9\big\{(4a^2+b^2)\mathrm{Var}\big(U^j_{(\lceil n_j\tau\rceil)}\big) + a^2\mathrm{Var}\big(U^j_{(\lceil n_j\tau\rceil)}\big)^2\big\}.
\end{aligned}$$

Therefore, it is enough to show that

$$\sup_{\tau\in\mathcal{T}} \mathrm{Var}\big(U^j_{(\lceil n_j\tau\rceil)}\big) \lesssim m/n. \tag{S.7.9}$$

Distinguish two cases. If m/n is bounded away from zero, this follows since $|U^j_{(\lceil n_j\tau\rceil)}| \le 1$ and thus $\mathrm{Var}\big(U^j_{(\lceil n_j\tau\rceil)}\big)$ is bounded by a constant independent of τ. We will thus without loss of generality assume that $m/n \to 0$. Observe the decomposition

$$\begin{aligned}
\mathrm{Var}(U^j_{(\lceil n_j\tau\rceil)}) &= \mathbb{E}\Big[\Big(U^j_{(\lceil n_j\tau\rceil)} - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}]\Big)^2\Big]\\
&= \mathbb{E}\Big[\Big(U^j_{(\lceil n_j\tau\rceil)} - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}|n_j]\Big)^2\Big] + \mathbb{E}\Big[\Big(\mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}|n_j] - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}]\Big)^2\Big].
\end{aligned}$$

First observe that

$$\begin{aligned}
&\mathbb{E}\Big[\Big(U^j_{(\lceil n_j\tau\rceil)} - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}|n_j]\Big)^2\Big]\\
&= \mathbb{E}\Big[\big(U^j_{(\lceil n_j\tau\rceil)} - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}|n_j]\big)^2\mathbf{1}\Big\{\Big|n_j - \frac{n}{m}\Big| \le 6\sqrt{\frac{n}{m}\log\Big(\frac{n}{m}\Big)}\Big\}\Big] + O\big((n/m)^{-2}\big)\\
&\le \max_{l:|l-(n/m)|\le 6\sqrt{(n/m)\log(n/m)}} \mathbb{E}\Big[\big(U^j_{(\lceil l\tau\rceil)} - \mathbb{E}[U^j_{(\lceil l\tau\rceil)}]\big)^2\Big] + O\big((n/m)^{-2}\big)\\
&= \max_{l:|l-(n/m)|\le 6\sqrt{(n/m)\log(n/m)}} \mathrm{Var}\big(U_{(\lceil l\tau\rceil)}\big) + O\big((n/m)^{-2}\big),\\
&\asymp m/n,
\end{aligned} \tag{S.7.10}$$

uniformly in $\tau \in \mathcal{T}$. Here, the first equality follows since, $\left|(U^j_{(\lceil n_j\tau\rceil)} - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}|n_j])\right|^2 \leq 1$ a.s., $n_j \sim Bin(n, 1/m)$ and the following result which is a consequence of Bernstein's inequality

$$P\Big(\Big|n_j - \frac{n}{m}\Big| \geq C\sqrt{\frac{n}{m}\log\Big(\frac{n}{m}\Big)}\Big) \leq 2\Big(\frac{n}{m}\Big)^{-C/3} \quad \text{for all } C \geq 1 \tag{S.7.11}$$

The final inequality is from (S.7.1).

Next, observe that, almost surely,

$$\Big|\mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}|n_j] - \tau\Big| \leq \frac{1}{n_j+1}$$

by (S.7.1). This implies $|\mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}] - \tau| \leq \mathbb{E}[(1+n_j)^{-1}]$, and thus

$$\mathbb{E}\Big[\Big(\mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}|n_j] - \mathbb{E}[U^j_{(\lceil n_j\tau\rceil)}]\Big)^2\Big] \lesssim \mathbb{E}\big[(1+n_j)^{-2}\big] + (\mathbb{E}[(1+n_j)^{-1}])^2 = O((m/n)^2) \tag{S.7.12}$$

by an application of (S.7.11). Combining (S.7.10) and (S.7.12) finishes the proof of (S.7.9).

To prove (S.3.25), note that by Hölder's inequality and Lemma S.7.3

$$\begin{aligned}
\mathbb{E}[|\check\beta_j(\tau) - \mathbb{E}[\check\beta_j(\tau)]|^2] &= \mathbb{E}\Big[\Big(\sum_{k=1}^K A_k(\tau)\{\widetilde\beta_j(\tau_k) - \mathbb{E}[\widetilde\beta_j(\tau_k)]\}\Big)^2\Big] \\
&\leq \mathbb{E}\Big[\Big(\sum_{k=1}^K |A_k(\tau)|\Big)\Big(\sum_{k=1}^K |A_k(\tau)||\widetilde\beta_j(\tau_k) - \mathbb{E}[\widetilde\beta_j(\tau_k)]|^2\Big)\Big] \\
&\lesssim \sup_{k=1,\dots,K} \mathrm{Var}(\widetilde\beta_j(\tau_k)) \\
&\lesssim 1/n,
\end{aligned}$$

where the last inequality is an application of (S.7.8). □

S.7.3. Auxiliary Lemmas.

LEMMA S.7.1. *Assume that $W_n \sim Bin(n, 1/m)$ and that $U_\tau|W_n = w \sim Beta(\lceil w\tau\rceil, w+1-\lceil w\tau\rceil)$ if $w > 0$ and $U_\tau = 0$ a.s. conditional on $W_n = 0$. Let $Q_{a,b}(x) := ax^2 + bx$. Then for any $a, b \geq 0$ and n, m such that $n \geq 2$, $n/m \geq 1$, we have for all $\tau \in (0,1)$ that*

$$\frac{n}{m}\big|\mathbb{E}[Q_{a,b}(U_\tau) - Q_{a,b}(\tau)]\big| \geq \frac{a\tau(1-\tau)}{200} - \frac{n}{m}b\big|\mathbb{E}[U_\tau - \tau]\big| - 2a\frac{n}{m}\big|\mathbb{E}[U_\tau - \tau]\big|. \tag{S.7.13}$$

Proof of Lemma S.7.1. For a proof of (S.7.13), note that

$$|(\mathbb{E}[U_\tau])^2 - \tau^2| \leq 2|\mathbb{E}[U_\tau] - \tau|.$$

Note also that

$$\mathrm{Var}(U_\tau) = \mathbb{E}[\mathrm{Var}(U_\tau|W)] + \mathrm{Var}(\mathbb{E}[U_\tau|W]) \geq \mathbb{E}[\mathrm{Var}(U_\tau|W)].$$

Thus with an application of triangle inequality, for $a, b \geq 0$,

$$\begin{aligned} \big|\mathbb{E}[Q_{a,b}(U_\tau) - Q_{a,b}(\tau)]\big| &= \big|a(\mathbb{E}[U_\tau^2 - \tau^2]) + b(\mathbb{E}[U_\tau - \tau])\big| \\ &\geq a\mathbb{E}[\mathrm{Var}(U_\tau|W)] - b\big|\mathbb{E}[U_\tau - \tau]\big| - 2a\big|\mathbb{E}[U_\tau - \tau]\big|. \end{aligned} \tag{S.7.14}$$

Next we will prove a lower bound on $\mathbb{E}[\mathrm{Var}(U_\tau|W)]$. The following tail bound on binomial probabilities follows from the Bernstein inequality

$$P\Big(W_n \geq (1+x)\frac{n}{m}\Big) \leq \exp\Big(-\frac{x^2}{2+x}\frac{n}{m}\Big) \quad \forall x \geq 0$$

In particular for $n/m \geq 1$ setting $C = x - 1$ yields for $C \geq 1$

$$P\Big(W_n \geq C\frac{n}{m}\Big) \leq \exp\Big(-\frac{(C-1)^2}{1+C}\Big) \tag{S.7.15}$$

Next note that

$$\mathbb{E}[\mathrm{Var}(U_\tau|W)] = \sum_{w=1}^{\infty} P(W_n = w)\frac{\lceil w\tau\rceil(w+1-\lceil w\tau\rceil)}{(w+1)^2(w+2)},$$

and moreover for any $w > 0$

$$\frac{\tau(1-\tau)}{4(w+2)} \leq \frac{\lceil w\tau\rceil(w+1-\lceil w\tau\rceil)}{(w+1)^2(w+2)} \leq \frac{1}{w+2}.$$

Thus from (S.7.15) we obtain for $n/m \geq 1$ by setting $C = 3$

$$\begin{aligned} \mathbb{E}[\mathrm{Var}(U_\tau|W)] &\geq \frac{\tau(1-\tau)}{4(3(n/m)+2)}\Big(P(W_n > 0) - e^{-1}\Big) \geq \frac{m}{n}\frac{\tau(1-\tau)}{20}\Big(\frac{1}{2} - e^{-1}\Big) \\ &\geq \frac{m}{n}\frac{\tau(1-\tau)}{200}. \end{aligned} \tag{S.7.16}$$

where we used the fact that for $n \geq 2, n/m \geq 1$ we have $P(W_n > 0) \geq 1/2$. The result follows by combining this with (S.7.14).

□

LEMMA S.7.2 (Localization). *Let all conditions in Lemma S.7.1 hold and assume that additionally $N \geq m$, $G \gg (N/m)^{\alpha}$ for some $\alpha > 0$ and $S \gg (N/m)^{1/2}$. Recall the definitions of $A_k, Q_{a,b}$ given in Section S.3.3.2 and assume that $r_\tau \geq 2$.*

1. *If $\limsup_{N\to\infty} n/m < \infty$ there exists a sequence τ_N in $\mathcal{T}$ such that*

(S.7.17)
$$\limsup_{N\to\infty} \frac{n}{m}\Big| \sum_{k:|\tau_k-\tau_N|\leq(\log G)^2/G} A_k(\tau_N)\mathbb{E}[Q_{a,b}(U_{\tau_k}) - Q_{a,b}(\tau_k)]\Big| \geq \limsup_{N\to\infty} (n/m)\big|\mathbb{E}[Q_{a,b}(U_{\tau_N}) - Q_{a,b}(\tau_N)]\big|$$

2. *If $\limsup_{N\to\infty} n/m = \infty$ there exists a sequence τ_N in $\mathcal{T}$ such that*

(S.7.18)
$$\limsup_{N\to\infty} \frac{n}{m}\Big| \sum_{k:|\tau_k-\tau_N|\leq(\log G)^2/G} A_k(\tau_N)\mathbb{E}[Q_{a,b}(U_{\tau_k}) - Q_{a,b}(\tau_k)]\Big| \geq \limsup_{N\to\infty} \frac{n}{m}\Big| a\frac{\tau_N(1-\tau_N)}{(n/m)+1} + (b+2a\tau_N) \sum_{k:|\tau_k-\tau_N|\leq(\log G)^2/G} A_k(\tau_N)\Big(\mathbb{E}[U_{\tau_k}]-\tau_N\Big)\Big|$$

PROOF OF LEMMA S.7.2. For a proof of (S.7.17), for arbitrary $w_0 \in \mathbb{N}$, observe that the sum

$$S_{w_0}(\tau) := \sum_{w=1}^{w_0} P(W_n = w)\big(a\mathbb{E}[U_\tau^2|W_n = w] + b\mathbb{E}[U_\tau|W_n = w]\big) \tag{S.7.19}$$

is a piecewise constant function of τ. This function is constant on intervals of the form $[l_1/w_1, l_2/w_2]$ for some positive integers $l_1, l_2, w_1, w_2 \leq w_0$ satisfying $\tau_L \leq l_1/w_1 < l_2/w_2 \leq \tau_U$. Thus, the length of each interval where S_{w_0} is constant can be bounded from below as follows

$$\frac{l_1}{w_1} - \frac{l_2}{w_2} = \frac{l_1w_2 - l_2w_1}{w_1w_2} \geq w_0^{-2}, \tag{S.7.20}$$

since $l_1w_2 - l_2w_1 \geq 1$ because both l_1w_2 and l_2w_1 are positive integers, and $w_1, w_2 \leq w_0$. Let $w_N = \log G$. Then

(S.7.21)
$$a(\mathbb{E}[U_\tau^2 - \tau^2]) + b(\mathbb{E}[U_\tau - \tau]) = S_{w_N}(\tau) - P(W_n \leq w_N)(b\tau + a\tau^2) + r_G(\tau)$$

where the remainder satisfies $\sup_{\tau\in\mathcal{T}} |r_G(\tau)| \leq 2(a+b)P(W_n > w_N)$, and by (S.7.15)

$$P(W_n > w_N) = o(1), \quad \sup_{\tau\in\mathcal{T}} |r_G(\tau)| = o(m/n) \tag{S.7.22}$$

since $w_N \to \infty$ and we are in the case n/m bounded. For each N, select τ_N as the midpoint of any open interval $\mathcal{J}_N \subset \mathcal{T}$ on which $S_{w_N}(\tau)$ is a constant. From the choice of τ_N and w_N, we have when N is sufficiently large,

$$S_{w_N}(\tau_k) = S_{w_N}(\tau_N) \text{ for all } k : |\tau_k - \tau_N| \le 1/(3w_N^2) \tag{S.7.23}$$

and in particular $S_{w_N}(\tau_k) = S_{w_N}(\tau_N)$ for $k : |\tau_k - \tau_N| \le (\log G)^2/G$ since $G \to \infty$. Hence, for each N, we get from (S.7.21) that

$$\begin{aligned}
&\sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N)\mathbb{E}[Q_{a,b}(U_{\tau_k}) - Q_{a,b}(\tau_k)] \\
&= \sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N)\big[a(\mathbb{E}[U_{\tau_k}^2 - \tau_k^2]) + b(\mathbb{E}[U_{\tau_k} - \tau_k])\big] \\
&= \sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N)\Big\{S_{w_N}(\tau_k) + \Big(- P(W_n \le w_N)(b\tau_k + a\tau_k^2) + r_G(\tau_k)\Big)\Big\}.
\end{aligned} \tag{S.7.24}$$

Note that by similar argument as (S.7.4) and (S.7.6), we have

$$\begin{aligned}
&\Big|\sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} \Big\{A_k(\tau_N)(b\tau_k + a\tau_k^2)\Big\} - (b\tau_N + a\tau_N^2)\Big| \\
&\overset{(S.7.6)}{=} \Big|\sum_{k=1}^{K} \Big\{A_k(\tau_N)(b\tau_k + a\tau_k^2)\Big\} - (b\tau_N + a\tau_N^2)\Big| + o(m/n) \overset{(S.7.4)}{=} o(m/n),
\end{aligned} \tag{S.7.25}$$

and

$$\Big|\sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N) r_G(\tau_k)\Big| \overset{(S.7.6)}{=} \Big|\sum_{k=1}^{K} A_k(\tau_N) r_G(\tau_k)\Big| + o(m/n) = o(m/n). \tag{S.7.26}$$

where the last identity follows from (S.7.21) and Lemma S.7.3. Hence, we

obtain from (S.7.24) that

$$\begin{aligned}
&\limsup_{N\to\infty}\frac{n}{m}\Big|\sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G} A_k(\tau_N)\mathbb{E}[Q_{a,b}(U_{\tau_k})-Q_{a,b}(\tau_k)]\Big| \\
&\overset{\substack{(S.7.25)\\(S.7.26)}}{=} \limsup_{N\to\infty}\frac{n}{m}\Big|\sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G}\Big\{A_k(\tau_N)S_{w_N}(\tau_k)\Big\}-P(W_n\le w_N)(b\tau_N+a\tau_N^2)\Big| \\
&\overset{(S.7.23)}{=} \limsup_{N\to\infty}\frac{n}{m}\Big|S_{w_N}(\tau_N)\sum_{k:|\tau_k-\tau_N|\le(\log G)^2/G}\Big\{A_k(\tau_N)\Big\}-P(W_n\le w_N)(b\tau_N+a\tau_N^2)\Big| \\
&\overset{(S.7.6)}{=} \limsup_{N\to\infty}\frac{n}{m}\Big|S_{w_N}(\tau_N)\sum_{k=1}^{K}\Big\{A_k(\tau_N)\Big\}-P(W_n\le w_N)(b\tau_N+a\tau_N^2)\Big| \\
&= \limsup_{N\to\infty}\frac{n}{m}\big|S_{w_N}(\tau_N)-P(W_n\le w_N)(b\tau_N+a\tau_N^2)+r_G(\tau_N)\big|.
\end{aligned}$$

Here the last line follows from Lemma S.7.3. Apply (S.7.21) to complete the proof of (S.7.17).

For a proof of (S.7.18), note that by the law of iterated expectation,

$$\mathbb{E}\big[Q_{a,b}(U_\tau)-Q_{a,b}(\tau)\big]=a(\mathbb{E}[v_{2,n}(\tau)]+\mathbb{E}[v_{1,n}^2(\tau)]-\tau^2)+b(\mathbb{E}[v_{1,n}(\tau)]-\tau) \tag{S.7.27}$$

where

$$\begin{aligned}
v_{1,n}(\tau)&:=\mathbb{E}[U_\tau|W_n]=\mathbf{1}\{W_n>0\}\frac{\lceil W_n\tau\rceil}{W_n+1};\\
v_{2,n}(\tau)&:=\mathrm{Var}(U_\tau|W_n)=\mathbf{1}\{W_n>0\}\frac{\lceil W_n\tau\rceil(W_n+1-\lceil W_n\tau\rceil)}{(W_n+1)^2(W_n+2)}.
\end{aligned} \tag{S.7.28}$$

The following bound for binomial random variables can be obtained by Bernstein's inequality for $n/m\ge 1$:

$$P\Big(|W_n-n/m|\ge C\sqrt{(n/m)\log(n/m)}\Big)\le 2(n/m)^{-C/3}\quad\text{for all } C\ge 1$$

Using this bound, straightforward calculations show that

$$\begin{aligned}
\mathbb{E}[v_{1,n}(\tau)]&=\tau+O(m/n)\\
\mathbb{E}[v_{2,n}(\tau)]&=\frac{\tau(1-\tau)}{(n/m)+1}+o(m/n)\\
\mathbb{E}[v_{1,n}^2(\tau)-\tau^2]&=2\tau\big(\mathbb{E}[v_{1,n}(\tau)]-\tau\big)+o(m/n)
\end{aligned}$$

where all remainder terms are uniform in $\tau \in \mathcal{T}$. Here, the first equation is a consequence of the fact that $|v_{1,n}(\tau) - \tau| \le (1 + W_n)^{-1}$ while the third equation follows since

$$\mathbb{E}[v_{1,n}^2(\tau) - \tau^2] - 2\tau\big(\mathbb{E}[v_{1,n}(\tau)] - \tau\big) = \mathbb{E}[(v_{1,n}(\tau) - \tau)^2].$$

Thus, from (S.7.27),

(S.7.29)
$$\mathbb{E}\big[Q_{a,b}(U_\tau) - Q_{a,b}(\tau)\big] = a\frac{\tau(1-\tau)}{(n/m)+1} + (b + 2a\tau)\big(\mathbb{E}[v_{1,n}(\tau)] - \tau\big) + r_{1,N}(\tau)$$

where $r_{1,N}(\tau) = o((n/m)^{-1})$ uniformly in $\tau \in \mathcal{T}$. Pick $\tau_N \equiv \tau$ for an arbitrary $\tau \in \mathcal{T}$. From Lemma S.7.3 we obtain

(S.7.30)
$$\Big|\sum_{k:|\tau_k - \tau_N| \le (\log G)^2/G} A_k(\tau_N) r_{1,N}(\tau_k)\Big| \le \sup_{k=1,\dots,K} |r_{1,N}(\tau_k)| \sum_{k=1}^K |A_k(\tau)| = o(m/n).$$

Hence, combine (S.7.30) and (S.7.29) to get

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{n}{m}\Big|\sum_{k:|\tau_k - \tau_N| \le (\log G)^2/G} A_k(\tau_N)\mathbb{E}\big[Q_{a,b}(U_{\tau_k}) - Q_{a,b}(\tau_k)\big]\Big| \\
&\overset{\substack{(S.7.29)\\(S.7.30)}}{=} \limsup_{N\to\infty} \frac{n}{m}\Big|\sum_{k:|\tau_k - \tau_N| \le (\log G)^2/G} A_k(\tau_N)\Big\{a\frac{\tau_k(1-\tau_k)}{(n/m)+1} + (b + 2a\tau_k)\big(\mathbb{E}[v_{1,n}(\tau_k)] - \tau_k\big)\Big\}\Big| \\
&\overset{\substack{(S.7.6)\\(S.7.4)}}{=} \limsup_{N\to\infty} \frac{n}{m}\Big|a\frac{\tau_N(1-\tau_N)}{(n/m)+1} + (b + 2a\tau_N)\sum_{k:|\tau_k - \tau_N| \le (\log G)^2/G} A_k(\tau_N)\big(\mathbb{E}[v_{1,n}(\tau_k)] - \tau_N\big)\Big|
\end{aligned}$$

where the last inequality follows since for all k in the sum $|\tau_k - \tau_N| = O((\log G)^2/G) = o(1)$ and $|\mathbb{E}[v_{1,n}(\tau_k)]| = \tau_k + O(m/n)$ uniformly in τ_k. This completes the proof of (S.7.18). □

LEMMA S.7.3. *Assume that $K \ge G$. Recall the definition of $A_k(\tau)$ in (S.3.23). There exist constants $c, C > 0$ independent of K, G such that*

$$\sup_{k,\tau\in\mathcal{T}} |A_k(\tau)| \le c, \quad \sup_{\tau\in\mathcal{T}} \sum_{k=1}^K |A_k(\tau)| \le C.$$

Moreover, for any $\tau \in \mathcal{T}$, $\sum_{k=1}^K A_k(\tau) = 1$.

PROOF OF LEMMA S.7.3. We start with proving $\sum_{k=1}^{K} A_k(\tau) = 1$. Observe that $\sum_{k=1}^{K} A_k(\tau) = \Pi_K g_1$, where Π_K is defined in Section S.3.4.5 and $g_1(x) \equiv 1$. Since the degree of the piecewise polynomials in $\mathbf{B}$ is greater or equal to 2 and g_1 is an order 0 polynomial, similar arguments as used in the proof of (S.7.4) show that $\|\Pi_K g_1 - 1\|_\infty = 0$. This implies $\sum_{k=1}^{K} A_k(\tau) = 1$ for all $\tau \in \mathcal{T}$.

To obtain a bound for $\sup_{k,\tau\in\mathcal{T}} |A_k(\tau)|$, note that

$$\sup_{k,\tau\in\mathcal{T}} \big|A_k(\tau)\big| = \sup_k \|\Pi_K \overline{e}_k\|_\infty \le \|\Pi_K\|_\infty \sup_k \|\overline{e}_k\|_\infty,$$

where $\overline{e}_k(\cdot) : \mathcal{T} \to [0,1]$ is the function interpolating $\big\{(\tau_1,0),...,\ (\tau_{k-1},0),\ (\tau_k,1),\ (\tau_{k+1},0)...,\ (\tau_K,0)\big\}$, so $\sup_k \|\overline{e}_k\|_\infty = 1$. In the first step in Section S.3.4.5 we have shown that $\|\Pi_K\|_\infty < C'$ for some constant $C' > 0$ independent of K. Hence, taking $c = \|\Pi_K\|_\infty$ finishes the first claim.

Next we show that $\sup_{\tau\in\mathcal{T}} \sum_{k=1}^{K} |A_k(\tau)| \le C$. For any $\tau \in \mathcal{T}$ and $\tau_k \in \mathcal{T}_K$, define the index sets

$$\mathcal{I}(\tau) := \big\{j \in \{1,...,q\} : B_j(\tau) \neq 0\big\};$$
$$\mathcal{I}_k := \big\{j \in \{1,...,q\} : B_j(\tau_k) \neq 0\big\}.$$

Note that $|\mathcal{I}(\tau)| = |\mathcal{I}(\tau_k)| = r_\tau + 1$ for all $\tau \in \mathcal{T}$ and k, and the elements in both sets are consecutive positive integers by the connective support of B_j.

By Lemma 6.3 of Zhou et al. (1998) (note that their $\mathbf{N}$, n and $G_{k,n}$ are our $\mathbf{B}$, K and $K^{-1}\sum_{k=1}^{K} \mathbf{B}(\tau_k)\mathbf{B}(\tau_k)^\top$, while their h is of order $1/G$ see page 1761, 1762 and 1765 of Zhou et al. (1998)), there exist constants $0 < \gamma < 1$ and $c_1 > 0$ independent of K, G, τ such that the ij'th element of the matrix satisfies

$$\Big(\sum_{k=1}^{K} \mathbf{B}(\tau_k)\mathbf{B}(\tau_k)^\top\Big)^{-1}_{ij} \le c_1 K^{-1} G \gamma^{|j-i|} \le c_1 \gamma^{|j-i|} \tag{S.7.31}$$

where the second inequality follows from our assumption that $K \ge G$.

Recall that our B_j is defined as given in Definition 4.19 on page 124 of Schumaker (1981). Thus, by equation (4.31) in the same reference,

$$\sup_{\tau\in\mathcal{T}} \max_{j\le q} |B_j(\tau)|^2 \le 1,$$

and by (S.7.31),

(S.7.32)
$$\big|A_k(\tau)\big| = \Big|\mathbf{B}(\tau)^\top \Big(\sum_{k=1}^{K} \mathbf{B}(\tau_k)\mathbf{B}(\tau_k)^\top\Big)^{-1} \mathbf{B}(\tau_k)\Big| \le c_1 \sum_{j\in\mathcal{I}(\tau), j'\in\mathcal{I}_k} \gamma^{|j-j'|}$$

(S.7.33)
$$\le c_1 \sum_{l=0}^{\infty} \big(\big|(\mathcal{I}(\tau)+l)\cap\mathcal{I}_k\big| + \big|(\mathcal{I}(\tau)-l)\cap\mathcal{I}_k\big|\big)\gamma^l,$$

where $\mathcal{I}(\tau)-l = \{j-l : j\in\mathcal{I}(\tau)\}$, $\mathcal{I}(\tau)+l = \{j+l : j\in\mathcal{I}(\tau)\}$. Now, we make a key observation that $\big(\big|(\mathcal{I}(\tau)+l)\cap\mathcal{I}_k\big| + \big|(\mathcal{I}(\tau)-l)\cap\mathcal{I}_k\big|\big) \neq 0$ for at most $6(r_\tau+1)$ many k at each fixed $l\in\mathbb{N}$. This is due to the fact that both $\mathcal{I}(\tau)$ and $\mathcal{I}_k$ consist of $r_\tau+1$ *consecutive* positive integers, and the same is true for both $\mathcal{I}(\tau)-l$ and $\mathcal{I}(\tau)+l$. Hence, from (S.7.33) and Fubini's theorem,

$$\sum_{k=1}^{K}\big|A_k(\tau)\big| \le c_1 \sum_{l=0}^{\infty}\sum_{k=1}^{K} \big(\big|(\mathcal{I}(\tau)+l)\cap\mathcal{I}_k\big| + \big|(\mathcal{I}(\tau)-l)\cap\mathcal{I}_k\big|\big)\gamma^l$$
$$\le 6c_1(r_\tau+1)\sum_{l=0}^{\infty}\gamma^l < \infty.$$

This completes the proof. □

APPENDIX S.8: TECHNICAL RESULTS FROM EMPIRICAL PROCESS THEORY

In this section, we collect some basic results from empirical process theory that we use throughout the proofs. Denote by $\mathcal{G}$ a class of functions that satisfies $|f(x)| \le F(x) \le U$ for every $f\in\mathcal{G}$ and let $\sigma^2 \ge \sup_{f\in\mathcal{G}} Pf^2$. Additionally, let for some $A>0, V>0$ and all $\varepsilon>0$

(S.8.1)
$$N(\varepsilon, \mathcal{G}, L_2(\mathbb{P}_n)) \le \Big(\frac{A\|F\|_{L^2(\mathbb{P}_n)}}{\varepsilon}\Big)^V.$$

In that case, the symmetrization inequality and inequality (2.2) from Koltchinskii (2006) yield

(S.8.2)
$$\mathbb{E}\|\mathbb{P}_n - P\|_{\mathcal{G}} \le c_0\Big[\sigma\Big(\frac{V}{n}\log\frac{A\|F\|_{L^2(P)}}{\sigma}\Big)^{1/2} + \frac{VU}{n}\log\frac{A\|F\|_{L^2(P)}}{\sigma}\Big]$$

for a universal constant $c_0>0$ provided that $1\ge\sigma^2 > \text{const}n^{-1}$ [in fact, the inequality in Koltchinskii (2006) is for $\sigma^2 = \sup_{f\in\mathcal{G}} Pf^2$. However, this

is not a problem since we can replace $\mathcal{G}$ by $\mathcal{G}\sigma/(\sup_{f\in\mathcal{G}} Pf^2)^{1/2}]$. The second inequality (a refined version of Talagrand's concentration inequality) states that for any countable class of measurable functions $\mathcal{F}$ with elements mapping into $[-M, M]$

(S.8.3)
$$P\Big\{\|\mathbb{P}_n - P\|_{\mathcal{F}} \geq 2\mathbb{E}\|\mathbb{P}_n - P\|_{\mathcal{F}} + c_1 n^{-1/2}\Big(\sup_{f\in\mathcal{F}} Pf^2\Big)^{1/2}\sqrt{t} + n^{-1}c_j M t\Big\} \leq e^{-t},$$

for all $t > 0$ and universal constants $c_1, c_j > 0$. This is a special case of Theorem 3 in Massart (2000) [in the notation of that paper, set $\varepsilon = 1$].

LEMMA S.8.1. *Under (A1)-(A3) and $\xi_m^2(\log n)^2 = o(N)$ we have for any $\delta > 0$ and $\mathbf{u}_N \in \mathbb{R}^m$,*

(S.8.4)
$$\lim_{\delta\to 0}\limsup_{N\to\infty} P\Big(\|\mathbf{u}_N\|_2^{-1} N^{1/2} \sup_{\tau_1,\tau_2\in\mathcal{T},|\tau_1-\tau_2|\leq\delta} \Big|\mathbf{u}_N^\top \boldsymbol{U}_N(\tau_1) - \mathbf{u}_N^\top \boldsymbol{U}_N(\tau_2)\Big| > \varepsilon\Big) = 0,$$

where $\boldsymbol{U}_N(\tau)$ is defined in (S.3.35).

Proof of Lemma S.8.1. See the proof for Lemma A.3 in Chao et al. (2017). □

LEMMA S.8.2. *Let p be a polynomial of degree q. Then for arbitrary intervals $[a, b]$ and $\alpha \in [0, 1)$ we have*

$$\Big|\int_a^b p^2(x)dx - \frac{b-a}{J}\sum_{j:j+\alpha\leq J} p^2\Big(a + \frac{j+\alpha}{J}(b-a)\Big)\Big| \leq \frac{C_q}{J-1}\int_a^b p^2(x)dx$$

for a constant $C_q > 1$ that depends only on q.

Proof of Lemma S.8.2. We begin by observing that it suffices to prove that for any polynomial p of degree q we have

$$\Big|\int_0^1 p^2(x)dx - \frac{1}{J}\sum_{j:j+\alpha\leq J} p^2\Big(\frac{j+\alpha}{J}\Big)\Big| \leq \frac{C_q}{J}\int_0^1 p^2(x)dx. \tag{S.8.5}$$

To see this, note that for any polynomial p of degree q the function $\widetilde{p}_{a,b}(x) := p(a + x(b-a))(b-a)^{1/2}$ is also a polynomial of degree q and we have $\int_0^1 \widetilde{p}_{a,b}^2(x)dx = \int_a^b p^2(x)dx$ by a change of variables and

$$\frac{b-a}{J}\sum_{j:j+\alpha\leq J} p^2\Big(a + \frac{j+\alpha}{J}(b-a)\Big) = \frac{1}{J}\sum_{j:j+\alpha\leq J} \widetilde{p}_{a,b}^2\Big(\frac{j+\alpha}{J}\Big).$$

To prove (S.8.5), consider the bound

$$\Big|\int_0^1 p^2(x)dx - \frac{1}{J}\sum_{j:j+\alpha\le J} p^2\Big(\frac{j+\alpha}{J}\Big)\Big| \le \frac{1}{J-1}\sup_{x\in[0,1]}\Big|\frac{dp^2(x)}{dx}\Big| + \frac{2}{J}\sup_{x\in[0,1]} p^2(x),$$

where the first term on the right is from the difference of the two integrals on the interior partitions, and the second term handles the boundary partitions which are not included in the summation on the left. Thus we need to bound $\sup_{x\in[0,1]}\left\{\left|\frac{dp^2(x)}{dx}\right| + p^2(x)\right\}$ by $C_q\int_0^1 p^2(x)dx$. To this end, denote by $\mathcal{P}_q$ the space of polynomials on $[0,1]$ with degree less or equal to q. This is a finite-dimensional vector space. Note that the operator $\Phi : f \mapsto f' + f$ is a linear operator between the finite-dimensional, normed spaces $(\mathcal{P}_{2q}, \|\cdot\|_1)$ and $(\mathcal{P}_{2q}, \|\cdot\|_\infty)$. Such an operator must be bounded, and hence there exists a constant C_q with $\|\Phi(p^2)\|_\infty \le C_q\|p^2\|_1$. This completes the proof. □

APPENDIX S.9: TECHNICAL DETAILS FOR SIMULATION AND ADDITIONAL SIMULATION RESULTS

This part of the supplement contains additional results and some technical details for the simulation study in Section 5. Section S.9.1 contains some technical details related to the simulation settings in Section 5, and additional figures (the case $m = 16$ and coverage probabilities of confidence intervals for $F_{Y|X}(y|x_0)$ and large values of S). In Section S.9.3, using the same linear model as in Section 5, we provide additional simulation results for the oracle properties of the divide and conquer estimator $\overline{\boldsymbol{\beta}}(\tau)$ and the distribution function estimator $\widehat{F}_{Y|X}(y|x)$; the emphasis of this section is on illustrating our oracle theory. In Section S.9.4, we consider a non-linear model and illustrate oracle properties of the divide and conquer estimator $\overline{\boldsymbol{\beta}}(\tau)$ and the distribution function estimator $\widehat{F}_{Y|X}(y|x)$. Some practical conclusions from the simulations are summarized in Section S.9.5

S.9.1. Details for Section 5 and additional simulation results. The simple pooled estimator for the asymptotic covariance matrix of $\boldsymbol{\beta}_{or}(\tau) - \boldsymbol{\beta}(\tau)$ was computed by first using the options `se='ker'`, `covariance=TRUE` in the `rq` function in `quantreg` (Koenker, 2016) for each sub-sample, and then averaging the results over sub-samples.

The values of the optimal constant $c^*(\tau)$ were determined from the theory developed in Kato (2012). More precisely, we use the following formula given on its page 264 of the latter reference,

$$c^*(\tau) = \sigma\left(\frac{4.5\sum_{j,k=1}^m \mathbb{E}[Z_j^2 Z_k^2]}{\alpha(\tau)\sum_{j,k=1}^m (\mathbb{E}[Z_j Z_k])^2}\right)^{1/5}, \tag{S.9.1}$$

where $\alpha(\tau) = (1-\Phi^{-1}(\tau)^2)^2\phi(\Phi^{-1}(\tau))$ (note that this is a corrected version of Kato's formula). Note that $c^*(\tau)$ is symmetric around $\tau = 0.5$. In order to save computation time in the simulation, we simulated an independent data set to compute $c^*(\tau)$, and used the same $c^*(\tau)$ (based on $m = 4$) for different m in all the simulations, because the exact value of $c^*(\tau)$ does not change much with m under our design of $\mathbf{Z}$; see Table S.9.1.

	$m = 4$	$m = 16$	$m = 32$
$\tau = 0.1$	0.242	0.252	0.254
0.5	0.173	0.179	0.180

TABLE S.9.1
Values of $c^*(\tau)$ based on (S.9.1).

In the following, we show additional simulation results. Figure S.9.1 shows the coverage probability of the confidence interval for $Q(x_0;\tau)$ based on $x_0^\top \bar{\boldsymbol{\beta}}(\tau)$ for dimension $m = 16$. For additional simulations for $F_{Y|X}(y|x_0)$

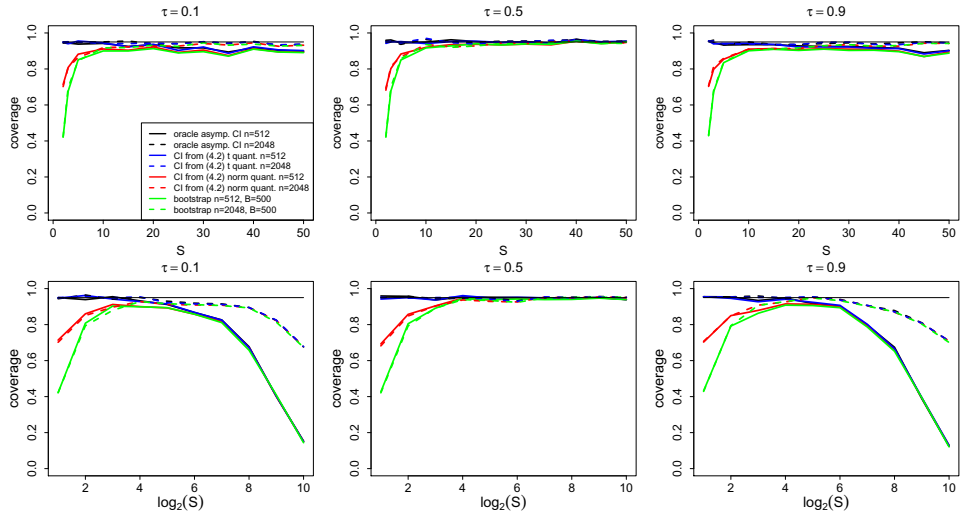


Fig S.9.1: Coverage probabilities for $Q(x_0;\tau)$ for $m = 16$ in homoskedastic model (5.1). Coverage probabilities for $x_0^\top \boldsymbol{\beta}(\tau)$ for different values of S and $\tau = 0.1, 0.5, 0.9$ (left, middle, right row). Solid lines: $n = 512$, dashed lines: $n = 2048$. Black: asymptotic oracle CI, blue: CI from (4.2) based on t distribution, red: CI from (4.1) based on normal distribution, green: bootstrap CI. Throughout $x_0 = (1, ..., 1)/m^{1/2}$, nominal coverage 0.95.

under the same settings as that of Figure 5.2 in Section 5, Figure S.9.2 shows

the coverage probability of the confidence interval with large S for $m = 4$ and 32 and Figure S.9.3 shows the coverage probability of the confidence interval for $m = 16$.

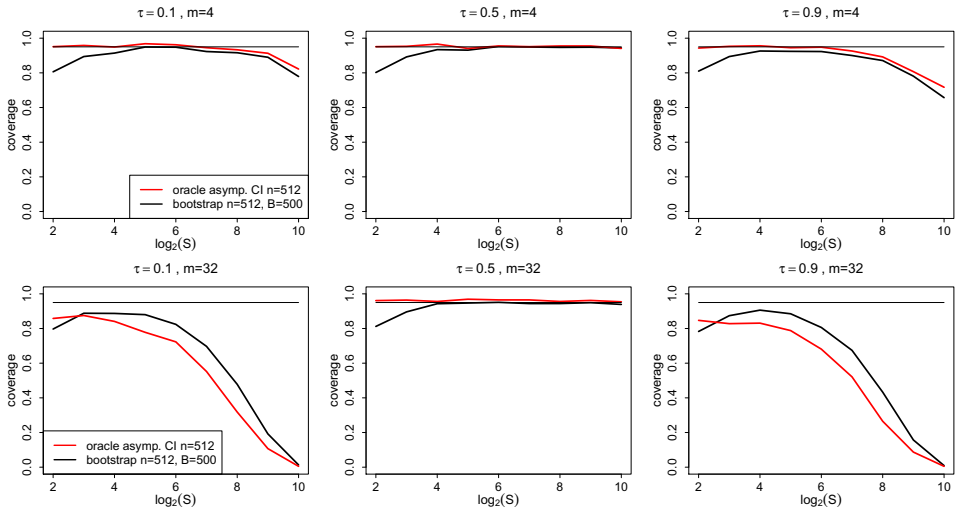


Fig S.9.2: Coverage probabilities for $F_{Y|X}(y|x_0)$ in homoskedastic model (5.1) with $m = 4, 32$ with large number of subsamples S, $n = 512$ and nominal coverage 0.95. Red: oracle asymptotic CI, black: bootstrap CI with $B = 500$.

S.9.2. Heteroskedastic model. We consider a linear location-scale shift model,

$$Y_i = 0.21 + \boldsymbol{\beta}_{m-1}^\top X_i + (1 + \mathbf{z}_{m-1}^\top X_i)\varepsilon_i, \quad i = 1, ..., N. \tag{S.9.2}$$

where $m \in \{4, 16, 32\}$, $\boldsymbol{\beta}_{m-1}$ is chosen as (S.9.3), X_i and ε_i are chosen as model (5.1), and the vector $\mathbf{z}_{m-1}$ takes the form

$$\begin{aligned} \mathbf{z}_3 &= (0.69, 0.56, 0.35)^\top; \\ \mathbf{z}_{15} &= (\mathbf{z}_3^\top, \mathbf{z}_3^\top, \mathbf{z}_3^\top, \mathbf{z}_3^\top, \mathbf{z}_3^\top)^\top; \\ \mathbf{z}_{31} &= (\boldsymbol{\beta}_{15}^\top, \boldsymbol{\beta}_{15}^\top, 0.69)^\top. \end{aligned} \tag{S.9.3}$$

S.9.2.1. *Results for the divide and conquer estimator $\bar{\boldsymbol{\beta}}(\tau)$.* We fix the sub-sample size n and consider the impact of S on the coverage probabilities of various 95% confidence intervals as in Section 5.1. We use the infeasible

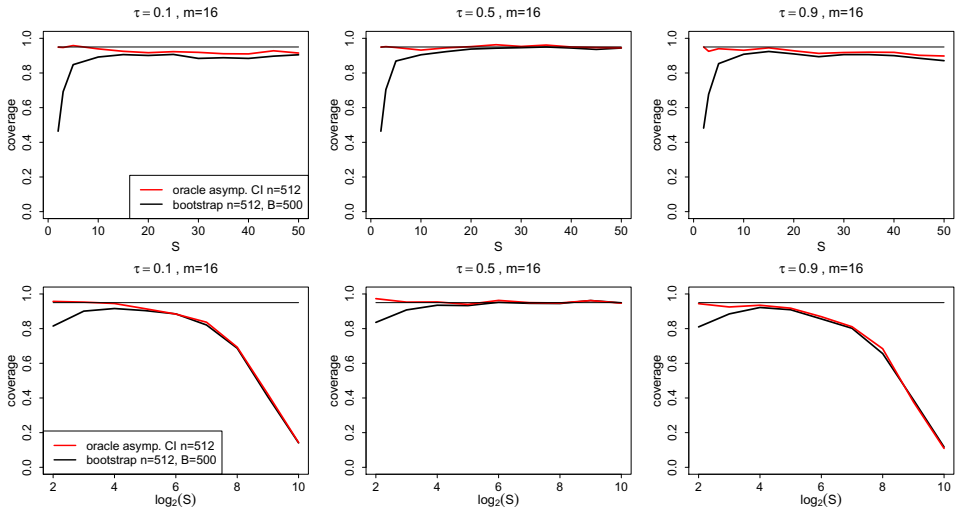


Fig S.9.3: Coverage probabilities in homoskedastic model (5.1) for $F_{Y|X}(y|x_0)$ for $m = 16$, $n = 512$ and nominal coverage 0.95. Red: oracle asymptotic CI, black: bootstrap CI with $B = 500$.

asymptotic confidence interval (5.3) as benchmark (note that the asymptotic variance is different in the homoskedastic and heteroskedastic model).

In the first step, as in Section 5.1, we consider the three types of confidence intervals in Section 4.1. The coverage probabilities of the corresponding confidence intervals are presented in Figure S.9.4. The general patterns are similar to those in Figure 5.1, but coverage starts to drop earlier for both $m = 4, 32$. The same conclusion holds for the coverage probabilities for dimension $m = 16$ when comparing Figure S.9.5 to Figure S.9.1 for the homoskedastic model.

Next, we analyse the asymptotic confidence intervals using the empirical asymptotic variance of $x_0^\top \boldsymbol{\beta}(\tau)$ from data, which is estimated with the three ways described on p.22 in Section 5.1. The constants $c^*(\tau)$ there are replaced by $c_g^*(\tau)$, which is computed by adapting the formula on page 263 of Kato (2012) to the $\mathcal{N}(0, \sigma^2)$ location-scale shift model (S.9.2). This yields the formula:

$$c_g^*(\tau) = \sigma\left(\frac{4.5\sum_{j,k=1}^m \mathbb{E}[(1+\mathbf{z}_{m-1}^\top X)^{-1} Z_j^2 Z_k^2]}{\alpha(\tau)\sum_{j,k=1}^m (\mathbb{E}[(1+\mathbf{z}_{m-1}^\top X)^{-3} Z_j Z_k])^2}\right)^{1/5}, \tag{S.9.4}$$

where $\alpha(\tau) = (1-\Phi^{-1}(\tau)^2)^2\phi(\Phi^{-1}(\tau))$. The values are listed in Table S.9.2. The results for the coverage probabilities are reported in Table S.9.3. The

	$m = 4$	$m = 16$	$m = 32$
$\tau = 0.1$	0.4206	1.2118	2.3222
0.5	0.2990	0.8614	1.6508

TABLE S.9.2

Values of $c_g^*(\tau)$ based on (S.9.4) of the linear Gaussian location-scale shift model (5.1) .

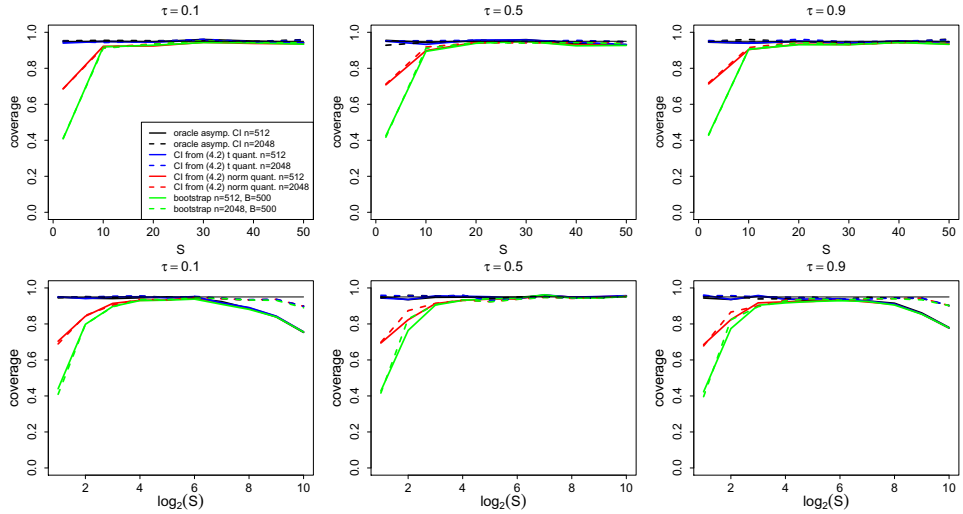
overall patterns are similar to those in Table 5.1, but the coverage probabilities start to drop earlier, this effect is similar to what we observed in Figure S.9.4.

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	94.1	95.0	93.9	95.1	94.0	92.7	90.5	94.3	91.6	83.0	73.8
def	92.8	92.7	93.8	92.6	94.9	93.5	91.9	89.8	94.6	90.4	80.2	71.3
nai	93.7	93.1	94.4	93.0	97.7	93.2	91.2	88.5	99.4	91.6	80.9	72.0
adj	93.7	93.6	94.9	93.8	97.7	94.1	92.9	90.9	99.4	93.4	83.7	75.6
	n = 512, m = 4, $\tau = 0.5$				n = 512, m = 16, $\tau = 0.5$				n = 512, m = 32, $\tau = 0.5$			
or	95.4	94.7	95.5	94.9	95.4	96.2	95.3	94.3	95.3	94.3	95.7	94.7
def	97.4	97.0	97.8	97.4	99.3	99.3	98.8	99.0	98.4	98.6	98.6	98.2
nai	96.0	95.6	96.1	95.8	98.6	98.0	97.3	97.1	99.0	97.5	97.2	97.4
adj	96.0	95.2	95.6	95.1	98.6	97.4	96.1	95.6	99.0	96.7	96.4	96.1
	n = 512, m = 4, $\tau = 0.9$				n = 512, m = 16, $\tau = 0.9$				n = 512, m = 32, $\tau = 0.9$			
or	95.0	95.2	93.9	94.5	95.3	94.2	92.3	91.6	94.7	90.9	80.1	72.7
def	92.9	94.5	92.5	93.6	94.0	93.8	91.6	90.8	94.0	88.5	79.2	69.8
nai	93.8	95.0	93.0	93.7	97.1	93.3	90.8	89.8	99.4	90.0	80.1	70.6
adj	93.8	95.2	93.6	94.5	97.1	94.6	92.4	91.8	99.4	91.6	83.2	73.9
	n = 2048, m = 4, $\tau = 0.1$				n = 2048, m = 16, $\tau = 0.1$				n = 2048, m = 32, $\tau = 0.1$			
or	95.1	96.3	94.9	95.1	95.6	94.1	94.8	93.8	95.2	95.0	90.5	88.2
def	94.0	95.8	94.2	94.8	94.6	93.5	93.7	92.8	94.7	94.3	89.3	87.0
nai	94.6	95.8	94.4	94.8	95.7	93.1	93.2	92.4	95.6	94.2	89.1	86.8
adj	94.6	96.0	94.6	95.0	95.7	93.8	94.4	93.4	95.6	94.9	90.2	87.9
	n = 2048, m = 4, $\tau = 0.5$				n = 2048, m = 16, $\tau = 0.5$				n = 2048, m = 32, $\tau = 0.5$			
or	95.1	94.4	95.6	94.6	94.8	95.0	95.4	95.1	95.2	95.1	95.2	95.0
def	95.9	95.5	96.6	95.7	97.6	97.8	98.0	98.0	97.5	97.1	97.6	97.4
nai	95.4	94.9	96.2	95.0	96.6	96.8	96.6	96.4	97.0	96.4	96.4	96.4
adj	95.4	94.6	95.8	94.8	96.6	95.9	95.8	95.6	97.0	96.0	95.4	95.6
	n = 2048, m = 4, $\tau = 0.9$				n = 2048, m = 16, $\tau = 0.9$				n = 2048, m = 32, $\tau = 0.9$			
or	95.1	94.9	95.6	94.6	95.0	94.8	94.2	93.2	95.3	94.5	91.0	89.4
def	94.2	94.4	95.1	94.1	94.0	94.2	93.4	92.1	94.8	93.8	90.1	88.2
nai	94.6	94.5	95.1	94.1	95.4	93.9	93.0	91.8	95.6	93.8	90.0	87.6
adj	94.6	94.7	95.6	94.4	95.4	94.4	94.0	93.2	95.6	94.8	91.1	89.4

TABLE S.9.3

Coverage probabilities under the linear Gaussian location-scale shift mode (S.9.2) 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.

(a) $m = 4$



(b) $m = 32$

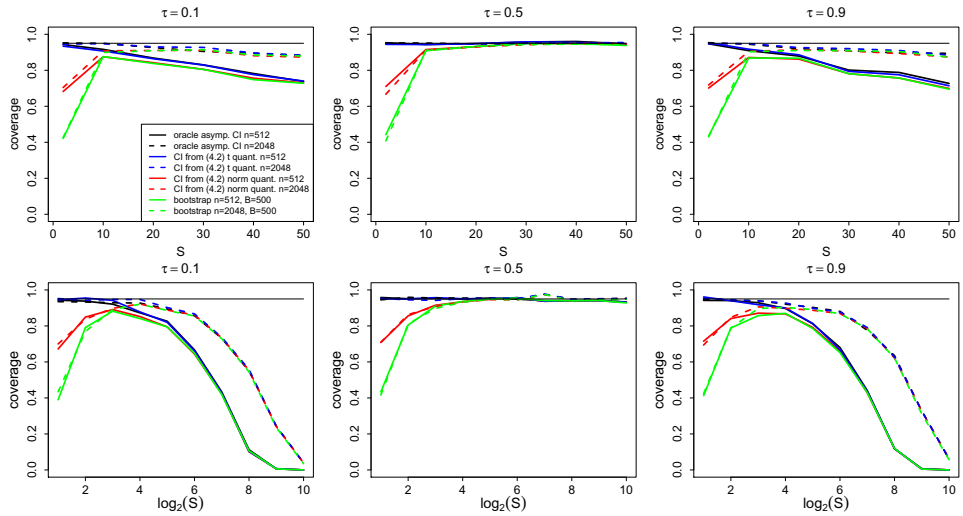


Fig S.9.4: Coverage probabilities under the linear Gaussian location-scale shift model (S.9.2) for $x_0^\top \boldsymbol{\beta}(\tau)$ for different values of S and $\tau = 0.1, 0.5, 0.9$ (left, middle, right row). Solid lines: $n = 512$, dashed lines: $n = 2048$. Black: asymptotic oracle CI, blue: CI from (4.2) based on t distribution, red: CI from (4.1) based on normal distribution, green: bootstrap CI. Throughout $x_0 = (1, ..., 1)/m^{1/2}$, nominal coverage 0.95.

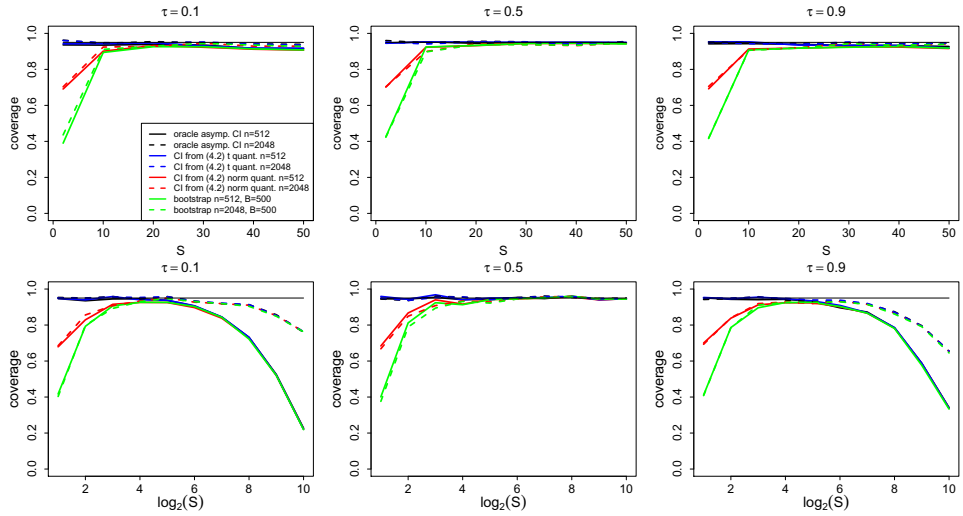


Fig S.9.5: Coverage probabilities for $Q(x_0;\tau)$ for $m = 16$ under the linear Gaussian location-scale shift model (S.9.2). Coverage probabilities for $x_0^\top \boldsymbol{\beta}(\tau)$ for different values of S and $\tau = 0.1, 0.5, 0.9$ (left, middle, right row). Solid lines: $n = 512$, dashed 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 $x_0 = (1, ..., 1)/m^{1/2}$, nominal coverage 0.95.

S.9.2.2. *Results for the estimator* $\widehat{F}_{Y|X}(y|x)$. In this section, we compare the coverage probabilities of the bootstrap confidence intervals for $F_{Y|X}(Q(x_0;\tau)|x_0) = \tau$ with the oracle asymptotic confidence interval (5.4). To estimate $\widehat{F}_{Y|X}$, we use the same number of quantile grid $K = 65$ and the same number of equidistant knots $G = 32$ for spline interpolation as in Section 5.2, which are sufficiently large to ensure the nominal coverage of the oracle confidence intervals. Coverage probabilities for $m = 4, 32$ are reported in Figure S.9.6. Results for large values of S and dimension $m = 16$ are presented in Figure S.9.7 and S.9.8. The general patterns are similar to the homoskedastic model with overall coverage starting to drop earlier.

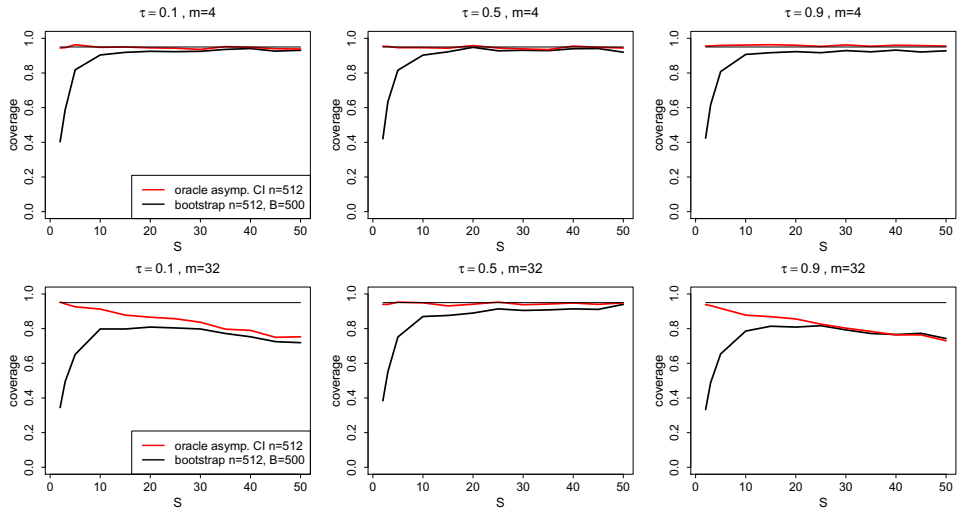


Fig S.9.6: Coverage probabilities for oracle confidence intervals (red) and bootstrap confidence intervals (black) for $F_{Y|X}(Q(x_0;\tau)|x_0) = \tau$ under the linear Gaussian location-scale shift model (S.9.2) for $x_0 = (1, ..., 1)/m^{1/2}$ and $y = Q(x_0;\tau)$, $\tau = 0.1, 0.5, 0.9$. $n = 512$ and nominal coverage 0.95.

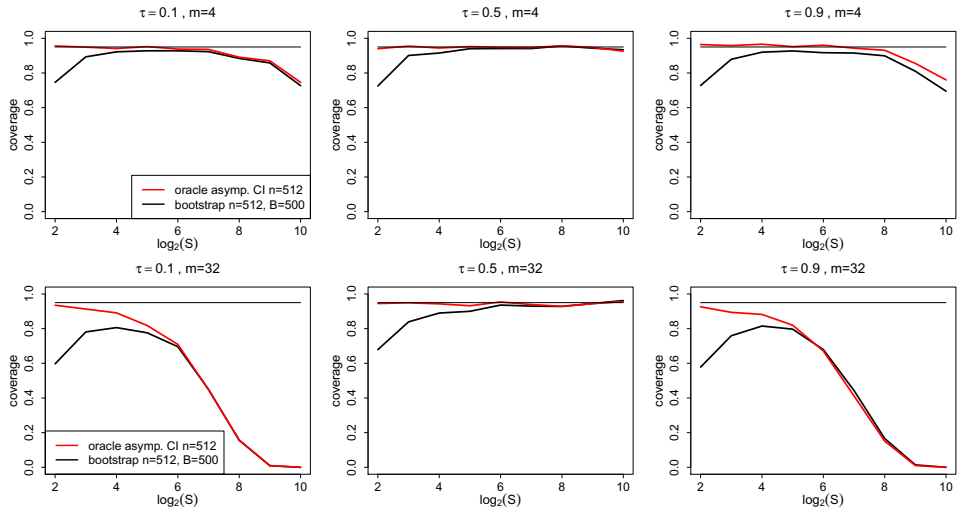


Fig S.9.7: Coverage probabilities for $F_{Y|X}(Q(x_0;\tau)|x_0) = \tau$ under the linear Gaussian location-scale shift model (S.9.2) for $m = 4, 32$ with large number of subsamples S, $n = 512$ and nominal coverage 0.95. Red: oracle asymptotic CI, black: bootstrap CI with $B = 500$.

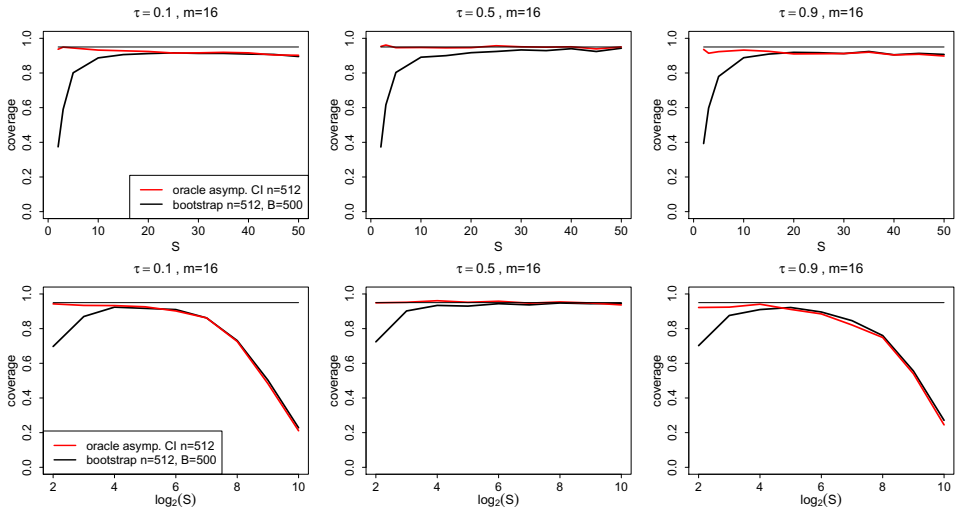


Fig S.9.8: Coverage probabilities for $F_{Y|X}(Q(x_0;\tau)|x_0) = \tau$ under the linear Gaussian location-scale shift model (S.9.2) for $m = 16$, $n = 512$ and nominal coverage 0.95. Red: oracle asymptotic CI, black: bootstrap CI with $B = 500$.

S.9.3. Additional simulations for the linear model. The purpose of the simulations in this section is to show the thresholds for the number of sub-samples S^* and the number of quantile grids K^* that guarantee the oracle properties of $\overline{\boldsymbol{\beta}}(\tau)$ and $\widehat{\boldsymbol{\beta}}(\tau)$ predicted by the sufficient and necessary conditions in the theorems in Section 3. Besides normal location shift models, we also consider models with exponential error distributions. Using the linear model (5.1) in Section 5, we show the coverage probabilities of the oracle asymptotic confidence intervals, i.e. the confidence intervals based on the true asymptotic variances. Similar to Section 5, we select $\tau = 0.1, 0.5, 0.9$, $m = 4, 16, 32$ and $\mathcal{T} = [0.05, 0.95]$. Section S.9.3.1 contains results for $\overline{\boldsymbol{\beta}}(\tau)$ defined in (2.3), while Section S.9.3.2 discusses $\widehat{F}_{Y|X}(y|x)$ defined in (2.8).

S.9.3.1. *Oracle rule for* $\overline{\boldsymbol{\beta}}(\tau)$. Recall that from Corollary 3.3, an asymptotic $1-\alpha$ confidence interval for $x_0^\top \boldsymbol{\beta}(\tau)$ is

$$\big[x_0^\top \overline{\boldsymbol{\beta}}(\tau) \pm N^{-1/2} f_{\varepsilon,\tau}^{-1} \sqrt{\tau(1-\tau) x_0^\top \Sigma_X^{-1} x_0}\, \Phi^{-1}(1-\alpha/2)\big], \tag{S.9.5}$$

where $f_{\varepsilon,\tau} = f_\epsilon(F_\epsilon^{-1}(\tau))$, f_ϵ is the error density and $\Sigma_X = \mathbb{E}[(1, X_i^\top)^\top (1, X_i^\top)]$. Here, we set $x_0 = (1, (m-1)^{-1/2} \mathbf{l}_{m-1}^\top)^\top$, where $\mathbf{l}_{m-1}$ is an $(m-1)$ vector with each entry equal to 1.

We verify the oracle rule by checking whether the empirical coverage probability of (5.3) equals $1-\alpha = 95\%$. Figure S.9.9 shows the coverage probabilities for $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. In all plots, the coverage probabilities reach the nominal level 95% for $S < S^*$, and then drop quickly to 0 for $S > S^*$ for some S^*. When N increases, S^* shifts toward $N^{1/2}$, which is the sharp upper bound in Corollary 3.3. Also, when N is fixed, a larger dimensionality m leads to a smaller S^*. For $\varepsilon \sim \text{Exp}(\lambda)$, Figure S.9.10 shows that the coverage probabilities are no longer symmetric in τ. When τ is small, N needs to be large enough to attain the same magnitude of S^* due to the skewness of $\text{Exp}(\lambda)$.

S.9.3.2. *Oracle rule for* $\widehat{F}_{Y|X}(y|x)$. We compute $\widehat{F}_{Y|X}(y_0|x_0)$ defined in (2.8) discretely with an equidistant partition of size 1000 on $[\tau_L, \tau_U]$. The involved $\widehat{\boldsymbol{\beta}}(\tau)$ is computed as in (2.5) with $\mathbf{B}$ being cubic B-spline with dimension q defined on $G = 4+q$ knots. The knots form a partition on $[\tau_L, \tau_U]$ with repetitions on the boundary; see Corollary 4.10 of Schumaker (1981). For simplicity, we set $y_0 = Q(x_0; \tau)$ so that $F_{Y|X}(y_0|x_0) = \tau$, where x_0 is chosen as in Section S.9.3.1. Recall from Corollary 3.5 that an asymptotic $1-\alpha$ confidence interval for $F_{Y|X}(Q(x_0;\tau)|x_0) = \tau$ is

$$\big[\widehat{F}_{Y|X}(Q(x_0;\tau)|x_0) \pm N^{-1/2}\sqrt{\tau(1-\tau)x_0^\top \Sigma_X^{-1} x_0}\,\Phi^{-1}(1-\alpha/2)\big]. \tag{S.9.6}$$

We fix $N = 2^{14}$ in this section.

We demonstrate here the impact of number of subsamples S, model dimension m, number of basis q and number of quantile grid points K on the coverage probability of (S.9.6). We note that q and K have similar impact on coverage probabilities given that other parameters are held fixed. In Figure S.9.11 with $\varepsilon \sim N(0, 0.1^2)$, we note that an increase in $q = \dim(\mathbf{B})$ improves the coverage probabilities given that m, S are held fixed. Also, at a fixed S, q can be chosen smaller if m is larger. This is consistent with Corollary 3.10 that requires $q \gg N^{1/(2\eta)}\|\mathbf{Z}(x_0)\|^{-\eta}$. Similar to Section S.9.3.1, an increase of the dimensionality m leads to a smaller S^* for any fixed q. For $\varepsilon \sim \text{Exp}(0.8)$ shown in Figure S.9.12, a difference to the normal case is that the performance of the coverage probability is better when τ is small.

S.9.4. Nonparametric model. In this section, we consider a nonlinear model

$$Y_i = 2.5 + \sin(2X_i) + 2\exp(-16X_i^2) + 0.7\epsilon_i, \quad X_i \sim \mathcal{U}(-1,1), \quad \epsilon_i \sim \mathcal{N}(0,1), \tag{S.9.7}$$

where the function $x \mapsto 2.5 + \sin(2x) + 2\exp(-16x^2)$ is plotted in Figure S.9.13. The basis $\mathbf{Z}$ is set as cubic B-spline defined at $m+4$ knots. The knots form a partition on $[-1, 1]$ with repetitions on the boundary; see Corollary 4.10 of Schumaker (1981). The model (S.9.7) implies the quantile of Y is $Q(x;\tau) = 2.5 + \sin(2x) + 2\exp(-16x^2) + 0.7\Phi^{-1}(\tau)$ for $0 < \tau < 1$.

Section S.9.4.1 concerns the coverage probabilities of the confidence intervals for $Q(x;\tau)$ for fixed τ. Section S.9.4.2 deals with the coverage probabilities of the confidence intervals for $F_{Y|X}(y|x)$. In both sections, we fix $N = 2^{16}$.

S.9.4.1. *Oracle rule for $\overline{\boldsymbol{\beta}}(\tau)$.* According to Corollary 3.9, an asymptotic $1-\alpha$ confidence interval for $Q(x_0;\tau)$ is

$$\big[\mathbf{Z}(x_0)^\top \overline{\boldsymbol{\beta}}(\tau) \pm N^{-1/2}\phi_\sigma^{-1}(\Phi_\sigma^{-1}(\tau))\sqrt{\tau(1-\tau)}\sigma_0(\mathbf{Z})\Phi^{-1}(1-\alpha/2)\big], \tag{S.9.8}$$

where $\sigma_0^2(\mathbf{Z}) = \mathbf{Z}(x_0)^\top \mathbb{E}[\mathbf{Z}(X)\mathbf{Z}^\top(X)]^{-1}\mathbf{Z}(x_0)$ and ϕ_σ is the density of $\mathcal{N}(0,\sigma^2)$.

Our simulation results show how S and m (the number of basis functions $m = \dim(\mathbf{Z})$) influence the coverage probabilities of (S.9.8). For all the plots in Figure S.9.14, the coverage probabilities corresponding to $m = 13$ at $x_0 = 0$ (the solid green curves) performs the worst. This is partly caused by the large bias from the large curvature of $Q(x_0;\tau)$ at $x_0 = 0$ (see Figure S.9.13). This bias can be reduced by setting greater m, but the increase of m results in a smaller S^*.

S.9.4.2. *Oracle rule for $\widehat{F}_{Y|X}(y|x)$.* We compute $\widehat{F}_{Y|X}(y|x)$ in a similar way as Section S.9.3.2 by taking $y_0 = Q(x_0;\tau)$, which gives $F_{Y|X}(y_0|x_0) = \tau$. By Corollary 3.10, an $1-\alpha$ confidence interval for τ is

$$\big[\widehat{F}_{Y|X}(Q(x_0;\tau)|x_0) \pm N^{-1/2}\sqrt{\tau(1-\tau)}\sigma_0(\mathbf{Z})\Phi^{-1}(1-\alpha/2)\big], \tag{S.9.9}$$

where $\sigma_0^2(\mathbf{Z})$ is defined as in (S.9.8).

We show the impact of number of subsamples S, model dimension m, number of basis q and number of quantile grid points K on the coverage probability of (S.9.9) in Figure S.9.15. We only show $\tau = 0.1$ and omit the results for $\tau = 0.5, 0.9$, as they do not show additional insights to Figure S.9.11. It can be seen from Figure S.9.15 that larger m lead to a smaller estimation bias, and thus improves performance under fixed S, q and K. However, such an increase of m also results in smaller S^*. Under a fixed m, the influence from S, q and K on the coverage probabilities is similar to the linear models with normal errors in Section S.9.3.2. These findings are consistent with Corollary 3.10.

S.9.5. Practical conclusions from simulations. In this section, we briefly comment on some practical recommendations regarding the choice of S, K. Our general recommendation would be to choose S as small as possible and K as large as possible since the only advantage of choosing large S or small K are computational savings. Both theory and simulations suggest that for models with more predictors, particular care in choosing S is necessary. Giving very specific default recommendations is difficult since, as seen from the simulations, the exact value of $\log_N(S)$ when coverage starts to drop below nominal values depends on the quantile at hand and (unknown) conditional density of response given predictors.

As a rule of thumb, we would recommend to ensure that the 'effective subsample size' n/m is at least 50 and that S does not exceed n/m. Additionally, if the spread of conditional quantiles as a function of τ changes rapidly, extra caution needs to be taken. This recommendation is based on Figure S.9.10 where the conditional density of response given predictors is exponential and coverage starts to drop earlier for smaller quantiles, which corresponds to regions where the second derivative of the conditional quantile function is larger. This is also in line with our derivation of lower bonds for the oracle rule, where the second derivative of the conditional quantile function enters as constant a (see Section S.3.3.2)

Regarding K, $K > N^{1/4}$ worked well in all simulations we considered. Again, the precise constant will depend on the roughness of the function $\tau \mapsto Q(x;\tau)$. The choice of G should depend on K. Although the formal statement of our theorem requires $K/G \to \infty$, a look at the proof reveals that for cubic splines $K/G > c$ for some constant c is sufficient. Simulations indicate that $G = K/2$ combined with cubic splines is a reasonable choice.

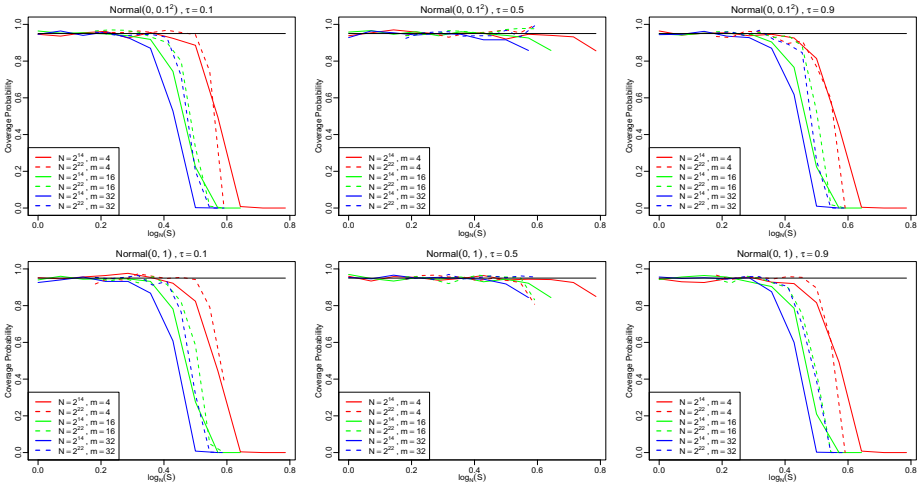


Fig S.9.9: Coverage probability of the $1-\alpha = 95\%$ confidence interval for $Q(x_0;\tau)$ in (5.3): $\left[x_0^\top \overline{\boldsymbol{\beta}}(\tau) \pm f_{\varepsilon,\tau}^{-1}(\tau(1-\tau)x_0^\top \Sigma_X^{-1} x_0/N)^{1/2}\Phi^{-1}(1-\alpha/2)\right]$ under the model $Y_i = 0.21 + \boldsymbol{\beta}_{m-1}^\top X_i + \varepsilon_i$ with $\boldsymbol{\beta}_{m-1}$ as (S.9.3) and $\varepsilon \sim \mathcal{N}(0,\sigma^2)$, where N is the total sample size, S is the number of subsamples and $m = 1 + \dim(\boldsymbol{\beta}_{m-1})$.

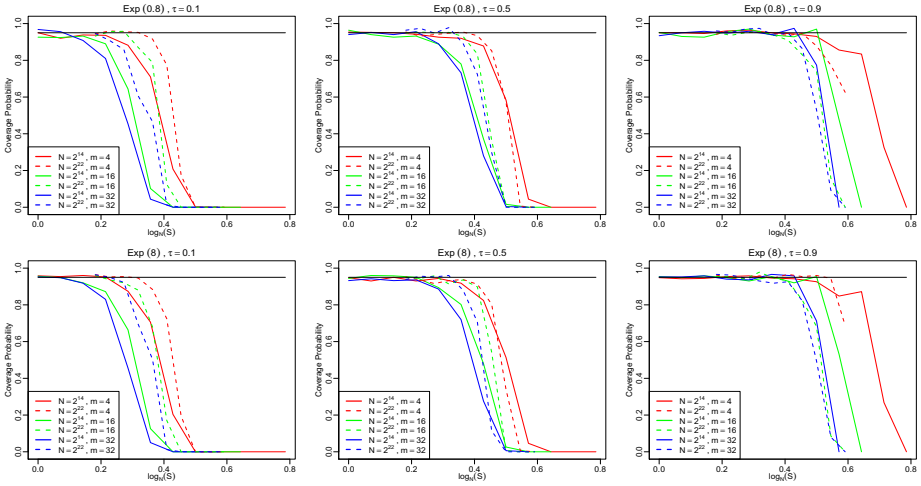


Fig S.9.10: Coverage probability under exactly the same setting as Figure S.9.9, only except for $\varepsilon \sim \text{Exp}(\lambda)$.

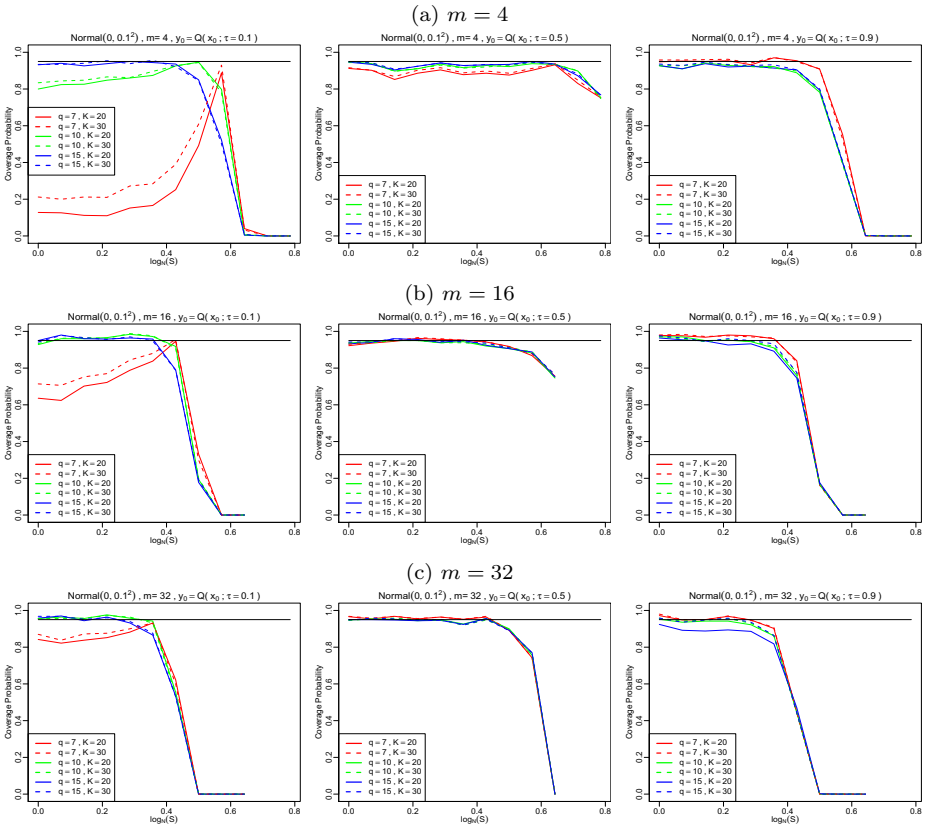


Fig S.9.11: Coverage probability of the $1-\alpha = 95\%$ confidence interval for $F(y_0|x_0)$ in (S.9.6): $\left[\widehat{F}_{Y|X}(Q(x_0;\tau)|x_0) \pm \left(\tau(1-\tau)x_0^\top \Sigma_X^{-1} x_0/N\right)^{1/2}\Phi^{-1}(1-\alpha/2)\right]$ for $Y_i = 0.21 + \boldsymbol{\beta}_{m-1}^\top X_i + \varepsilon_i$ with $\boldsymbol{\beta}_{m-1}$ as (S.9.3) and $\varepsilon \sim \mathcal{N}(0, 0.1^2)$, where the total number of samples is $N = 2^{14}$, S is the number of subsamples, $m = 1 + \dim(\boldsymbol{\beta}_{m-1})$, $q = \dim(\mathbf{B})$ (corresponds to projection in τ direction) and K is the number of the quantile grid points.

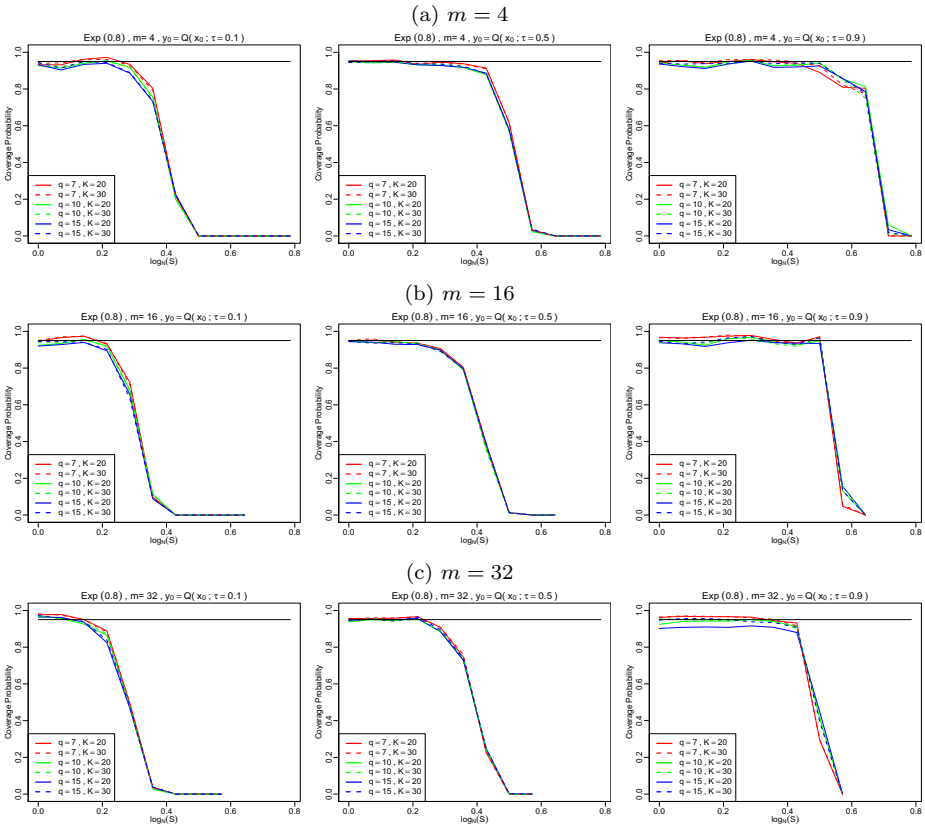


Fig S.9.12: Coverage probability under exactly the same setting as Figure S.9.11, only except for $\varepsilon \sim \text{Exp}(\lambda)$.

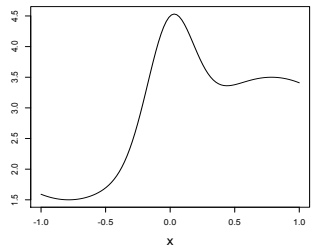


Fig S.9.13: The plot for the function $x \mapsto 2.5+\sin(2x)+2\exp(-16x^2)$, which is used in model (S.9.7).

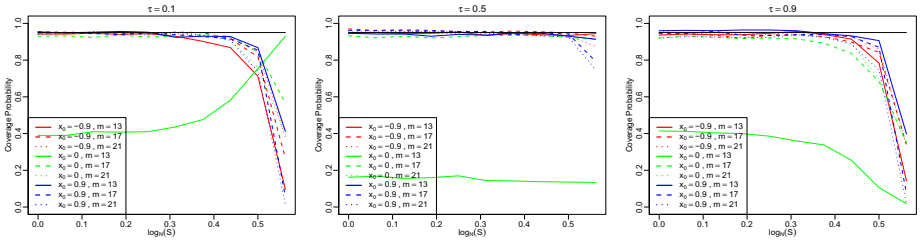


Fig S.9.14: Coverage probability of the $1-\alpha = 95\%$ confidence interval for $Q(x_0;\tau)$ in (S.9.8): $\left[\mathbf{Z}(x_0)^\top \overline{\boldsymbol{\beta}}(\tau) \pm N^{-1/2}\phi_\sigma^{-1}(\Phi_\sigma^{-1}(\tau))\sqrt{\tau(1-\tau)}\sigma_0(\mathbf{Z})\Phi^{-1}(1-\alpha/2)\right]$ under the nonlinear model (S.9.7), where $m = \dim(\mathbf{Z})$ (corresponds to basis expansion in x) and $N = 2^{16}$.

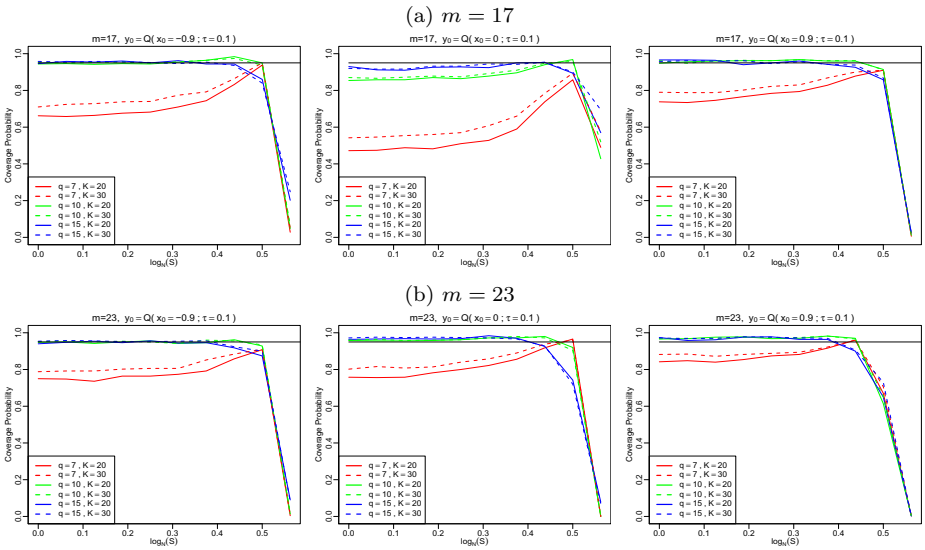


Fig S.9.15: Coverage probability of the $1-\alpha = 95\%$ confidence interval for $F(y_0|x_0)$ in (S.9.9): $\left[\widehat{F}_{Y|X}(Q(x_0;\tau)|x_0) \pm N^{-1/2}\sqrt{\tau(1-\tau)}\sigma_0(\mathbf{Z})\Phi^{-1}(1-\alpha/2)\right]$ under the nonlinear model (S.9.7), where $m = \dim(\mathbf{Z})$ (corresponds to basis expansion in x), S is the number of subsamples, $N = 2^{16}$, $q = \dim(\mathbf{B})$ (corresponds to projection in τ direction) and K is the number of the quantile grid points.