Supplement to "Uniform Estimation and Inference for Nonparametric Partitioning-Based M-Estimators"

Matias D. Cattaneo[*] Yingjie Feng[†] Boris Shigida[‡]

August 20, 2025

Abstract

This supplement is self-contained. It presents more general theoretical results than those reported in the paper, as well as their proofs. In particular, a larger class of loss functions is allowed for, and weaker regularity conditions are employed when possible. Additional results not reported in the paper and their proofs are given, and some of the technical results presented herein may be of independent theoretical interest. Omitted details on the motivating examples are also provided. Finally, simulation evidence is reported.

[*]Department of Operations Research and Financial Engineering, Princeton University.
[†]School of Economics and Management, Tsinghua University.
[‡]Department of Operations Research and Financial Engineering, Princeton University.

Contents

List of important statements

List of lemmas

SA1 Introduction

Let $\mathcal{Q} \subset \mathbb{R}^{d_{\mathcal{Q}}}$, $\mathcal{X} \subset \mathbb{R}^d$ be fixed compact sets, where $d_{\mathcal{Q}}$ and d are positive integers. (In the paper, $d_{\mathcal{Q}} = 1$ was set only for simplicity.) Suppose that $\big((y_i, \boldsymbol{x}_i)\big)_{1\le i\le n}$ is a random sample, where $y_i \in \mathcal{Y} \subset \mathbb{R}$ is a scalar response variable, $\boldsymbol{x}_i$ is a d-dimensional covariate with values in $\mathcal{X}$. Let $\rho(\cdot,\cdot;\boldsymbol{q})$ be a loss function parametrized by $\boldsymbol{q} \in \mathcal{Q}$ (Borel-measurable in all three arguments), and let $\eta(\cdot)\colon \mathbb{R} \to \mathcal{E}$ be a strictly monotonic continuously differentiable transformation function. (More detailed assumptions on $\rho(\cdot,\cdot;\boldsymbol{q})$ and $\eta(\cdot)$ will be provided below.) We fix a function $\mu_0(\cdot,\cdot)\colon \mathcal{X}\times\mathcal{Q} \to \mathbb{R}$, Borel-measurable in both arguments and such that

$$\mu_0(\cdot,\boldsymbol{q}) \in \operatorname*{arg\,min}_{\mu} \mathbb{E}[\rho(y_1, \eta(\mu(\boldsymbol{x}_1)); \boldsymbol{q})], \tag{SA1.1}$$

where the argmin is over the space of Borel functions $\mathcal{X} \to \mathbb{R}$. In particular, we assume that the minimum is finite, and such a minimizer exists.

Our main goal is to conduct uniform (over $\mathcal{X}\times\mathcal{Q}$) estimation and inference for μ_0, and transformations thereof, employing the partitioning-based series M-estimator

$$\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}) = \boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\boldsymbol{q}), \qquad \widehat{\boldsymbol{\beta}}(\boldsymbol{q}) \in \operatorname*{arg\,min}_{\boldsymbol{b}\in\mathcal{B}} \sum_{i=1}^{n} \rho(y_i, \eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{b}); \boldsymbol{q}), \tag{SA1.2}$$

where $\mathcal{B} \subseteq \mathbb{R}^K$ is the feasible set of the optimization problem, and

$$\boldsymbol{x} \mapsto \boldsymbol{p}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x};\Delta,m) = \big(p_1(\boldsymbol{x};\Delta,m),\dots,p_K(\boldsymbol{x};\Delta,m)\big)^{\mathsf{T}}$$

is a dictionary of K locally supported basis functions of order m based on a quasi-uniform partition $\Delta = \{\delta_l : 1 \le l \le \kappa\}$ containing a collection of open disjoint polyhedra in $\mathcal{X}$ such that the closure of their union covers $\mathcal{X}$. The m parameter controls how well μ_0 can be approximated by linear combinations of the basis (Assumption SA2.6 below); the partition being quasi-uniform intuitively means that the largest size of a cell cannot get asymptotically bigger than the smallest one (Assumption SA2.1 below). We consider large sample approximations where d and m are fixed constants, and $\kappa \to \infty$ (and thus $K \to \infty$) as $n \to \infty$. Prior literature is discussed in the paper.

SA1.1 Notation

For any real function f depending on d variables $(t_1,\dots,t_d)$ and any vector $\boldsymbol{v}$ of nonnegative integers, denote

$$f^{(\boldsymbol{v})} := \partial^{\boldsymbol{v}} f := \frac{\partial^{|\boldsymbol{v}|}}{\partial t_1^{v_1}\dots\partial t_d^{v_d}} f.$$

the multi-indexed partial derivative of f, where $|\boldsymbol{v}| = \sum_{k=1}^d v_j$. A derivative of order zero is the function itself, so if $v_i = 0$, the ith partial differentiation is ignored. For functions that depend on $(\boldsymbol{x},\boldsymbol{q})$, the multi-index derivative notation is taken with respect to the first argument $\boldsymbol{x}$, unless otherwise noted.

We will denote $N(\mathcal{F},\rho,\varepsilon)$ the ε-covering number of a class $\mathcal{F}$ with respect to a semi-metric ρ defined on it.

For a function $f\colon S \to \mathbb{R}$ the set $\{(x,t) \in S\times\mathbb{R} : t < f(x)\}$ is called the *subgraph* of f. A class $\mathcal{F}$ of measurable functions from $\mathcal{S}$ to $\mathbb{R}$ is called a *VC-subgraph class* or *VC-class* if the collections of all subgraphs of functions in $\mathcal{F}$ is a VC-class of sets in $S\times\mathbb{R}$, which means that for some finite

m no set of size m is shattered by it. In this case, the smallest such m is called the VC-index of $\mathcal{F}$. See [18] for details.

For a measurable function $f\colon S \to \mathbb{R}$ on a measurable space $(S, \mathcal{S})$, a probability measure $\mathbb{Q}$ on this space and some $q \geq 1$, define the $(\mathbb{Q}, q)$-*norm* of f as $\|f\|_{\mathbb{Q},q}^q = \mathbb{E}_{X\sim\mathbb{Q}}[f(X)^q]$.

We will say that a class of measurable functions $\mathcal{F}$ from any set S to $\mathbb{R}$ has a measurable *envelope* F if $F\colon S \to \mathbb{R}$ is such a measurable function that $|f(s)| \leq F(s)$ for all $s \in S$ and all $f \in \mathcal{F}$. We will say that this class satisfies the *uniform entropy bound* with envelope F and real constants $A \geq e$ and $V \geq 1$ if

$$\sup_{\mathbb{Q}} N(\mathcal{F}, \|\cdot\|_{\mathbb{Q},2}, \varepsilon\|F\|_{\mathbb{Q},2}) \leq \left(\frac{A}{\varepsilon}\right)^V \tag{SA1.3}$$

for all $0 < \varepsilon \leq 1$, where the supremum is taken over all finite discrete probability measures $\mathbb{Q}$ with $\|F\|_{\mathbb{Q},2} > 0$, $\|\cdot\|_{\mathbb{Q},2}$ denotes the $(\mathbb{Q}, 2)$-norm.

We will say that an $\mathbb{R}$-valued random variable ξ is σ^2-*sub-Gaussian*, where $\sigma^2 > 0$, if

$$\mathbb{P}\{|\xi| \geq t\} \leq 2\exp\{-t^2/\sigma^2\} \quad \text{for all } t \geq 0. \tag{SA1.4}$$

We will denote by $\boldsymbol{D}_n$ the random vector of all the data $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$, $\mathbf{X}_n$ for $\{\boldsymbol{x}_i\}_{i=1}^n$ and $\boldsymbol{y}_n$ for $\{y_i\}_{i=1}^n$.

For two random elements Z_1 and Z_2, the notation $Z_1 \perp\!\!\!\perp Z_2$ means they are independent, and the notation $Z_1 \underset{\xi}{\perp\!\!\!\perp} Z_2$ means they are independent given a third random element ξ.

The notation $Z_1 \overset{d}{=} Z_2$ means the two random elements Z_1 and Z_2 have the same laws.

If we say the probability space is *rich enough*, it means that it admits a randomization variable, that is, a random variable distributed uniformly on $[0, 1]$ independent of the data. Since this one random variable can be replicated (see, e. g., Lemma 4.21 in [14]), we can find a random variable distributed uniformly on $[0, 1]$ independent of the data and such random variables previously used, whenever the argument requires.

Finally, we will use the following notations:

$$\bar{\mathbf{Q}}_{\boldsymbol{q}} := \mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\Psi_1(\boldsymbol{x}_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))^2] \tag{SA1.5}$$

$$\mathbf{Q}_{0,\boldsymbol{q}} := \mathbb{E}[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\Psi_1(\boldsymbol{x}_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))^2] \tag{SA1.6}$$

$$\boldsymbol{\Sigma}_{0,\boldsymbol{q}} := \mathbb{E}[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})^2\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))^2] \tag{SA1.7}$$

$$\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}} := \mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\sigma_{\boldsymbol{q}}^2(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))^2] \tag{SA1.8}$$

$$\Omega_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q}) := \boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^\mathsf{T}\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\boldsymbol{\Sigma}_{0,\boldsymbol{q}}\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) \tag{SA1.9}$$

$$\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q}) := \boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^\mathsf{T}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}), \tag{SA1.10}$$

where $\psi(\cdot, \cdot; \cdot)$, $\Psi_1(\cdot, \cdot; \cdot)$, $\sigma_\cdot^2(\cdot)$ are defined in Assumption SA2.4.

Many mathematical notations are clickable and link to their definitions; for example, clicking on $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}$ should lead to (SA1.8).

SA1.2 Organization

Section SA2 collects the general assumptions used through this supplemental appendix. In subsequent sections, we will list which assumptions are required for each lemma, proposition, or theorem. Note that Assumption SA2.4 is weaker than the assumptions on the loss function described in Section 3 of the main paper. In particular, a complexity condition on $\{\psi(\cdot, \cdot; \boldsymbol{q})\}$ is absent in Section 3 because the main paper contains a stronger but simpler Assumption E. Thus, the setup in this supplement is more general that in the main paper; this fact is formally shown by Proposition SA2.9. The correspondence of assumptions given in the paper and the versions given in the supplement is as follows:

Assumption A $\Leftrightarrow$ Assumption SA2.1,
Assumption B $\Leftrightarrow$ Assumption SA2.2,
Assumption C $\Leftrightarrow$ Assumption SA2.3,
Assumption D $\Leftrightarrow$ Assumption SA2.4,
Assumption E $\Rightarrow$ Assumptions SA2.5 and SA2.8 (by Proposition SA2.9),
Assumption F $\Leftrightarrow$ Assumption SA2.6.

Section SA3 records some well-known results and tools in probability and stochastic process theory that will be useful for our theoretical analysis, and also presents a collections of lemmas that are used repeatedly in many subsequent arguments throughout this supplement.

The rest of the supplement presents our main theoretical results, which encompass and generalize the simplified results presented in the main paper (to improve exposition). In the rest of this subsection, we explain how the two set of theoretical results relate to each other. All equivalences below are given up to the fact that the supplement contains weaker complexity Assumptions SA2.5 and SA2.8, which are not present in the paper because the simplified setup (Assumption E) is used.

Section SA4 presents all our consistency results, categorized based on whether the objective function is convex or not. Additional results of theoretical and methodological interest, such as the consistency of partitioning-based M-estimators in special cases (e.g., an unconnected basis or a strongly convex and smooth loss function), which were not formally reported in the paper, are also presented. The matching of consistency results given in the paper with the versions given in the supplement is as follows:

Lemma 1 $\Leftrightarrow$ Lemma SA4.1,
Lemma 2 $\Leftrightarrow$ Lemma SA4.3.

Section SA5 proves our main Bahadur representation result and its corollaries, matched as follows to the results in the paper:

Theorem 1 $\Leftrightarrow$ Theorem SA5.1,
Corollary 1 $\Leftarrow$ Corollary SA5.16,
Corollary 2 $\Leftrightarrow$ Corollary SA5.17.

Section SA6 develops strong approximation results using a generalized conditional Yurinskii's coupling approach. First, Section SA6.1 gives general results that may be of independent theoretical interest: they provide generalizations of, and in some cases complement, prior coupling results established in [2], [7], [8], and references therein. Second, Section SA6.2 deploys those results to the setting of interest in our paper to verify our main result Theorem SA6.4 that confirms Theorem 2 in the paper:

Theorem 2 $\Leftarrow$ Theorem SA6.4 (see also Remark SA6.5).

Section SA7 discusses results related to the implementation of uniform inference, formally showing the validity of the plug-in approximation method and confidence bands described in the paper. In particular, the following correspondence holds:

Theorem 3 $\Leftarrow$ Theorem SA7.1 (see also Remark SA7.2).

Section SA8 discusses in detail the verification of our high-level assumptions for each of the four motivating examples in the paper, proving Propositions 1 through 4 in the paper along with verifying more general complexity assumptions discussed above. The correspondence between the results in the paper and the results in this section is as follows:

$$\text{Proposition 1} \Leftarrow \text{Proposition SA8.1},$$
$$\text{Proposition 2} \Leftarrow \text{Proposition SA8.3},$$
$$\text{Proposition 3} \Leftarrow \text{Proposition SA8.5},$$
$$\text{Proposition 4} \Leftarrow \text{Proposition SA8.6}.$$

Section SA9 discusses other parameters of potential interest such as the level and marginal effect functions, which formalizes the claims made in Section 9 of the paper.

Section SA10 is devoted to simulation runs with synthetic data, covering pointwise and uniform inference.

SA1.3 Comparison to prior literature

Our general M-estimation theoretical results are on par or improve over prior literature partitioning-based nonparametric least squares regression [3, 5, 6, 4, 9, 13, 19], and on series nonparametric quantile regression [2]. Both strands of the literature assume an identity transformation function $\eta(\cdot)$. In addition, we also compare to the closely related setting of nonparametric smoothing spline nonlinear regression [17].

The main paper gives a detailed discussion of our improvements when discussing the theoretical results for consistency (Section 4), Bahadur representation (Section 5), and strong approximation (Section 6). To complement that discussion, we present three tables that summarize the comparison between prior literature and this paper.

Table 1 compares the best known results for mean-square (L_2) and uniform (L_∞) consistency of series and partitioning-based estimators. In each case considered in this table, rate-optimal L_2 and L_∞ convergence rates are achievable, but with different side rate conditions on h, m, and d. The table shows that our results either achieve the best (minimal) restrictions in the literature for the special case of square loss and identity transformation functions, or improve upon prior results in the general M-estimation case.

Table 1: Comparison to prior literature: Consistency (Section 4)

	Side rate restriction	m v.s. d restriction	Uniformity
Mean regression	$nh^d \to \infty$	None	$\boldsymbol{x}$ only
Quantile regression, L_2-consistency	$nh^{2d} \to \infty$	$m > d$	$\boldsymbol{x}$, and q (scalar)
Quantile regression, L_∞-consistency	$nh^{4d} \to \infty$	$m > d$	$\boldsymbol{x}$, and q (scalar)
This paper, L_2- and L_∞-consistency			
General partitioning basis	$nh^{2d} \to \infty$	$m > d/2$	$\boldsymbol{x}$, and $\boldsymbol{q}$
Strongly convex and smooth loss function	$nh^d \to \infty$	None	$\boldsymbol{x}$, and $\boldsymbol{q}$
Unconnected partitioning basis	$nh^d \to \infty$	None	$\boldsymbol{x}$, and $\boldsymbol{q}$

Notes: (i) m is the smoothness of $\boldsymbol{x} \mapsto \mu_0(\boldsymbol{x}, \boldsymbol{q})$ and order of the basis; (ii) d is the dimension of $\boldsymbol{x}_i$; and (iii) poly-log-n and similar additional terms are not reported to simplify the exposition.

Table 2 compares the best known results for uniform Bahadur representation for series and partitioning-based estimators. It again demonstrates that the results in our paper either achieve the best (minimal) restrictions in the literature for the special case of square loss and identity transformation functions, or improve upon prior literature in the general M-estimation case.

Table 2: Comparison to prior literature: Uniform Bahadur representation (Section 5)

	Side rate restriction	m v.s. d restriction	remainder	Uniformity
Mean regression	$nh^d \to \infty$	None	$\frac{1}{nh^d}$ (optimal)	$\boldsymbol{x}$ only
Quantile regression	$nh^{4d} \to \infty$	$m > d$	$\frac{1}{(nh^d)^{3/4}h^{d/2}}$	$\boldsymbol{x}$, and q (scalar)
Smooth loss + smoothing spline	$nh^{2d} \to \infty$	$d = 1$	$\frac{1}{nh^{10m-1}}$	x, and q (scalar)
This paper (assuming consistency)				
Smooth weak derivative ($\alpha = 1$)	$nh^d \to \infty$	None	$\frac{1}{(nh^d)^{3/4}}$ (optimal)	$\boldsymbol{x}$, and $\boldsymbol{q}$
Non-smooth week derivative ($\alpha \in (0, 1)$)	$nh^d \to \infty$	None	See Theorem 1	$\boldsymbol{x}$, and $\boldsymbol{q}$

Notes: (i) m is the smoothness of $\boldsymbol{x} \mapsto \mu_0(\boldsymbol{x}, \boldsymbol{q})$ and order of the basis; (ii) d is the dimension of $\boldsymbol{x}_i$; and (iii) poly-log-n and similar additional terms are not reported to simplify the exposition.

Finally, Table 3 compares the best known results for uniform strong approximation for series and partitioning-based estimators. Our results achieve the best (minimal) restrictions in the literature for the special case of square loss and identity transformation functions when $d > 1$, and improve upon prior literature in the general M-estimation case.

Table 3: Comparison to prior literature: Strong approximation (Section 6)

	Side rate restriction	m v.s. d restriction	Uniformity
Mean regression			
$d = 1$	$nh^d \to \infty$	None	$\boldsymbol{x}$ only
$d \geq 1$	$nh^{3d} \to \infty$	None	$\boldsymbol{x}$ only
Quantile regression ($d \geq 1$)	$nh^{4d \vee (2+3d)} \to \infty$	$m > d$	$\boldsymbol{x}$, and q (scalar)
This paper (assuming consistency, $d \geq 1$)	$nh^{3d} \to \infty$	None	$\boldsymbol{x}$, and $\boldsymbol{q}$

Notes: (i) m is the smoothness of $\boldsymbol{x} \mapsto \mu_0(\boldsymbol{x}, \boldsymbol{q})$ and order of the basis; (ii) d is the dimension of $\boldsymbol{x}_i$; and (iii) poly-log-n and similar additional terms are not reported to simplify the exposition.

SA2 Assumptions

This section collects the assumptions used throughout the supplemental appendix. These assumptions are weaker than (i.e, implied by) the assumptions imposed in the paper.

Assumption SA2.1 (Quasi-uniform partition)**.** *The ratio of the sizes of inscribed and circumscribed balls of each $\delta \in \Delta$ is bounded away from zero uniformly in $\delta \in \Delta$, and*

$$\frac{\max\{\operatorname{diam}(\delta) : \delta \in \Delta\}}{\min\{\operatorname{diam}(\delta) : \delta \in \Delta\}} \lesssim 1$$

where $\operatorname{diam}(\delta)$ denotes the diameter of δ. Further, for $h = \max\{\operatorname{diam}(\delta) : \delta \in \Delta\}$, assume $h = o(1)$ and $\log(1/h) \lesssim \log(n)$.

Assumption SA2.2 (Local basis).

(i) For each basis function p_k, $k = 1, \dots, K$, the union of elements of Δ on which p_k is active is a connected set, denoted by $\mathcal{H}_k$. For all $k = 1, \dots, K$, both the number of elements of $\mathcal{H}_k$ and the number of basis functions which are active on $\mathcal{H}_k$ are bounded by a constant.

(ii) For any $\boldsymbol{a} = (a_1, \dots, a_K)^\mathsf{T} \in \mathbb{R}^K$

$$\boldsymbol{a}^\mathsf{T} \int_{\mathcal{H}_k} \boldsymbol{p}(\boldsymbol{x}; \Delta, m) \boldsymbol{p}(\boldsymbol{x}; \Delta, m)^\mathsf{T} d\boldsymbol{x} \; \boldsymbol{a} \gtrsim a_k^2 h^d, \quad k = 1, \dots, K.$$

(iii) Let $|\boldsymbol{v}| < m$. There exists an integer $\varsigma \in [|\boldsymbol{v}|, m)$ such that, for all $\boldsymbol{\varsigma}, |\boldsymbol{\varsigma}| \leq \varsigma$,

$$h^{-|\boldsymbol{\varsigma}|} \lesssim \inf_{\delta \in \Delta} \inf_{\boldsymbol{x} \in \mathrm{cl}(\delta)} \|\boldsymbol{p}^{(\boldsymbol{\varsigma})}(\boldsymbol{x}; \Delta, m)\| \leq \sup_{\delta \in \Delta} \sup_{\boldsymbol{x} \in \mathrm{cl}(\delta)} \|\boldsymbol{p}^{(\boldsymbol{\varsigma})}(\boldsymbol{x}; \Delta, m)\| \lesssim h^{-|\boldsymbol{\varsigma}|}$$

where $\mathrm{cl}(\delta)$ is the closure of δ.

Assumption SA2.2 implicitly relates the number of basis functions and the maximum mesh size: $K \asymp h^{-d} = J^d$.

For the following assumption and throughout the document, when speaking of the conditional distribution of y_1 given $\boldsymbol{x}_1$, or its functionals (like conditional moments or quantiles), we mean one fixed regular variant of such a distribution satisfying all the assumptions listed.

Assumption SA2.3 (Data generating process).

(i) $\big((y_i, \boldsymbol{x}_i)\big)_{1 \leq i \leq n}$ is a random sample satisfying (SA1.1).

(ii) The distribution of $\boldsymbol{x}_i$ admits a Lebesgue density $f_X(\cdot)$ which is continuous and bounded away from zero on support $\mathcal{X} \subset \mathbb{R}^d$, where $\mathcal{X}$ is the closure of an open, connected and bounded set.

(iii) The conditional distribution of y_i given $\boldsymbol{x}_i$ admits a conditional density $f_{Y|X}(y|\boldsymbol{x})$ with support $\mathcal{Y}_{\boldsymbol{x}}$ with respect to some (sigma-finite) measure $\mathfrak{M}$, and $\sup_{\boldsymbol{x} \in \mathcal{X}} \sup_{y \in \mathcal{Y}_{\boldsymbol{x}}} f_{Y|X}(y|\boldsymbol{x}) < \infty$.

(iv) $\boldsymbol{x} \mapsto \mu_0(\boldsymbol{x}, \boldsymbol{q})$ is $m \geq 1$ times continuously differentiable for every $\boldsymbol{q} \in \mathcal{Q}$, $\boldsymbol{x} \mapsto \mu_0(\boldsymbol{x}, \boldsymbol{q})$ and its derivatives of order no greater than m are bounded uniformly over $(\boldsymbol{x}, \boldsymbol{q}) \in \mathcal{X} \times \mathcal{Q}$, and

$$\sup_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{q}_1 \neq \boldsymbol{q}_2} \frac{|\mu_0(\boldsymbol{x}, \boldsymbol{q}_1) - \mu_0(\boldsymbol{x}, \boldsymbol{q}_2)|}{\|\boldsymbol{q}_1 - \boldsymbol{q}_2\|} \lesssim 1.$$

For the following assumption and throughout the document, we fix some (small enough) $r > 0$ and denote

$$B_{\boldsymbol{q}}(\boldsymbol{x}) := \{\zeta \colon |\zeta - \mu_0(\boldsymbol{x}, \boldsymbol{q})| \leq r\}, \tag{SA2.1}$$

i. e. we will work with a fixed neighborhood of $\mu_0(\boldsymbol{x}, \boldsymbol{q})$.

Assumption SA2.4 (Loss function).

(i) Let $\mathcal{Q} \subset \mathbb{R}^{d_\mathcal{Q}}$ be a connected compact set. For each $\boldsymbol{q} \in \mathcal{Q}$ and $y \in \mathcal{Y}$, and some open connected subset $\mathcal{E}$ of $\mathbb{R}$ not depending on y, $\eta \mapsto \rho(y, \eta; \boldsymbol{q})$ is absolutely continuous on closed bounded intervals within $\mathcal{E}$, and admits an a. e. derivative $\psi(y, \eta; \boldsymbol{q})$, Borel measurable as a function of $(y, \eta, \boldsymbol{q})$.

(ii) The first-order optimality condition $\mathbb{E}[\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q}) \,|\, \boldsymbol{x}_i] = 0$ holds; the function

$$\sigma_{\boldsymbol{q}}^2(\boldsymbol{x}) := \mathbb{E}\big[\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})^2 \,\big|\, \boldsymbol{x}_i = \boldsymbol{x}\big]$$

is continuous in both arguments $(\boldsymbol{x}, \boldsymbol{q}) \in \mathcal{X} \times \mathcal{Q}$, bounded away from zero, and Lipschitz in $\boldsymbol{q}$ uniformly in $\boldsymbol{x}$; there is a positive measurable envelope function $\overline{\psi}(\boldsymbol{x}_i, y_i)$ such that $\sup_{\boldsymbol{q} \in \mathcal{Q}} |\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})| \leq \overline{\psi}(\boldsymbol{x}_i, y_i)$ with

$$\sup_{\boldsymbol{x} \in \mathcal{X}} \mathbb{E}\big[\overline{\psi}(\boldsymbol{x}_i, y_i)^\nu \,\big|\, \boldsymbol{x}_i = \boldsymbol{x}\big] < \infty \quad \text{for some } \nu > 2.$$

(iii) The function $\eta(\cdot)\colon \mathbb{R} \to \mathcal{E}$ is strictly monotonic and twice continuously differentiable; for some fixed constant $\alpha \in (0,1]$, for any $(\boldsymbol{x},\boldsymbol{q}) \in \mathcal{X} \times \mathcal{Q}$, and a pair of points $\zeta_1, \zeta_2 \in B_{\boldsymbol{q}}(\boldsymbol{x})$, $\psi(\cdot,\cdot;\boldsymbol{q})$ satisfies the following (constants hidden in $\lesssim$ do not depend on $\boldsymbol{x},\boldsymbol{q},\zeta_1,\zeta_2$):

- *if $\mathfrak{M}$ in Assumption SA2.3(iii) is the Lebesgue measure, then*

$$\begin{aligned}
&\sup_{\lambda\in[0,1]}\sup_{y\notin[\eta(\zeta_1)\wedge\eta(\zeta_2),\eta(\zeta_1)\vee\eta(\zeta_2)]}\big|\psi(y,\eta(\zeta_1+\lambda(\zeta_2-\zeta_1));\boldsymbol{q})-\psi(y,\eta(\zeta_2);\boldsymbol{q})\big| \lesssim |\zeta_1-\zeta_2|^{\alpha},\\
&\sup_{\lambda\in[0,1]}\sup_{y\in[\eta(\zeta_1)\wedge\eta(\zeta_2),\eta(\zeta_1)\vee\eta(\zeta_2)]}\big|\psi(y,\eta(\zeta_1+\lambda(\zeta_2-\zeta_1));\boldsymbol{q})-\psi(y,\eta(\zeta_2);\boldsymbol{q})\big| \lesssim 1;
\end{aligned} \tag{SA2.2}$$

- *if $\mathfrak{M}$ is* not *the Lebesgue measure, then*

$$\sup_{\lambda\in[0,1]}\sup_{y\in\mathcal{Y}}\big|\psi(y,\eta(\zeta_1+\lambda(\zeta_2-\zeta_1));\boldsymbol{q})-\psi(y,\eta(\zeta_2);\boldsymbol{q})\big| \lesssim |\zeta_1-\zeta_2|^{\alpha}. \tag{SA2.3}$$

(iv) $\Psi(\boldsymbol{x},\eta;\boldsymbol{q}) := \mathbb{E}[\psi(y_i,\eta;\boldsymbol{q})\,|\,\boldsymbol{x}_i=\boldsymbol{x}]$ is twice continuously differentiable with respect to η,

$$\sup_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{q}\in\mathcal{Q}}\sup_{\zeta\in B_{\boldsymbol{q}}(\boldsymbol{x})}|\Psi_k(\boldsymbol{x},\eta(\zeta);\boldsymbol{q})|<\infty, \quad \Psi_k(\boldsymbol{x},\eta;\boldsymbol{q}) := \frac{\partial^k}{\partial\eta^k}\Psi(\boldsymbol{x},\eta;\boldsymbol{q}), \quad k=1,2, \tag{SA2.4}$$

and

$$\inf_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{q}\in\mathcal{Q}}\inf_{\zeta\in B_{\boldsymbol{q}}(\boldsymbol{x})}\Psi_1(\boldsymbol{x},\eta(\zeta);\boldsymbol{q})\eta^{(1)}(\zeta)^2>0.$$

In addition, we make the following assumption on the complexity of $\{\psi(\cdot,\cdot;\boldsymbol{q})\}$. It is much more general than Assumption E in the main paper (Proposition SA2.9).

Assumption SA2.5 (Complexity of $\{\psi(\cdot,\cdot;\boldsymbol{q})\}$.)**.** *For any fixed $r>0$ and $c>0$, $l\in\{1,\dots,K\}$, the classes of functions*

$$\begin{aligned}
\mathcal{G}_1 &:= \big\{\mathcal{X}\times\mathcal{Y}\ni(\boldsymbol{x},y)\mapsto\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta});\boldsymbol{q})-\psi(y,\eta(\mu_0(\boldsymbol{x},\boldsymbol{q}));\boldsymbol{q})\colon\\
&\qquad \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty\le r,\boldsymbol{q}\in\mathcal{Q}\big\}\\
\mathcal{G}_2 &:= \big\{\mathcal{X}\times\mathcal{Y}\ni(\boldsymbol{x},y)\mapsto\psi(y,\eta(\mu_0(\boldsymbol{x},\boldsymbol{q}));\boldsymbol{q})\colon\boldsymbol{q}\in\mathcal{Q}\big\},\\
\mathcal{G}_3 &:= \big\{\mathcal{X}\times\mathcal{Y}\ni(\boldsymbol{x},y)\mapsto\\
&\qquad [\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta});\boldsymbol{q})-\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})]\mathbb{1}(\boldsymbol{x}\in\mathcal{N}_{[c\log n]}(\delta))\colon\\
&\qquad \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty\le r,\delta\in\Delta,\boldsymbol{q}\in\mathcal{Q}\big\},\\
\mathcal{G}_4 &:= \big\{\mathcal{X}\ni\boldsymbol{x}\mapsto\Psi_1(\boldsymbol{x},\eta(\mu_0(\boldsymbol{x},\boldsymbol{q}));\boldsymbol{q}) : \boldsymbol{q}\in\mathcal{Q}\big\},\\
\mathcal{G}_5 &:= \big\{\mathcal{X}\times\mathcal{Y}\ni(\boldsymbol{x},y)\mapsto\\
&\qquad p_l(\boldsymbol{x})[\psi(y,\eta(\mu_0(\boldsymbol{x},\boldsymbol{q}));\boldsymbol{q})-\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})] : \boldsymbol{q}\in\mathcal{Q}\big\}
\end{aligned}$$

satisfy the uniform entropy bound (SA1.3) with respective envelopes and constants as follows:

$$\begin{aligned}
\bar{G}_1 &\lesssim 1, & A_1 &\lesssim 1, & V_1 &\lesssim K \asymp h^{-d};\\
\bar{G}_2(\boldsymbol{x},y) &\lesssim \overline{\psi}(\boldsymbol{x},y), & A_2 &\lesssim 1, & V_2 &\lesssim 1;\\
\bar{G}_3 &\lesssim 1, & A_3 &\lesssim 1, & V_3 &\lesssim \log^d n;\\
\bar{G}_4 &\lesssim 1, & A_4 &\lesssim 1, & V_4 &\lesssim 1;\\
\bar{G}_5 &\lesssim 1, & A_5 &\lesssim 1, & V_5 &\lesssim 1,
\end{aligned}$$

where for $s\in\mathbb{Z}\cap[0,\infty)$, $\mathcal{N}_s(\delta)$ denotes the s-neighborhood of cell $\delta\in\Delta$ which is the union of all cells $\delta'\in\Delta$ reachable from some point of δ in no more than s steps (following a continuous path).

Assumption SA2.6 (Approximation error)**.** *There exists a vector of coefficients $\boldsymbol{\beta}_0(\boldsymbol{q}) \in \mathbb{R}^K$ such that for all $\boldsymbol{\varsigma}$ satisfying $|\boldsymbol{\varsigma}| \le \varsigma$ in Assumption SA2.2,*

$$\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{x}\in\mathcal{X}} \big|\mu_0^{(\boldsymbol{\varsigma})}(\boldsymbol{x},\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{\varsigma})}(\boldsymbol{x};\Delta,m)\big| \lesssim h^{m-|\boldsymbol{\varsigma}|}.$$

In particular, this requires $\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x};\Delta,m)\big| \lesssim h^{m-|\boldsymbol{v}|}$.

Assumption SA2.7 (Estimator of the Gram matrix)**.** *$\widehat{\mathbf{Q}}_{\boldsymbol{q}}$ is an estimator of the matrix $\bar{\mathbf{Q}}_{\boldsymbol{q}}$ such that $\|\widehat{\mathbf{Q}}_{\boldsymbol{q}} - \bar{\mathbf{Q}}_{\boldsymbol{q}}\|_\infty \lesssim_{\mathbb{P}} h^d r_{\mathtt{Q}}$ and $\|\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_\infty \lesssim_{\mathbb{P}} h^{-d}$, where $r_{\mathtt{Q}} = o(1)$.*

Finally, we state the following technical addition to Assumption SA2.5 needed for Theorem SA5.1. As proven in Proposition SA2.9, it is also automatically true under Assumption E in the main paper.

Assumption SA2.8 (Additional complexity assumption)**.** *For any fixed $c > 0$, $\gamma > 0$, and a positive sequence $\varepsilon_n \to 0$ the class of functions*

$$\Big\{(\boldsymbol{x},y) \mapsto \int_{-\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{v}}^{0} \big[\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta})+t);\boldsymbol{q}) - \psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\big] \\ \times \eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta})+t)\,\mathrm{d}t \cdot \mathbb{1}(\boldsymbol{x}\in\mathcal{N}_{[c\log n]}(\delta)) : \\ \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty \le \gamma, \|\boldsymbol{v}\|_\infty \le \varepsilon_n, \delta\in\Delta, \boldsymbol{q}\in\mathcal{Q}\Big\}$$

with envelope ε_n multiplied by a large enough constant (not depending on n), satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1/\varepsilon_n$, $V \lesssim \log^d n$, where the constants in $\lesssim$ do not depend on n.

SA2.1 Simplifying assumption on loss function

As discussed in the paper, it is possible to impose a simplifying assumption on the loss function, which is motivated by the examples considered. More specifically, we aim at simplifying the general complexity Assumptions SA2.5 and SA2.8.

Proposition SA2.9 (Simplified setup)**.** *Assume the following conditions.*

(i) $q \mapsto \mu_0(\boldsymbol{x},q)$ is non-decreasing.

(ii) $\rho(y,\eta;q) = \sum_{j=1}^4 \rho_j(y,\eta;q)$, where the functions $\rho_j(y,\eta;q)$ are of the following types:

- *Type I: $\rho_1(y,\eta;q) = (f_1(y) + D_1\eta)\mathbb{1}\{y \le \eta\}$,*
- *Type II: $\rho_2(y,\eta;q) = (f_2(y) + D_2\eta)\mathbb{1}\{y \le q\}$,*
- *Type III: $\rho_3(y,\eta;q) = (f_3(y) + D_3\eta)q$,*
- *Type IV: $\rho_4(y,\eta;q) = \mathcal{T}(y,\eta)$,*

where $f_j(\cdot)$ are fixed continuous functions, D_j are universal constants, and $\mathcal{T}(y;\eta) : \mathcal{Y}\times\mathcal{E} \to \mathbb{R}$ is a measurable function, differentiable in η for any fixed y with a derivative $\tau(y,\eta) := \frac{\partial}{\partial\eta}\mathcal{T}(y,\eta)$.

(iii) $\mathbb{E}[\tau(y_i,\eta)\,|\,\boldsymbol{x}_i = \boldsymbol{x}]$ is also differentiable in η for any $\boldsymbol{x}$. The functions

$$(y,\eta) \mapsto \tau(y,\eta), \text{ and} \\ (\boldsymbol{x},\eta) \mapsto \frac{\partial}{\partial\eta}\mathbb{E}[\tau(y_i,\eta)\,|\,\boldsymbol{x}_i = \boldsymbol{x}]$$

are continuous in their arguments, and α-Hölder continuous (with $0 < \alpha \le 1$) in η on $\mathcal{Y}\times\mathcal{K}$ and $\mathcal{X}\times\mathcal{K}$ respectively, where $\mathcal{K}$ is any fixed compact subset of $\mathcal{E}$ (with the Hölder constant possibly depending on $\mathcal{K}$, but on y or $\boldsymbol{x}$).

(iv) Assumptions SA2.1 to SA2.4 hold with $\overline{\psi}(\boldsymbol{x}, y) = \overline{\tau}(\boldsymbol{x}, y) + |D_1| + |D_2| + |D_3| \max_{q \in \mathcal{Q}} |q|$, *where* $\overline{\tau}(\boldsymbol{x}, y)$ *is a measurable envelope of* $\{(\boldsymbol{x}, y) \mapsto \tau(y, \eta(\mu_0(\boldsymbol{x}, q)))\}$.

(v) If D_1 *is nonzero,* $F_{Y|X}$ *is differentiable and* $f_{Y|X}$ *is its derivative (in particular,* $\mathfrak{M}$ *is Lebesgue measure),* $(\boldsymbol{x}, \eta) \mapsto f_{Y|X}(\eta|\boldsymbol{x})$ *is continuous in both arguments and continuously differentiable in* η.

Then Assumptions SA2.5 and SA2.8 also hold.

We will prove this now.

In this case

$$\psi(y, \eta; q) = D_1 \mathbb{1}\{y \le \eta\} + D_2 \mathbb{1}\{y \le q\} + D_3 q + \tau(y, \eta).$$

SA2.1.1 Verifying Assumption SA2.5

Lemma SA2.10 (Class G_1)**.** *The class* $\mathcal{G}_1$ *described in Assumption SA2.5 with a large enough constant envelope satisfies the uniform entropy bound* (SA1.3) *with* $A \lesssim 1$, $V \lesssim K$.

Proof of Lemma SA2.10. It is shown in the proof of Proposition SA8.1 (replacing $<$ with $\le$ does not change the argument) that for any fixed $r > 0$ the class

$$\left\{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y \le \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})\} - \mathbb{1}\{y \le \eta(\mu_0(\boldsymbol{x}, q))\} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, q \in \mathcal{Q}\right\}$$

with envelope 2 satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$.

Next, assume that the infinity-norms of $\boldsymbol{\beta}$ and $\widetilde{\boldsymbol{\beta}}$ lie in a fixed bounded interval, and let $q, \tilde{q} \in \mathcal{Q}$. Note that by α-Hölder continuity of $\tau(y, \cdot)$ on compacta

$$\begin{aligned}
&\left|\tau(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})) - \tau(y; \eta(\mu_0(\boldsymbol{x}, q))) - \tau(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\widetilde{\boldsymbol{\beta}})) + \tau(y, \eta(\mu_0(\boldsymbol{x}, \tilde{q})))\right| \\
&\quad \lesssim \left|\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}) - \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\widetilde{\boldsymbol{\beta}})\right|^\alpha + \left|\eta(\mu_0(\boldsymbol{x}, q)) - \eta(\mu_0(\boldsymbol{x}, \tilde{q}))\right|^\alpha \\
&\quad \lesssim \|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\|_\infty^\alpha + |q - \tilde{q}|^\alpha,
\end{aligned}$$

where the constants in $\lesssim$ do not depend on $\boldsymbol{\beta}$, $\widetilde{\boldsymbol{\beta}}$, q, $\tilde{q}$, and we used that $\eta(\cdot)$ on a fixed bounded interval is Lipschitz, and $q \mapsto \mu_0(\boldsymbol{x}, q)$ is Lipschitz (uniformly in $\boldsymbol{x}$). Again by α-Hölder continuity for any fixed $r > 0$ the class

$$\left\{(\boldsymbol{x}, y) \mapsto \tau\big(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})\big) - \tau\big(y, \eta(\mu_0(\boldsymbol{x}, q))\big) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, q \in \mathcal{Q}\right\}$$

has a constant envelope. It follows that it satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim K$.

Combining these results concludes the proof of Lemma SA2.10 by Lemma SA3.4. $\square$

Lemma SA2.11 (Class G_2)**.** *The class* $\mathcal{G}_2$ *described in Assumption SA2.5 with envelope* $\overline{\psi}(\boldsymbol{x}, y)$ *satisfies the uniform entropy bound* (SA1.3) *with* $A \lesssim 1$, $V \lesssim 1$.

Proof of Lemma SA2.11. The class

$$\left\{\mathcal{X} \times \mathcal{Y} \ni (\boldsymbol{x}, y) \mapsto \tau\big(y, \eta(\mu_0(\boldsymbol{x}, q))\big) : q \in \mathcal{Q}\right\}$$

with envelope $\overline{\tau}(\boldsymbol{x}, y)$ satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim 1$ by α-Hölder continuity and since $\eta(\cdot)$ on a fixed bounded interval is Lipschitz, $q \mapsto \mu_0(\boldsymbol{x}, q)$ is Lipschitz (uniformly in $\boldsymbol{x}$).

The class

$$\left\{\mathcal{X} \times \mathcal{Y} \ni (\boldsymbol{x}, y) \mapsto \mathbb{1}\{y \le \eta(\mu_0(\boldsymbol{x}, q))\} : q \in \mathcal{Q}\right\}$$

with envelope 1 satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim 1$ because $\mu_0(\boldsymbol{x}, q)$ is nondecreasing in q, see the proof of Proposition SA8.1.

The class

$$\{\mathcal{X} \times \mathcal{Y} \ni (\boldsymbol{x}, y) \mapsto \mathbb{1}\{y \le q\} \colon q \in \mathcal{Q}\}$$

with envelope 1 satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim 1$ because it is VC-subgraph with a constant index, (cf. the proof of Proposition SA8.3).

The class

$$\{\mathcal{X} \times \mathcal{Y} \ni (\boldsymbol{x}, y) \mapsto q \colon q \in \mathcal{Q}\}$$

with envelope $\max_{q \in \mathcal{Q}} |q|$ satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim 1$ because it is VC-subgraph with a constant index (as a subclass of a one-dimensional space of functions, namely constants in $\boldsymbol{x}, y$).

It is left to apply Lemma SA3.4, concluding the proof of Lemma SA2.11. □

Lemma SA2.12 (Class G_3)**.** *The class $\mathcal{G}_3$ described in Assumption SA2.5 with a large enough constant envelope satisfies the uniform entropy bound* (SA1.3) *with $A \lesssim 1$, $V \lesssim \log^d n$.*

Proof of Lemma SA2.12. Fix $\delta \in \Delta$ and some large enough $R > 0$. Note that if $\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\}$ and $\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}$ are both nonzero, $\boldsymbol{\beta}$ must lie in a vector subspace $\mathcal{B}_\delta$ of $\mathbb{R}^K$ of dimension $O(\log^d n)$. For any positive and small enough ε, the class of vectors $\{\boldsymbol{\beta} \in \mathcal{B}_\delta, \|\boldsymbol{\beta}\|_\infty < R\}$ has an infinity-norm ε-net $\bar{\mathcal{B}}_\delta^\varepsilon$ such that

$$\log|\bar{\mathcal{B}}_\delta^\varepsilon| \lesssim \log^d n \log(C/\varepsilon),$$

where C is some positive constant.

By α-Hölder continuity of $\tau(y, \cdot)$ on compacta and since $\eta(\cdot)$ on a compact is Lipschitz, this means that the class of bounded (by a constant not depending on n) functions

$$\begin{aligned}\mathcal{G}_{3,\delta} := \{&(\boldsymbol{x}, y) \mapsto [\tau(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})) - \tau(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(q)))]\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\} : \\ &\|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, q \in \mathcal{Q}\}\end{aligned}$$

has a covering number bound

$$\log N(\mathcal{G}_{3,\delta}, \text{sup-norm}, C'\varepsilon^\alpha) \lesssim \log^d n \log(C/\varepsilon),$$

where C' is some other positive constant. This means that also

$$\log N(\mathcal{G}_{3,\delta}, \text{sup-norm}, \varepsilon) \lesssim \log^d n \log(C''/\varepsilon)$$

for some other positive constant C''. Finally, from this we can conclude that $\mathcal{G}_{3,\delta}$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$. Therefore, the union of $O(h^{-d})$ such classes

$$\begin{aligned}\{&(\boldsymbol{x}, y) \mapsto \{\tau(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})) - \tau(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(q)))\}\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\} : \\ &\|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, q \in \mathcal{Q}, \delta \in \Delta\}\end{aligned}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$, see (SA5.34).

Next, since $\eta(\cdot)$ is monotonic, the functions $\boldsymbol{x} \mapsto \boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}$ with $\boldsymbol{\beta} \in \mathcal{B}_\delta$ form a vector space of $O(\log^d n)$ dimension, and $\boldsymbol{x} \mapsto \mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\}$ is one fixed function, the class

$$\{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y \le \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})\}\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\} : \boldsymbol{\beta} \in \mathbb{R}^K\}$$

is VC with an index $O(\log^d n)$. Therefore, it satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$. By Lemma SA3.4, a subclass of the difference of two such classes

$$\big\{(\boldsymbol{x}, y) \mapsto \big(\mathbb{1}\{y \le \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\} - \mathbb{1}\{y \le \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q))\}\big)\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\} : \\ \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, q \in \mathcal{Q}\big\}$$

for fixed $\delta \in \Delta$ also satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$. Therefore, the union of $O(h^{-d})$ such classes

$$\big\{(\boldsymbol{x}, y) \mapsto \{\mathbb{1}\{y \le \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\} - \mathbb{1}\{y \le \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q))\}\}\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\} : \\ \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, \delta \in \Delta, q \in \mathcal{Q}\big\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$, see (SA5.34).

It is left to apply Lemma SA3.4 once again, concluding the proof of Lemma SA2.12. □

Lemma SA2.13 (Class G_4)**.** *The class $\mathcal{G}_4$ described in Assumption SA2.5 with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim 1$.*

Proof of Lemma SA2.13. In this case

$$\Psi(\boldsymbol{x}, \eta; q) = \mathbb{E}[\tau(y_i, \eta) \,|\, \boldsymbol{x}_i = \boldsymbol{x}] + D_1 F_{Y|X}(\eta|\boldsymbol{x}) + D_2 F_{Y|X}(q|\boldsymbol{x}) + D_3 q$$

and

$$\Psi_1(\boldsymbol{x}, \eta; q) = \frac{\partial}{\partial\eta}\mathbb{E}[\tau(y_i, \eta) \,|\, \boldsymbol{x}_i = \boldsymbol{x}] + D_1 f_{Y|X}(\eta|\boldsymbol{x}).$$

By assumption, if η lies in a fixed compact, this function of $\boldsymbol{x}, \eta$ is bounded (by continuity). Moreover,

$$\begin{aligned}&\big|\Psi_1(\boldsymbol{x}, \eta(\mu_0(\boldsymbol{x}, q)); q) - \Psi_1(\tilde{q}, \boldsymbol{x}; q)\eta(\mu_0(\boldsymbol{x}, \tilde{q}))\big| \\ &\quad\lesssim |\eta(\mu_0(\boldsymbol{x}, q)) - \eta(\mu_0(\boldsymbol{x}, \tilde{q}))|^\alpha + |\eta(\mu_0(\boldsymbol{x}, q)) - \eta(\mu_0(\boldsymbol{x}, \tilde{q}))| \overset{\text{(a)}}{\lesssim} |q - \tilde{q}|^\alpha + |q - \tilde{q}|,\end{aligned}$$

where in (a) we used that $\eta(\cdot)$ on compacta is Lipschitz and $q \mapsto \mu_0(\boldsymbol{x}, q)$ is uniformly Lipschitz in q. The result of Lemma SA2.13 follows. □

Lemma SA2.14 (Class G_5)**.** *The class $\mathcal{G}_5$ described in Assumption SA2.5 with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim 1$.*

Proof of Lemma SA2.14. Take $R > 0$ fixed and large enough so that $\|\boldsymbol{\beta}_{0,q}\|_\infty \le R$ for all q and n. Note that for $p_l(\boldsymbol{x})$ and $\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}$ to be nonzero at the same time, $\boldsymbol{\beta}$ must lie in a fixed vector subspace $\mathcal{B}_l$ of $\mathbb{R}^K$ of bounded dimension. For $0 < \varepsilon < 1$, the class of vectors $\{\boldsymbol{\beta} \in \mathcal{B}_l, \|\boldsymbol{\beta}\|_\infty \le R\}$ has an infinity-norm ε-net $\bar{\mathcal{B}}_l^\varepsilon$ such that

$$\log\big|\bar{\mathcal{B}}_l^\varepsilon\big| \lesssim \log(C/\varepsilon),$$

where C is some positive constant.

By α-Hölder continuity of $\tau(y, \cdot)$ on compacta, since $\eta(\cdot)$ on a compact is Lipschitz and $\mu_0(\cdot, q)$ is uniformly Lipschitz in q, this means that the class of bounded (by a constant not depending on n) functions

$$\mathcal{G}_5' := \big\{(\boldsymbol{x}, y) \mapsto p_l(\boldsymbol{x})\big(\tau(y, \eta(\mu_0(\boldsymbol{x}, q))) - \tau(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q)))\big) : q \in \mathcal{Q}\big\}$$

has a covering number bound

$$\log N(\mathcal{G}_5', \text{sup-norm}, C'\varepsilon^\alpha) \lesssim \log(C/\varepsilon),$$

where C' is some other positive constant. It follows that this class with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim 1$.

As in the proof of Lemma SA2.11, the class

$$\big\{\mathcal{X} \times \mathcal{Y} \ni (\boldsymbol{x}, y) \mapsto \mathbb{1}\{y \leq \eta(\mu_0(\boldsymbol{x}, q))\} \colon q \in \mathcal{Q}\big\}$$

with envelope 1 satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim 1$. Since $\eta(\cdot)$ is monotonic, the functions $\boldsymbol{x} \mapsto \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x})$ with $\boldsymbol{\beta} \in \mathcal{B}_l$ form a vector space, and $p_l(\boldsymbol{x})$ is one fixed function, we have that the class

$$\big\{\mathcal{X} \times \mathcal{Y} \ni (\boldsymbol{x}, y) \mapsto p_l(\boldsymbol{x})\mathbb{1}\{y \leq \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})\} : \boldsymbol{\beta} \in \mathbb{R}^K\big\}$$

is VC with a bounded index. Then it also satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim 1$.

It is left to apply Lemma SA3.4, concluding the proof of Lemma SA2.14. □

SA2.1.2 Verifying Assumption SA2.8

Suppose $\theta_1, \theta_2, \theta \in \mathbb{R}$ and $\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta} \in \mathbb{R}$ all lie in a fixed compact interval. Then

$$\begin{aligned}
&\mathcal{T}(y, \eta(\theta_1)) - \mathcal{T}(y, \eta(\theta_2)) - [\eta(\theta_1) - \eta(\theta_2)]\tau(y, \eta(\theta)) \\
&\quad - \mathcal{T}(y, \eta(\tilde{\theta}_1)) + \mathcal{T}(y, \eta(\tilde{\theta}_2)) + [\eta(\tilde{\theta}_1) - \eta(\tilde{\theta}_2)]\tau(y, \eta(\tilde{\theta})) \\
&= \tau(y, \zeta_{1,y})[\eta(\theta_1) - \eta(\tilde{\theta}_1)] - \tau(y, \zeta_{2,y})[\eta(\theta_2) - \eta(\tilde{\theta}_2)] \\
&\quad - [\eta(\theta_1) - \eta(\tilde{\theta}_1) + \eta(\tilde{\theta}_1) - \eta(\theta_2)]\tau(y, \eta(\theta)) + [\eta(\tilde{\theta}_1) - \eta(\theta_2) + \eta(\theta_2) - \eta(\tilde{\theta}_2)]\tau(y, \eta(\tilde{\theta})) \\
&= [\tau(y, \zeta_{1,y}) - \tau(y, \eta(\theta))] \cdot [\eta(\theta_1) - \eta(\tilde{\theta}_1)] - [\tau(y, \zeta_{2,y}) - \tau(y, \eta(\tilde{\theta}))] \cdot [\eta(\theta_2) - \eta(\tilde{\theta}_2)] \\
&\quad - [\eta(\tilde{\theta}_1) - \eta(\theta_2)]\tau(y, \eta(\theta)) + [\eta(\tilde{\theta}_1) - \eta(\theta_2)]\tau(y, \eta(\tilde{\theta})) \\
&= \underbrace{[\tau(y, \zeta_{1,y}) - \tau(y, \eta(\theta))]}_{\lesssim 1} \cdot \underbrace{[\eta(\theta_1) - \eta(\tilde{\theta}_1)]}_{\lesssim |\theta_1 - \tilde{\theta}_1|} - \underbrace{[\tau(y, \zeta_{2,y}) - \tau(y, \eta(\tilde{\theta}))]}_{\lesssim 1} \cdot \underbrace{[\eta(\theta_2) - \eta(\tilde{\theta}_2)]}_{\lesssim |\theta_2 - \tilde{\theta}_2|} \\
&\quad + \underbrace{[\eta(\tilde{\theta}_1) - \eta(\theta_2)]}_{\lesssim 1} \underbrace{[\tau(y, \eta(\tilde{\theta})) - \tau(y, \eta(\theta))]}_{\lesssim |\theta - \tilde{\theta}|^\alpha} \lesssim |\theta_1 - \tilde{\theta}_1| + |\theta_2 - \tilde{\theta}_2| + |\theta - \tilde{\theta}|^\alpha
\end{aligned}$$

for some $\zeta_{1,y}$ between $\eta(\theta_1)$ and $\eta(\tilde{\theta}_1)$, $\zeta_{2,y}$ between $\eta(\theta_2)$ and $\eta(\tilde{\theta}_2)$, where we used the α-Hölder continuity of $\tau(y, \cdot)$ on a fixed compact and the Lipschitzness of $\eta(\cdot)$ on a compact. This means that the class of functions

$$\begin{aligned}
\mathcal{G}' := \Big\{(\boldsymbol{x}, y) \mapsto \big(&\mathcal{T}\big(y, \eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})\big)\big) - \mathcal{T}\big(y, \eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})\big)\big) \\
&- \big[\eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})\big) - \eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})\big)\big]\tau\big(y, \eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(q)\big)\big)\big) \\
&\times \mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\} : q \in \mathcal{Q}, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \leq r, \|\boldsymbol{v}\|_\infty \leq \varepsilon_n\Big\}
\end{aligned}$$

has a covering number bound

$$\log N(\mathcal{G}', \text{sup-norm}, \varepsilon) \lesssim \log^d n \log\left(\frac{C_1}{\varepsilon}\right), \tag{SA2.5}$$

for all small enough positive ε, and C_1 is some positive constant (not depending on n), where we used that all $\boldsymbol{\beta}_0(q)$, $\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}-\boldsymbol{v}$ must lie in a vector space of dimension $O(\log^d n)$ if $\mathbb{1}\{\boldsymbol{x}\in\mathcal{N}_{[c\log n]}(\delta)\}$ is not zero. Applying the mean-value theorem to

$$\mathcal{T}\big(y,\eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta})\big)\big)-\mathcal{T}\big(y,\eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}-\boldsymbol{v})\big)\big)$$

and α-Hölder continuity again, we see that class $\mathcal{G}'$ has an envelope which is ε_n multiplied by a large enough constant C_2. Replacing ε with $C_2\varepsilon\varepsilon_n$ (for large enough n this is small enough) in (SA2.5), we get a covering number bound

$$\log N(\mathcal{G}',\text{sup-norm},C_2\varepsilon\varepsilon_n)\lesssim\log^d n\log\bigg(\frac{C_1}{C_2\varepsilon\varepsilon_n}\bigg).$$

It follows that class $\mathcal{G}'$ satisfies the uniform entropy bound (SA1.3) with $A\lesssim 1/\varepsilon_n$ and $V\lesssim\log^d n$.

Therefore, the union of $O(h^{-d})$ such classes

$$\begin{aligned}\mathcal{G}':=\Big\{\Big(&\mathcal{T}\big(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}))\big)-\mathcal{T}\big(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}-\boldsymbol{v}))\big)\\&-\big[\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}))-\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}-\boldsymbol{v}))\big]\tau(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(q)))\Big)\\&\times\mathbb{1}\{\boldsymbol{x}\in\mathcal{N}_{[c\log n]}(\delta)\}:q\in\mathcal{Q},\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(q)\|_\infty\le r,\|\boldsymbol{v}\|_\infty\le\varepsilon_n,\delta\in\Delta\Big\},\end{aligned}$$

also with envelope ε_n multiplied by a large enough constant, satisfies the uniform entropy bound (SA1.3) with $A\lesssim 1/\varepsilon_n$ and $V\lesssim\log^d n$, by the same argument as (SA5.34).

Next,

$$\begin{aligned}&[f_1(y)+D_1\eta(\theta_1)]\mathbb{1}\{y\le\eta(\theta_1)\}-[f_1(y)+D_1\eta(\theta_2)]\mathbb{1}\{y\le\eta(\theta_2)\}\\&\quad-[\eta(\theta_1)-\eta(\theta_2)]D_1\mathbb{1}\{y\le\eta(\theta)\}\\&=f_1(y)[\mathbb{1}\{y\le\eta(\theta_1)\}-\mathbb{1}\{y\le\eta(\theta_2)\}]+D_1\eta(\theta_1)\mathbb{1}\{y\le\eta(\theta_1)\}-D_1\eta(\theta_2)\mathbb{1}\{y\le\eta(\theta_2)\}\\&\quad-D_1[\eta(\theta_1)-\eta(\theta_2)]\mathbb{1}\{y\le\eta(\theta)\}.\end{aligned}$$

It is proven by the same argument as in the proof of Proposition SA8.1 that the class

$$\begin{aligned}\Big\{(\boldsymbol{x},y)\mapsto\Big(&\big[f_1(y)+D_1\eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta})\big)\big]\mathbb{1}\big\{y\le\eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta})\big)\big\}\\&-\big[f_1(y)+D_1\eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}-\boldsymbol{v})\big)\big]\mathbb{1}\{y\le\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}-\boldsymbol{v}))\}\\&-\big[\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}))-\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q)+\boldsymbol{\beta}-\boldsymbol{v}))\big]D_1\mathbb{1}\{y\le\eta(\theta)\}\Big)\\&\times\mathbb{1}\{\boldsymbol{x}\in\mathcal{N}_{[c\log n]}(\delta)\}:q\in\mathcal{Q},\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(q)\|_\infty\le r,\|\boldsymbol{v}\|_\infty\le\varepsilon_n,\delta\in\Delta\Big\}\end{aligned}$$

satisfies the uniform entropy bound (SA1.3) with $A\lesssim 1/\varepsilon_n$ and $V\lesssim\log^d n$.

The terms $(f_2(y)+D_2\eta)\mathbb{1}\{y\le q\}$ and $(f_3(y)+D_3\eta)q$ play no role in this verification because they cancel out in the class described in Assumption SA2.8.

It is left to apply Lemma SA3.4.

The proof of Proposition SA2.9 is finished.

SA3 Frequently used lemmas

We collect several lemmas that will be used multiple times throughout this supplemental appendix. Lemmas SA3.1 to SA3.9 are well-known facts, so we provide either brief proofs or references to the literature.

Lemma SA3.1 (Second moment bound of the max of sub-Gaussian random variables). *Let $n \geq 3$ and $\xi_1, \dots, \xi_n$ be σ^2-sub-Gaussian random variables (not necessarily independent). Then*

$$\mathbb{E}\Big[\max_{1\leq i\leq n} \xi_i^2\Big]^{1/2} \leq C_U \sigma \sqrt{\log n},$$

where C_U is a universal constant.

Proof. If p is an even positive integer, $\mathbb{E}[\xi_i^p] \leq 3\sigma^p p (p/2)^{p/2}$. Then

$$\begin{aligned}\mathbb{E}\Big[\max_{1\leq i\leq n} \xi_i^2\Big]^{1/2} &\leq \mathbb{E}\Big[\max_{1\leq i\leq n} \xi_i^p\Big]^{1/p} \leq \Big(\sum_{i=1}^n \mathbb{E}\big[\xi_i^p\big]\Big)^{1/p} \lesssim n^{1/p} \cdot \sigma \cdot p^{1/p}\sqrt{p} \\ &\lesssim \sigma n^{1/p}\sqrt{p} \qquad \text{using } p^{1/p} \leq 2.\end{aligned}$$

It is left to take $p = p_n$ such that $\ln n \leq p \leq 2\ln n$. $\square$

Lemma SA3.2 (Boundedness of conditional expectation in probability implies unconditional boundedness in probability). *Let X_n be a sequence of integrable random variables, $\boldsymbol{D}_n$ a sequence of random vectors, r_n a sequence of positive numbers. If $\mathbb{E}[|X_n| \mid \boldsymbol{D}_n] \lesssim_{\mathbb{P}} r_n$, then $|X_n| \lesssim_{\mathbb{P}} r_n$.*

Proof. Take any sequence of positive numbers $\gamma_n \to \infty$. By Markov's inequality,

$$\mathbb{P}\{|X_n| > \gamma_n r_n \mid \boldsymbol{D}_n\} \leq \frac{\mathbb{E}[|X_n| \mid \boldsymbol{D}_n]}{\gamma_n r_n} \lesssim_{\mathbb{P}} \frac{1}{\gamma_n} = o(1).$$

In other words, the sequence of random variables $\mathbb{P}\{|X_n| > \gamma_n r_n \mid \boldsymbol{D}_n\}$ converges to zero in probability. By dominated convergence (in probability), the sequence of numbers $\mathbb{P}\{|X_n| > \gamma_n r_n\}$ converges to zero. Since it is true for any positive sequence $\gamma_n \to \infty$, this means $|X_n| = O_{\mathbb{P}}(r_n)$. $\square$

Lemma SA3.3 (Converging to zero in conditional probability is the same as converging to zero in probability). *Let X_n be a sequence of random variables, $\boldsymbol{D}_n$ a sequence of random vectors. The following are equivalent:*

(i) for any $\varepsilon > 0$, we have $\mathbb{P}\{|X_n| > \varepsilon \mid \boldsymbol{D}_n\} = o_{\mathbb{P}}(1)$;

(ii) $|X_n| = o_{\mathbb{P}}(1)$.

Proof. The implication (i) $\Rightarrow$ (ii) follows from dominated convergence in probability. To prove the converse, take any $\varepsilon, \gamma > 0$. By Markov's inequality,

$$\mathbb{P}\{\mathbb{P}\{|X_n| > \varepsilon \mid \boldsymbol{D}_n\} > \gamma\} \leq \frac{\mathbb{P}\{|X_n| > \varepsilon\}}{\gamma} \to 0,$$

so by definition $\mathbb{P}\{|X_n| > \varepsilon \mid \boldsymbol{D}_n\} = o_{\mathbb{P}}(1)$. $\square$

Lemma SA3.4 (Permanence properties of the uniform entropy bound). *Let $\mathcal{F}$ and $\mathcal{G}$ be two classes of measurable functions from $S \to \mathbb{R}$ on a measurable space $(S, \mathcal{S})$ with strictly positive measurable envelopes F and G respectively. Then the uniform entropy numbers of $\mathcal{F}\mathcal{G} = \{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$ satisfy*

$$\sup_{\mathbb{Q}} \log N\big(\mathcal{F}\mathcal{G}, \|\cdot\|_{\mathbb{Q},2}, \varepsilon\|FG\|_{\mathbb{Q},2}\big)$$

$$\leq \sup_{\mathbb{Q}} \log N\Big(\mathcal{F}, \|\cdot\|_{\mathbb{Q},2}, \frac{\varepsilon\|F\|_{\mathbb{Q},2}}{2}\Big) + \sup_{\mathbb{Q}} \log N\Big(\mathcal{G}, \|\cdot\|_{\mathbb{Q},2}, \frac{\varepsilon\|G\|_{\mathbb{Q},2}}{2}\Big)$$

for all $\varepsilon > 0$. Also, the uniform entropy numbers of $\mathcal{F} + \mathcal{G} = \{f + g : f \in \mathcal{F}, g \in \mathcal{G}\}$ satisfy

$$\begin{aligned}
&\sup_{\mathbb{Q}} \log N\big(\mathcal{F} + \mathcal{G}, \|\cdot\|_{\mathbb{Q},2}, \varepsilon\|F + G\|_{\mathbb{Q},2}\big) \\
&\quad \leq \sup_{\mathbb{Q}} \log N\Big(\mathcal{F}, \|\cdot\|_{\mathbb{Q},2}, \frac{\varepsilon\|F\|_{\mathbb{Q},2}}{2}\Big) + \sup_{\mathbb{Q}} \log N\Big(\mathcal{G}, \|\cdot\|_{\mathbb{Q},2}, \frac{\varepsilon\|G\|_{\mathbb{Q},2}}{2}\Big)
\end{aligned}$$

for all $\varepsilon > 0$. In both cases, $\mathbb{Q}$ ranges over all finitely-discrete probability measures.

Proof. This lemma is well-known. See, for example, [18]. □

Lemma SA3.5 (Maximal inequality for Gaussian vectors)**.** *Take $n \geq 2$. Let $X_i \sim \mathcal{N}(0, \sigma_i^2)$ for $1 \leq i \leq n$ (not necessarily independent), with $\sigma_i^2 \leq \sigma^2$. Then*

$$\mathbb{E}\Big[\max_{1\leq i\leq n} X_i\Big] \leq \sigma\sqrt{2\log n},$$
$$\mathbb{E}\Big[\max_{1\leq i\leq n} |X_i|\Big] \leq 2\sigma\sqrt{\log n}.$$

If $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are positive semi-definite $n \times n$ matrices and $\boldsymbol{n} \sim \mathcal{N}(0, \mathbf{I}_n)$, then

$$\mathbb{E}[\|\mathbf{\Sigma}_1^{1/2}\boldsymbol{n} - \mathbf{\Sigma}_2^{1/2}\boldsymbol{n}\|_\infty] \leq 2\sqrt{\log n}\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_2^{1/2}.$$

If further $\mathbf{\Sigma}_1$ is positive definite, then

$$\mathbb{E}[\|\mathbf{\Sigma}_1^{1/2}\boldsymbol{n} - \mathbf{\Sigma}_2^{1/2}\boldsymbol{n}\|_\infty] \leq \sqrt{\log n}\lambda_{\min}(\mathbf{\Sigma}_1)^{-1/2}\|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_2.$$

Proof. See Lemma SA31 in [7]. □

Lemma SA3.6 (A maximal inequality for i.n.i.d. empirical processes)**.** *Let $X_1, \ldots, X_n$ be independent but not necessarily identically distributed (i.n.i.d.) random variables taking values in a measurable space $(S, \mathcal{S})$. Denote the joint distribution of $X_1, \ldots, X_n$ by $\mathbb{P}$ and the marginal distribution of X_i by $\mathbb{P}_i$, and let $\overline{\mathbb{P}} = n^{-1}\sum_i \mathbb{P}_i$.*

Let $\mathcal{F}$ be a class of Borel measurable functions from S to $\mathbb{R}$ which is pointwise measurable (i.e. it contains a countable subclass which is dense under pointwise convergence), and satisfying the uniform entropy bound (SA1.3) with parameters A and V. Let F be a strictly positive measurable envelope function for $\mathcal{F}$ (i.e. $|f(s)| \leq |F(s)|$ for all $f \in \mathcal{F}$ and $s \in S$). Suppose that $\|F\|_{\overline{\mathbb{P}},2} < \infty$. Let $\sigma > 0$ satisfy $\sup_{f\in\mathcal{F}} \|f\|_{\overline{\mathbb{P}},2} \leq \sigma \leq \|F\|_{\overline{\mathbb{P}},2}$ and $M = \max_{1\leq i\leq n} F(X_i)$.

For $f \in \mathcal{F}$ define the empirical process

$$G_n(f) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]).$$

Then we have

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}} |G_n(f)|\Big] \lesssim \sigma\sqrt{V\log\big(A\|F\|_{\overline{\mathbb{P}},2}/\sigma\big)} + \frac{\|M\|_{\mathbb{P},2}V\log\big(A\|F\|_{\overline{\mathbb{P}},2}/\sigma\big)}{\sqrt{n}},$$

where $\lesssim$ is up to a universal constant.

Proof. See Lemmas SA24 and SA25 in [7]. ☐

Lemma SA3.7 (Maximal inequalities for Gaussian processes)**.** *Let Z be a separable mean-zero Gaussian process indexed by $x \in \mathcal{X}$. Recall that Z is separable for example if $\mathcal{X}$ is Polish and Z has continuous trajectories. Define its covariance structure on $\mathcal{X} \times \mathcal{X}$ by $\Sigma(x, x') := \mathbb{E}[Z(x)Z(x')]$, and the corresponding semimetric on $\mathcal{X}$ by*

$$\rho(x, x') := \mathbb{E}[(Z(x) - Z(x'))^2]^{1/2} = \big(\Sigma(x, x) - 2\Sigma(x, x') + \Sigma(x', x')\big)^{1/2}.$$

Let $N(\mathcal{X}, \rho, \varepsilon)$ denote the ε-covering number of $\mathcal{X}$ with respect to the semimetric ρ. Define $\sigma := \sup_x \Sigma(x, x)^{1/2}$.

Then there exists a universal constant $C > 0$ such that for any $\delta > 0$,

$$\mathbb{E}\Big[\sup_{x\in\mathcal{X}} |Z(x)|\Big] \le C\sigma + C \int_0^{2\sigma} \sqrt{\log N(\mathcal{X}, \rho, \varepsilon)}\, \mathrm{d}\varepsilon,$$
$$\mathbb{E}\Big[\sup_{\rho(x,x')\le\delta} |Z(x) - Z(x')|\Big] \le C \int_0^{\delta} \sqrt{\log N(\mathcal{X}, \rho, \varepsilon)}\, \mathrm{d}\varepsilon.$$

Proof. This lemma is well-known. See, for example, [18]. ☐

Lemma SA3.8 (Closeness in probability implies closeness of conditional quantiles)**.** *Let X_n and Y_n be random variables and $\boldsymbol{D}_n$ be a random vector. Let $F_{X_n}(x|\boldsymbol{D}_n)$ and $F_{Y_n}(x|\boldsymbol{D}_n)$ denote the conditional distribution functions, and $F^{-1}_{X_n}(x|\boldsymbol{D}_n)$ and $F^{-1}_{Y_n}(x|\boldsymbol{D}_n)$ denote the corresponding conditional quantile functions. If $|X_n - Y_n| = o_{\mathbb{P}}(r_n)$, then there exists a sequence of positive numbers $\nu_n \to 0$, depending on r_n, such that w. p. a. 1*

$$F^{-1}_{X_n}(p|\boldsymbol{D}_n) \le F^{-1}_{Y_n}(p + \nu_n|\boldsymbol{D}_n) + r_n \quad \text{and} \quad F^{-1}_{Y_n}(p|\boldsymbol{D}_n) \le F^{-1}_{X_n}(p + \nu_n|\boldsymbol{D}_n) + r_n$$

for all $p \in (\nu_n, 1 - \nu_n)$.

Proof. See Lemma 13 in [2]. ☐

Lemma SA3.9 (Anti-concentration for suprema of separable Gaussian processes)**.** *Let $X = (X_t)_{t\in T}$ be a mean-zero separable Gaussian process indexed by a semimetric space T such that $\mathbb{E}[X_t^2] = 1$ for all $t \in T$. Then for any $\varepsilon > 0$,*

$$\sup_{u\in\mathbb{R}} \mathbb{P}\bigg\{\bigg|\sup_{t\in T} |X_t| - u\bigg| \le \varepsilon\bigg\} \le 4\varepsilon\bigg(\mathbb{E}\bigg[\sup_{t\in T} |X_t|\bigg] + 1\bigg).$$

Proof. See Corollary 2.1 in [11]. ☐

The following lemma appears to be new to the literature at the level of generality considered. It guarantees the existence and gives some properties of the main estimand considered in this paper.

Lemma SA3.10 (The existence of $\mu(\cdot)$)**.** *Suppose Assumptions SA2.3(i) and SA2.3(iii) hold. We will suppress the dependence of $\rho(\cdot, \cdot)$ on $\boldsymbol{q}$ in this lemma because the result can be applied separately for each $\boldsymbol{q}$. Assume $\eta \mapsto \rho(y, \eta)$ is convex on $\mathcal{E}$, $\mathcal{E}$ is an open connected subset of $\mathbb{R}$, $\psi(y, \cdot)$ is the left or right derivative of $\rho(y, \cdot)$ (in particular, it is a subgradient: $(\eta_1 - \eta_0)\psi(y, \eta_0) \le \rho(y, \eta_1) - \rho(y, \eta_0)$), and $\psi(y, \eta)$ is strictly increasing in η for any fixed $y \in \mathcal{Y}$. Assume the real inverse link function $\eta(\cdot)\colon \mathbb{R} \to \mathcal{E}$ is strictly monotonic and two times continuously differentiable.*

Denoting a_l and a_r the left and right ends of $\mathcal{E}$ respectively (possibly $\pm\infty$), assume that for each $\boldsymbol{x} \in \mathcal{X}$ the expectation $\mathbb{E}[\psi(y_i, \zeta) \,|\, \boldsymbol{x}_i = \boldsymbol{x}]$ is negative for (real deterministic) ζ in a neighborhood of a_l, positive for ζ in a neighborhood of a_r, and continuous in ζ (in particular, it crosses zero).

Then for each $\boldsymbol{x} \in \mathcal{X}$ the number $\min\{\zeta : \mathbb{E}[\psi(y_i, \zeta) \,|\, \boldsymbol{x}_i = \boldsymbol{x}] \geq 0\}$ exists and belongs to $\mathcal{E}$. Moreover,

$$\begin{aligned}\mu_0(\boldsymbol{x}) &:= \eta^{-1}\big(\min\big\{\zeta : \mathbb{E}[\psi(y_i, \zeta) \,|\, \boldsymbol{x}_i = \boldsymbol{x}] \geq 0\big\}\big) \\ &= \eta^{-1}\big(\min\big\{\zeta : \mathbb{E}[\psi(y_i, \zeta) \,|\, \boldsymbol{x}_i = \boldsymbol{x}] = 0\big\}\big)\end{aligned} \tag{SA3.1}$$

defines a Borel-measurable function such that for all $\boldsymbol{x} \in \mathcal{X}$

$$\mu_0(\boldsymbol{x}) \in \operatorname*{arg\,min}_{\zeta \in \mathbb{R}} \mathbb{E}[\rho(y_i, \eta(\zeta)) \,|\, \boldsymbol{x}_i = \boldsymbol{x}].$$

If $\mathcal{Q}$ is not a singleton, applying this result for each $\boldsymbol{q} \in \mathcal{Q}$ gives a function $\mu_0(\boldsymbol{x}, \boldsymbol{q})$ which is Borel in $\boldsymbol{x}$ for each fixed $\boldsymbol{q}$. Measurability in $\boldsymbol{q}$ is not asserted by this lemma.

Proof. The conditions ensure that $\min\big\{\zeta : \mathbb{E}[\psi(y_i, \zeta) \,|\, \boldsymbol{x}_i = \boldsymbol{x}] \geq 0\big\}$ exists and belongs to $\mathcal{E}$ by continuity.

So the function $\mu_0(\boldsymbol{x})$ is well-defined. It is Borel because η^{-1} is continuous and

$$\{\boldsymbol{x} : \eta(\mu_0(\boldsymbol{x})) > a\} = \{\boldsymbol{x} : \mathbb{E}[\psi(y_i, a) \,|\, \boldsymbol{x}_i = \boldsymbol{x}] < 0\}$$

is a Borel set (the equality of the two sets is true because $\zeta \mapsto \psi(y, \zeta)$ is strictly increasing).

For any $\zeta \in \mathbb{R}$, using $(\eta(\zeta) - \eta(\mu_0(\boldsymbol{x})))\psi(y, \eta(\mu_0(\boldsymbol{x}))) \leq \rho(y, \eta(\zeta)) - \rho(y, \eta(\mu_0(\boldsymbol{x})))$, we have

$$\begin{aligned}0 &= (\eta(\zeta) - \eta(\mu_0(\boldsymbol{x})))\,\mathbb{E}\big[\psi(y_i, a) \,\big|\, \boldsymbol{x}_i = \boldsymbol{x}\big]\big|_{a = \eta(\mu_0(\boldsymbol{x}))} \\ &\leq \mathbb{E}[\rho(y, \eta(\zeta)) \,|\, \boldsymbol{x}_i = \boldsymbol{x}] - \mathbb{E}[\rho(y, a) \,|\, \boldsymbol{x}_i = \boldsymbol{x}]\big|_{a = \eta(\mu_0(\boldsymbol{x}))},\end{aligned}$$

so $\mu_0(\boldsymbol{x})$ is indeed the argmin. □

The following lemma establishes basic properties of the "Gram" (or Hessian, depending on the perspective) matrix generated by the partitioning-based M-estimator.

Lemma SA3.11 (Gram matrix)**.** *Suppose Assumptions SA2.1 to SA2.6 hold. Then*

$$h^d \lesssim \inf_{\boldsymbol{q}} \lambda_{\min}(\mathbf{Q}_{0,\boldsymbol{q}}) \leq \sup_{\boldsymbol{q}} \lambda_{\max}(\mathbf{Q}_{0,\boldsymbol{q}}) \lesssim h^d, \tag{SA3.2}$$

$$\sup_{\boldsymbol{q}} \big\|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\big\|_\infty \lesssim h^{-d}. \tag{SA3.3}$$

If, in addition, $\frac{\log(1/h)}{nh^d} = o(1)$, then uniformly over $\boldsymbol{q} \in \mathcal{Q}$

$$\sup_{\boldsymbol{q}} \big\{\big\|\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}\big\|_\infty \vee \big\|\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}\big\|\big\} \lesssim_{\mathbb{P}} h^d \bigg(\frac{\log(1/h)}{nh^d}\bigg)^{1/2}, \tag{SA3.4}$$

$$h^d \lesssim \inf_{\boldsymbol{q}} \lambda_{\min}(\bar{\mathbf{Q}}_{\boldsymbol{q}}) \leq \sup_{\boldsymbol{q}} \lambda_{\max}(\bar{\mathbf{Q}}_{\boldsymbol{q}}) \lesssim h^d \quad \textit{w.p.a. 1}, \tag{SA3.5}$$

$$\sup_{\boldsymbol{q}} \big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\big\|_\infty \lesssim h^{-d} \quad \textit{w.p.a. 1}, \tag{SA3.6}$$

$$\sup_{\boldsymbol{q}} \big\{\big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \mathbf{Q}_{0,\boldsymbol{q}}^{-1}\big\|_\infty \vee \big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \mathbf{Q}_{0,\boldsymbol{q}}^{-1}\big\|\big\} \lesssim_{\mathbb{P}} h^{-d} \bigg(\frac{\log(1/h)}{nh^d}\bigg)^{1/2}. \tag{SA3.7}$$

For some positive integer $L_n \lesssim 1/h$, the rows and columns of $\bar{\mathbf{Q}}_{\boldsymbol{q}}$ and its inverse can be numbered by multi-indices $(\boldsymbol{\alpha}, a) = (\alpha_1, \dots, \alpha_d, a)$ and $(\boldsymbol{\beta}, b) = (\beta_1, \dots, \beta_d, b)$, where

$$\boldsymbol{\alpha}, \boldsymbol{\beta} \in \{1, \dots, L_n\}^d, \quad a \in \{1, \dots, T_{n,\boldsymbol{\alpha}}\}, b \in \{1, \dots, T_{n,\boldsymbol{\beta}}\}, \quad T_{n,\boldsymbol{\alpha}}, T_{n,\boldsymbol{\beta}} \lesssim 1,$$

in the following way. First, $\bar{\mathbf{Q}}_{\boldsymbol{q}}$ has a multi-banded structure:

$$[\bar{\mathbf{Q}}_{\boldsymbol{q}}]_{(\boldsymbol{\alpha},a),(\boldsymbol{\beta},b)} = 0 \quad \text{if } \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_\infty > C, \tag{SA3.8}$$

for some constant $C > 0$ (not depending on n). Second, with probability approaching one

$$\sup_{\boldsymbol{q}} \big|[\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}]_{(\boldsymbol{\alpha},a),(\boldsymbol{\beta},b)}\big| \lesssim h^{-d} \varrho^{\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_\infty} \tag{SA3.9}$$

for some constant $\varrho \in (0, 1)$ (not depending on n).

The same results hold with $\mathbf{Q}_{0,\boldsymbol{q}}$ replaced by $\boldsymbol{\Sigma}_{0,\boldsymbol{q}}$ and $\bar{\mathbf{Q}}_{\boldsymbol{q}}$ replaced by $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}$.

Finally, the same results hold with $\mathbf{Q}_{0,\boldsymbol{q}}$ replaced by $\mathbb{E}[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}]$ and $\bar{\mathbf{Q}}_{\boldsymbol{q}}$ replaced by $\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}]$.

Proof. The last claim of the lemma, corresponding to the case

$$\Psi_1(\boldsymbol{x}_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q}) \eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))^2 \equiv 1,$$

is Lemma SA-2.1 in [6]. The properties (SA3.8) and (SA3.9) are not explicitly stated but follow from the proof.

In the general case, by Assumption SA2.4(iv) $\Psi_1(\boldsymbol{x}_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q}) \eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))^2$ is bounded and bounded away from zero uniformly over i, n and $\boldsymbol{q}$, so (SA3.2) and (SA3.5) follow from the previous case. The additional $\Psi_1(\boldsymbol{x}_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q}) \eta^{(1)}(\mu_0(\boldsymbol{x}_i))^2$ term does not influence the multi-banded structure of the matrices, so (SA3.3), (SA3.6), (SA3.8), (SA3.9) remain true by the same argument as in the previous case. The inequalities

$$\begin{aligned}
\big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \mathbf{Q}_{0,\boldsymbol{q}}^{-1}\big\|_\infty &\le \big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\big\|_\infty \cdot \big\|\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}\big\|_\infty \cdot \big\|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\big\|_\infty, \\
\big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \mathbf{Q}_{0,\boldsymbol{q}}^{-1}\big\| &\le \big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\big\| \cdot \big\|\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}\big\| \cdot \big\|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\big\|
\end{aligned}$$

show that (SA3.7) follows from norm bounds (SA3.2), (SA3.3), (SA3.5), (SA3.6) and concentration (SA3.4).

Now we prove Eq. (SA3.4).

Define the class of functions

$$\mathcal{G} := \big\{\boldsymbol{x} \mapsto p_k(\boldsymbol{x}) p_l(\boldsymbol{x}) \Psi_1(\boldsymbol{x}, \eta(\mu_0(\boldsymbol{x}, \boldsymbol{q})); \boldsymbol{q}) \eta^{(1)}(\mu_0(\boldsymbol{x}, \boldsymbol{q}))^2 \colon 1 \le k, l \le K, \boldsymbol{q} \in \mathcal{Q}\big\}.$$

We will now prove that the class $\mathcal{G}$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim h^{-2d}$ and $V \lesssim 1$. By Assumption SA2.5, the class

$$\{\boldsymbol{x} \mapsto \Psi_1(\boldsymbol{x}, \eta(\mu_0(\boldsymbol{x}, \boldsymbol{q})); \boldsymbol{q}) \colon \boldsymbol{q} \in \mathcal{Q}\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim 1$. Since it is also true of the class $\big\{\eta^{(1)}\big(\mu_0(\boldsymbol{x}, \boldsymbol{q})\big)^2 \colon \boldsymbol{q} \in \mathcal{Q}\big\}$ because $\eta^{(1)}\big(\mu_0(\boldsymbol{x}, \boldsymbol{q})\big)^2$ is Lipschitz is $\boldsymbol{q}$, by Lemma SA3.4 the class

$$\big\{\boldsymbol{x} \mapsto \Psi_1(\boldsymbol{x}, \eta(\mu_0(\boldsymbol{x}, \boldsymbol{q})); \boldsymbol{q}) \eta^{(1)}(\mu_0(\boldsymbol{x}, \boldsymbol{q}))^2 \colon \boldsymbol{q} \in \mathcal{Q}\big\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim 1$. The class $\{\boldsymbol{x} \mapsto p_k(\boldsymbol{x}) p_l(\boldsymbol{x}) \colon 1 \le k, l \le K\}$ with a large enough constant envelope just contains K^2 functions, so it also satisfies the uniform entropy bound (SA1.3) with $A \lesssim h^{-2d}$ and $V = 1$, where we used $K \asymp h^{-d}$. By Lemma SA3.4, combining these facts proves the claim about the complexity of $\mathcal{G}$.

Moreover, class $\mathcal{G}$ satisfies the following variance bound:

$$\sup_{g\in\mathcal{G}} \mathbb{E}\big[g(\boldsymbol{x}_i)^2\big] \lesssim h^d,$$

which follows from the fact that the class is bounded by a large enough constant and the Lebesgue measure of the support of $p_k(\boldsymbol{x})p_l(\boldsymbol{x})$ shrinks (uniformly over k, l) at the rate h^d.

Applying Lemma SA3.6, we see that

$$\begin{aligned}
\sup_{g\in\mathcal{G}}\bigg|\frac{1}{n}\sum_{i=1}^n g(\boldsymbol{x}_i) - \mathbb{E}[g(\boldsymbol{x}_i)]\bigg| &\lesssim_{\mathbb{P}} h^d\bigg(\frac{\log(1/h)}{nh^d}\bigg)^{1/2} + \frac{\log(1/h)}{n}\\
&\lesssim h^d\bigg(\frac{\log(1/h)}{nh^d}\bigg)^{1/2} && \text{since } \tfrac{\log(1/h)}{nh^d} = o(1).
\end{aligned}$$

So we have shown $\max_{k,l}\big|(\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}})_{k,l}\big| \lesssim_{\mathbb{P}} h^d\Big(\frac{\log(1/h)}{nh^d}\Big)^{1/2}$. Since each row and column of $\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}$ has a bounded number of nonzero entries, this implies

$$\|\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}\|_\infty \asymp \|\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}\|_1 \lesssim_{\mathbb{P}} h^d\bigg(\frac{\log(1/h)}{nh^d}\bigg)^{1/2}.$$

To conclude the proof of Eq. (SA3.4), it is left to use the inequality

$$\|\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}\| \le \sqrt{\|\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}\|_1\|\bar{\mathbf{Q}}_{\boldsymbol{q}} - \mathbf{Q}_{0,\boldsymbol{q}}\|_\infty}.$$

The claim about $\boldsymbol{\Sigma}_{0,\boldsymbol{q}}$ and $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}$ is proven analogously, using that $\sigma^2_{\boldsymbol{q}}(\boldsymbol{x})$ is bounded and bounded away from zero and Lipschitz in $\boldsymbol{q}$ by Assumption SA2.4(ii).

Lemma SA3.11 is proven. □

The following Lemmas SA3.12 and SA3.14 are needed for the proof of the Bahadur representation (Theorem SA5.1), Corollaries SA5.16 and SA5.17 and a version of the consistency result (Lemma SA4.6).

Lemma SA3.12 (Uniform convergence: variance)**.** *Suppose Assumptions SA2.1 to SA2.6 hold. If*

$$\frac{\big[\log(1/h)\big]^{\nu/(\nu-2)}}{nh^{\nu d/(\nu-2)}} = o(1), \text{ or} \tag{SA3.10}$$

$$\overline{\psi}(\boldsymbol{x}_i, y_i) \text{ is } \sigma^2\text{-sub-Gaussian conditionally on } \boldsymbol{x}_i \text{ and } \frac{\log n\log(1/h)}{nh^d} = o(1), \tag{SA3.11}$$

then

$$\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{x}\in\mathcal{X}}\big|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\mathbb{E}_n\big[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\big]\big| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|}\bigg(\frac{\log(1/h)}{nh^d}\bigg)^{1/2};$$

and the same inequality is true with $\bar{\mathbf{Q}}_{\boldsymbol{q}}$ replaced by $\mathbf{Q}_{0,\boldsymbol{q}}$.

Proof. By Assumption SA2.2, $\sup_{\boldsymbol{x}\in\mathcal{X}}\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\| \lesssim h^{-|\boldsymbol{v}|}$; by Lemma SA3.11, $\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_\infty + \|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\|_\infty \lesssim_{\mathbb{P}} h^{-d}$. Therefore, it is enough to show

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\mathbb{E}_n\big[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\big]\big\|_\infty \lesssim_{\mathbb{P}} \sqrt{\frac{h^d\log(1/h)}{n}}. \tag{SA3.12}$$

Define the function class

$$\mathcal{G} := \big\{(\boldsymbol{x},y)\mapsto p_l(\boldsymbol{x})\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\psi(y,\eta(\mu_0(\boldsymbol{x},\boldsymbol{q}));\boldsymbol{q}) : 1\le l\le K, \boldsymbol{q}\in\mathcal{Q}\big\}.$$

We will now control the complexity of $\mathcal{G}$. Introduce some more classes of functions:

$$\begin{aligned}
\mathcal{W}_1 &:= \{(\boldsymbol{x},y)\mapsto p_l(\boldsymbol{x}) : 1\le l\le K\},\\
\mathcal{W}_2 &:= \{(\boldsymbol{x},y)\mapsto \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q})) : \boldsymbol{q}\in\mathcal{Q}\},\\
\mathcal{W}_3 &:= \{(\boldsymbol{x},y)\mapsto \psi(y,\eta(\mu_0(\boldsymbol{x},\boldsymbol{q}));\boldsymbol{q}) : \boldsymbol{q}\in\mathcal{Q}\}.
\end{aligned}$$

$\mathcal{W}_1$ with a large enough constant envelope contains K fixed measurable functions, so it satisfies the uniform entropy bound (SA1.3) with $A\lesssim h^{-d}$ and $V=1$. $\mathcal{W}_2$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A,V\lesssim 1$ because $\mu_0(\boldsymbol{x},\boldsymbol{q})$ is bounded uniformly over $\boldsymbol{x},\boldsymbol{q}$ and Lipschitz in $\boldsymbol{q}$, $\eta^{(1)}$ on a fixed bounded interval is Lipschitz. $\mathcal{W}_3$ with envelope $\overline{\psi}(\boldsymbol{x},y)$ satisfies the uniform entropy bound (SA1.3) with $A,V\lesssim 1$ by Assumption SA2.5. By Lemma SA3.4, $\mathcal{G}$ with envelope $\overline{\psi}(\boldsymbol{x},y)$ multiplied by a large enough constant satisfies the uniform entropy bound (SA1.3) with $A\lesssim h^{-d}$ and $V\lesssim 1$.

Moreover, class $\mathcal{G}$ satisfies the following variance bound:

$$\sup_{g\in\mathcal{G}}\mathbb{E}\big[g(\boldsymbol{x}_i,y_i)^2\big]\lesssim h^d.$$

Indeed, for a fixed $i\in\{1,\dots,n\}$

$$\sup_{g\in\mathcal{G}}\mathbb{E}\big[g(\boldsymbol{x}_i,y_i)^2\big]\lesssim \sup_l \mathbb{E}\Big[p_l(\boldsymbol{x}_i)^2\mathbb{E}\Big[\overline{\psi}(\boldsymbol{x}_i,y_i)^2\,\Big|\,\boldsymbol{x}_i\Big]\Big]\lesssim \sup_l\mathbb{E}\big[p_l(\boldsymbol{x}_i)^2\big]\lesssim h^d.$$

Finally, under (SA3.10)

$$\begin{aligned}
\mathbb{E}\Big[\max_{1\le i\le n}|\overline{\psi}(\boldsymbol{x}_i,y_i)|^2\Big]^{1/2} &\le \mathbb{E}\Big[\max_{1\le i\le n}|\overline{\psi}(\boldsymbol{x}_i,y_i)|^\nu\Big]^{1/\nu}\le \mathbb{E}\Big[\sum_{i=1}^n|\overline{\psi}(\boldsymbol{x}_i,y_i)|^\nu\Big]^{1/\nu}\\
&= \Big(\sum_{i=1}^n\mathbb{E}\big[\overline{\psi}(\boldsymbol{x}_i,y_i)^\nu\big]\Big)^{1/\nu}\lesssim\Big(\sum_{i=1}^n 1\Big)^{1/\nu}=n^{1/\nu},
\end{aligned}$$

and under (SA3.11)

$$\mathbb{E}\Big[\max_{1\le i\le n}|\overline{\psi}(\boldsymbol{x}_i,y_i)|^2\Big]^{1/2}\lesssim\sqrt{\log n}$$

by Lemma SA3.1.

Applying Lemma SA3.6, we obtain (SA3.12) since

$$\begin{aligned}
&\frac{n^{1/\nu}\log(1/h)}{n}=\sqrt{\frac{h^d\log(1/h)}{n}}\cdot\sqrt{\frac{\log(1/h)}{n^{1-2/\nu}h^d}}=\sqrt{\frac{h^d\log(1/h)}{n}}\cdot o(1),\ \text{and}\\
&\frac{\sqrt{\log n}\log(1/h)}{n}=\sqrt{\frac{h^d\log(1/h)}{n}}\cdot\sqrt{\frac{\log n\log(1/h)}{nh^d}}=\sqrt{\frac{h^d\log(1/h)}{n}}\cdot o(1).
\end{aligned}$$

Lemma SA3.12 is proven. □

Remark SA3.13. Using

$$\frac{n^{1/\nu}\log(1/h)}{n} = o(h^d) \Leftrightarrow \frac{[\log(1/h)]^{\nu/(\nu-1)}}{nh^{\nu d/(\nu-1)}} = o(1) \tag{SA3.13}$$

instead of

$$\frac{n^{1/\nu}\log(1/h)}{n} = \sqrt{\frac{h^d\log(1/h)}{n}}\cdot o(1)$$

in the argument leads to the bound

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\mathbb{E}_n\big[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\big]\big\|_\infty = o_{\mathbb{P}}(h^d)$$

under just the condition (SA3.13) which is slightly weaker than (SA3.10).

The following lemma gives control on the projection approximation error.

Lemma SA3.14 (Projection of approximation error)**.** *Suppose Assumptions SA2.1 to SA2.6 hold. If*

$$\frac{\big[\log(1/h)\big]^{\nu/(\nu-2)}}{nh^{\nu d/(\nu-2)}} = o(1),\ \ \textit{or} \tag{SA3.14}$$

$$\overline{\psi}(\boldsymbol{x}_i,y_i)\ \textit{is}\ \sigma^2\textit{-sub-Gaussian conditionally on}\ \boldsymbol{x}_i\ \textit{and}\ \frac{\log n\log(1/h)}{nh^d} = o(1),$$

then

$$\begin{aligned}\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{x}\in\mathcal{X}}&\big|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^\mathsf{T}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\mathbb{E}_n\big[\boldsymbol{p}(\boldsymbol{x}_i)\big\{\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\\&\qquad-\psi(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\big\}\big]\big|\\&\lesssim_{\mathbb{P}} h^{m-|\boldsymbol{v}|}+h^{(\alpha\wedge(1/2))m-|\boldsymbol{v}|}\left(\frac{\log(1/h)}{nh^d}\right)^{1/2}+\frac{\log(1/h)}{nh^{|\boldsymbol{v}|+d}}.\end{aligned}$$

Proof. Denote

$$\begin{aligned}A_{1,\boldsymbol{q}}(\boldsymbol{x}_i,y_i) &:= \psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\big\{\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))-\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\big\},\\A_{2,\boldsymbol{q}}(\boldsymbol{x}_i,y_i) &:= \big\{\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})-\psi(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\big\}\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})).\end{aligned}$$

By Assumption SA2.2, $\sup_{\boldsymbol{x}\in\mathcal{X}}\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|\lesssim h^{-|\boldsymbol{v}|}$; by Lemma SA3.11, $\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_\infty\lesssim_{\mathbb{P}} h^{-d}$. Therefore, it is enough to show

$$\begin{aligned}\sup_{\boldsymbol{q}\in\mathcal{Q}}\Big\|\mathbb{E}_n\Big[&\boldsymbol{p}(\boldsymbol{x}_i)\Big\{\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\\&-\psi(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\Big\}\Big]\Big\|_\infty\\&\lesssim_{\mathbb{P}} h^{d+m}+h^{\frac{d}{2}+(\alpha\wedge(1/2))m}\left(\frac{\log(1/h)}{n}\right)^{1/2}+\frac{\log(1/h)}{n}.\end{aligned}$$

We will do this by showing the three bounds

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)A_{1,\boldsymbol{q}}(\boldsymbol{x}_i,y_i)]\big\|_\infty\lesssim_{\mathbb{P}} h^{\frac{d}{2}+m}\left(\frac{\log(1/h)}{n}\right)^{1/2}, \tag{SA3.15}$$

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\mathbb{E}_n\big[\mathbb{E}\big[\boldsymbol{p}(\boldsymbol{x}_i)A_{2,\boldsymbol{q}}(\boldsymbol{x}_i,y_i)\,\big|\,\boldsymbol{x}_i\big]\big]\big\|_\infty \lesssim_{\mathbb{P}} h^{d+m}, \tag{SA3.16}$$

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\mathbb{E}_n\big[\boldsymbol{p}(\boldsymbol{x}_i)(A_{2,\boldsymbol{q}}(\boldsymbol{x}_i,y_i)-\mathbb{E}[A_{2,\boldsymbol{q}}(\boldsymbol{x}_i,y_i)\,|\,\boldsymbol{x}_i])\big]\big\|_\infty \lesssim_{\mathbb{P}} h^{\frac{d}{2}+(\alpha\wedge(1/2))m}\sqrt{\frac{\log(1/h)}{n}}+\frac{\log(1/h)}{n}. \tag{SA3.17}$$

To show (SA3.15), consider the class of functions

$$\{(\boldsymbol{x},y)\mapsto p_l(\boldsymbol{x})A_{1,\boldsymbol{q}}(\boldsymbol{x},y):1\le l\le K,\boldsymbol{q}\in\mathcal{Q}\}.$$

Note that $\sup_{\boldsymbol{q},\boldsymbol{x}}|\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))-\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))|\lesssim h^m$ by Assumption SA2.6. (SA3.15) follows by the same concentration argument as in Lemma SA3.12.

To show (SA3.16), note that

$$\begin{aligned}
&\mathbb{E}\Big[\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})-\psi(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\,\Big|\,\boldsymbol{x}_i\Big]\\
&\quad=-\Psi(\boldsymbol{x}_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})=\\
&\quad=\Psi_1(\boldsymbol{x}_i,\zeta;\boldsymbol{q})\eta^{(1)}(\tilde{\zeta})\{\mu_0(\boldsymbol{x}_i,\boldsymbol{q})-\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})\},
\end{aligned}$$

where ζ is between $\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))$ and $\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))$, $\tilde{\zeta}$ is between $\mu_0(\boldsymbol{x}_i,\boldsymbol{q})$ and $\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})$. By Assumption SA2.4(iv) and SA2.6, it follows that a. s.

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}|\mathbb{E}[\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})-\psi(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\,|\,\boldsymbol{x}_i]|\lesssim h^m.$$

Since $\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))$ is bounded, (SA3.16) follows by applying Lemma SA3.6 to the class $\{\boldsymbol{x}\mapsto |p_l(\boldsymbol{x})|, 1\le l\le K\}$.

It is left to show (SA3.17).

Consider the class of functions

$$\mathcal{G}:=\{(\boldsymbol{x},y)\mapsto p_l(\boldsymbol{x})A_{2,\boldsymbol{q}}(\boldsymbol{x},y):1\le l\le K,\boldsymbol{q}\in\mathcal{Q}\}.$$

We will now control the complexity of $\mathcal{G}$. Introduce some more classes of functions:

$$\begin{aligned}
\mathcal{W}_{1,l}&:=\{(\boldsymbol{x},y)\mapsto p_l(\boldsymbol{x})[\psi(y,\eta(\mu_0(\boldsymbol{x},\boldsymbol{q}));\boldsymbol{q})-\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})]:\boldsymbol{q}\in\mathcal{Q}\},\\
\mathcal{W}_1&:=\bigcup_{l=1}^K\mathcal{W}_{1,l},\\
\mathcal{W}_{2,l}&:=\{(\boldsymbol{x},y)\mapsto\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\mathbb{1}\{\boldsymbol{x}\in\operatorname{supp}p_l\}:\boldsymbol{q}\in\mathcal{Q}\},\\
\mathcal{W}_2&:=\bigcup_{l=1}^K\mathcal{W}_{2,l}=\{(\boldsymbol{x},y)\mapsto\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\mathbb{1}\{\boldsymbol{x}\in\operatorname{supp}p_l\}:1\le l\le K,\boldsymbol{q}\in\mathcal{Q}\}.
\end{aligned}$$

By Assumption SA2.5, for l fixed $\mathcal{W}_{1,l}$ with a large enough constant envelope (not depending on l) satisfies the uniform entropy bound (SA1.3) with $A,V\lesssim 1$ (not depending on l). This immediately implies that $\mathcal{W}_1$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A\lesssim h^{-d}$ and $V\lesssim 1$.

For l fixed, $\mathcal{W}_{2,l}$ is a product of a (bounded) subclass of $\{\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}):\boldsymbol{\beta}\in\mathcal{B}_l\}$, where $\mathcal{B}_l$ is a vector space of dimension $O(1)$ (not depending on l), and a fixed function. By Lemma 2.6.15 in [18], $\{\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}:\boldsymbol{\beta}\in\mathcal{B}_l\}$ is VC with a bounded index. Therefore, since $\eta^{(1)}$ on a bounded interval is Lipschitz, $\mathcal{W}_{2,l}$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A, V \lesssim 1$. This immediately implies that $\mathcal{W}_2$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim h^{-d}$ and $V \lesssim 1$.

By Lemma SA3.4, it follows from the above that $\mathcal{G}$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim h^{-d}$ and $V \lesssim 1$.

Next, we will show that class $\mathcal{G}$ satisfies the following variance bound:

$$\sup_{g \in \mathcal{G}} \mathbb{E}_n[\mathbb{V}[g(\boldsymbol{x}_i, y_i) \,|\, \boldsymbol{x}_i]] \lesssim h^{d+(2\alpha \wedge 1)m} \quad \text{w. p. a. 1.} \tag{SA3.18}$$

We will prove (SA3.18) under the assumption that $\mathfrak{M}$ is Lebesgue measure, so (SA2.2) holds; the argument under (SA2.3) is similar (and leads to an even stronger variance bound), so it is omitted. For y outside the closed segment between $\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))$ and $\eta(\mu_0(\boldsymbol{x}, \boldsymbol{q}))$,

$$\begin{aligned}
&|\psi(y, \eta(\mu_0(\boldsymbol{x}, \boldsymbol{q})); \boldsymbol{q}) - \psi(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})| \\
&\quad \lesssim (\overline{\psi}(\boldsymbol{x}, y) + 1) \cdot |\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}) - \mu_0(\boldsymbol{x}, \boldsymbol{q})|^\alpha \\
&\quad \lesssim (\overline{\psi}(\boldsymbol{x}, y) + 1) h^{\alpha m},
\end{aligned} \tag{SA3.19}$$

where in (SA3.19) we used Assumption SA2.6. For y between $\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))$ and $\eta(\mu_0(\boldsymbol{x}, \boldsymbol{q}))$ inclusive,

$$\begin{aligned}
&|\psi(y, \eta(\mu_0(\boldsymbol{x}, \boldsymbol{q})); \boldsymbol{q}) - \psi(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})| \lesssim \overline{\psi}(\boldsymbol{x}, y) \cdot |\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}) - \mu_0(\boldsymbol{x}, \boldsymbol{q})| + 1 \\
&\quad \lesssim \overline{\psi}(\boldsymbol{x}, y) h^m + 1,
\end{aligned} \tag{SA3.20}$$

where in (SA3.20) we again used Assumption SA2.6.

In the chain below, to avoid cluttering notation we will use $[a, b]$ to denote the closed segment between a and b regardless of their ordering (a more standard notation is $[a \wedge b, a \vee b]$). Using that $\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))$ is also bounded uniformly over $\boldsymbol{x} \in \mathcal{X}$, we have a. s.

$$\begin{aligned}
&\mathbb{E}\big[A_{2,\boldsymbol{q}}(\boldsymbol{x}_i, y_i)^2 \,\big|\, \boldsymbol{x}_i\big] \\
&\quad = \mathbb{E}\big[A_{2,\boldsymbol{q}}(\boldsymbol{x}_i, y_i)^2 \mathbb{1}\big\{y_i \notin \big[\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})), \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\big]\big\} \,\big|\, \boldsymbol{x}_i\big] \\
&\qquad + \mathbb{E}\big[A_{2,\boldsymbol{q}}(\boldsymbol{x}_i, y_i)^2 \mathbb{1}\big\{y_i \in \big[\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})), \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\big]\big\} \,\big|\, \boldsymbol{x}_i\big] \\
&\quad \lesssim h^{2\alpha m}\mathbb{E}\big[(\overline{\psi}(\boldsymbol{x}_i, y_i)^2 + 1)\mathbb{1}\big\{y_i \notin \big[\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})), \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\big]\big\} \,\big|\, \boldsymbol{x}_i\big] \\
&\qquad + h^{2m}\mathbb{E}\big[\overline{\psi}(\boldsymbol{x}_i, y_i)^2 \mathbb{1}\big\{y_i \in \big[\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})), \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\big]\big\} \,\big|\, \boldsymbol{x}_i\big] \\
&\qquad + \mathbb{P}\big\{y_i \in \big[\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})), \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\big] \,\big|\, \boldsymbol{x}_i\big\} \\
&\quad \le h^{2\alpha m}\mathbb{E}\Big[(\overline{\psi}(\boldsymbol{x}_i, y_i)^2 + 1) \,\Big|\, \boldsymbol{x}_i\Big] + h^{2m}\mathbb{E}\Big[\overline{\psi}(\boldsymbol{x}_i, y_i)^2 \,\Big|\, \boldsymbol{x}_i\Big] \\
&\qquad + \mathbb{P}\Big\{y_i \in \big[\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})), \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\big] \,\Big|\, \boldsymbol{x}_i\Big\} \\
&\quad \lesssim h^{2\alpha m} + h^{2m} + |\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}) - \mu_0(\boldsymbol{x}_i, \boldsymbol{q})| && \text{(SA3.21)} \\
&\quad \lesssim h^{2\alpha m} + h^{2m} + h^m && \text{(SA3.22)} \\
&\quad \lesssim h^{(2\alpha \wedge 1)m},
\end{aligned}$$

where in (SA3.21) we used that by Assumption SA2.3(iii) the conditional density of $y_i | \boldsymbol{x}_i$ is bounded and Assumption SA2.4(ii), in (SA3.22) we used Assumption SA2.6.

Therefore, uniformly over l and $\boldsymbol{q}$

$$\mathbb{E}_n[\mathbb{V}[g(\boldsymbol{x}_i, y_i) \,|\, \boldsymbol{x}_i]] \le \mathbb{E}_n\big[\mathbb{E}[g(\boldsymbol{x}_i, y_i)^2 \,|\, \boldsymbol{x}_i]\big] = \mathbb{E}_n\big[p_l(\boldsymbol{x}_i)^2 \mathbb{E}[A_{2,\boldsymbol{q}}(\boldsymbol{x}_i, y_i)^2 \,|\, \boldsymbol{x}_i]\big]$$

$$
\begin{aligned}
&\lesssim h^{(2\alpha\wedge 1)m}\mathbb{E}_n[p_l(\boldsymbol{x}_i)^2] \\
&\leq h^{(2\alpha\wedge 1)m}\|\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}]\| \\
&\lesssim h^{d+(2\alpha\wedge 1)m} \quad \text{w. p. a. } 1,
\end{aligned} \tag{SA3.23}
$$

where in (SA3.23) we used Lemma SA3.11. We have proven (SA3.18).

Applying Lemma SA3.6 conditionally on $\{\boldsymbol{x}_i\}_{i=1}^n$, on an event with probability approaching one, we get (SA3.17), and the proof of Lemma SA3.14 is finished. □

Remark SA3.15. Inspecting the argument, one can note that instead of (SA3.14) the condition

$$\frac{\log n}{nh^d} = o(1)$$

is enough for Eqs. (SA3.16) and (SA3.17).

SA4 Consistency

We first study the convex case, and then move on to the non-convex case.

SA4.1 Convex case

The following lemma gives our most general result for a convex objective function. This is Lemma 1 in the main paper.

Lemma SA4.1 (Consistency, convex case)**.** *Suppose Assumptions SA2.1 to SA2.6 hold, $\rho(y, \eta(\theta); \boldsymbol{q})$ is convex with respect to θ with left or right derivative $\psi(y, \eta(\theta); \boldsymbol{q})\eta^{(1)}(\theta)$, $\mathcal{B} = \mathbb{R}^K$ in (SA1.2), and $m > d/2$. Furthermore, assume that one of the following two conditions holds:*

$$\frac{\big[\log(1/h)\big]^{\frac{\nu}{\nu-1}}}{nh^{\frac{2\nu}{\nu-1}d}} = o(1), \text{ or} \tag{SA4.1}$$

$$\overline{\psi}(\boldsymbol{x}_i, y_i) \text{ is } \sigma^2\text{-sub-Gaussian conditionally on } \boldsymbol{x}_i \text{ and } \frac{\sqrt{\log n}\log(1/h)}{nh^{2d}} = o(1). \tag{SA4.2}$$

Then

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\| = o_\mathbb{P}(1), \tag{SA4.3}$$

$$\sup_{\boldsymbol{x}\in\mathcal{X}}\sup_{\boldsymbol{q}\in\mathcal{Q}}\big|\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big| = o_\mathbb{P}(h^{-|\boldsymbol{v}|}), \tag{SA4.4}$$

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\Big(\int_\mathcal{X}\big(\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big)^2 f_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\Big)^{1/2} = o_\mathbb{P}\big(h^{d/2-|\boldsymbol{v}|}\big). \tag{SA4.5}$$

Proof. First, note that (SA4.4) follows from (SA4.3) since uniformly over $\boldsymbol{x}\in\mathcal{X}$ and $\boldsymbol{q}\in\mathcal{Q}$

$$
\begin{aligned}
&\big|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})^\mathsf{T}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big| \\
&\quad\leq \big|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})^\mathsf{T}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \boldsymbol{\beta}_0(\boldsymbol{q})^\mathsf{T}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\big| + \big|\boldsymbol{\beta}_0(\boldsymbol{q})^\mathsf{T}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big| \\
&\quad\lesssim \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\|_\infty \cdot \big\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\big\|_1 + \big|\boldsymbol{\beta}_0(\boldsymbol{q})^\mathsf{T}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big| \\
&\quad\lesssim \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\|_\infty \cdot h^{-|\boldsymbol{v}|} + \big|\boldsymbol{\beta}_0(\boldsymbol{q})^\mathsf{T}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big| \qquad \text{by Assumption SA2.2}
\end{aligned}
$$

$$\lesssim \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\|_\infty \cdot h^{-|\boldsymbol{v}|} + h^{m-|\boldsymbol{v}|} \qquad \text{by Assumption SA2.6},$$

where we used that only a bounded number of elements in $\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})$ are nonzero. Similarly, (SA4.5) follows from (SA4.3) since

$$\begin{aligned}
&\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big\|_{L_2(X)} \\
&\quad\le \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\big\|_{L_2(X)} + \big\|\boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},q)\big\|_{L_2(X)} \\
&\quad\le \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\big\|_{L_2(X)} + \sup_{\boldsymbol{x}\in\mathcal{X}}\big|\boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big| \\
&\quad\lesssim \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}) - \boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\big\|_{L_2(X)} + h^{m-|\boldsymbol{v}|} \qquad \text{by Assumption SA2.6} \\
&\quad= \Big(\big(\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big)^{\mathsf{T}}\mathbb{E}\big[\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}_i)\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}_i)^{\mathsf{T}}\big]\big(\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big)\Big)^{1/2} + h^{m-|\boldsymbol{v}|} \\
&\quad\le \lambda_{\max}\big(\mathbb{E}\big[\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}_i)\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}_i)^{\mathsf{T}}\big]\big)^{1/2}\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\| + h^{m-|\boldsymbol{v}|} \\
&\quad\le h^{d/2-|\boldsymbol{v}|}\cdot\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\| + h^{m-|\boldsymbol{v}|} = o_{\mathbb{P}}\big(h^{d/2-|\boldsymbol{v}|}\big) + h^{m-|\boldsymbol{v}|} = o_{\mathbb{P}}\big(h^{d/2-|\boldsymbol{v}|}\big).
\end{aligned}$$

uniformly over $\boldsymbol{q}\in\mathcal{Q}$, where by $\|g(\boldsymbol{x})\|_{L_2(X)}$ we denote $\big(\int_{\mathcal{X}} g(\boldsymbol{x})^2 f_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\big)^{1/2}$ for simplicity. In the last equality we used $m > d/2$ again. We also used that the largest eigenvalue of $\mathbb{E}\big[\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}_i)\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}_i)^{\mathsf{T}}\big]$ is bounded from above by $h^{d-2|\boldsymbol{v}|}$ up to a multiplicative coefficient, which is proven by the same argument as for $\boldsymbol{v}=0$ in Lemma SA3.11 in combination with Assumption SA2.2.

It is left to prove (SA4.3). Fix a sufficiently small $\gamma > 0$. Denote for $i\in\{1,\dots,n\}$ and $\boldsymbol{\alpha}\in\mathcal{S}^{K-1}$

$$\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha}) := \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\psi(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\gamma\boldsymbol{\alpha}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\gamma\boldsymbol{\alpha})).$$

Since $\mathbb{E}_n\big[\psi(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta});\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta})\boldsymbol{p}(\boldsymbol{x}_i)\big]$ is a subgradient of the convex (by Assumption SA2.4(iii)) objective function $\mathbb{E}_n[\rho(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta});\boldsymbol{q})]$ of $\boldsymbol{\beta}$, the strategy is to show that

$$\inf_{\boldsymbol{q},\boldsymbol{\alpha}}\mathbb{E}_n[\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})] > 0 \quad \text{with probability approaching 1,} \tag{SA4.6}$$

which is enough to prove Lemma SA4.1 by convexity.

To implement this, we will show

$$\inf_{\boldsymbol{q},\boldsymbol{\alpha}}\mathbb{E}_n[\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})\,|\,\boldsymbol{x}_i]] \gtrsim \inf_{\boldsymbol{\alpha}}\mathbb{E}_n[\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\alpha}] + o_{\mathbb{P}}(h^d) \quad \text{and} \tag{SA4.7}$$

$$\sup_{\boldsymbol{q},\boldsymbol{\alpha}}\big|\mathbb{E}_n[\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha}) - \mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})\,|\,\boldsymbol{x}_i]]\big| \lesssim_{\mathbb{P}} \begin{cases}\sqrt{\frac{\log(1/h)}{n}} + \frac{\log(1/h)}{n^{1-1/\nu}h^d} = o(h^d), \\ \sqrt{\frac{\log(1/h)}{n}} + \frac{\sqrt{\log n}\log(1/h)}{nh^d} = o(h^d)\end{cases} \tag{SA4.8}$$

under (SA4.1) and (SA4.2) respectively (proof below), and conclude

$$\begin{aligned}
&\inf_{\boldsymbol{q},\boldsymbol{\alpha}}\mathbb{E}_n[\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})] \ge \\
&\quad\ge \inf_{\boldsymbol{q},\boldsymbol{\alpha}}\mathbb{E}_n\big[\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})\,|\,\mathbf{X}_n]\big] - \sup_{\boldsymbol{q},\boldsymbol{\alpha}}\big|\mathbb{E}_n\big[\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha}) - \mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})\,|\,\mathbf{X}_n]\big]\big| \\
&\quad\gtrsim \inf_{\boldsymbol{\alpha}}\mathbb{E}_n[\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\alpha}] + o_{\mathbb{P}}(h^d),
\end{aligned}$$

which gives (SA4.6) by Lemma SA3.11.

We will now prove (SA4.7). By Assumption SA2.4(iv),

$$\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})\,|\,\boldsymbol{x}_i] = \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\Psi(\boldsymbol{x}_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\gamma\boldsymbol{\alpha}));\boldsymbol{q})$$

$$
\begin{aligned}
&\times \eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \gamma\boldsymbol{\alpha}))\\
&= \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\Psi_1(\boldsymbol{x}_i, \xi_{\boldsymbol{q},i}; \boldsymbol{q})(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \gamma\boldsymbol{\alpha}) - \mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\\
&\quad \times \eta^{(1)}(\zeta_{\boldsymbol{q},i})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \gamma\boldsymbol{\alpha}))\\
&\gtrsim \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\alpha} - C \sup_{\boldsymbol{q}\in\mathcal{Q}, \boldsymbol{x}\in\mathcal{X}} |\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}) - \mu_0(\boldsymbol{x}, \boldsymbol{q})| \cdot |\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)|
\end{aligned}
$$

almost surely uniformly over $\boldsymbol{q}$, where C is some positive constant not depending on n or i, $\xi_{\boldsymbol{q},i}$ is between $\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\gamma\boldsymbol{\alpha}))$ and $\eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))$, $\zeta_{\boldsymbol{q},i}$ between $\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\gamma\boldsymbol{\alpha})$ and $\mu_0(\boldsymbol{x}_i, \boldsymbol{q})$. We used that $\Psi(\boldsymbol{x}_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q}) = 0$, γ is small enough, $\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})$ is (for large enough n) uniformly close to $\mu_0(\boldsymbol{x}, \boldsymbol{q})$ by Assumption SA2.6 and $\eta(\cdot)$ is strictly monotonic by Assumption (SA2.4)(iii) giving the positivity of the product $\eta^{(1)}(\zeta_{\boldsymbol{q},i})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \gamma\boldsymbol{\alpha}))$.

Again using the uniform approximation bound $\sup_{\boldsymbol{q}\in\mathcal{Q}, \boldsymbol{x}\in\mathcal{X}}|\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}) - \mu_0(\boldsymbol{x}, \boldsymbol{q})| \lesssim h^m$ by Assumption SA2.6, we obtain

$$
\sup_{\boldsymbol{q}\in\mathcal{Q}, \boldsymbol{x}\in\mathcal{X}} |\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}) - \mu_0(\boldsymbol{x}, \boldsymbol{q})| \cdot \mathbb{E}_n\big[|\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)|\big] \lesssim_{\mathbb{P}} h^{m+d/2} \tag{SA4.9}
$$

since $\mathbb{E}_n\big[|\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)|\big] \lesssim \mathbb{E}_n\big[(\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i))^2\big]^{1/2} \lesssim_{\mathbb{P}} h^{d/2}$ by Lyapunov's inequality and Lemma SA3.11. Note that since $m > d/2$, $h^{m+d/2} = o(h^d)$. (SA4.7) is proven.

We will now prove (SA4.8). Define the function class

$$
\mathcal{G}_1 := \{(\boldsymbol{x}_i, y_i) \mapsto \delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})\colon \|\boldsymbol{\alpha}\| = 1, \boldsymbol{q} \in \mathcal{Q}\}.
$$

By Assumption SA2.4(iii), SA2.2(iii) and Assumption SA2.6, for γ small enough

$$
\big|\psi(y_i, \eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \gamma\boldsymbol{\alpha})); \boldsymbol{q}) - \psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})\big| \lesssim 1 + \overline{\psi}(\boldsymbol{x}_i, y_i).
$$

Recalling the envelope condition in Assumption SA2.4(ii) and that $\mu_0(\cdot, \boldsymbol{q})$ is bounded by Assumption SA2.3(iv), we see that $\sup_{g\in\mathcal{G}_1}|g| \lesssim 1 + \overline{\psi}(\boldsymbol{x}_i, y_i)$, which means that under (SA4.1)

$$
\begin{aligned}
\mathbb{E}\Big[\max_{1\le i\le n}|\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})|^2 \,\Big|\, \mathbf{X}_n\Big]^{1/2} &\le \mathbb{E}\Big[\max_{1\le i\le n}|\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})|^\nu \,\Big|\, \mathbf{X}_n\Big]^{1/\nu}\\
&\le \mathbb{E}\Big[\sum_{i=1}^n|\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})|^\nu \,\Big|\, \mathbf{X}_n\Big]^{1/\nu} \lesssim \Big(\sum_{i=1}^n \mathbb{E}\big[\big(1 + \overline{\psi}(\boldsymbol{x}_i, y_i)\big)^\nu \,\big|\, \mathbf{X}_n\big]\Big)^{1/\nu}\\
&\lesssim \Big(\sum_{i=1}^n 1\Big)^{1/\nu} = n^{1/\nu} \quad \text{a. s.}
\end{aligned}
$$

and under (SA4.2) by Lemma SA3.1

$$
\mathbb{E}\Big[\max_{1\le i\le n}|\delta_{\boldsymbol{q},i}(\boldsymbol{\alpha})|^2 \,\Big|\, \mathbf{X}_n\Big]^{1/2} \lesssim \sqrt{\log n} \quad \text{a. s.}
$$

By similar considerations $\sup_{g\in\mathcal{G}_1}\mathbb{E}_n\big[\mathbb{E}[g^2 \,|\, \boldsymbol{x}_i]\big] \lesssim \mathbb{E}_n[\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\alpha}] \lesssim h^d$ w. p. a. 1, where the last inequality holds by Lemma SA3.11.

By Assumption SA2.5, the class

$$
\{(\boldsymbol{x}_i, y_i) \mapsto \psi(y_i, \eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta}); \boldsymbol{q})\colon \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(\boldsymbol{q})\| \le \gamma, \boldsymbol{q} \in \mathcal{Q}\}
$$

with envelope $1 + \overline{\psi}(\boldsymbol{x}_i, y_i)$ multiplied by a constant has a uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim K \asymp h^{-d}$. Moreover, the class

$$
\{(\boldsymbol{x}_i, y_i) \mapsto \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\colon \|\boldsymbol{\alpha}\| = 1\}
$$

has a constant envelope and is VC with index no more than $K+2$ by Lemma 2.6.15 in [18], which means it satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim K \asymp h^{-d}$. Similarly, the same is true of

$$\{(\boldsymbol{x}_i, y_i) \mapsto \boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta} \colon \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(\boldsymbol{q})\| \le \gamma, \boldsymbol{q} \in \mathcal{Q}\}$$

and therefore of

$$\{(\boldsymbol{x}_i, y_i) \mapsto \eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta}) \colon \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(\boldsymbol{q})\| \le \gamma, \boldsymbol{q} \in \mathcal{Q}\}$$

since $\eta^{(1)}(\cdot)$ on a bounded interval is Lipschitz. By Lemma SA3.4, we conclude that $\mathcal{G}_1$ satisfies the uniform entropy bound (SA1.3) with envelope $1 + \overline{\psi}(\boldsymbol{x}_i, y_i)$ multiplied by a constant, $A \lesssim 1$ and $V \lesssim K \asymp h^{-d}$.

Applying the maximal inequality Lemma SA3.6, we obtain (SA4.8). □

SA4.2 Nonconvex case

Our next goal is to prove the consistency result Lemma SA4.3 for the nonconvex case. We will need the following lemma.

Lemma SA4.2 (Preparation for consistency in the nonconvex case)**.** *Suppose Assumptions SA2.1 to SA2.4 hold. Then the infinity norm of $\boldsymbol{\beta}_0(\boldsymbol{q})$ is bounded:*

$$\sup_{\boldsymbol{q} \in \mathcal{Q}} \|\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty \lesssim 1. \tag{SA4.10}$$

Moreover, for any $R > 0$, there is a positive constant $C_1 = C_1(R)$ depending only on R such that for any $\boldsymbol{x} \in \mathcal{X}$

$$\sup_{\boldsymbol{q} \in \mathcal{Q}} \sup_{\|\boldsymbol{\beta}\|_\infty \le R} \left|\rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}); \boldsymbol{q}) - \rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\right| \le C_1(1 + \overline{\psi}(\boldsymbol{x}, y)). \tag{SA4.11}$$

Proof. We prove (SA4.10) first. By Assumption SA2.6 ($\boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x})$ is close to $\mu_0(\boldsymbol{x}, \boldsymbol{q})$) and Assumption SA2.3(iv) ($\mu_0(\boldsymbol{x}, \boldsymbol{q})$ is uniformly bounded), $|\boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x})|$ is bounded uniformly over $\boldsymbol{q} \in \mathcal{Q}$ and $\boldsymbol{x} \in \mathcal{X}$. By Assumption SA2.2, we can bound the kth coordinate of $\boldsymbol{\beta}_0(\boldsymbol{q})$

$$\begin{aligned}
|(\boldsymbol{\beta}_0(\boldsymbol{q}))_k| &\lesssim h^{-d/2}\left(\int_{\mathcal{H}_k} \left(\boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x})\right)^2 \mathrm{d}\boldsymbol{x}\right)^{1/2} \\
&\le h^{-d/2} \sup_{\boldsymbol{x} \in \mathcal{X}} |\boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x})| \cdot (\operatorname{Leb} \mathcal{H}_k)^{1/2} \lesssim \sup_{\boldsymbol{x} \in \mathcal{X}} |\boldsymbol{\beta}_0(\boldsymbol{q})^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x})| \lesssim 1,
\end{aligned}$$

where the constants in $\lesssim$ do not depend on k.

Now we prove (SA4.11). Note that

$$\begin{aligned}
&\rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}); \boldsymbol{q}) - \rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q}) \\
&\quad = \int_0^{\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0(\boldsymbol{q}))} (\psi(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}) + t); \boldsymbol{q}) - \psi(y, \eta(\mu_0(\boldsymbol{x}, \boldsymbol{q})); \boldsymbol{q})) \\
&\qquad\quad \times \eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}) + t)\,\mathrm{d}t \\
&\qquad + \psi(y, \eta(\mu_0(\boldsymbol{x}, \boldsymbol{q})); \boldsymbol{q}) \int_0^{\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0(\boldsymbol{q}))} \eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}) + t)\,\mathrm{d}t
\end{aligned}$$

By Assumption SA2.4(iii), we have a bound

$$|\psi(y, \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}) + t); \boldsymbol{q}) - \psi(y, \eta(\mu_0(\boldsymbol{x}, \boldsymbol{q})); \boldsymbol{q})| \lesssim \overline{\psi}(\boldsymbol{x}, y) + 1.$$

Since both $\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})$ and $\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}$ lie in a fixed compact interval (not depending on $\boldsymbol{x}$ or $\boldsymbol{q}$), $\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})+t)$ is uniformly bounded in absolute value. This means that for any $\boldsymbol{x}\in\mathcal{X}$, $\boldsymbol{q}\in\mathcal{Q}$, $\|\boldsymbol{\beta}\|_\infty\le R$, we have for some positive constants C_2 and C_1 depending only on R

$$
\begin{aligned}
&|\rho(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta});\boldsymbol{q})-\rho(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})|\le C_2(1+\overline{\psi}(\boldsymbol{x},y))\cdot|\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))|\\
&\quad\le C_2(1+\overline{\psi}(\boldsymbol{x},y))\cdot\|\boldsymbol{p}(\boldsymbol{x})\|_1\cdot\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty\\
&\quad\le C_2(1+\overline{\psi}(\boldsymbol{x},y))\cdot\|\boldsymbol{p}(\boldsymbol{x})\|_1\cdot(\|\boldsymbol{\beta}\|_\infty+\|\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty)\\
&\quad\le C_1(1+\overline{\psi}(\boldsymbol{x},y)),
\end{aligned}
$$

concluding the proof. □

We are now ready to prove a general consistency result for an estimator under constraints $\|\boldsymbol{\beta}\|_\infty\le R$ for some large enough constant R. This is Lemma 2 in the paper.

Lemma SA4.3 (Consistency, nonconvex case)**.** *Assume the following conditions.*

(i) Assumptions SA2.1 to SA2.4 and SA2.6 hold.

(ii) $m>d/2$.

(iii) The following rate condition holds:

$$
\frac{\big[\log(1/h)\big]^{\frac{\nu}{\nu-1}}}{nh^{\frac{2\nu}{\nu-1}d}}=o(1),\ \ \textit{or} \tag{SA4.12}
$$

$$
\overline{\psi}(\boldsymbol{x}_i,y_i)\text{ is }\sigma^2\text{-sub-Gaussian conditionally on }\boldsymbol{x}_i\text{ and }\frac{\sqrt{\log n}\log(1/h)}{nh^{2d}}=o(1). \tag{SA4.13}
$$

(iv) $\mathcal{B}=\{\boldsymbol{\beta}\in\mathbb{R}^K:\|\boldsymbol{\beta}\|_\infty\le R\}$ *in (SA1.2), where* $R>0$ *is a fixed number (not depending on* n*) such that* $\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty\le R/2$ *(existing by Lemma SA4.2).*

(v) There is a positive constant c *such that we have* $\inf\Psi_1(\boldsymbol{x},\zeta;\boldsymbol{q})>c$*, where the infimum is over* $\boldsymbol{x}\in\mathcal{X}$*,* $\boldsymbol{q}\in\mathcal{Q}$*,* ζ *between* $\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})$ *and* $\eta(\mu_0(\boldsymbol{x},\boldsymbol{q}))$*, and* $\boldsymbol{\beta}\in\mathcal{B}$*.*

(vi) The class of functions

$$
\{(\boldsymbol{x},y)\mapsto\rho(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta});\boldsymbol{q})-\rho(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q}):\|\boldsymbol{\beta}\|_\infty\le R,\boldsymbol{q}\in\mathcal{Q}\}
$$

with envelope $C_1(1+\overline{\psi}(\boldsymbol{x},y))$ *(by Lemma SA4.2) satisfies the uniform entropy bound (SA1.3) with* $A\lesssim 1$ *and* $V\lesssim K$*.*

Then Eqs. (SA4.3) to (SA4.5) hold.

Proof. For $\boldsymbol{\beta}$ satisfying the constraint $\|\boldsymbol{\beta}\|_\infty\le R$, define

$$
\begin{aligned}
\delta_{\boldsymbol{q},i}(\boldsymbol{\beta})&:=\rho(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta});\boldsymbol{q})-\rho(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\\
&=\int_0^{\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))}\psi(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})+t);\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})+t)\,\mathrm{d}t.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta})\,|\,\boldsymbol{x}_i]&=\int_0^{\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))}\Psi(\boldsymbol{x}_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})+t);\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})+t)\,\mathrm{d}t\\
&\overset{\text{(a)}}{=}\int_0^{\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))}\Psi_1(\boldsymbol{x}_i,\xi_{\boldsymbol{q},i,t};\boldsymbol{q})\\
&\quad\times\eta^{(1)}(\zeta_{\boldsymbol{q},i,t})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})+t)\{\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})-\mu_0(\boldsymbol{x}_i,\boldsymbol{q})+t\}\,\mathrm{d}t
\end{aligned}
$$

$$\overset{\text{(b)}}{\geq} C_4\{\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))\}^2 - C_3 \sup_{\boldsymbol{q},\boldsymbol{x}}|\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}) - \mu_0(\boldsymbol{x},\boldsymbol{q})| \cdot |\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))|$$

$$\overset{\text{(c)}}{\geq} C_4\{\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))\}^2 - C_5 h^m |\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))|,$$

with some positive constants C_4 and C_5 (depending on R), where in (a) we used

$$\begin{aligned}
\Psi(\boldsymbol{x}_i, \eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})+t);\boldsymbol{q}) &= \Psi(\boldsymbol{x}_i, \eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})+t);\boldsymbol{q}) - \Psi(\boldsymbol{x}_i, \eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q}) \\
&= \Psi_1(\boldsymbol{x}_i, \xi_{\boldsymbol{q},i,t};\boldsymbol{q})\{\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})+t) - \eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\} \\
&= \Psi_1(\boldsymbol{x}_i, \xi_{\boldsymbol{q},i,t};\boldsymbol{q})\eta^{(1)}(\zeta_{\boldsymbol{q},i,t})\{\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}) - \mu_0(\boldsymbol{x}_i,\boldsymbol{q}) + t\}
\end{aligned}$$

for some $\xi_{\boldsymbol{q},i,t}$ between $\eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})+t)$ and $\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))$, and some $\zeta_{\boldsymbol{q},i,t}$ between $\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})+t$ and $\mu_0(\boldsymbol{x}_i,\boldsymbol{q})$ by the mean-value theorem applied twice; in (b) we used Condition (v), Assumption SA2.4, in particular that $\eta(\cdot)$ is strictly monotonic giving the positivity of the product $\eta^{(1)}(\zeta_{\boldsymbol{q},i,t})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})+t)$; in (c) we used Assumption SA2.6. By Lyapunov's inequality, $\mathbb{E}_n[|\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))|] \leq \mathbb{E}_n[(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})))^2]^{1/2}$. We conclude

$$\begin{aligned}
&\mathbb{E}_n\big[\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}) \,|\, \boldsymbol{x}_i]\big] \\
&\quad\geq C_4(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))^\mathsf{T}\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}](\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})) - C_5 h^m \mathbb{E}_n[(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})))^2]^{1/2} \\
&\quad\overset{\text{(a)}}{\geq} C_6 h^d \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|^2 - C_7 h^{m+d/2}\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|
\end{aligned}$$

with probability approaching one for some other positive constants C_6 and C_7 (depending on R), where (a) is by Lemma SA3.11.

Fix $\varepsilon > 0$ smaller than $R/2$. In this case

$$\{\boldsymbol{\beta} : \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\| \leq \varepsilon\} \subset \{\boldsymbol{\beta} : \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty \leq \varepsilon\} \subset \{\boldsymbol{\beta} : \|\boldsymbol{\beta}\|_\infty \leq R\}$$

because

$$\|\boldsymbol{\beta}\|_\infty \leq \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty + \|\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty \leq \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty + R/2.$$

Define the class of functions

$$\begin{aligned}
\mathcal{G} := \Big\{&(\boldsymbol{x}, y) \mapsto \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|^{-1}(\rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta});\boldsymbol{q}) - \rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})) : \\
&\|\boldsymbol{\beta}\|_\infty \leq R, \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\| > \varepsilon, \boldsymbol{q} \in \mathcal{Q}\Big\}.
\end{aligned}$$

It is a product of a subclass of the class

$$\{(\boldsymbol{x}, y) \mapsto a : 0 < a < 1/\varepsilon\}$$

with envelope $1/\varepsilon$, obviously satisfying the uniform entropy bound (SA1.3) with $A \lesssim 1$, $V \lesssim 1$, and a subclass of the class

$$\{(\boldsymbol{x}, y) \mapsto \rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})) - \rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))) : \|\boldsymbol{\beta}\|_\infty \leq R, \boldsymbol{q} \in \mathcal{Q}\}$$

with envelope $C_1(1+\overline{\psi}(\boldsymbol{x}, y))$, satisfying the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$ by the conditions of Lemma SA4.3.

By Lemma SA3.4, class $\mathcal{G}$ with envelope $C_1/\varepsilon(1+\overline{\psi}(\boldsymbol{x}, y))$ satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$.

Next, under (SA4.12)

$$
\begin{aligned}
&\mathbb{E}\big[\max_{1\le i\le n}(C_1/\varepsilon)^2(1+\overline{\psi}(\boldsymbol{x},y))^2 \,\big|\, \mathbf{X}_n\big]^{1/2}\\
&\quad\le \mathbb{E}\big[\max_{1\le i\le n}\big(C_1/\varepsilon\big)^{\nu}\big(1+\overline{\psi}(\boldsymbol{x},y)\big)^{\nu} \,\big|\, \mathbf{X}_n\big]^{1/\nu}\\
&\quad\le \mathbb{E}\big[\sum_{i=1}^{n}\big(C_1/\varepsilon\big)^{\nu}\big(1+\overline{\psi}(\boldsymbol{x},y)\big)^{\nu} \,\big|\, \mathbf{X}_n\big]^{1/\nu}\\
&\quad\lesssim \frac{1}{\varepsilon}\bigg(\sum_{i=1}^{n}\mathbb{E}\big[(1+\overline{\psi}(\boldsymbol{x}_i,y_i))^{\nu} \,\big|\, \mathbf{X}_n\big]\bigg)^{1/\nu}\\
&\quad\lesssim \frac{1}{\varepsilon}\bigg(\sum_{i=1}^{n}1\bigg)^{1/\nu} = \frac{n^{1/\nu}}{\varepsilon} \quad \text{a. s.}
\end{aligned}
$$

with constants in $\lesssim$ depending on R but not on n or ε, and under (SA4.13) by Lemma SA3.1

$$
\mathbb{E}\bigg[\max_{1\le i\le n}(C_1/\varepsilon)^2(1+\overline{\psi}(\boldsymbol{x},y))^2 \,\bigg|\, \mathbf{X}_n\bigg]^{1/2} \lesssim \frac{\sqrt{\log n}}{\varepsilon} \quad \text{a. s.}
$$

Moreover,

$$
\begin{aligned}
&\mathbb{E}_n\big[\mathbb{E}[g(\boldsymbol{x}_i,y_i)^2 \mid \boldsymbol{x}_i]\big]\\
&\quad\le \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|^{-2}C_2^2\mathbb{E}_n\big[\big(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))\big)^2\mathbb{E}[(1+\overline{\psi}(\boldsymbol{x}_i,y_i))^2 \mid \boldsymbol{x}_i]\big]\\
&\quad\lesssim \|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|^{-2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))^{\mathsf{T}}\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}](\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q}))\\
&\quad\overset{\text{(a)}}{\lesssim} h^d,
\end{aligned}
$$

where (a) is by Lemma SA3.11.

By Lemma SA3.6, we have

$$
\sup_{g\in\mathcal{G}}\big|\mathbb{E}\big[g(\boldsymbol{x}_i,y_i)-\mathbb{E}[g(\boldsymbol{x}_i,y_i)\mid\boldsymbol{x}_i]\big]\big| \lesssim_{\mathbb{P}} \begin{cases}\sqrt{\frac{\log(1/h)}{n}}+\frac{\log(1/h)}{n^{1-1/\nu}h^d}=o(h^d),\\ \sqrt{\frac{\log(1/h)}{n}}+\frac{\sqrt{\log n}\log(1/h)}{nh^d}=o(h^d)\end{cases}
$$

under (SA4.12) and (SA4.13) respectively (since ε is fixed).

Combining, we infer from the previous results that with probability approaching one for all $\|\boldsymbol{\beta}\|_\infty\le R$, $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|>\varepsilon$, $\boldsymbol{q}\in\mathcal{Q}$

$$
\begin{aligned}
\mathbb{E}_n[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta})] &\ge C_6h^d\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|^2-C_7h^{m+d/2}\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|+\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|\cdot o(h^d)\\
&=\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|\cdot\{C_6h^d\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|-C_7h^{m+d/2}+o(h^d)\}\\
&>\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|\cdot\{C_6\varepsilon h^d-C_7h^{m+d/2}+o(h^d)\}\\
&\overset{\text{(a)}}{=}\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|\cdot h^d\{C_6\varepsilon+o(1)\}>0,
\end{aligned}
$$

where in (a) we used $m>d/2$.

It follows that the constrained minimizer under the constraint $\|\boldsymbol{\beta}\|_\infty\le R$ has to lie inside the ball $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(\boldsymbol{q})\|\le\varepsilon$ for all $\boldsymbol{q}\in\mathcal{Q}$ with probability approaching one. Since ε was arbitrary smaller than $R/2$, it is equivalent to $\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})\|=o_{\mathbb{P}}(1)$. Equations (SA4.4) and (SA4.5) follow (as in Lemma SA4.1). □

SA4.3 Weaker conditions for special cases

Lemmas SA4.4 and SA4.5 consider the special case of unconnected basis functions.

Lemma SA4.4 (Consistency, convex case, unconnected basis functions)**.** *Assume the following.*

(i) Assumptions SA2.1 to SA2.6 hold.

(ii) $\rho(y, \eta(\theta); \boldsymbol{q})$ *is convex with respect to* θ*, and* $\psi(y, \eta(\theta); \boldsymbol{q})\eta^{(1)}(\theta)$ *is its left or right derivative, and* $\mathcal{B} = \mathbb{R}^K$ *in* (SA1.2)*.*

(iii) For all $k \in \{1, \dots, K\}$ *the kth basis function* $p_k(\cdot)$ *is only active on one of the cells of* Δ*.*

(iv) The rate of convergence of h *to zero is restricted by*

$$\frac{\big[\log(1/h)\big]^{\frac{\nu}{\nu-1}}}{nh^{\frac{\nu}{\nu-1}d}} = o(1), \text{ or}$$

$$\overline{\psi}(\boldsymbol{x}_i, y_i) \text{ is } \sigma^2\text{-sub-Gaussian conditionally on } \boldsymbol{x}_i \text{ and } \frac{\sqrt{\log n}\log(1/h)}{nh^d} = o(1).$$

Then Eqs. (SA4.3) *to* (SA4.5) *hold.*

Proof. As in Lemma SA4.1, Eqs. (SA4.4) and (SA4.5) follow from Eq. (SA4.3).

For $l \in \{1, \dots, \kappa\}$, the number M_l of basis functions in $\boldsymbol{p}(\cdot)$ which are active on the lth cell of Δ is bounded by a constant. Denote the vector of such basis functions $\boldsymbol{p}_l := (p_{l,1}, \dots, p_{l,M_l})^\mathsf{T}$. Define the matrices $\mathbf{Q}_{0,\boldsymbol{q},l}$ and $\bar{\mathbf{Q}}_{\boldsymbol{q},l}$ as before with $\boldsymbol{p}$ replaced by $\boldsymbol{p}_l$ (for different l, the dimensions of these square matrices may vary but are bounded from above). By a simple modification of the argument in Lemma SA3.11, the analogues of (SA3.2) and (SA3.5) continue to hold: uniformly over $\boldsymbol{q} \in \mathcal{Q}$ and l

$$\begin{aligned} &h^d \lesssim \lambda_{\min}(\mathbf{Q}_{0,\boldsymbol{q},l}) \le \lambda_{\max}(\mathbf{Q}_{0,\boldsymbol{q},l}) \lesssim h^d, \\ &h^d \lesssim \lambda_{\min}(\bar{\mathbf{Q}}_{\boldsymbol{q},l}) \le \lambda_{\max}(\bar{\mathbf{Q}}_{\boldsymbol{q},l}) \lesssim h^d \quad \text{w. p. a. 1.} \end{aligned} \tag{SA4.14}$$

By the assumption of the lemma, we can write $\boldsymbol{\beta}_0(\boldsymbol{q}) = (\boldsymbol{\beta}_{0,\boldsymbol{q},1}, \dots, \boldsymbol{\beta}_{0,\boldsymbol{q},\kappa})^\mathsf{T}$, where $\boldsymbol{\beta}_{0,\boldsymbol{q},l}$ is a subvector of dimension M_l corresponding to the elements in $\boldsymbol{p}$ active on the lth cell.

Fix a sufficiently small $\gamma > 0$. Denote for $l \in \{1, \dots, \kappa\}$, $i \in \{1, \dots, n\}$ and $\boldsymbol{\alpha}_l \in \mathcal{S}^{M_l - 1}$

$$\delta_{\boldsymbol{q},i,l}(\boldsymbol{\alpha}_l) := \boldsymbol{\alpha}_l^\mathsf{T}\boldsymbol{p}_l(\boldsymbol{x}_i)\psi(y_i, \eta(\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}_{0,\boldsymbol{q},l} + \gamma\boldsymbol{\alpha}_l)); \boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}_{0,\boldsymbol{q},l} + \gamma\boldsymbol{\alpha}_l)).$$

Proceeding in the same way as in Lemma SA4.1, we will show

$$\inf_{\boldsymbol{q},\boldsymbol{\alpha}_l,l} \mathbb{E}_n[\delta_{\boldsymbol{q},i,l}(\boldsymbol{\alpha}_l)] > 0 \quad \text{with probability approaching 1,} \tag{SA4.15}$$

which is again enough to prove the lemma by convexity. It will follow that with probability approaching one the minimizer $\widehat{\boldsymbol{\beta}}_{\boldsymbol{q},l}$ of $\mathbb{E}_n\big[\rho(y_i, \eta(\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_l); \boldsymbol{q})\big]$ with respect to $\boldsymbol{\beta}_l$ has to lie inside the ball $\|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\| \le \gamma$, and in particular inside the cube $\|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\|_\infty \le \gamma$. But note that $\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) = \big(\widehat{\boldsymbol{\beta}}_{\boldsymbol{q},1}^\mathsf{T}, \dots, \widehat{\boldsymbol{\beta}}_{\boldsymbol{q},\bar{k}}^\mathsf{T}\big)^\mathsf{T}$. So, with probability approaching one, for all $\boldsymbol{q} \in \mathcal{Q}$, we have $\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\|_\infty \le \gamma$. Since γ was arbitrary (small enough), it is equivalent to $\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\|_\infty = o_\mathbb{P}(1)$.

Equation (SA4.15) is proven analogously to the corresponding argument in Lemma SA4.1. The class of functions

$$\mathcal{G}_1 := \big\{(\boldsymbol{x}_i, y_i) \mapsto \delta_{\boldsymbol{q},i,l}(\boldsymbol{\alpha}_l) \colon \boldsymbol{\alpha}_l \in \mathcal{S}^{M_l - 1}, l \in \{1, \dots, \kappa\}, \boldsymbol{q} \in \mathcal{Q}\big\}$$

now satisfies the uniform entropy bound (SA1.3) with $A \lesssim \kappa \asymp h^{-d}$ (since there are κ different values of l) and $V \lesssim 1$ (since the vectors $\boldsymbol{\alpha}_l$ are of bounded dimensions). The bound

$\sup_{l,\boldsymbol{\alpha}_l} \mathbb{E}_n\big[|\boldsymbol{\alpha}_l^\mathsf{T}\boldsymbol{p}_l(\boldsymbol{x}_i)|\big] \lesssim h^d$ with probability approaching one can be proven without assuming $m > d/2$ by using

$$\sup_{l,\boldsymbol{\alpha}_l} \mathbb{E}_n\big[|\boldsymbol{\alpha}_l^\mathsf{T}\boldsymbol{p}_l(\boldsymbol{x}_i)|\big] \le \sup_{l,\boldsymbol{\alpha}_l} \|\boldsymbol{\alpha}_l\|_\infty \mathbb{E}_n\big[\|\boldsymbol{p}_l(\boldsymbol{x}_i)\|_1\big] \lesssim h^d \quad \text{w. p. a. 1},$$

since the dimension of $\boldsymbol{p}_l(\cdot)$ is uniformly bounded. □

Lemma SA4.5 (Consistency in the nonconvex case: unconnected basis)**.** *Assume that for all $k \in \{1, \ldots, K\}$ the kth basis function $p_k(\cdot)$ is only active on one of the cells of Δ, and define $\boldsymbol{p}_l(\cdot)$, M_l, $\boldsymbol{\beta}_{0,\boldsymbol{q},l}$ as in the proof of Lemma SA4.4. Further, assume the conditions of Lemma SA4.3 with Condition (ii) removed, Condition (iii) replaced by*

$$\frac{\big[\log(1/h)\big]^{\frac{\nu}{\nu-1}}}{nh^{\frac{\nu}{\nu-1}d}} = o(1), \text{ or} \tag{SA4.16}$$

$$\overline{\psi}(\boldsymbol{x}_i, y_i) \text{ is } \sigma^2\text{-sub-Gaussian conditionally on } \boldsymbol{x}_i \text{ and } \frac{\sqrt{\log n}\log(1/h)}{nh^d} = o(1), \tag{SA4.17}$$

and Condition (vi) replaced by the following: the class

$$\{(\boldsymbol{x}, y) \mapsto \rho(y, \eta(\boldsymbol{p}_l(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_l)) - \rho(y, \eta(\boldsymbol{p}_l(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_{0,\boldsymbol{q},l})) : \|\boldsymbol{\beta}_l\|_\infty \le R, \boldsymbol{q} \in \mathcal{Q}, l \in \{1, \ldots, \kappa\}\}$$

satisfies the uniform entropy bound (SA1.3) *with $A \lesssim \kappa \asymp h^{-d}$ and $V \lesssim 1$. Then Eqs.* (SA4.3) *to* (SA4.5) *hold.*

Proof. As in Lemma SA4.1, Eqs. (SA4.4) and (SA4.5) follow from Eq. (SA4.3).

Define matrices $\mathbf{Q}_{0,\boldsymbol{q},l}$ and $\bar{\mathbf{Q}}_{\boldsymbol{q},l}$ as in the proof of Lemma SA4.4, and recall that the asymptotic bounds on their eigenvalues are the same as in the general (not restricted to one cell) case, i. e. (SA4.14) holds.

For any M_l-dimensional vector $\boldsymbol{\beta}_l$ satisfying the constraint $\|\boldsymbol{\beta}_l\|_\infty \le R$, define

$$\begin{aligned}\delta_{\boldsymbol{q},i,l}(\boldsymbol{\beta}_l) &:= \rho(y_i, \eta(\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_l); \boldsymbol{q}) - \rho(y_i, \eta(\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_{0,\boldsymbol{q},l}); \boldsymbol{q}) \\ &= \int_0^{\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l})} \psi(y_i, \eta(\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_{0,\boldsymbol{q},l} + t); \boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_{0,\boldsymbol{q},l} + t)\,\mathrm{d}t.\end{aligned}$$

By the same argument as in the proof of Lemma SA4.3, we have

$$\begin{aligned}&\mathbb{E}_n\big[\mathbb{E}[\delta_{\boldsymbol{q},i,l}(\boldsymbol{\beta}_l)|\boldsymbol{x}_i]\big] \\ &\quad\ge C_8(\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l})^\mathsf{T}\mathbb{E}_n[\boldsymbol{p}_l(\boldsymbol{x}_i)\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}](\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}) - C_9 h^m \mathbb{E}_n\big[\big|\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l})\big|\big] \\ &\quad\overset{\text{(a)}}{\ge} C_8 h^d\|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\|^2 - C_{10}h^{m+d}\|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\|\end{aligned}$$

with probability approaching one for some positive constants C_8, C_9 and C_{10} (depending on R, but not on $\boldsymbol{q}$ or l), where in (a) we used $\mathbb{E}_n\big[\big|\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l})\big|\big] \le \|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\|_\infty \mathbb{E}_n\big[\|\boldsymbol{p}_l(\boldsymbol{x}_i)\|_1\big] \lesssim \|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\|_\infty h^d \le \|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\|_\infty h^d$ with probability approaching one since the dimension of $\boldsymbol{p}_l(\cdot)$ is bounded.

Next, proceeding with the same concentration argument as in Lemma SA4.3, we will obtain that with probability approaching one for all $l \in \{1, \ldots, \kappa\}$, $\boldsymbol{q} \in \mathcal{Q}$, $\|\boldsymbol{\beta}_l\|_\infty \le R$, $\|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\| > \varepsilon$,

$$\begin{aligned}\mathbb{E}_n[\delta_{\boldsymbol{q},i,l}(\boldsymbol{\beta}_l)] &\ge C_8 h^d\|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\|^2 - C_{10}h^{m+d}\|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\| + \|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\| \cdot o(h^d) \\ &= \|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\| \cdot \big\{C_8 h^d\|\boldsymbol{\beta} - \boldsymbol{\beta}_0(\boldsymbol{q})\| - C_{10}h^{m+d} + o(h^d)\big\}\end{aligned}$$

$$
\begin{aligned}
&> \|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\| \cdot \bigl\{C_8 \varepsilon h^d - C_{10} h^{m+d} + o(h^d)\bigr\} \\
&\overset{(a)}{=} \|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\| \cdot h^d \{C_8 \varepsilon + o(1)\} > 0.
\end{aligned}
$$

It follows that the constrained minimizer $\widehat{\boldsymbol{\beta}}_{\boldsymbol{q},\mathrm{constr},l}$ of $\mathbb{E}_n\bigl[\rho(y_i, \eta(\boldsymbol{p}_l(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_l); \boldsymbol{q})\bigr]$ with respect to $\boldsymbol{\beta}_l$ under the constraint $\|\boldsymbol{\beta}_l\|_\infty \le R$ has to lie inside the ball $\|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\| \le \varepsilon$, and in particular inside the cube $\|\boldsymbol{\beta}_l - \boldsymbol{\beta}_{0,\boldsymbol{q},l}\|_\infty \le \varepsilon$. But this optimization can be solved separately for all l, i. e. $\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) = \bigl(\widehat{\boldsymbol{\beta}}_{\boldsymbol{q},\mathrm{constr},1}, \dots, \widehat{\boldsymbol{\beta}}_{\boldsymbol{q},\mathrm{constr},\bar{k}}\bigr)^\mathsf{T}$. So, with probability approaching one, for all $\boldsymbol{q} \in \mathcal{Q}$, we have $\bigl\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\bigr\|_\infty \le \varepsilon$. Since ε was arbitrary smaller than $R/2$, it is equivalent to $\sup_{\boldsymbol{q}\in\mathcal{Q}}\bigl\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\bigr\|_\infty = o_{\mathbb{P}}(1)$. □

The following lemma considers the special case of strongly convex and strongly smooth loss function.

Lemma SA4.6 (Consistency: strongly convex and strongly smooth loss case)**.** *Assume the following conditions.*

(i) Assumptions SA2.1 to SA2.6 hold.

(ii) The rate of convergence of h to zero is restricted by

$$
\frac{\bigl[\log(1/h)\bigr]^{\frac{\nu}{\nu-1}}}{n h^{\frac{\nu}{\nu-1}d}} = o(1), \quad \text{or}
$$

$$
\overline{\psi}(\boldsymbol{x}_i, y_i) \text{ is } \sigma^2\text{-sub-Gaussian conditionally on } \boldsymbol{x}_i \text{ and } \frac{\log n \log(1/h)}{n h^d} = o(1).
$$

(iii) The function $\eta \mapsto \psi(y, \eta; \boldsymbol{q})$ is continuously differentiable on $\mathbb{R}$ (for all $y, \boldsymbol{q}$), and there exist fixed (not depending on n, $\boldsymbol{q}$ or θ) numbers λ, Λ such that

$$
0 < \lambda \le \frac{\partial}{\partial \theta}\bigl(\psi(y_i, \eta(\theta); \boldsymbol{q})\eta^{(1)}(\theta)\bigr) \le \Lambda,
$$

and $\mathcal{B} = \mathbb{R}^K$.

Then Eqs. (SA4.3) to (SA4.5) hold.

Proof. As in Lemma SA4.1, Eqs. (SA4.4) and (SA4.5) follow from Eq. (SA4.3).

Denote for $\boldsymbol{\beta} \in \mathbb{R}^K$

$$
G_n(\boldsymbol{\beta}) := \mathbb{E}_n[\psi(y_i, \eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}); \boldsymbol{q})\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta})\boldsymbol{p}(\boldsymbol{x}_i)],
$$

which is the gradient of the convex (by Assumption SA2.4(iii)) function $\mathbb{E}_n[\rho(y_i, \eta(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}); \boldsymbol{q})]$ of $\boldsymbol{\beta}$. By definition of $\widehat{\boldsymbol{\beta}}(\boldsymbol{q})$ and differentiability, $G_n\bigl(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})\bigr) = 0$. By the mean value theorem,

$$
G_n(\boldsymbol{\beta}_0(\boldsymbol{q})) = G_n(\boldsymbol{\beta}_0(\boldsymbol{q})) - G_n\bigl(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})\bigr) = \mathbb{E}_n[\mu_i \boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}]\bigl(\boldsymbol{\beta}_0(\boldsymbol{q}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{q})\bigr), \tag{SA4.18}
$$

where

$$
\mu_i := \frac{\partial}{\partial \theta}\bigl(\psi(y_i, \eta(\theta); \boldsymbol{q})\eta^{(1)}(\theta)\bigr)\Big|_{\theta = \widetilde{\theta}_i} \quad \text{for some } \widetilde{\theta}_i \text{ between } \boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}) \text{ and } \boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\widehat{\boldsymbol{\beta}}(\boldsymbol{q}).
$$

By the assumption of the lemma, $0 < \lambda \le \mu_i \le \Lambda$. Therefore, for any vector $\boldsymbol{a} \in \mathbb{R}^K$

$$
\begin{aligned}
\lambda \cdot \boldsymbol{a}^\mathsf{T}\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}]\boldsymbol{a} &\le \boldsymbol{a}^\mathsf{T}\mathbb{E}_n[\mu_i \boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}]\boldsymbol{a} \\
&= \mathbb{E}_n[\mu_i(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{a})^2] \le \Lambda \cdot \boldsymbol{a}^\mathsf{T}\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}]\boldsymbol{a}.
\end{aligned}
$$

Moreover, the matrix

$$\mathbb{E}_n[\mu_i \boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}]$$

has the same multi-banded structure as

$$\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}].$$

That means that by the same argument as that in Lemma SA3.11 we have

$$\left\|\mathbb{E}_n[\mu_i \boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}]^{-1}\right\|_\infty \lesssim_{\mathbb{P}} h^{-d}. \tag{SA4.19}$$

It is shown in the proofs of Lemma SA3.12 and Lemma SA3.14 (see also Remarks SA3.13 and SA3.15) that

$$\left\|G_n(\boldsymbol{\beta}_0(\boldsymbol{q}))\right\|_\infty = o_{\mathbb{P}}(h^d). \tag{SA4.20}$$

From (SA4.18)

$$\|\boldsymbol{\beta}_0(\boldsymbol{q}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{q})\|_\infty \le \left\|\mathbb{E}_n[\mu_i \boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}]^{-1}\right\|_\infty \cdot \|G_n(\boldsymbol{\beta}_0(\boldsymbol{q}))\|_\infty,$$

which in combination with (SA4.19) and (SA4.20) gives

$$\|\boldsymbol{\beta}_0(\boldsymbol{q}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{q})\|_\infty = o_{\mathbb{P}}(1)$$

uniformly over $\boldsymbol{q} \in \mathcal{Q}$. □

SA5 Bahadur representation

We will now prove our first main result, the novel Bahadur representation which is Theorem 1 in the paper. Recall the notation

$$\mathsf{L}^{(\boldsymbol{v})}(\boldsymbol{x}, \boldsymbol{q}) := -\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}} \mathbf{Q}_{0,\boldsymbol{q}}^{-1} \mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})]. \tag{SA5.1}$$

Theorem SA5.1 (Bahadur representation)**.** *Suppose Assumptions SA2.1 to SA2.6 hold. Furthermore, assume the following conditions:*

(i) $\sup_{\boldsymbol{q}\in\mathcal{Q}} \left\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\right\|_\infty = o_{\mathbb{P}}(1)$;

(ii) there exists a constant $c > 0$ *such that* $\{\boldsymbol{\beta} \in \mathbb{R}^K : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty \le c, \boldsymbol{q} \in \mathcal{Q}\} \subseteq \mathcal{B}$;

(iii) $\frac{\log^{d+2} n}{nh^d} = o(1)$;

(iv) either $\frac{(h^{-1}\log n)^{\frac{\nu}{\nu-2}d}}{n} = o(1)$, *or* $\overline{\psi}(\boldsymbol{x}_i, y_i)$ *is* σ^2*-sub-Gaussian conditionally on* $\boldsymbol{x}_i$;

(v) either $\theta \mapsto \rho(y, \eta(\theta); \boldsymbol{q})$ *is convex with left or right derivative* $\psi(y, \eta(\theta); \boldsymbol{q})\eta^{(1)}(\theta)$ *and* $\mathcal{B} = \mathbb{R}^K$, *or the additional complexity Assumption SA2.8 holds.*

(a) Then

$$\sup_{\boldsymbol{q}\in\mathcal{Q}, \boldsymbol{x}\in\mathcal{X}} \left|\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x}, \boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x}, \boldsymbol{q}) - \mathsf{L}^{(\boldsymbol{v})}(\boldsymbol{x}, \boldsymbol{q})\right| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|} r_{\mathsf{BR}} \tag{SA5.2}$$

with

$$r_{\mathsf{BR}} := \left(\frac{\log^d n}{nh^d}\right)^{1/2 + (\alpha/2 \wedge 1/4)} \log n + h^{(\alpha\wedge 1/2)m} \left(\frac{\log^{d+1} n}{nh^d}\right)^{1/2} + h^m. \tag{SA5.3}$$

(b) If, in addition to the previous conditions, (SA2.3) *holds (without any restrictions on* y*), then*

$$\sup_{\boldsymbol{q}\in\mathcal{Q}, \boldsymbol{x}\in\mathcal{X}} \left|\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x}, \boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x}, \boldsymbol{q}) - \mathsf{L}^{(\boldsymbol{v})}(\boldsymbol{x}, \boldsymbol{q})\right| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|} \tilde{r}_{\mathsf{BR}} \tag{SA5.4}$$

with

$$\tilde{r}_{\mathsf{BR}} := \left(\frac{\log^d n}{nh^d}\right)^{(1+\alpha)/2} \log n + h^{\alpha m} \left(\frac{\log^{d+1} n}{nh^d}\right)^{1/2} + h^m. \tag{SA5.5}$$

Remark SA5.2. As will be clear in the proof, the matrix $\mathbf{Q}_{0,\boldsymbol{q}}^{-1}$ in Eqs. (SA5.2) and (SA5.4) can be replaced by $\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}$.

SA5.1 Proof: convex case

We will now prove Theorem SA5.1 under the assumption that $\theta \mapsto \rho(y, \eta(\theta); \boldsymbol{q})$ is convex with left or right derivative $\psi(y, \eta(\theta); \boldsymbol{q})\eta^{(1)}(\theta)$ and $\mathcal{B} = \mathbb{R}^K$. We only show the (a) part, since the argument for (b) is very similar with minor changes in obvious places.

Notation In this proof, we will denote

$$\mathbb{G}_n^i[g(\boldsymbol{x}_i, y_i)] := \sqrt{n}\mathbb{E}_n\big[g(\boldsymbol{x}_i, y_i) - \mathbb{E}[g(\boldsymbol{x}_i, y_i) \mid \boldsymbol{x}_i]\big].$$

SA5.1.1 Strategy

By $\sup_{\boldsymbol{x}}|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})| \lesssim h^{-|\boldsymbol{v}|}$ (Assumption SA2.2), $|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x}, \boldsymbol{q})| \lesssim h^{m-|\boldsymbol{v}|}$ (Assumption SA2.6), and Lemma SA3.14, the inequality

$$\begin{aligned}
\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{x}\in\mathcal{X}} \Big|&\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x}, \boldsymbol{q}) \\
&+ \boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})]\Big| \\
&\lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|}r_{\mathtt{BR}}
\end{aligned} \tag{SA5.6}$$

is implied by

$$\sup_{\boldsymbol{q}}\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\big\|_\infty \lesssim_{\mathbb{P}} \left(\frac{\log^d n}{nh^d}\right)^{1/2+(\alpha/2\wedge 1/4)}\log n + h^{(\alpha\wedge 1/2)m}\left(\frac{\log^{d+1} n}{nh^d}\right)^{1/2}, \tag{SA5.7}$$

where we put

$$\begin{aligned}
&\boldsymbol{p}_i := \boldsymbol{p}(\boldsymbol{x}_i), \\
&\bar{\boldsymbol{\beta}}_{\boldsymbol{q}} := -\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\mathbb{E}_n\big[\boldsymbol{p}_i\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}))\psi(y_i, \eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\big]
\end{aligned} \tag{SA5.8}$$

to declutter notation, and noted that for $h < 1$

$$h^{(\alpha\wedge 1/2)m}\left(\frac{\log^{d+1} n}{nh^d}\right)^{1/2} \lesssim \left(\frac{\log^d n}{nh^d}\right)^{1/2+(\alpha/2\wedge 1/4)}\log n + h^m. \tag{SA5.9}$$

Indeed, if $\alpha \leq 1/2$, either $\big(\frac{\log^{d+1} n}{nh^d}\big)^{1/2} \leq h^m$, in which case (for $h < 1$)

$$h^{\alpha m}\left(\frac{\log^{d+1} n}{nh^d}\right)^{1/2} \leq \left(\frac{\log^{d+1} n}{nh^d}\right)^{1/2} \leq h^m,$$

or $\big(\frac{\log^{d+1} n}{nh^d}\big)^{1/2} > h^m$, in which case

$$h^{\alpha m}\left(\frac{\log^{d+1} n}{nh^d}\right)^{1/2} \leq \left(\frac{\log^{d+1} n}{nh^d}\right)^{(1+\alpha)/2} \leq \left(\frac{\log^d n}{nh^d}\right)^{(1+\alpha)/2}\log n.$$

If $\alpha > 1/2$, (SA5.9) follows from $ab \lesssim a^2 + b^2$.

From (SA5.6), (SA5.2) follows, because $\sup_{\boldsymbol{x}}\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|_1 \lesssim h^{-|\boldsymbol{v}|}$ and

$$\begin{aligned}
&\sup_{\boldsymbol{q}}\big\|(\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}-\mathbf{Q}_{0,\boldsymbol{q}}^{-1})\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})]\big\|_\infty \\
&\quad\lesssim \sup_{\boldsymbol{q}}\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}-\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\|_\infty \sup_{\boldsymbol{q}}\big\|\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})]\big\|_\infty \\
&\quad\overset{\text{(a)}}{\lesssim_{\mathbb{P}}} h^{-d}\Big(\frac{\log(1/h)}{nh^d}\Big)^{1/2}\Big(\frac{h^d\log(1/h)}{n}\Big)^{1/2} = \frac{\log(1/h)}{nh^d} = o(r_{\mathtt{BR}}),
\end{aligned} \tag{SA5.10}$$

where (a) is by Lemma SA3.11 and Eq. (SA3.12). So, we will be showing (SA5.7).

Note that for any vector $\boldsymbol{\alpha}$

$$\mathbb{E}_n\big[\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}))\big] \le 0, \tag{SA5.11}$$

because $g(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}) := \mathbb{E}_n\big[\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}))\boldsymbol{p}_i\big]$ is a subgradient of the function $f(\boldsymbol{\beta}) := \mathbb{E}_n\big[\rho(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta});\boldsymbol{q})\big]$ at $\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}$, and $\widehat{\boldsymbol{\beta}}(\boldsymbol{q})$ is the minimizer of this function, giving $\boldsymbol{\alpha}^\mathsf{T}g\big(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}\big) \le f(\widehat{\boldsymbol{\beta}}(\boldsymbol{q}))-f(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}) \le 0$. Apply (SA5.11) with a particular vector $\boldsymbol{\alpha}_{\boldsymbol{q}}$ that will be chosen later, and decompose

$$\begin{aligned}
0 &\ge \mathbb{E}_n\big[\boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T}\boldsymbol{p}_i\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}_{\boldsymbol{q}}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}_{\boldsymbol{q}}))\big] \\
&= \mathbb{E}_n\big[\boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T}\boldsymbol{p}_i\{\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}_{\boldsymbol{q}}));\boldsymbol{q})-\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\}\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}_{\boldsymbol{q}}))\big] \\
&\quad + T_1 + \mathbb{E}_n\big[\boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T}\boldsymbol{p}_i\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\big],
\end{aligned} \tag{SA5.12}$$

where

$$\begin{aligned}
T_1 &:= \mathbb{E}_n\big[\boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T}\boldsymbol{p}_i\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\alpha}_{\boldsymbol{q}}))\big] \\
&\quad - \mathbb{E}_n\big[\boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T}\boldsymbol{p}_i\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\big]
\end{aligned} \tag{SA5.13}$$

will be bounded later. Define for simplicity

$$\begin{aligned}
\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha}) &:= \boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\big\{\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}));\boldsymbol{q})-\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\big\} \\
&\qquad \times \eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha})),
\end{aligned}$$

so that

$$\text{(SA5.12)} = \mathbb{E}_n\big[\delta_{\boldsymbol{q},i}(\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\boldsymbol{\alpha}_{\boldsymbol{q}})\big].$$

Now, add and subtract the conditional mean of this term, continuing

$$0 \ge n^{-1/2}\mathbb{G}_n^i\big[\delta_{\boldsymbol{q},i}(\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\boldsymbol{\alpha}_{\boldsymbol{q}})\big] \tag{SA5.14}$$

$$+\,\mathbb{E}_n\big[\mathbb{E}\big[\delta_{\boldsymbol{q},i}(\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\boldsymbol{\alpha}_{\boldsymbol{q}})\,\big|\,\boldsymbol{x}_i\big]\big] \tag{SA5.15}$$

$$+\,T_1 + \mathbb{E}_n\big[\boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T}\boldsymbol{p}_i\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\big].$$

The difference (SA5.14) will be shown to be small by the usual concentration argument. Using Taylor expansion, we will show that the term in (SA5.15) is close to $\boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q}))$ with the matrix

$$\widetilde{\mathbf{Q}}_{\boldsymbol{q}} := \mathbb{E}_n\big[\boldsymbol{p}_i\boldsymbol{p}_i^\mathsf{T}\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))^2\big], \tag{SA5.16}$$

that is, we will bound

$$T_2 := \mathbb{E}_n\big[\mathbb{E}\big[\delta_{\boldsymbol{q},i}(\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\boldsymbol{\alpha}_{\boldsymbol{q}})\,\big|\,\boldsymbol{x}_i\big]\big] - \boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})). \tag{SA5.17}$$

To deal with the remaining term $\boldsymbol{\alpha}_{\boldsymbol{q}}^{\mathsf{T}}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q}))$, start with replacing the approximation $\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})$ in the definition of $\widetilde{\mathbf{Q}}_{\boldsymbol{q}}$ with the true function $\mu_0(\boldsymbol{x}_i,\boldsymbol{q})$ leaving us with $\bar{\mathbf{Q}}_{\boldsymbol{q}}$; the error introduced by this operation

$$T_3 := \boldsymbol{\alpha}_{\boldsymbol{q}}^{\mathsf{T}}(\widetilde{\mathbf{Q}}_{\boldsymbol{q}}-\bar{\mathbf{Q}}_{\boldsymbol{q}})(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})) \tag{SA5.18}$$

will be bounded. Next, write

$$\begin{aligned}&\boldsymbol{\alpha}_{\boldsymbol{q}}^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q}))+\mathbb{E}_n\big[\boldsymbol{\alpha}_{\boldsymbol{q}}^{\mathsf{T}}\boldsymbol{p}_i\psi(y_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}))\big]\\ &\quad=\boldsymbol{\alpha}_{\boldsymbol{q}}^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}).\end{aligned}$$

We obtained

$$0\geq n^{-1/2}\mathbb{G}_n^i\big[\delta_{\boldsymbol{q},i}(\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}},\boldsymbol{\alpha}_{\boldsymbol{q}})\big]+T_1+T_2+T_3+\boldsymbol{\alpha}_{\boldsymbol{q}}^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}). \tag{SA5.19}$$

At this point, it would be convenient if we could choose $\boldsymbol{\alpha}_{\boldsymbol{q}}$ so that the last term $\boldsymbol{\alpha}_{\boldsymbol{q}}^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}})$ is proportional (with a positive coefficient) to $\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty$, because then in combination with all the other terms being small in absolute value, we can conclude that $\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty$ is small in absolute value as well (otherwise it is impossible to obtain a nonnegative quantity). This is essentially what we will do, with one caveat: for the bounds on the other terms to work out, it is helpful if $\boldsymbol{\alpha}_{\boldsymbol{q}}$ is very sparse. Following these considerations, we introduce another vector $\bar{\boldsymbol{\alpha}}_{\boldsymbol{q}}$ proportional to $\big[\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\big]_{\boldsymbol{k},k}\operatorname{sign}\big((\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}})_{\boldsymbol{k},k}\big)$ (with a positive coefficient), where $(\boldsymbol{k},k)$ is the index of the largest component of $\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}$ (in the blockwise index notation as in Lemma SA3.11), and $\big[\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\big]_{\boldsymbol{k},k}$ is the $(\boldsymbol{k},k)$th row of $\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}$: this way, $\bar{\boldsymbol{\alpha}}_{\boldsymbol{q}}^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}})$ is indeed proportional to $\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty$. Then we choose $\boldsymbol{\alpha}_{\boldsymbol{q}}$ to be a sparse vector that is close enough to $\bar{\boldsymbol{\alpha}}_{\boldsymbol{q}}$ for an appropriate bound on

$$T_4 := (\boldsymbol{\alpha}_{\boldsymbol{q}}-\bar{\boldsymbol{\alpha}}_{\boldsymbol{q}})^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})-\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}), \tag{SA5.20}$$

which can be done because of the structure of the rows of $\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}$, specifically the exponential decay in Eq. (SA3.9).

SA5.1.2 Main argument

We will now give some specifics. To show a precise bound in probability in (SA5.2) (as opposed to a $o_{\mathbb{P}}(\cdot)$ bound), it will be convenient to multiply $r_{\mathtt{BR}}$ by another positive sequence γ_n that arbitrarily slowly diverges to infinity:

$$\mathfrak{r}_{2,n} := r_{\mathtt{BR}}\gamma_n. \tag{SA5.21}$$

We will also put

$$r_{1,n} := \left[\left(\frac{\log^d n}{nh^d}\right)^{1/2}+h^m\right]\gamma_n^{1/(1\,+\,\alpha\,\wedge\,(1/2))}, \tag{SA5.22}$$

so that in particular

$$\left(\frac{\log^d n}{nh^d}\right)^{1/2}+h^m=o(r_{1,n}) \tag{SA5.23}$$

and by Lemmas SA3.12 and SA3.14, we have on the event $\mathcal{A}_0$ of $1-o(1)$ probability

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty\leq r_{1,n}. \tag{SA5.24}$$

Let γ_n diverge slowly enough so that $r_{1,n} + \mathfrak{r}_{2,n} = o(1)$; in particular, $\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty$ is smaller than any positive constant with probability approaching one.

Fix two small enough constants $c_1, c_2 > 0$ (the restrictions on which will be discussed later). Since c_1 is a constant, with probability approaching one both $\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty \le c_1/2$ (by consistency) and $\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty \le c_1/2$ (as just discussed); hence, with probability approaching one $\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty \le c_1$. This means that the probability of the event $\bigcap_{\boldsymbol{q}\in\mathcal{Q}}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) \in \mathcal{O}_{\boldsymbol{q}}\}$ approaches one, where we use the partitioning

$$\mathcal{O}_{\boldsymbol{q}} := \bigcup_{\ell=-\infty}^{\bar{L}_n} \mathcal{O}_{\boldsymbol{q},\ell}, \quad \mathcal{O}_{\boldsymbol{q},\ell} := \left\{\boldsymbol{\beta} \in \mathbb{R}^K : 2^{\ell-1}\mathfrak{r}_{2,n} \le \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty < 2^\ell \mathfrak{r}_{2,n}\right\}$$

with $\bar{L}_n$ defined as the smallest integer such that $2^{\bar{L}_n}\mathfrak{r}_{2,n} \ge c_1$. (The sequence $\bar{L}_n$ diverges to infinity.)

Put

$$\bar{\boldsymbol{\alpha}}_{\boldsymbol{q}} = c_2 2^L \mathfrak{r}_{2,n} h^d \left[\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\right]_{\boldsymbol{k},k} \operatorname{sign}\left((\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}})_{\boldsymbol{k},k}\right),$$

where $(\boldsymbol{k}, k)$ is such an index that $|(\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}})_{\boldsymbol{k},k}| = \|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty$ (as we already discussed above) and L is a constant integer chosen later. The vector $\bar{\boldsymbol{\alpha}}_{\boldsymbol{q}}$ is not sparse, but the components decay exponentially with the "distance" to the multi-index $(\boldsymbol{k}, k)$ according to Eq. (SA3.9). Therefore, we can zero out all components except a logarithmic neighborhood around $(\boldsymbol{k}, k)$ of this vector and control the error introduced by this operation: specifically, take $\boldsymbol{\alpha}_{\boldsymbol{q}} \in \mathbb{R}^K$ with components $v_{\boldsymbol{q},\boldsymbol{j},j} = \bar{v}_{\boldsymbol{q},\boldsymbol{j},j}$ for $\|\boldsymbol{j} - \boldsymbol{k}\|_\infty \le c_3 \log n$ and zero otherwise, where c_3 is some constant. On the event

$$\mathcal{A}_1 := \left\{\sup_{\delta\in\Delta} \mathbb{E}_n[\mathbb{1}(\boldsymbol{x}_i \in \delta)] \le C_{11} h^d\right\} \tag{SA5.25}$$

of $1 - o(1)$ probability (for some large enough constant $C_{11} > 0$), we have

$$\|\bar{\mathbf{Q}}_{\boldsymbol{q}}\|_\infty \lesssim h^d,$$

so on $\mathcal{A}_1 \cap \left\{\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) \in \mathcal{O}_{\boldsymbol{q},\ell}\right\}$

$$|T_4| \lesssim (2^L \mathfrak{r}_{2,n}) h^d (2^\ell \mathfrak{r}_{2,n}) n^{-c_4} = o(h^d 2^{L+\ell} \mathfrak{r}_{2,n}^2),$$

where $c_4 > 0$ is a constant (depending on the constant controlling the neighborhood size c_3).

Bounding T_1, T_2, T_3 is deferred to Lemmas SA5.6 to SA5.8. Specifically, on an event $\mathcal{A} := \mathcal{A}_0 \cap \mathcal{A}_1 \cap \mathcal{A}_2$ with probability $1 - o(1)$, using Lemma SA5.6 (where $\mathcal{A}_2$ is defined) and the restriction

$$r_{1,n}^2 = o(\mathfrak{r}_{2,n}), \tag{SA5.26}$$

we bound

$$|T_1| = O(h^d r_{1,n}(r_{1,n} + 2^\ell \mathfrak{r}_{2,n} + 2^L \mathfrak{r}_{2,n}) 2^L \mathfrak{r}_{2,n}) = o(h^d 2^{L+\ell} \mathfrak{r}_{2,n}^2).$$

Using Lemma SA5.7, we bound also on $\left\{\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) \in \mathcal{O}_{\boldsymbol{q},\ell}\right\} \cap \mathcal{A}$ with $L \le \ell \le \bar{L}_n$:

$$|T_2| \le \frac{c_2}{8} h^d 2^{L+\ell} \mathfrak{r}_{2,n}^2. \tag{SA5.27}$$

Indeed,

$$2^L \mathfrak{r}_{2,n} h^d \left(r_{1,n} + 2^\ell \mathfrak{r}_{2,n} + 2^L \mathfrak{r}_{2,n}\right)^2 \lesssim 2^L \mathfrak{r}_{2,n} h^d r_{1,n}^2 + 2^L \mathfrak{r}_{2,n} h^d 2^{2\ell} \mathfrak{r}_{2,n}^2$$

$$\stackrel{\text{(a)}}{=} 2^L \mathfrak{r}_{2,n} h^d 2^{2\ell} \mathfrak{r}_{2,n}^2 + o(h^d 2^L \mathfrak{r}_{2,n}^2),$$

where in (a) we used $r_{1,n}^2 = o(\mathfrak{r}_{2,n})$ again. Since $2^\ell \mathfrak{r}_{2,n} \le 2c_1$, we can therefore ensure

$$\left|T_2 + \boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T} \widetilde{\mathbf{Q}}_{\boldsymbol{q}} \boldsymbol{\alpha}_{\boldsymbol{q}}\right| \le \frac{c_2}{16} h^d 2^{L+\ell} \mathfrak{r}_{2,n}^2$$

by taking c_1 sufficiently small relative to c_2. Next, taking c_2 sufficiently small (in particular, making c_2^2 much smaller than c_2), we can upper-bound $\lambda_{\max}(\widetilde{\mathbf{Q}}_{\boldsymbol{q}})\|\boldsymbol{\alpha}_{\boldsymbol{q}}\|^2$ by $(c_2/16)h^d 2^{L+\ell}\mathfrak{r}_{2,n}^2$ as well, giving (SA5.27).

Finally, using Lemma SA5.8, we bound on $\mathcal{A}$

$$|T_3| = O(h^{m+d}(2^L \mathfrak{r}_{2,n})(r_{1,n} + 2^\ell \mathfrak{r}_{2,n})) \stackrel{\text{(a)}}{=} o(h^d 2^{L+\ell} \mathfrak{r}_{2,n}^2),$$

where in (a) we used

$$h^m r_{1,n} = o(\mathfrak{r}_{2,n}). \tag{SA5.28}$$

Combining, we have

$$|T_1| + |T_2| + |T_3| + |T_4| \le \frac{c_2}{4} h^d 2^{L+\ell} \mathfrak{r}_{2,n}^2.$$

Since

$$\bar{\boldsymbol{\alpha}}_{\boldsymbol{q}}^\mathsf{T} \bar{\boldsymbol{Q}}_{\boldsymbol{q}} (\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}) = c_2 2^L \mathfrak{r}_{2,n} h^d \|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\|_\infty \ge c_2 2^L \mathfrak{r}_{2,n} h^d 2^{\ell-1} \mathfrak{r}_{2,n},$$

we conclude from (SA5.19) that on $\{\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) \in \mathcal{O}_{\boldsymbol{q},\ell}\} \cap \mathcal{A}$, with $L \le \ell \le \bar{L}_n$,

$$\begin{aligned} -n^{-1/2}\mathbb{G}_n^i\left[\delta_{\boldsymbol{q},i}(\bar{\boldsymbol{\beta}}_{\boldsymbol{q}}, \widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}, \boldsymbol{\alpha}_{\boldsymbol{q}})\right] &\ge c_2 2^L \mathfrak{r}_{2,n} h^d 2^{\ell-1} \mathfrak{r}_{2,n} - \frac{c_2}{4} h^d 2^{L+\ell} \mathfrak{r}_{2,n}^2 \\ &= \frac{c_2}{4} h^d 2^{L+\ell} \mathfrak{r}_{2,n}^2. \end{aligned}$$

We will prove that the probability of this event is small enough, by using a concentration argument.

Lemma SA5.3 (Uniform concentration)**.** *Define*

$$\begin{aligned} \mathcal{V} &:= \big\{\boldsymbol{\alpha} \in \mathbb{R}^K : \exists\ d\text{-dimensional multi-index } \boldsymbol{k}, \\ &\qquad |v_{\boldsymbol{\ell},l}| \le \varrho^{\|\boldsymbol{k}-\boldsymbol{l}\|_\infty} 2^L \mathfrak{r}_{2,n} \text{ for } \|\boldsymbol{k}-\boldsymbol{l}\|_\infty \le c_3 \log n \text{ and } v_{\boldsymbol{\ell},l} = 0 \text{ otherwise}\big\}, \\ \mathcal{H}_1 &:= \{\boldsymbol{\beta} \in \mathbb{R}^K : \|\boldsymbol{\beta}\|_\infty \le r_{1,n}\}, \\ \mathcal{H}_{2,\ell} &:= \{\boldsymbol{\beta} \in \mathbb{R}^K : \|\boldsymbol{\beta}\|_\infty \le 2^\ell \mathfrak{r}_{2,n}\}, \end{aligned}$$

where ϱ is the constant from Lemma SA3.11. On the event $\mathcal{A}_1$ defined in Eq. (SA5.25) *we have*

$$\begin{aligned} &\mathbb{E}\Big[\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{\beta}_1\in\mathcal{H}_1,\boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell},\boldsymbol{\alpha}\in\mathcal{V}} \left|\mathbb{E}_n[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha}) - \mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})|\boldsymbol{x}_i]]\right| \Big| \{\boldsymbol{x}_i\}_{i=1}^n\Big] \\ &\quad \le C_{12}\left(\frac{h^{d/2} 2^L \mathfrak{r}_{2,n}(r_{1,n} + 2^\ell \mathfrak{r}_{2,n} + 2^L \mathfrak{r}_{2,n})^{\alpha\wedge(1/2)}}{\sqrt{n}} \log^{(d+1)/2} n + \frac{2^L \mathfrak{r}_{2,n} \log^{d+1} n}{n}\right), \end{aligned} \tag{SA5.29}$$

where the constant C_{12} does not depend on L or ℓ.

Section SA5.1.3 is devoted to the proof of this fact.

On $\{\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) \in \mathcal{O}_{\boldsymbol{q},\ell}\} \cap \mathcal{A}$, we have (as long as c_2 is small enough)

$$\boldsymbol{\alpha}_{\boldsymbol{q}} \in \mathcal{V}, \quad \bar{\boldsymbol{\beta}}_{\boldsymbol{q}} \in \mathcal{H}_1, \quad \widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}} \in \mathcal{H}_{2,\ell}.$$

Therefore,

$$\begin{aligned}
&\mathbb{P}\Big\{\sup_{\boldsymbol{q}\in\mathcal{Q}} \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\big\|_\infty \geq 2^L \mathfrak{r}_{2,n} \,\Big|\, \{\boldsymbol{x}_i\}_{i=1}^n\Big\} \\
&\quad = \mathbb{P}\Big\{\bigcap_{\boldsymbol{q}\in\mathcal{Q}} \bigcup_{\ell=L+1}^{\infty} \{\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) \in \mathcal{O}_{\boldsymbol{q},\ell}\} \,\Big|\, \{\boldsymbol{x}_i\}_{i=1}^n\Big\} \\
&\quad = \mathbb{P}\Big\{\bigcap_{\boldsymbol{q}\in\mathcal{Q}} \bigcup_{\ell=L+1}^{\bar{L}_n} \{\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) \in \mathcal{O}_{\boldsymbol{q},\ell}\} \cap \mathcal{A} \,\Big|\, \{\boldsymbol{x}_i\}_{i=1}^n\Big\} + o_{\mathbb{P}}(1) \\
&\quad \leq \mathbb{P}\Big\{\bigcup_{\ell=L+1}^{\bar{L}_n} \Big\{\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{\beta}_1\in\mathcal{H}_1,\boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell},\boldsymbol{\alpha}\in\mathcal{V}} \big|n^{-1/2}\mathbb{G}_n^i\big[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})\big]\big| \geq \frac{c_2}{4} h^d 2^{L+\ell}\mathfrak{r}_{2,n}^2\Big\} \,\Big|\, \{\boldsymbol{x}_i\}_{i=1}^n\Big\} \mathbb{1}(\mathcal{A}_1) + o_{\mathbb{P}}(1) \\
&\quad \leq \sum_{\ell=L+1}^{\bar{L}_n} \mathbb{P}\Big\{\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{\beta}_1\in\mathcal{H}_1,\boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell},\boldsymbol{\alpha}\in\mathcal{V}} \big|n^{-1/2}\mathbb{G}_n^i\big[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})\big]\big| \geq \frac{c_2}{4} h^d 2^{L+\ell}\mathfrak{r}_{2,n}^2 \,\Big|\, \{\boldsymbol{x}_i\}_{i=1}^n\Big\} \mathbb{1}(\mathcal{A}_1) + o_{\mathbb{P}}(1) \\
&\quad \leq \sum_{\ell=L+1}^{\bar{L}_n} \Big(\frac{c_2}{4} h^d 2^{L+\ell}\mathfrak{r}_{2,n}^2\Big)^{-1} \mathbb{E}\Big[\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{\beta}_1\in\mathcal{H}_1,\boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell},\boldsymbol{\alpha}\in\mathcal{V}} \big|n^{-1/2}\mathbb{G}_n^i\big[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})\big]\big| \,\Big|\, \{\boldsymbol{x}_i\}_{i=1}^n\Big] \mathbb{1}(\mathcal{A}_1) + o_{\mathbb{P}}(1) \\
&\quad \leq \sum_{\ell=L+1}^{\bar{L}_n} C_{12}\Big(\frac{c_2}{4} h^d 2^{L+\ell}\mathfrak{r}_{2,n}^2\Big)^{-1} \Big(\frac{h^{d/2}2^L\mathfrak{r}_{2,n}(r_{1,n} + 2^\ell \mathfrak{r}_{2,n} + 2^L\mathfrak{r}_{2,n})^{\alpha\wedge(1/2)}}{\sqrt{n}} \log^{(d+1)/2} n + \frac{2^L\mathfrak{r}_{2,n}\log^{d+1} n}{n}\Big) \\
&\quad \lesssim \sum_{\ell=L+1}^{\bar{L}_n} 2^{-\ell}\mathfrak{r}_{2,n}^{-1} \Big(\frac{r_{1,n}^{\alpha\wedge(1/2)} + \big(2^\ell \mathfrak{r}_{2,n}\big)^{\alpha\wedge(1/2)}}{\sqrt{nh^d}} \log^{(d+1)/2} n + \frac{\log^{d+1} n}{nh^d}\Big), \\
&\quad \overset{\text{(a)}}{\lesssim} \sum_{\ell=L+1}^{\bar{L}_n} \big(2^{-\ell} + 2^{\alpha\wedge(1/2)-\ell}\big) \lesssim \sum_{\ell=L+1}^{\infty} 2^{-\ell} = 2^{-L},
\end{aligned}$$

where in (a) we used

$$\frac{r_{1,n}^{\alpha\wedge(1/2)}}{\sqrt{nh^d}} \log^{(d+1)/2} n \lesssim \mathfrak{r}_{2,n}, \tag{SA5.30}$$

$$\frac{\log^{(d+1)/2} n}{\sqrt{nh^d}} \lesssim \mathfrak{r}_{2,n}^{1-\alpha\wedge(1/2)}, \tag{SA5.31}$$

$$\frac{\log^{d+1} n}{nh^d} \lesssim \mathfrak{r}_{2,n}. \tag{SA5.32}$$

We conclude that

$$\sup_{\boldsymbol{q}\in\mathcal{Q}} \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\big\|_\infty \lesssim_{\mathbb{P}} \mathfrak{r}_{2,n}.$$

Since γ_n was arbitrarily slowly diverging, it follows that in fact

$$\sup_{\boldsymbol{q}\in\mathcal{Q}} \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) - \bar{\boldsymbol{\beta}}_{\boldsymbol{q}}\big\|_\infty \lesssim_{\mathbb{P}} r_{\mathtt{BR}}.$$

SA5.1.3 Concentration argument (proof of Lemma SA5.3)

We will first argue that $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ can be considered effectively low-dimensional (of polylog n dimension).

Note that $\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha}) \neq 0$ only if $\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i \neq 0$. For each $\boldsymbol{\alpha} \in \mathcal{V}$, let $\mathcal{J}_{\boldsymbol{\alpha}} := \{j : v_j \neq 0\}$. By construction, the cardinality of $\mathcal{J}_{\boldsymbol{\alpha}}$ is bounded by $(2c_3 \log n + 1)^d$. We have $\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha}) \neq 0$ only if $p_j(\boldsymbol{x}_i) \neq 0$ for some $j \in \mathcal{J}_{\boldsymbol{\alpha}}$, which happens only if $\boldsymbol{x}_i \in \mathcal{I}_{\boldsymbol{\alpha}}$ where

$$\mathcal{I}_{\boldsymbol{\alpha}} := \bigcup\{\delta \in \Delta \colon \delta \cap \operatorname{supp} p_j \neq \varnothing \text{ for some } j \in \mathcal{J}_{\boldsymbol{\alpha}}\}.$$

The family $\mathcal{I}_{\boldsymbol{\alpha}}$ includes at most $c_5(c_3 \log n)^d$ cells. Moreover, at most $c_6(c_3 \log n)^d$ basis functions in $\boldsymbol{p}$ have supports overlapping with $\mathcal{I}_{\boldsymbol{\alpha}}$. Denote the set of indices of such basis functions by $\bar{\mathcal{J}}_{\boldsymbol{\alpha}}$. Based on the above observations, we have $\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha}) = \delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_{1,\bar{\mathcal{J}}_{\boldsymbol{\alpha}}},\boldsymbol{\beta}_{2,\bar{\mathcal{J}}_{\boldsymbol{\alpha}}},\boldsymbol{\alpha})$, where

$$\begin{aligned}
&\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_{1,\bar{\mathcal{J}}_{\boldsymbol{\alpha}}},\boldsymbol{\beta}_{2,\bar{\mathcal{J}}_{\boldsymbol{\alpha}}},\boldsymbol{\alpha}) \\
&\quad := \sum_{j\in\mathcal{J}_{\boldsymbol{\alpha}}} p_{i,j}v_j\Bigg[\psi\Bigg(y_i, \eta\Bigg(\sum_{l\in\bar{\mathcal{J}}_{\boldsymbol{\alpha}}} p_{i,l}(\beta_{0,\boldsymbol{q},l}+\beta_{1,l}+\beta_{2,l}) - \sum_{j\in\mathcal{J}_{\boldsymbol{\alpha}}} p_{i,j}v_j\Bigg);\boldsymbol{q}\Bigg) \\
&\qquad - \psi\Bigg(y_i, \eta\Bigg(\sum_{l\in\bar{\mathcal{J}}_{\boldsymbol{\alpha}}} p_{i,l}\beta_{0,\boldsymbol{q},l}\Bigg);\boldsymbol{q}\Bigg)\Bigg] \\
&\qquad \times \eta^{(1)}\Bigg(\sum_{l\in\bar{\mathcal{J}}_{\boldsymbol{\alpha}}} p_{i,l}(\beta_{0,\boldsymbol{q},l}+\beta_{1,l}+\beta_{2,l}) - \sum_{j\in\mathcal{J}_{\boldsymbol{\alpha}}} p_{i,j}v_j\Bigg)\mathbb{1}\{\boldsymbol{x}_i \in \mathcal{I}_{\boldsymbol{\alpha}}\}.
\end{aligned} \tag{SA5.33}$$

Accordingly, for $\widetilde{\boldsymbol{\beta}}_1 \in \mathbb{R}^{c_6(c_3 \log n)^d}$, $\widetilde{\boldsymbol{\beta}}_2 \in \mathbb{R}^{c_6(c_3 \log n)^d}$, define the function class

$$\mathcal{G} := \big\{(\boldsymbol{x}_i,y_i) \mapsto \delta_{\boldsymbol{q},i}\big(\widetilde{\boldsymbol{\beta}}_1,\widetilde{\boldsymbol{\beta}}_2,\boldsymbol{\alpha}\big) : \boldsymbol{q} \in \mathcal{Q}, \boldsymbol{\alpha} \in \mathcal{V}, \big\|\widetilde{\boldsymbol{\beta}}_1\big\|_\infty \le r_{1,n}, \big\|\widetilde{\boldsymbol{\beta}}_2\big\|_\infty \le 2^\ell \mathfrak{r}_{2,n}\big\}.$$

We will now bound $\sup_{g\in\mathcal{G}}|\mathbb{E}_n[g(\boldsymbol{x}_i,y_i)] - \mathbb{E}\big[g(\boldsymbol{x}_i,y_i) \,\big|\, \boldsymbol{x}_i\big]|$. As usual, the strategy is to check conditions of Lemma SA3.6.

Lemma SA5.4 (Bonding variance)**.** *On $\mathcal{A}_1$, class $\mathcal{G}$ satisfies the following variance bound:*

$$\sup_{g\in\mathcal{G}} \mathbb{E}_n[\mathbb{V}[g(\boldsymbol{x}_i,y_i) \mid \boldsymbol{x}_i]] \lesssim 2^{2L}\mathfrak{r}_{2,n}^2 h^d (r_{1,n} + 2^\ell \mathfrak{r}_{2,n} + 2^L \mathfrak{r}_{2,n})^{(2\alpha)\wedge 1}.$$

Proof. We will now proceed under the assumption that $\mathfrak{M}$ is Lebesgue measure, so (SA2.2) holds; the argument under (SA2.3) is similar (and leads to an even stronger variance bound), so it is omitted.

By the same argument as in the proof of Lemma SA3.14, using $\big|\psi(y,\eta(p(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\big| \lesssim \overline{\psi}(\boldsymbol{x},y)+1$, for y_i *outside* the closed segment between the two points

$$\eta\Bigg(\sum_{l\in\bar{\mathcal{J}}_v} p_{i,l}\beta_{0,\boldsymbol{q},l}\Bigg), \quad \text{and} \quad \eta\Bigg(\sum_{l\in\bar{\mathcal{J}}_v} p_{i,l}(\beta_{0,\boldsymbol{q},l}+\beta_{1,l}+\beta_{2,l}) - \sum_{j\in\mathcal{J}_v} p_{i,j}v_j\Bigg),$$

we have

$$\begin{aligned}
&\Bigg|\psi\Bigg(y_i, \eta\Bigg(\sum_{l\in\bar{\mathcal{J}}_v} p_{i,l}(\beta_{0,\boldsymbol{q},l}+\beta_{1,l}+\beta_{2,l}) - \sum_{j\in\mathcal{J}_v} p_{i,j}v_j\Bigg);\boldsymbol{q}\Bigg) - \psi\Bigg(y_i, \eta\Bigg(\sum_{l\in\bar{\mathcal{J}}_v} p_{i,l}\beta_{0,\boldsymbol{q},l}\Bigg);\boldsymbol{q}\Bigg)\Bigg| \\
&\quad \lesssim \big(\overline{\psi}(\boldsymbol{x}_i,y_i)+1\big)\big(r_{1,n} + 2^\ell \mathfrak{r}_{2,n} + 2^L \mathfrak{r}_{2,n}\big)^\alpha,
\end{aligned}$$

and for y_i *in* this segment we have

$$\left|\psi\Big(y_i, \eta\Big(\sum_{l\in\bar{\mathcal{J}}_v} p_{i,l}(\beta_{0,\boldsymbol{q},l}+\beta_{1,l}+\beta_{2,l}) - \sum_{j\in\mathcal{J}_v} p_{i,j}v_j\Big); \boldsymbol{q}\Big) - \psi\Big(y_i, \eta\Big(\sum_{l\in\bar{\mathcal{J}}_v} p_{i,l}\beta_{0,\boldsymbol{q},l}\Big); \boldsymbol{q}\Big)\right|$$
$$\lesssim (\overline{\psi}(\boldsymbol{x}_i, y_i)+1)(r_{1,n}+2^{\ell}\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})+1$$

uniformly over $\boldsymbol{q}$.

By construction, for each $\boldsymbol{\alpha}\in\mathcal{V}$, there exists some $\boldsymbol{k}_{\boldsymbol{\alpha}}$ such that $|v_{\boldsymbol{l},l}|\le \varrho^{\|\boldsymbol{l}-\boldsymbol{k}_{\boldsymbol{\alpha}}\|_\infty}2^L\mathfrak{r}_{2,n}$ if $\|\boldsymbol{l}-\boldsymbol{k}_{\boldsymbol{\alpha}}\|_\infty\le M_n$, and otherwise $v_{\boldsymbol{l},l}=0$. The above facts imply(cf. the proof of Lemma SA3.14) that for any $\boldsymbol{x}_i\in\delta\subset\mathcal{I}_{\boldsymbol{\alpha}}$,

$$\mathbb{V}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})\,|\,\boldsymbol{x}_i]\lesssim 2^{2L}\mathfrak{r}_{2,n}^2(r_{1,n}+2^{\ell}\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^{(2\alpha)\wedge 1}\sum_{(\boldsymbol{l},l)\in\mathcal{L}_\delta}\varrho^{2\|\boldsymbol{l}-\boldsymbol{k}_{\boldsymbol{\alpha}}\|_\infty}$$
$$\text{for}\quad \mathcal{L}_\delta:=\{(\boldsymbol{l},l)\colon\ \operatorname{supp}p_{\boldsymbol{l},l}\cap\delta\neq\varnothing\}.$$

In addition, since $\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})\neq 0$ only if $\boldsymbol{x}_i\in\mathcal{I}_{\boldsymbol{\alpha}}$, for all $g\in\mathcal{G}$,

$$\mathbb{E}_n[\mathbb{V}[g(\boldsymbol{x}_i,y_i)\,|\,\boldsymbol{x}_i]]\lesssim 2^{2L}\mathfrak{r}_{2,n}^2(r_{1,n}+2^{\ell}\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^{(2\alpha)\wedge 1}\sum_{\delta\subset\mathcal{I}_{\boldsymbol{\alpha}}}\mathbb{E}_n[\mathbb{1}(\boldsymbol{x}_i\in\delta)]\sum_{(\boldsymbol{l},l)\in\mathcal{L}_\delta}\varrho^{2\|\boldsymbol{l}-\boldsymbol{k}_{\boldsymbol{\alpha}}\|_\infty}$$
$$=2^{2L}\mathfrak{r}_{2,n}^2(r_{1,n}+2^{\ell}\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^{(2\alpha)\wedge 1}\sum_{\boldsymbol{l},l}\varrho^{2\|\boldsymbol{l}-\boldsymbol{k}_{\boldsymbol{\alpha}}\|_\infty}\sum_{\delta\in\mathcal{L}^\star_{\boldsymbol{l},l}}\mathbb{E}_n[\mathbb{1}(\boldsymbol{x}_i\in\delta)]$$
$$\text{for}\quad \mathcal{L}^\star_{\boldsymbol{l},l}:=\{\delta\subset\mathcal{I}_{\boldsymbol{\alpha}}\colon\ \operatorname{supp}p_{\boldsymbol{l},l}\cap\delta\neq\varnothing\}.$$

Note that $\mathcal{L}^\star_{\boldsymbol{l},l}$ contains a bounded number of elements. Then on $\mathcal{A}_1$,

$$\sup_{g\in\mathcal{G}}\mathbb{E}_n[\mathbb{V}[g(\boldsymbol{x}_i,y_i)\,|\,\boldsymbol{x}_i]]\lesssim 2^{2L}\mathfrak{r}_{2,n}^2h^d(r_{1,n}+2^{\ell}\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^{(2\alpha)\wedge 1}\sum_{\boldsymbol{l},l}\varrho^{2\|\boldsymbol{l}-\boldsymbol{k}_{\boldsymbol{\alpha}}\|_\infty}$$
$$\lesssim 2^{2L}\mathfrak{r}_{2,n}^2h^d(r_{1,n}+2^{\ell}\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^{(2\alpha)\wedge 1}\sum_{\boldsymbol{l}}\varrho^{2\|\boldsymbol{l}-\boldsymbol{k}_{\boldsymbol{\alpha}}\|_\infty}\qquad\text{since } l \text{ is bounded}$$
$$\lesssim 2^{2L}\mathfrak{r}_{2,n}^2h^d(r_{1,n}+2^{\ell}\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^{(2\alpha)\wedge 1}\sum_{\boldsymbol{t}\in\mathbb{Z}^d}\varrho^{2\|\boldsymbol{t}\|_\infty}\lesssim 2^{2L}\mathfrak{r}_{2,n}^2h^d(r_{1,n}+2^{\ell}\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^{(2\alpha)\wedge 1},$$

concluding the proof of Lemma SA5.4. □

Lemma SA5.5 (Complexity of class $\mathcal{G}$)**.** *Class $\mathcal{G}$ with envelope $2^L\mathfrak{r}_{2,n}$ multiplied by a large enough constant satisfies the uniform entropy bound* (SA1.3) *with $A\lesssim 1$ and $V\lesssim\log^d n$.*

Proof of Lemma SA5.5. First, indeed $\sup_{\boldsymbol{x},y}\sup_{g\in\mathcal{G}}|g(\boldsymbol{x},y)|\lesssim 2^L\mathfrak{r}_{2,n}$.

Next, the class of functions $\mathcal{W}_1:=\{(\boldsymbol{x}_i,y_i)\mapsto\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\colon\boldsymbol{\alpha}\in\mathcal{V}\}$ is a union of $O(h^{-d})$ classes $\mathcal{W}_{1,\boldsymbol{k}}:=\{(\boldsymbol{x}_i,y_i)\mapsto\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\colon\boldsymbol{\alpha}\in\mathcal{V}_{\boldsymbol{k}}\}$, where

$$\mathcal{V}_{\boldsymbol{k}}=\big\{\boldsymbol{\alpha}\in\mathbb{R}^K: |v_{\boldsymbol{\ell},l}|\le\varrho^{\|\boldsymbol{k}-\boldsymbol{\ell}\|_\infty}2^L\mathfrak{r}_{2,n}\text{ for }\|\boldsymbol{k}-\boldsymbol{\ell}\|_\infty\le M_n\text{ and }v_{\boldsymbol{\ell},l}=0\text{ otherwise}\big\}.$$

Since $\mathcal{W}_{1,\boldsymbol{k}}$ is a subclass of a vector space of functions of dimension $O(\log^d n)$, by Lemma 2.6.15 in [18] it is VC with index $O(\log^d n)$. This implies that $\mathcal{W}_{1,\boldsymbol{k}}$ with envelope $O(2^L\mathfrak{r}_{2,n})$ satisfies the uniform entropy bound (SA1.3) with $A\lesssim 1$ and $V\lesssim\log^d n$. Since there are $O(h^{-d})$ such classes and $\log(1/h)\lesssim\log n$, using the chain

$$O(h^{-d})\Big(\frac{A}{\varepsilon}\Big)^{O(\log^d n)}=e^{O(\log n)}\Big(\frac{A}{\varepsilon}\Big)^{O(\log^d n)}\le\Big(\frac{A}{\varepsilon}\Big)^{O(\log n)+O(\log^d n)}=\Big(\frac{A}{\varepsilon}\Big)^{O(\log^d n)}\qquad\text{(SA5.34)}$$

(recall that $A \geq e$), we get that $\mathcal{W}_1$ also satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$.

By Assumption SA2.5, the class of functions

$$\begin{aligned}\mathcal{W}_2 := \big\{&(\boldsymbol{x}_i, y_i) \mapsto \\ &[\psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 - \boldsymbol{\alpha})); \boldsymbol{q}) - \psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})]\mathbb{1}\{\boldsymbol{x}_i \in \mathcal{I}_{\boldsymbol{\alpha}}\}\colon \\ &\boldsymbol{\beta}_1 \in \mathcal{H}_1, \boldsymbol{\beta}_2 \in \mathcal{H}_{2,\ell}, \boldsymbol{\alpha} \in \mathcal{V}, \boldsymbol{q} \in \mathcal{Q}\big\}\end{aligned}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$.

The class of functions

$$\begin{aligned}\mathcal{W}_3 := \big\{(\boldsymbol{x}_i, y_i) \mapsto \eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 - \boldsymbol{\alpha}))\mathbb{1}\{\boldsymbol{x}_i \in \mathcal{I}_{\boldsymbol{\alpha}}\}\colon \\ \boldsymbol{\beta}_1 \in \mathcal{H}_1, \boldsymbol{\beta}_2 \in \mathcal{H}_{2,\ell}, \boldsymbol{\alpha} \in \mathcal{V}, \boldsymbol{q} \in \mathcal{Q}\big\}\end{aligned}$$

is a subset of the union over $\delta \in \Delta$ of classes (for some fixed positive constants c and r, n large enough)

$$\mathcal{W}_{3,\delta} := \big\{(\boldsymbol{x}_i, y_i) \mapsto \eta^{(1)}(\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\boldsymbol{\beta})\mathbb{1}\{\boldsymbol{x}_i \in \mathcal{N}_{[c\log n]}(\delta)\}\colon \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(\boldsymbol{q})\|_\infty \leq r, \boldsymbol{q} \in \mathcal{Q}\big\}.$$

Note that $\boldsymbol{\beta}$ can be assumed to lie in a fixed vector space $\mathcal{B}_\delta$ of dimension $\dim \mathcal{B}_\delta = O(\log^d n)$. Again applying Lemma 2.6.15 in [18] and noting that $\eta^{(1)}$ on a fixed (bounded) interval is Lipschitz, we have that $\mathcal{W}_{3,\delta}$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$. Similarly to the argument for $\mathcal{W}_1$, this implies that the same is true for $\mathcal{W}_3$.

Applying Lemma SA3.4 concludes the proof of Lemma SA5.5. □

Proof of Lemma SA5.3. Apply Lemma SA3.6 conditionally on $\{\boldsymbol{x}_i\}_{i=1}^n$ on $\mathcal{A}_1$. □

SA5.1.4 Bounding T_1, T_2, T_3

Lemma SA5.6 (Bounding T_1)**.** *There exists an event $\mathcal{A}_2$ whose probability converges to one such that on $\mathcal{A}_2$*

$$\begin{aligned}\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{\beta}_1\in\mathcal{H}_1,\boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell},\boldsymbol{\alpha}\in\mathcal{V}} &\big|\mathbb{E}_n[\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 - \boldsymbol{\alpha}))] \\ &- \mathbb{E}_n[\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))]\big| \lesssim h^d r_{1,n}(r_{1,n} + 2^\ell \mathfrak{r}_{2,n} + 2^L\mathfrak{r}_{2,n})2^L\mathfrak{r}_{2,n}.\end{aligned}$$

Proof of Lemma SA5.6. Note that

$$\begin{aligned}&\mathbb{E}_n\big[\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 - \boldsymbol{\alpha}))\big] \\ &\quad - \mathbb{E}_n[\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))] \\ &= \boldsymbol{\alpha}^\mathsf{T}\mathbb{E}_n[\psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\eta^{(2)}(\xi_{\boldsymbol{q},i})\boldsymbol{p}_i\boldsymbol{p}_i^\mathsf{T}](\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 - \boldsymbol{\alpha}) \\ &\lesssim h^d r_{1,n}(r_{1,n} + 2^\ell\mathfrak{r}_{2,n} + 2^L\mathfrak{r}_{2,n})2^L\mathfrak{r}_{2,n},\end{aligned}$$

where $\xi_{\boldsymbol{q},i}$ is between $\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 - \boldsymbol{\alpha})$ and $\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})$. The bound holds on the event

$$\mathcal{A}_2 := \big\{\sup\big\|\mathbb{E}_n[\psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\eta^{(2)}(\xi_{\boldsymbol{q},i})\boldsymbol{p}_i\boldsymbol{p}_i^\mathsf{T}]\big\|_\infty \leq h^d r_{1,n}\big\},$$

where the supremum is over $\boldsymbol{\beta}_1 \in \mathcal{H}_1$, $\boldsymbol{\beta}_2 \in \mathcal{H}_{2,\ell}$, $\boldsymbol{\alpha} \in \mathcal{V}$, $\boldsymbol{q} \in \mathcal{Q}$ and $\xi_{\boldsymbol{q},i}$ between $\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 - \boldsymbol{\alpha})$ and $\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})$. By the same argument as Lemmas SA3.12 and SA3.14, $\mathbb{P}\{\mathcal{A}_2\} \to 1$. □

Lemma SA5.7 (Bounding T_2)**.** *On $\mathcal{A}_1$, we have*

$$\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{\beta}_1\in\mathcal{H}_1,\boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell},\boldsymbol{\alpha}\in\mathcal{V}}\left|\mathbb{E}_n[\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})|\boldsymbol{x}_i]]-\boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+\boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\boldsymbol{\alpha}\right|$$
$$\lesssim 2^L\mathfrak{r}_{2,n}h^d(r_{1,n}+2^\ell\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^2,$$

and in addition for all $\boldsymbol{q}\in\mathcal{Q}$, $\boldsymbol{\alpha}\in\mathcal{V}$

$$|\boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\boldsymbol{\alpha}|\lesssim h^d\|\boldsymbol{\alpha}\|^2.$$

Proof of Lemma SA5.7. First, on $\mathcal{A}_1$ the largest eigenvalue of $\widetilde{\mathbf{Q}}_{\boldsymbol{q}}$ is bounded by h^d up to a constant factor (uniformly in $\boldsymbol{q}$):

$$\begin{aligned}
\lambda_{\max}\big(\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\big)&=\sup_{\|\boldsymbol{\alpha}\|=1}\boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\boldsymbol{\alpha}=\sup_{\|\boldsymbol{\alpha}\|=1}\mathbb{E}_n\big[(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i)^2\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))^2\big]\\
&\le\sup_{\|\boldsymbol{\alpha}\|=1}\mathbb{E}_n\big[(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i)^2\big|\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\big|\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))^2\big]\\
&\lesssim\sup_{\|\boldsymbol{\alpha}\|=1}\mathbb{E}_n\big[(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i)^2\big]
\end{aligned}$$

(because $\big|\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\big|\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))^2\lesssim 1$ by Assumptions SA2.4 and SA2.6)

$$\begin{aligned}
&=\sup_{\|\boldsymbol{\alpha}\|=1}\sum_{l=1}^{\kappa}\mathbb{E}_n\big[(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i)^2\mathbb{1}\{\boldsymbol{x}_i\in\delta_l\}\big]\\
&\lesssim\sup_{\|\boldsymbol{\alpha}\|=1}\sum_{l=1}^{\kappa}\left(\sum_{k=1}^{M_l}\alpha_{l,k}^2\right)\mathbb{E}_n\big[\mathbb{1}\{\boldsymbol{x}_i\in\delta_l\}\big]
\end{aligned}$$

(because $\sup_{\boldsymbol{x}\in\delta_l}(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}(\boldsymbol{x}))^2\lesssim\sum_{k=1}^{M_l}\alpha_{l,k}^2$, where $\{\alpha_{l,k}\}_{k=1}^{M_l}$ are the components of $\boldsymbol{\alpha}$ corresponding to the M_l basis functions supported on δ_l)

$$\begin{aligned}
&\le\sup_{\delta\in\Delta}\mathbb{E}_n\big[\mathbb{1}\{\boldsymbol{x}_i\in\delta\}\big]\sup_{\|\boldsymbol{\alpha}\|=1}\sum_{l=1}^{\kappa}\left(\sum_{k=1}^{M_l}\alpha_{l,k}^2\right)\\
&\lesssim\sup_{\delta\in\Delta}\mathbb{E}_n\big[\mathbb{1}\{\boldsymbol{x}_i\in\delta\}\big]\sup_{\|\boldsymbol{\alpha}\|=1}\|\boldsymbol{\alpha}\|^2=\sup_{\delta\in\Delta}\mathbb{E}_n\big[\mathbb{1}\{\boldsymbol{x}_i\in\delta\}\big].
\end{aligned}$$

Next, by Taylor expansion,

$$\begin{aligned}
&\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})|\boldsymbol{x}_i]\\
&\quad=\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i[\Psi(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}));\boldsymbol{q})-\Psi(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})]\\
&\qquad\times\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))\\
&\quad=\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\big[\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\big\{\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha})+\\
&\qquad\quad+(1/2)\eta^{(2)}(\xi_{\boldsymbol{q},i})(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))^2\big\}\\
&\qquad+(1/2)\Psi_2(\boldsymbol{x}_i,\tilde{\xi}_{\boldsymbol{q},i};\boldsymbol{q})\big\{\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))-\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\big\}^2\big]\\
&\qquad\times\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))
\end{aligned}$$

for some $\xi_{\boldsymbol{q},i}$ between $\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})$ and $\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha})$, $\tilde{\xi}_{\boldsymbol{q},i}$ is between $\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))$ and $\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))$. This gives

$$\mathbb{E}_n[\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})|\boldsymbol{x}_i]]=\boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)-\boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\boldsymbol{\alpha}+\mathrm{I}+\mathrm{II}+\mathrm{III},$$

where for some $\check{\xi}_{\boldsymbol{q},i}$ between $\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})$ and $\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha})$

$$\begin{aligned}
\mathrm{I} &:= \mathbb{E}_n\big[\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\eta^{(2)}(\check{\xi}_{\boldsymbol{q},i})(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))^2\big],\\
\mathrm{II} &:= \frac{1}{2}\mathbb{E}_n\big[\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(2)}(\xi_{\boldsymbol{q},i})\\
&\quad\times\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))^2\big],\\
\mathrm{III} &:= \frac{1}{2}\mathbb{E}_n\big[\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{p}_i\Psi_2(\boldsymbol{x}_i,\tilde{\xi}_{\boldsymbol{q},i};\boldsymbol{q})\{\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))-\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\}^2\\
&\quad\times\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))\big]
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{I} &\lesssim 2^L\mathfrak{r}_{2,n}h^d(r_{1,n}+2^\ell\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^2,\\
\mathrm{II} &\lesssim 2^L\mathfrak{r}_{2,n}h^d(r_{1,n}+2^\ell\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^2,\\
\mathrm{III} &\lesssim 2^L\mathfrak{r}_{2,n}h^d(r_{1,n}+2^\ell\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^2
\end{aligned}$$

on the event $\mathcal{A}_1$. □

Lemma SA5.8 (Bounding T_3)**.** *For the matrix $\widetilde{\mathbf{Q}}_{\boldsymbol{q}}$ defined in Eq. (SA5.16), we have the following bound on the event $\mathcal{A}_1$:*

$$\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{\beta}_1\in\mathcal{H}_1,\boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell},\boldsymbol{\alpha}\in\mathcal{V}}\big|\boldsymbol{\alpha}^\mathsf{T}(\widetilde{\mathbf{Q}}_{\boldsymbol{q}}-\bar{\mathbf{Q}}_{\boldsymbol{q}})(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)\big|\lesssim h^{m+d}2^L\mathfrak{r}_{2,n}(r_{1,n}+2^\ell\mathfrak{r}_{2,n}). \tag{SA5.35}$$

Proof of Lemma SA5.8. By the same logic as in Lemma SA3.11, we have on $\mathcal{A}_1$

$$\|\bar{\mathbf{Q}}_{\boldsymbol{q}}-\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\|_\infty\vee\|\bar{\mathbf{Q}}_{\boldsymbol{q}}-\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\|\lesssim h^m h^d$$

uniformly over $\boldsymbol{q}$ with probability approaching one. This gives (SA5.35), proving Lemma SA5.8. □

We have now proved the deferred lemmas, and the proof of Theorem SA5.1 is concluded.

Remark SA5.9 (Rate restrictions). The rates in the proof are determined by four restrictions: Eqs. (SA5.23), (SA5.26), (SA5.30) and (SA5.31). Equation (SA5.32) follows from Eqs. (SA5.23) and (SA5.30). Equation (SA5.28) follows from Eq. (SA5.23) and Eq. (SA5.26).

SA5.2 Proof: general case

We will state another version of the Bahadur Representation theorem, where $\theta\mapsto\rho(y,\eta(\theta);\boldsymbol{q})$ is not assumed to be convex. Here, we will use the additional complexity Assumption SA2.8.

The argument is almost the same as for Theorem SA5.1, so we will only describe the changes that need to be made.

The setup is the same as in the convex case except the definition of $\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})$ is replaced with

$$\begin{aligned}
&\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})\\
&\quad:=\rho(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2));\boldsymbol{q})-\rho(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}));\boldsymbol{q})\\
&\qquad-\big[\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2))-\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha}))\big]\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\\
&\quad=\int_{-\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\alpha}}^{0}\big[\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t);\boldsymbol{q})-\psi(y_i,\eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\big]
\end{aligned}$$

$$\times\, \eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}(\boldsymbol{\beta}_0(\boldsymbol{q}) + \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2) + t)\, \mathrm{d}t,$$

and (SA5.33) is changed to fit this definition.

Instead of (SA5.11), we have for any vector $\boldsymbol{\alpha}$

$$\mathbb{E}_n\big[\rho(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\widehat{\boldsymbol{\beta}}(\boldsymbol{q})); \boldsymbol{q}) - \rho(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}(\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\alpha})); \boldsymbol{q})\big] \leq 0$$

by the definition of $\widehat{\boldsymbol{\beta}}(\boldsymbol{q})$ as long as $\|\boldsymbol{\alpha}\|_\infty$ is small enough (so that $\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\alpha}$ satisfies the constraints). Accordingly, the definition of T_1 becomes

$$\begin{aligned}T_1 := \mathbb{E}_n\big[&(\eta(\boldsymbol{p}_i^\mathsf{T}\widehat{\boldsymbol{\beta}}(\boldsymbol{q})) - \eta(\boldsymbol{p}_i(\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\alpha}_{\boldsymbol{q}})))\psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\big]\\ &- \mathbb{E}_n\big[\boldsymbol{\alpha}_{\boldsymbol{q}}^\mathsf{T}\boldsymbol{p}_i\psi(y_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))\big].\end{aligned}$$

Lemma SA5.10 (Bounding variance)**.** *On $\mathcal{A}_1$ as in Lemma SA5.4, class $\mathcal{G}$ satisfies the following variance bound:*

$$\sup_{g \in \mathcal{G}} \mathbb{E}_n[\mathbb{V}[g(\boldsymbol{x}_i, y_i) \mid \boldsymbol{x}_i]] \lesssim 2^{2L}\mathfrak{r}_{2,n}^2 h^d (r_{1,n} + 2^\ell \mathfrak{r}_{2,n} + 2^L \mathfrak{r}_{2,n})^{(2\alpha)\wedge 1}.$$

Proof of Lemma SA5.10. This is proven by the same argument as in the proof of Theorem SA5.1. □

Lemma SA5.11 (Complexity of class $\mathcal{G}$)**.** *Class $\mathcal{G}$ with envelope $2^L\mathfrak{r}_{2,n}$ multiplied by a large enough constant satisfies the uniform entropy bound (SA1.3) with $A \lesssim (2^L\mathfrak{r}_{2,n})^{-1}$ and $V \lesssim \log^d n$.*

Note that A is not constant in this statement but it will not matter since $\log((2^L\mathfrak{r}_{2,n})^{-1}) \lesssim \log n$.

Proof of Lemma SA5.11. This is a directly assumed in Assumption SA2.8. □

Lemma SA5.12 (Uniform concentration in $\mathcal{G}$)**.** *On the event $\mathcal{A}_1$, (SA5.29) holds.*

Proof of Lemma SA5.12. This is proven by the same argument as in the proof of Theorem SA5.1. □

Lemma SA5.13. *For $\widetilde{\mathbf{Q}}_{\boldsymbol{q}} := \mathbb{E}_n\big[\boldsymbol{p}_i\boldsymbol{p}_i^\mathsf{T}\Psi_1(\boldsymbol{x}_i, \eta(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^\mathsf{T}\boldsymbol{\beta}_0(\boldsymbol{q}))^2\big]$, (SA5.35) holds.*

Proof of Lemma SA5.13. This is proven by the same argument as in the proof of Theorem SA5.1. □

Lemma SA5.14. *On $\mathcal{A}_1$, we have*

$$\begin{aligned}\sup_{\boldsymbol{q}\in\mathcal{Q}, \boldsymbol{\beta}_1\in\mathcal{H}_1, \boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell}, \boldsymbol{\alpha}\in\mathcal{V}} &\big|\mathbb{E}_n\big[\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha})|\boldsymbol{x}_i]\big] - \boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2) + \boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\boldsymbol{\alpha}\big|\\ &\lesssim 2^L\mathfrak{r}_{2,n}h^d(r_{1,n} + 2^\ell\mathfrak{r}_{2,n} + 2^L\mathfrak{r}_{2,n})^2,\end{aligned}$$

and in addition for all $\boldsymbol{q} \in \mathcal{Q}$, $\boldsymbol{\alpha} \in \mathcal{V}$

$$|\boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\boldsymbol{\alpha}| \lesssim h^d\|\boldsymbol{\alpha}\|^2.$$

Proof of Lemma SA5.14. First, on $\mathcal{A}_1$ the largest eigenvalue of $\widetilde{\mathbf{Q}}_{\boldsymbol{q}}$ is bounded by h^d up to a constant factor (uniformly in $\boldsymbol{q}$), which is proven in Lemma SA5.7.

Next, by the Taylor expansion,

$$\begin{aligned}
&\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})|\boldsymbol{x}_i]\\
&\quad=\int_{-\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\alpha}}^{0}\big[\Psi(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t);\boldsymbol{q})-\Psi(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\big]\\
&\qquad\times\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)\,\mathrm{d}t\\
&\quad=\int_{-\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\alpha}}^{0}\big[\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\{\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}))(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)\\
&\qquad\quad+\frac{1}{2}\eta^{(2)}(\xi_{\boldsymbol{q},i,t})(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)^2\}\\
&\qquad+\frac{1}{2}\Psi_2(\boldsymbol{x}_i,\tilde{\xi}_{\boldsymbol{q},i,t};\boldsymbol{q})\{\eta(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)-\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}))\}^2\big]\\
&\qquad\times\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)\,\mathrm{d}t
\end{aligned}$$

for some $\xi_{\boldsymbol{q},i,t}$ between $\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})$ and $\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t$, $\tilde{\xi}_{\boldsymbol{q},i,t}$ between $\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}))$ and $\eta(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)$. This gives

$$\mathbb{E}_n[\mathbb{E}[\delta_{\boldsymbol{q},i}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\alpha})|\boldsymbol{x}_i]]=\boldsymbol{\alpha}^{\mathsf{T}}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)-\boldsymbol{\alpha}^{\mathsf{T}}\widetilde{\mathbf{Q}}_{\boldsymbol{q}}\boldsymbol{\alpha}+\mathrm{I}+\mathrm{II}+\mathrm{III},$$

where for some $\check{\xi}_{\boldsymbol{q},i,t}$ between $\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})$ and $\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t$ again

$$\begin{aligned}
\mathrm{I}&:=\mathbb{E}_n\Big[\int_{-\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\alpha}}^{0}\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}))\eta^{(2)}(\check{\xi}_{\boldsymbol{q},i})(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)^2\,\mathrm{d}t\Big],\\
\mathrm{II}&:=\frac{1}{2}\mathbb{E}_n\Big[\int_{-\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\alpha}}^{0}\Psi_1(\boldsymbol{x}_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(2)}(\xi_{\boldsymbol{q},i,t})\\
&\qquad\times\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)^2\Big]\,\mathrm{d}t,\\
\mathrm{III}&:=\frac{1}{2}\mathbb{E}_n\Big[\int_{-\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\alpha}}^{0}\Psi_2(\boldsymbol{x}_i,\tilde{\xi}_{\boldsymbol{q},i,t};\boldsymbol{q})\{\eta(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)-\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}))\}^2\\
&\qquad\times\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)+t)\,\mathrm{d}t\Big]
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{I}&\lesssim 2^L\mathfrak{r}_{2,n}h^d(r_{1,n}+2^\ell\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^2,\\
\mathrm{II}&\lesssim 2^L\mathfrak{r}_{2,n}h^d(r_{1,n}+2^\ell\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^2,\\
\mathrm{III}&\lesssim 2^L\mathfrak{r}_{2,n}h^d(r_{1,n}+2^\ell\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})^2
\end{aligned}$$

on the event $\mathcal{A}_1$. □

Lemma SA5.15. *There exists an event $\mathcal{A}_2$ whose probability converges to one such that on $\mathcal{A}_2$*

$$\begin{aligned}
\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{\beta}_1\in\mathcal{H}_1,\boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell},\boldsymbol{\alpha}\in\mathcal{V}}&\big|\mathbb{E}_n[\psi(y_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\\
&\times(\eta(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2))-\eta(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha})))]\\
&-\mathbb{E}_n[\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}_i\psi(y_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}))]\big|\lesssim h^dr_{1,n}(r_{1,n}+2^\ell\mathfrak{r}_{2,n}+2^L\mathfrak{r}_{2,n})2^L\mathfrak{r}_{2,n}.
\end{aligned}$$

Proof of Lemma SA5.15. By the Taylor expansion,

$$\begin{aligned}
&\big|\mathbb{E}_n[\psi(y_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})(\eta(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2))-\eta(\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha})))]\\
&\quad-\mathbb{E}_n[\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}_i\psi(y_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}))]\big|\\
&\quad=\big|\mathbb{E}_n\big[\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\alpha}\psi(y_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\{\eta^{(2)}(\xi_{\boldsymbol{q},i})\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha})+(1/2)\eta^{(2)}(\tilde{\xi}_{\boldsymbol{q},i})\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\alpha}\}\big]\big|\\
&\quad\lesssim h^d r_{1,n}(r_{1,n}+2^{\ell}\mathfrak{r}_{2,n}+2^{L}\mathfrak{r}_{2,n})2^{L}\mathfrak{r}_{2,n},
\end{aligned}$$

where $\xi_{\boldsymbol{q},i}$ is between $\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha})$ and $\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})$, $\tilde{\xi}_{\boldsymbol{q},i}$ between $\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)$ and $\boldsymbol{p}_i^{\mathsf{T}}(\boldsymbol{\beta}_0(\boldsymbol{q})+\boldsymbol{\beta}_1+\boldsymbol{\beta}_2-\boldsymbol{\alpha})$. The bound holds on the event $\mathcal{A}_2:=\mathcal{A}_2'\cap\mathcal{A}_2''$, where

$$\mathcal{A}_2':=\Big\{\sup_{\boldsymbol{\beta}_1\in\mathcal{H}_1,\boldsymbol{\beta}_2\in\mathcal{H}_{2,\ell},\boldsymbol{\alpha}\in\mathcal{V},\boldsymbol{q}\in\mathcal{Q}}\big\|\mathbb{E}_n[\psi(y_i,\eta(\boldsymbol{p}_i^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q}));\boldsymbol{q})\eta^{(2)}(\xi_{\boldsymbol{q},i})\boldsymbol{p}_i\boldsymbol{p}_i^{\mathsf{T}}]\big\|_\infty\le h^d r_{1,n}\Big\},$$

and $\mathcal{A}_2''$ is defined the same way as $\mathcal{A}_2'$ with $\xi_{\boldsymbol{q},i}$ replaced by $\tilde{\xi}_{\boldsymbol{q},i}$.

By the same argument as Lemma SA3.12, $\mathbb{P}\{\mathcal{A}_2\}\to 1$. □

SA5.3 Rates of convergence

Corollary SA5.16 (Uniform rate of convergence)**.**

(a) If the conditions of Theorem SA5.1(a) hold, then

$$\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{x}\in\mathcal{X}}\big|\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})-\mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big|\lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|}\bigg[\Big(\frac{\log^d n}{nh^d}\Big)^{1/2}\log n+h^m\bigg] \tag{SA5.36}$$

(b) If the conditions of Theorem SA5.1(a) hold and

$$[\log n]^{(d+1)/(\alpha\wedge 1/2)+d}=O(nh^d),\quad h^{(\alpha\wedge1/2)m}\log^{d/2}n=O(1), \tag{SA5.37}$$

then

$$\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{x}\in\mathcal{X}}\big|\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})-\mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big|\lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|}\bigg[\Big(\frac{\log n}{nh^d}\Big)^{1/2}+h^m\bigg]. \tag{SA5.38}$$

(c) If the conditions of Theorem SA5.1(b) hold and

$$[\log n]^{(d+1)/\alpha+d}=O(nh^d),\quad h^{\alpha m}\log^{d/2}n=O(1),$$

then (SA5.38) is also true.

Proof. By Theorem SA5.1, Lemma SA3.12 and triangle inequality,

$$\begin{aligned}
&\sup_{\boldsymbol{q}\in\mathcal{Q},\boldsymbol{x}\in\mathcal{X}}\big|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big|\\
&\lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|}\bigg[\Big(\frac{\log(1/h)}{nh^d}\Big)^{1/2}+\Big(\frac{\log^d n}{nh^d}\Big)^{1/2+\alpha/2\wedge1/4}\log n+h^{(\alpha\wedge1/2)m}\Big(\frac{\log^{d+1}n}{nh^d}\Big)^{1/2}+h^m\bigg].
\end{aligned}$$

Using $\log(1/h)\lesssim\log n$ and simplifying the right-hand size, we obtain (SA5.36). Additional restrictions (SA5.37) allow us to get a slightly stronger result (SA5.38). □

Corollary SA5.17 (Mean square rate of convergence)**.**

(a) *If the conditions of Theorem SA5.1(a) hold and*

$$[\log n]^{(d+2)/(\alpha\wedge 1/2)+d} = o(nh^d), \quad h^{(\alpha\wedge 1/2)m}\log^{(d+1)/2} n = o(1), \tag{SA5.39}$$

then

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\Big(\int_{\mathcal{X}}\big|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big|^2 f_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\Big)^{1/2} \lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|}\Big[\frac{1}{\sqrt{nh^d}} + h^m\Big]. \tag{SA5.40}$$

(b) *If the conditions of Theorem SA5.1(b) hold and*

$$[\log n]^{(d+2)/\alpha+d} = o(nh^d), \quad h^{\alpha m}\log^{(d+1)/2} n = o(1),$$

then (SA5.40) *is also true.*

Proof. To prove (SA5.40), note that

$$\begin{aligned}
&\sup_{\boldsymbol{q}}\int_{\mathcal{X}}\big|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(\boldsymbol{q})\big|^2 f_X(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&\quad= \sup_{\boldsymbol{q}}\big(\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big)^{\mathsf{T}}\mathbb{E}[\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}]\big(\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big) \overset{\text{(a)}}{\lesssim} h^{d-2|\boldsymbol{v}|}\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\|^2,
\end{aligned}$$

where inequality (a) is true because the largest eigenvalue of $\mathbb{E}[\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}_i)\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x}_i)^{\mathsf{T}}]$ is bounded from above by $h^{d-2|\boldsymbol{v}|}$ up to a multiplicative coefficient, which is proven by the same argument as for $\boldsymbol{v} = 0$ in Lemma SA3.11 in combination with Assumption SA2.2.

It is left to prove

$$\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\| \lesssim_{\mathbb{P}} \frac{1}{h^d\sqrt{n}}. \tag{SA5.41}$$

By the triangle inequality,

$$\begin{aligned}
\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q})\big\| &\le \big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) + \mathbf{Q}_{0,\boldsymbol{q}}^{-1}\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})]\big\|\\
&\quad+ \big\|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})]\big\|.
\end{aligned} \tag{SA5.42}$$

To bound the second term in (SA5.42) on the right-hand side, consider the expectation

$$\begin{aligned}
&\mathbb{E}\big[\big\|\mathbb{E}_n\big[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\big]\big\|^2\big]\\
&\quad= \frac{1}{n^2}\sum_{i=1}^n\mathbb{E}\big[\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))^2\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})^2\|\boldsymbol{p}(\boldsymbol{x}_i)\|^2\big]\\
&\quad= \frac{1}{n^2}\sum_{i=1}^n\mathbb{E}\big[\sigma_{\boldsymbol{q}}^2(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))^2\|\boldsymbol{p}(\boldsymbol{x}_i)\|^2\big] \overset{\text{(a)}}{\lesssim} \frac{1}{n^2}\sum_{i=1}^n 1 = \frac{1}{n},
\end{aligned}$$

where in (a) we used uniform boundedness of $\sigma_{\boldsymbol{q}}^2(\boldsymbol{x})$ by Assumption SA2.4(ii), uniform boundedness of $\mu_0(\boldsymbol{x},\boldsymbol{q})$ and $\|\boldsymbol{p}(\boldsymbol{x})\|$. By Markov's inequality and Lemma SA3.11, this immediately implies

$$\big\|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\mathbb{E}_n\big[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\big]\big\| \lesssim_{\mathbb{P}} \|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\|\cdot\frac{1}{\sqrt{n}} \lesssim \frac{1}{h^d\sqrt{n}}.$$

Concerning the first term in (SA5.42), it is proven in Theorem SA5.1 (see Eq. (SA5.10)) that

$$\big\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \boldsymbol{\beta}_0(\boldsymbol{q}) + \mathbf{Q}_{0,\boldsymbol{q}}^{-1}\mathbb{E}_n\big[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\big]\big\|_\infty \lesssim_{\mathbb{P}} r_{\mathtt{BR}}$$

for $r_{\mathtt{BR}}$ defined in (SA5.3), so that

$$\begin{aligned}&\left\|\widehat{\boldsymbol{\beta}}(\boldsymbol{q})-\boldsymbol{\beta}_0(\boldsymbol{q})+\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\mathbb{E}_n\left[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\right]\right\|\\&\lesssim_{\mathbb{P}}\sqrt{K}r_{\mathtt{BR}}\lesssim\frac{r_{\mathtt{BR}}}{h^{d/2}}\overset{\text{(a)}}{=}o\left(\frac{1}{h^d\sqrt{n}}\right),\end{aligned}$$

where in equality (a) we used $r_{\mathtt{BR}}=o\big(1/\sqrt{nh^d}\big)$ under the assumptions. This concludes the proof of (SA5.41). □

SA6 Strong approximation

Section SA6.1 collects results that may be of independent theoretical interest. Specifically, we use the conditional Strassen's theorem Theorem SA6.1 to prove conditional Yurinskii coupling in d-norm, Theorem SA6.2. We then use it to prove Lemma SA6.3, a version of Gaussian approximation for a K-dimensional empirical process. Section SA6.2 proves the main result of this section, Theorem SA6.4, which is essentially a corollary of Lemma SA6.3 after some additional technical work done in Lemmas SA6.6 to SA6.8.

SA6.1 Yurinskii coupling

The three theorems, and their proofs, in this subsection are self-contained, and hence all variables, functions, and stochastic processes, should be treated as defined within each of the theorems and their proofs, and independently of all other statements elsewhere in the supplemental appendix.

The following theorem is due to [15]. We use the statement from [10] making it explicit that the supremum over Borel sets may not be a random variable (a direct proof may also be found in that work). Let (S,d) be a Polish space (where d is its metric), and $\mathcal{B}(S)$ its Borel sigma-algebra.

Theorem SA6.1 (Conditional Strassen's theorem)**.** *Let X be a random variable defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ and with values in some Polish (S,d). Let $\mathcal{J}$ be a contably generated sub-sigma algebra of $\mathcal{F}$ and assume that this probability space is rich enough: there exists a random variable U that is independent of the sigma-algebra $\mathcal{J}\vee\sigma(X)$. Let $\mathcal{B}(S)\times\Omega\ni(A,\omega)\mapsto G(A|\mathcal{J})(\omega)$ be a regular conditional distribution on $\mathcal{B}(S)$, i. e., for each $A\in\mathcal{B}(S)$, $G(A|\mathcal{J})$ is $\mathcal{J}$-measurable, and for each $\omega\in\Omega$, $G(\cdot|\mathcal{J})(\omega)$ is a probability measure on $\mathcal{B}(S)$. Suppose that for some nonnegative numbers α and β*

$$\mathbb{E}^*\Big[\sup_{A\in\mathcal{B}}\{\mathbb{P}\{X\in A\,|\,\mathcal{J}\}-G(A^\alpha|\mathcal{J})\}\Big]\leq\beta,$$

where $\mathbb{E}^*$ denotes outer expectation. Then on this probability space there exists an S-valued random element Y such that $G(\cdot|\mathcal{J})$ is its regular conditional distribution given $\mathcal{J}$ and $\mathbb{P}\{d(X,Y)>\alpha\}\leq\beta$.

The following theorem is a conditional version of Lemma 39 in [2]; its proof carefully leverages Theorem SA6.1. See also [8] for a related, but different, conditional Yurinskii's coupling result.

Theorem SA6.2 (Conditional Yurinskii coupling, d-norm)**.** *Let a random vector $\boldsymbol{C}$ with values in $\mathbb{R}^m$, sequence of random vectors $\{\boldsymbol{\xi}_i\}_{i=1}^n$ with values in $\mathbb{R}^k$ and sequence of random vectors $\{\boldsymbol{g}_i\}_{i=1}^n$ with values in $\mathbb{R}^k$ be defined on the same probability space and be such that the all the $2n$ vectors $\{\boldsymbol{\xi}_i\}_{i=1}^n\cup\{\boldsymbol{g}_i\}_{i=1}^n$ are independent conditionally on $\boldsymbol{C}$, are mean zero conditionally on $\boldsymbol{C}$ and for each $i\in\{1,\dots,n\}$ the distribution of $\boldsymbol{g}_i$ conditionally on $\boldsymbol{C}$ is $\mathcal{N}(0,\mathbb{V}[\boldsymbol{\xi}_i\,|\,\boldsymbol{C}])$. Assume* *that this probability space is rich enough: there exists a random variable U that is independent of $\{\boldsymbol{\xi}_i\}_{i=1}^n \cup \{\boldsymbol{g}_i\}_{i=1}^n$. Denote*

$$\boldsymbol{S} := \boldsymbol{\xi}_1 + \ldots + \boldsymbol{\xi}_n, \quad \boldsymbol{T} := \boldsymbol{g}_1 + \ldots + \boldsymbol{g}_n,$$

and let

$$\beta := \sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{\xi}_i\|^2 \|\boldsymbol{\xi}_i\|_d\big] + \sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{g}_i\|^2 \|\boldsymbol{g}_i\|_d\big]$$

be finite. Then for each $\delta > 0$, on this probability space there exists a random vector $\boldsymbol{T}'$ such that $\mathbb{P}_{\boldsymbol{T}|\sigma(\boldsymbol{C})}(\cdot)$ is its regular conditional distribution given $\sigma(\boldsymbol{C})$, and

$$\mathbb{P}\big\{\|\boldsymbol{S} - \boldsymbol{T}'\|_d > 3\delta\big\} \leq \min_{t \geq 0} \left(2\mathbb{P}\{\|\boldsymbol{Z}\|_d > t\} + \frac{\beta}{\delta^3} t^2\right),$$

where $\boldsymbol{Z} \sim \mathcal{N}(0, I_k)$.

Proof. By the conditional Strassen's theorem, it is enough to show

$$\mathbb{E}^*\Big[\sup_{A \in \mathcal{B}(\mathbb{R}^k)} \big(\mathbb{P}\{\boldsymbol{S} \in A \,|\, \sigma(\boldsymbol{C})\} - \mathbb{P}_{\boldsymbol{T}|\sigma(\boldsymbol{C})}(A^{3\delta})\big)\Big] \leq \min_{t \geq 0} \left(2\mathbb{P}\{\|\boldsymbol{Z}\|_d > t\} + \frac{\beta}{\delta^3} t^2\right),$$

or, equivalently, for any $t > 0$

$$\mathbb{E}^*\Big[\sup_{A \in \mathcal{B}(\mathbb{R}^k)} \big(\mathbb{P}\{\boldsymbol{S} \in A \,|\, \boldsymbol{C}\} - \mathbb{P}\{\boldsymbol{T} \in A^{3\delta} \,|\, \boldsymbol{C}\}\big)\Big] \leq 2\mathbb{P}\{\|bZ\|_d > t\} + \frac{\beta}{\delta^3} t^2. \tag{SA6.1}$$

Fix $t > 0$ and $A \in \mathcal{B}(\mathbb{R}^k)$. Let $f\colon \mathbb{R}^k \to \mathbb{R}$ be the same as in the proof of Lemma 39 in [2], namely, it is such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^k$,

$$\begin{aligned}
&\big|f(\boldsymbol{x} + \boldsymbol{y}) - f(\boldsymbol{x}) - \boldsymbol{y}^\mathsf{T} \nabla f(\boldsymbol{x}) - (1/2)\boldsymbol{y}^\mathsf{T} \nabla^2 f(\boldsymbol{x}) \boldsymbol{y}\big| \leq \frac{\|\boldsymbol{y}\|^2 \|\boldsymbol{y}\|_d}{\sigma^2 \delta},\\
&(1 - \epsilon) 1\{\boldsymbol{x} \in A\} \leq f(\boldsymbol{x}) \leq \epsilon + (1 - \epsilon) 1\{\boldsymbol{x} \in A^{3\delta}\},
\end{aligned}$$

with $\sigma := \delta/t$ and $\epsilon := \mathbb{P}\{\|\boldsymbol{Z}\|_d > t\}$.

Then note that

$$\begin{aligned}
\mathbb{P}\{\boldsymbol{S} \in A \,|\, \boldsymbol{C}\} &= \mathbb{E}\big[\mathbb{1}\{\boldsymbol{S} \in A\} - f(\boldsymbol{S}) \,\big|\, \boldsymbol{C}\big] + \mathbb{E}\big[f(\boldsymbol{S}) - f(\boldsymbol{T}) \,\big|\, \boldsymbol{C}\big] + \mathbb{E}[f(\boldsymbol{T}) \,|\, \boldsymbol{C}]\\
&\leq \epsilon \mathbb{E}\big[\mathbb{1}\{\boldsymbol{S} \in A\} \,\big|\, \boldsymbol{C}\big] + \mathbb{E}\big[f(\boldsymbol{S}) - f(\boldsymbol{T}) \,\big|\, \boldsymbol{C}\big] + \epsilon + (1 - \epsilon)\mathbb{E}\big[\mathbb{1}\{\boldsymbol{T} \in A^{3\delta}\} \,\big|\, \boldsymbol{C}\big]\\
&\leq 2\epsilon + \mathbb{E}\big[\mathbb{1}\{\boldsymbol{T} \in A^{3\delta}\} \,\big|\, \boldsymbol{C}\big] + \mathbb{E}\big[f(\boldsymbol{S}) - f(\boldsymbol{T}) \,\big|\, \boldsymbol{C}\big].
\end{aligned} \tag{SA6.2}$$

Now we bound $\mathbb{E}[f(\boldsymbol{S}) - f(\boldsymbol{T}) \,|\, \boldsymbol{C}]$:

$$\begin{aligned}
\mathbb{E}[f(\boldsymbol{S}) - f(\boldsymbol{T}) \,|\, \boldsymbol{C}] &= \sum_{i=1}^n \mathbb{E}[f(\boldsymbol{X}_i + \boldsymbol{Y}_i) - f(\boldsymbol{X}_i + \boldsymbol{W}_i) \,|\, \boldsymbol{C}]\\
&\leq \sum_{i=1}^n \mathbb{E}\left[f(\boldsymbol{X}_i) + \boldsymbol{Y}_i^\mathsf{T} \nabla f(\boldsymbol{X}_i) + \frac{1}{2} \boldsymbol{Y}_i^\mathsf{T} \nabla^2 f(\boldsymbol{X}_i) \boldsymbol{Y}_i + \frac{\|\boldsymbol{Y}_i\|^2 \|\boldsymbol{Y}_i\|_d}{\sigma^2 \delta} \,\middle|\, \boldsymbol{C}\right]\\
&\quad - \sum_{i=1}^n \mathbb{E}\left[f(\boldsymbol{X}_i) + \boldsymbol{W}_i^\mathsf{T} \nabla f(\boldsymbol{X}_i) + \frac{1}{2} \boldsymbol{W}_i^\mathsf{T} \nabla^2 f(\boldsymbol{X}_i) \boldsymbol{W}_i - \frac{\|\boldsymbol{W}_i\|^2 \|\boldsymbol{W}_i\|_d}{\sigma^2 \delta} \,\middle|\, \boldsymbol{C}\right]
\end{aligned}$$

$$\overset{(a)}{=} \sum_{i=1}^{n} \mathbb{E}[\frac{\|\boldsymbol{Y}_i\|^2\|\boldsymbol{Y}_i\|_d + \|\boldsymbol{W}_i\|^2\|\boldsymbol{W}_i\|_d}{\sigma^2\delta} \,|\, \boldsymbol{C}] \text{ a. s.}$$

for $\boldsymbol{X}_i := \boldsymbol{\xi}_1 + \ldots + \boldsymbol{\xi}_{i-1} + \boldsymbol{g}_{i+1} + \ldots + \boldsymbol{g}_n$, $\boldsymbol{Y}_i := \boldsymbol{\xi}_i$, $\boldsymbol{W}_i := \boldsymbol{g}_i$, in (a) we used the conditional independence of the family $\{\boldsymbol{\xi}_i\}_{i=1}^n \cup \{\boldsymbol{g}_i\}_{i=1}^n$, that they are conditionally mean zero and the equality of the corresponding conditional second moments.

We conclude that almost surely

$$\sup_{A\in\mathcal{B}(\mathbb{R}^k)} \big(\mathbb{P}\{\boldsymbol{S} \in A \,|\, \boldsymbol{C}\} - \mathbb{P}\{\boldsymbol{T} \in A^{3\delta} \,|\, \boldsymbol{C}\}\big) \le 2\epsilon + \sum_{i=1}^{n} \mathbb{E}\Big[\frac{\|\boldsymbol{\xi}_i\|^2\|\boldsymbol{\xi}_i\|_d + \|\boldsymbol{g}_i\|^2\|\boldsymbol{g}_i\|_d}{\sigma^2\delta} \,\Big|\, \boldsymbol{C}\Big].$$

By the definition of outer expectation, this implies

$$\mathbb{E}^*\Big[\sup_{A\in\mathcal{B}(\mathbb{R}^k)} \big(\mathbb{P}\{\boldsymbol{S} \in A \,|\, \boldsymbol{C}\} - \mathbb{P}\{\boldsymbol{T} \in A^{3\delta} \,|\, \boldsymbol{C}\}\big)\Big] \le 2\epsilon + \frac{\beta}{\sigma^2\delta},$$

which is (SA6.1). □

The following lemma generalizes Lemma 36 in [2], and also builds on the argument for Lemma SA27 in [7].

Lemma SA6.3 (Yurinskii coupling: K-dimensional process)**.** *Let $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$ be a random sample, where $\boldsymbol{x}_i$ has compact support $\mathcal{X} \subset \mathbb{R}^d$, $y_i \in \mathcal{Y} \subset \mathbb{R}$ is a scalar. Also let $\mathcal{Q} \subseteq \mathbb{R}^{d_{\mathcal{Q}}}$ be a fixed compact set.*

Let $A_n\colon \mathcal{Q} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a Borel measurable function satisfying $\sup_{\boldsymbol{q}\in\mathcal{Q}}|A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i)| \le \bar{A}_n(\boldsymbol{x}_i, y_i)$, where $\bar{A}_n(\boldsymbol{x}_i, y_i)$ is a Borel measurable envelope, $\mathbb{E}[A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i) \,|\, \boldsymbol{x}_i] = 0$ for all $\boldsymbol{q} \in \mathcal{Q}$, $\sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}[|\bar{A}(\boldsymbol{x}_i, y_i)|^\nu \,|\, \boldsymbol{x}_i = \boldsymbol{x}] \le \mu_n < \infty$ for some $\nu \ge 3$ with $\mu_n \gtrsim 1$ and $\log \mu_n \lesssim \log n$, which satisfies the Lipschitz condition

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}\big[|A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i) - A_n(\tilde{\boldsymbol{q}}, \boldsymbol{x}_i, y_i)|^2 \,\big|\, \boldsymbol{x}_i = \boldsymbol{x}\big] \lesssim \|\boldsymbol{q} - \tilde{\boldsymbol{q}}\|$$

for all $\boldsymbol{q}, \tilde{\boldsymbol{q}} \in \mathcal{Q}$. Also, the (regular) conditional variance $\mathbb{E}[A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i)^2 \,|\, \boldsymbol{x}_i = \boldsymbol{x}]$ is continuous in $\boldsymbol{x} \in \mathcal{X}$. Moreover, assume that the class of functions $\{(\boldsymbol{x}, y) \mapsto A_n(\boldsymbol{q}, \boldsymbol{x}, y)\colon \boldsymbol{q} \in \mathcal{Q}\}$ is VC-subgraph with an index bounded from above by a constant not depending on n.

Let $\boldsymbol{b}(\cdot)$ be a Borel measurable function $\mathcal{X} \to \mathbb{R}^K$ (where $K = K_n$ is some sequence of positive integers tending to infinity and satisfying $\log K \lesssim \log n$) such that $\sup_{\boldsymbol{x}\in\mathcal{X}}\|\boldsymbol{b}(\boldsymbol{x})\| \le \zeta_K$ and the probability of the event $\mathcal{A} := \{\sup_{\|\boldsymbol{\alpha}\|=1} \mathbb{E}_n\big[(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{b}(\boldsymbol{x}_i))^2\big] \le C_{\mathrm{Gr}}\}$ approaches one, where C_{Gr} is some positive constant. Assume ζ_K satisfies $1/\zeta_K \lesssim 1$, $|\log \zeta_K| \lesssim \log n$.

Let $r_{n,\mathrm{yur}} = r_{\mathtt{YUR}} \to 0$ be a sequence of positive numbers satisfying

$$\Big(\frac{\zeta_K^3 \mu_n^{3/\nu}}{\sqrt{n}}\Big)^{\frac{1}{3+2d_{\mathcal{Q}}}} \sqrt{\log n} + \frac{\zeta_K \mu_n^{1/\nu}}{n^{1/2-1/\nu}} \log n = o(r_{\mathtt{YUR}}). \tag{SA6.3}$$

Assume also that the probability space is rich enough, and denote

$$\boldsymbol{G}_n(\boldsymbol{q}) := \sqrt{n}\mathbb{E}_n[A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i)\boldsymbol{b}(\boldsymbol{x}_i)].$$

Then, on the same probability space, there exists a K-dimensional, conditionally on $\mathbf{X}_n$ mean-zero Gaussian, process $\boldsymbol{Z}_n(\boldsymbol{q})$ on $\mathcal{Q}$ with a. s. continuous trajectories, satisfying

$$\mathbb{E}\big[\boldsymbol{G}_n(\boldsymbol{q})\boldsymbol{G}_n(\tilde{\boldsymbol{q}})^\mathsf{T} \,\big|\, \mathbf{X}_n\big] = \mathbb{E}\big[\boldsymbol{Z}_n(\boldsymbol{q})\boldsymbol{Z}_n(\tilde{\boldsymbol{q}})^\mathsf{T} \,\big|\, \mathbf{X}_n\big], \quad \boldsymbol{q}, \tilde{\boldsymbol{q}} \in \mathcal{Q};$$

$$\sup_{\boldsymbol{q}\in\mathcal{Q}} \|\boldsymbol{G}_n(\boldsymbol{q}) - \boldsymbol{Z}_n(\boldsymbol{q})\|_\infty = o_{\mathbb{P}}(r_{\mathtt{YUR}}).$$

Moreover, if $\bar{A}_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i)$ is σ^2-sub-Gaussian, then (SA6.3) can be replaced with

$$\left(\frac{\zeta_K^3}{\sqrt{n}}\right)^{\frac{1}{3+2d_{\mathcal{Q}}}} \sqrt{\log n} + \frac{\zeta_K}{\sqrt{n}} \log^{3/2} n = o(r_{\mathtt{YUR}}).$$

Proof. Let $\mathcal{Q}_n^\delta := \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{|\mathcal{Q}_n^\delta|}\}$ be an internal δ_n-covering of $\mathcal{Q}$ with respect to the 2-norm $\|\cdot\|$ of cardinality $|\mathcal{Q}_n^\delta| \lesssim 1/\delta_n^{d_{\mathcal{Q}}}$, where δ_n is chosen later. Denote $\pi_n^\delta \colon \mathcal{Q} \to \mathcal{Q}$ a sequence of projections associated with this covering: it maps each point in $\mathcal{Q}$ to the center of the ball containing this point (if such a ball is not unique, choose one by an arbitrary rule).

Strategy The plan of attack is to

1. show that $\boldsymbol{G}_n(\boldsymbol{q})$ does not deviate too much in sup-norm from its projected version, i. e. bound the tails of $\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\boldsymbol{G}_n(\boldsymbol{q}) - \boldsymbol{G}_n \circ \pi_n^\delta(\boldsymbol{q})\|_\infty$,
2. apply Yurinskii coupling to the finite-dimensional vector $\big(\boldsymbol{G}_n \circ \pi_n^\delta(\boldsymbol{q})\big)_{\boldsymbol{q}\in\mathcal{Q}_n^\delta}$ and obtain a conditionally on $\mathbf{X}_n$ Gaussian vector $\boldsymbol{Z}_n^\delta$ with the right structure that is close enough, i. e. with a bound on the tails of $\|\boldsymbol{G}_n \circ \pi_n^\delta - \boldsymbol{Z}_n^\delta\|_\infty$,
3. extend this conditionally Gaussian vector to a K-dimensional conditionally Gaussian process $\boldsymbol{Z}_n$,
4. and finally show that $\boldsymbol{Z}_n(\boldsymbol{q})$ does not deviate too much from its projected version, i. e. bound the tails of $\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\boldsymbol{Z}_n(\boldsymbol{q}) - \boldsymbol{Z}_n \circ \pi_n^\delta(\boldsymbol{q})\|_\infty$.

If we complete these steps, it will prove the theorem by the triangle inequality.

Discretization of G_n Consider the class of functions

$$\mathcal{G}_n' := \{\mathcal{X} \times \mathcal{Y} \ni (\boldsymbol{x}, y) \mapsto A_n(\boldsymbol{q}, \boldsymbol{x}, y) b_l(\boldsymbol{x}) \colon 1 \le l \le K, \boldsymbol{q} \in \mathcal{Q}\}$$

with envelope $\zeta_K \bar{A}_n(X)$. Since $\{A_n(\boldsymbol{q}, \boldsymbol{x}, y)\}$ is a VC class with $O(1)$ index and envelope $\bar{A}_n(\boldsymbol{x}, y)$, $\mathcal{G}_n'$ satisfies the uniform entropy bound (SA1.3) with $A \lesssim K$ and $V \lesssim 1$.

Next, consider the class of functions

$$\mathcal{G}_n^\delta := \{\mathcal{X} \times \mathcal{Y} \ni (\boldsymbol{x}, y) \mapsto \big(A_n(\boldsymbol{q}, \boldsymbol{x}, y) - A_n(\tilde{\boldsymbol{q}}, \boldsymbol{x}, y)\big) b_l(\boldsymbol{x}) \colon 1 \le l \le K, \boldsymbol{q}, \tilde{\boldsymbol{q}} \in \mathcal{Q}, \|\boldsymbol{q} - \tilde{\boldsymbol{q}}\| \lesssim \delta_n\}$$

with envelope $(\boldsymbol{x}, y) \mapsto 2\zeta_K \bar{A}_n(\boldsymbol{x}, y)$. Using Lemma SA3.4, we get that this class satisfies the uniform entropy bound (SA1.3) with $A \lesssim K$ and $V \lesssim 1$.

Now we apply Lemma SA3.6 conditionally on $\mathbf{X}_n$ on $\mathcal{A}$ with $\|F\|_{\bar{\mathbb{P}},2} \le 2\zeta_K \mu_n^{1/\nu}$ since

$$\|\bar{A}_n(\boldsymbol{x}, y)\|_{\bar{\mathbb{P}},2}^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\bar{A}_n(\boldsymbol{x}_i, y_i)^2 \,|\, \boldsymbol{x}_i] \le \frac{1}{n}\sum_{i=1}^n \mu_n^{2/\nu} = \mu_n^{2/\nu},$$

$\|M\|_{\mathbb{P},2} \le 2\zeta_K (\mu_n n)^{1/\nu}$ since

$$\mathbb{E}\Big[\Big(\max_{1\le i\le n} \bar{A}_n(\boldsymbol{x}_i, y_i)\Big)^2 \,\Big|\, \mathbf{X}_n\Big] \le \mathbb{E}\Big[\Big(\max_{1\le i\le n} \bar{A}_n(\boldsymbol{x}_i, y_i)\Big)^\nu \,\Big|\, \mathbf{X}_n\Big]^{2/\nu}$$

$$
\leq \mathbb{E}\Big[\sum_{i=1}^n \bar{A}_n(\boldsymbol{x}_i, y_i)^\nu \,\Big|\, \mathbf{X}_n\Big]^{2/\nu} \leq (\mu_n n)^{2/\nu},
$$

and $\sigma \lesssim \sqrt{\delta_n}$ since on $\mathcal{A}$

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\big(A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i) - A_n(\tilde{\boldsymbol{q}}, \boldsymbol{x}_i, y_i)\big)^2 b_l(\boldsymbol{x}_i)^2 \,\big|\, \mathbf{X}_n\big] \\
&\quad = \frac{1}{n}\sum_{i=1}^n b_l(\boldsymbol{x}_i)^2 \mathbb{E}\big[\big(A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i) - A_n(\tilde{\boldsymbol{q}}, \boldsymbol{x}_i, y_i)\big)^2 \,\big|\, \mathbf{X}_n\big] \\
&\quad \lesssim \|\boldsymbol{q} - \tilde{\boldsymbol{q}}\| \frac{1}{n}\sum_{i=1}^n b_l(\boldsymbol{x}_i)^2 \lesssim \|\boldsymbol{q} - \tilde{\boldsymbol{q}}\|.
\end{aligned}
$$

This gives that on $\mathcal{A}$

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\leq\delta_n} \|\boldsymbol{G}_n(\boldsymbol{q}) - \boldsymbol{G}_n(\tilde{\boldsymbol{q}})\|_\infty \,\Big|\, \mathbf{X}_n\Big] \\
&\quad \lesssim \sqrt{\delta_n \log \frac{K\zeta_K \mu_n^{1/\nu}}{\sqrt{\delta_n}}} + \frac{\zeta_K(\mu_n n)^{1/\nu}}{\sqrt{n}} \log \frac{K\zeta_K\mu_n^{1/\nu}}{\sqrt{\delta_n}}.
\end{aligned}
$$

By Markov's inequality and since $\mathbb{P}\{\mathcal{A}\} \to 1$, for any sequence $t_n > 0$

$$
\begin{aligned}
&\mathbb{P}\Bigg\{\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\leq\delta_n} \|\boldsymbol{G}_n(\boldsymbol{q}) - \boldsymbol{G}_n(\tilde{\boldsymbol{q}})\|_\infty > t_n\Bigg\} \\
&\quad \lesssim \frac{1}{t_n}\sqrt{\delta_n \log \frac{K\zeta_K \mu_n^{1/\nu}}{\sqrt{\delta_n}}} + \frac{\zeta_K(\mu_n n)^{1/\nu}}{t_n\sqrt{n}} \log \frac{K\zeta_K\mu_n^{1/\nu}}{\sqrt{\delta_n}} + o(1),
\end{aligned}
\tag{SA6.4}
$$

where the constant in $\lesssim$ does not depend on n.

Coupling Define a $K|\mathcal{Q}_n^\delta|$-dimensional vector $\boldsymbol{\xi}_i := \big(A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i) b_l(\boldsymbol{x}_i)/\sqrt{n}\big)_{1\leq l\leq K, \boldsymbol{q}\in\mathcal{Q}_n^\delta}$, so that we have $\boldsymbol{G}_n \circ \pi_n^\delta = \sum_{i=1}^n \boldsymbol{\xi}_i$. We make some preparations before applying Theorem SA6.2.

Firstly, we bound $\mathbb{E}\big[\|\boldsymbol{\xi}_i\|^2 \|\boldsymbol{\xi}_i\|_\infty \,\big|\, \mathbf{X}_n\big]$.

$$
\begin{aligned}
&\sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{\xi}_i\|^2\|\boldsymbol{\xi}_i\|_\infty \,\big|\, \mathbf{X}_n\big] \\
&\quad = \frac{1}{n^{3/2}} \sum_{i=1}^n \|\boldsymbol{b}(\boldsymbol{x}_i)\|^2 \|\boldsymbol{b}(\boldsymbol{x}_i)\|_\infty \mathbb{E}\Big[\sum_{\boldsymbol{q}\in\mathcal{Q}_n^\delta} A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i)^2 \cdot \max_{\boldsymbol{q}\in\mathcal{Q}_n^\delta} |A_n(\boldsymbol{q}, \boldsymbol{x}_i, y_i)| \,\Big|\, \boldsymbol{x}_i\Big] \\
&\quad \leq \frac{1}{n^{3/2}} |\mathcal{Q}_n^\delta| \sum_{i=1}^n \|\boldsymbol{b}(\boldsymbol{x}_i)\|^2 \|\boldsymbol{b}(\boldsymbol{x}_i)\|_\infty \mathbb{E}\big[\bar{A}_n(\boldsymbol{x}_i, y_i)^3 \,\big|\, \boldsymbol{x}_i\big] \\
&\quad \leq \frac{|\mathcal{Q}_n^\delta| \mu_n^{3/\nu}}{n^{3/2}} \sum_{i=1}^n \|\boldsymbol{b}(\boldsymbol{x}_i)\|^2 \|\boldsymbol{b}(\boldsymbol{x}_i)\|_\infty \lesssim \frac{|\mathcal{Q}_n^\delta| \mu_n^{3/\nu}}{n^{1/2}} \zeta_K^3.
\end{aligned}
$$

Secondly, for $i \in \{1, \ldots, n\}$ let $\boldsymbol{g}_i \sim \mathcal{N}(0, \Sigma_i)$ be independent vectors, where $\Sigma_i = \mathbb{V}[\boldsymbol{\xi}_i \,|\, \boldsymbol{x}_i]$. Since there is an independent random variable U_1 distributed uniformly on $[0, 1]$, we can construct the family $\{\boldsymbol{g}_i\}$ on the same probability space. Then by Jensen's inequality for any $\lambda > 0$ we have

$$\begin{aligned}
\mathbb{E}\big[\|\boldsymbol{g}_i\|_\infty^2 \,\big|\, \boldsymbol{x}_i\big] &\le \frac{1}{\lambda}\log \mathbb{E}\big[e^{\lambda\|\boldsymbol{g}_i\|_\infty^2} \,\big|\, \boldsymbol{x}_i\big] \le \frac{1}{\lambda}\log \sum_{t=1}^{K|\mathcal{Q}_n^\delta|} \mathbb{E}\big[e^{\lambda(g_{it})^2} \,\big|\, \boldsymbol{x}_i\big] \\
&\le \frac{-\frac{1}{2}\log\Big(1 - \frac{2\lambda}{n}\zeta_K^2\mu_n^{2/\nu}\Big) + \log K + \log|\mathcal{Q}_n^\delta|}{\lambda} \lesssim \frac{\zeta_K^2\mu_n^{2/\nu}}{n}\big(\log K + \log|\mathcal{Q}_n^\delta|\big),
\end{aligned}$$

where we used the moment-generating function of χ_1^2: $\mathbb{E}[\exp\{\alpha\chi_1^2\}] = (1-2\alpha)^{-1/2}$ for $\alpha < 1/2$, the bound $\mathbb{V}[\xi_{it} \,|\, \boldsymbol{x}_i] \le \zeta_K^2\mu_n^{2/\nu}\Big/n$, and put $\lambda := \big(4\zeta_K^2\mu_n^{2/\nu}/n\big)^{-1}$. Also,

$$\begin{aligned}
\mathbb{E}\big[\|\boldsymbol{g}_i\|^4 \,\big|\, \boldsymbol{x}_i\big]^{1/2} &= \mathbb{E}\Bigg[\Bigg(\sum_{t=1}^{K|\mathcal{Q}_n^\delta|} g_{it}^2\Bigg)^2 \,\Bigg|\, \boldsymbol{x}_i\Bigg]^{1/2} \le \sum_{t=1}^{K|\mathcal{Q}_n^\delta|} \mathbb{E}[g_{it}^4 \,|\, \boldsymbol{x}_i]^{1/2} \\
&= \sum_{t=1}^{K|\mathcal{Q}_n^\delta|} \sqrt{3}\mathbb{E}[g_{it}^2 \,|\, \boldsymbol{x}_i] \lesssim \mathbb{E}\big[\|\boldsymbol{g}_i\|^2 \,\big|\, \boldsymbol{x}_i\big] = \mathbb{E}[\|\boldsymbol{\xi}_i\|^2 \,|\, \boldsymbol{x}_i],
\end{aligned}$$

which gives

$$\sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{g}_i\|^4 \,\big|\, \boldsymbol{x}_i\big]^{1/2} \lesssim \sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{\xi}_i\|^2 \,\big|\, \boldsymbol{x}_i\big] \le \mu_n^{2/\nu}|\mathcal{Q}_n^\delta|\mathbb{E}_n\big[\|\boldsymbol{b}(\boldsymbol{x}_i)\|^2\big] \lesssim \zeta_K^2\mu_n^{2/\nu}|\mathcal{Q}_n^\delta|.$$

Therefore, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{g}_i\|^2\|\boldsymbol{g}_i\|_\infty \,\big|\, \boldsymbol{x}_i\big] &\le \sum_{i=1}^n \mathbb{E}\big[\|g_i\|^4 \,\big|\, \boldsymbol{x}_i\big]^{1/2}\mathbb{E}\big[\|g_i\|_\infty^2 \,\big|\, \boldsymbol{x}_i\big]^{1/2} \\
&\lesssim \frac{\zeta_K^3\mu_n^{3/\nu}}{\sqrt{n}}|\mathcal{Q}_n^\delta|\sqrt{\log\big(K|\mathcal{Q}_n^\delta|\big)}.
\end{aligned}$$

Now, since there exists a random variable U_2 independent of $\{\boldsymbol{\xi}_i\}_{i=1}^n \cup \{\boldsymbol{g}_i\}_{i=1}^n$, applying Theorem SA6.2 with

$$\begin{aligned}
\beta &:= \sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{\xi}_i\|^2\|\boldsymbol{\xi}_i\|_\infty\big] + \sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{g}_i\|^2\|\boldsymbol{g}_i\|_\infty\big] \\
&\lesssim \frac{\zeta_K^3\mu_n^{3/\nu}}{\sqrt{n}}|\mathcal{Q}_n^\delta|\sqrt{\log\big(K|\mathcal{Q}_n^\delta|\big)} \\
&\lesssim \frac{\zeta_K^3\mu_n^{3/\nu}}{n^{1/2}\delta_n^{d_\mathcal{Q}}}\sqrt{\log\frac{K}{\delta_n^{d_\mathcal{Q}}}}
\end{aligned}$$

gives that for any $t_n > 0$, on the same probability space there exists a vector $\boldsymbol{Z}_n^\delta \sim \mathcal{N}\big(0, \mathbb{V}[\boldsymbol{G}_n \circ \pi_n^\delta \,|\, \mathbf{X}_n]\big)$, generally different for different t_n, such that

$$\mathbb{P}\big\{\big\|\boldsymbol{G}_n \circ \pi_n^\delta - \boldsymbol{Z}_n^\delta\big\|_\infty > 3t_n\big\} \le \min_{s \ge 0}\big\{2\mathbb{P}\{\|\boldsymbol{N}\|_\infty > s\} + \frac{\beta}{t_n^3}s^2\big\},$$

where $\boldsymbol{N}$ is a $K|\mathcal{Q}_n^\delta|$-dimensional standard Gaussian vector. By the union bound,

$$\mathbb{P}\{\|\boldsymbol{N}\|_\infty > s\} \le 2K|\mathcal{Q}_n^\delta|e^{-s^2/2},$$

so by taking $s := C\sqrt{\log\bigl(K|\mathcal{Q}_n^\delta|\bigr)}$ for a positive constant C not depending on n (chosen later), we have

$$\begin{aligned}\mathbb{P}\bigl\{\bigl\|\boldsymbol{G}_n \circ \pi_n^\delta - \boldsymbol{Z}_n^\delta\bigr\|_\infty > 3t_n\bigr\} &\lesssim \bigl(K|\mathcal{Q}_n^\delta|\bigr)^{1-C^2/2} + \frac{\beta}{t_n^3} C^2 \log\bigl(K|\mathcal{Q}_n^\delta|\bigr)\\ &\lesssim \left(\frac{K}{\delta_n^{d_\mathcal{Q}}}\right)^{1-C^2/2} + \frac{\beta}{t_n^3} C^2 \log \frac{K}{\delta_n^{d_\mathcal{Q}}}\\ &\lesssim \left(\frac{K}{\delta_n^{d_\mathcal{Q}}}\right)^{1-C^2/2} + C^2 \frac{\zeta_K^3 \mu_n^{3/\nu}}{t_n^3\sqrt{n}\delta_n^{d_\mathcal{Q}}} \log^{3/2} \frac{K}{\delta_n^{d_\mathcal{Q}}},\end{aligned}$$

where the constant in $\lesssim$ does not depend on n.

Embedding a conditionally Gaussian vector into a conditionally Gaussian process For a fixed vector in $\boldsymbol{X} \in \mathcal{X}^n$, by standard existence results for Gaussian processes, there exists a mean-zero K-dimensional Gaussian process whose covariance structure is the same as that of $\boldsymbol{G}_n(\boldsymbol{q})$ given $\mathbf{X}_n = \boldsymbol{X}$. It follows from Kolmogorov's continuity criterion that this process can be defined on $\mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$. The laws of such processes define a family of Gaussian probability measures on the Borel σ-algebra $\mathcal{B}(\mathcal{C}(\mathcal{Q}, \mathbb{R}^K))$ of the space $\mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$, denoted $\{\mathbb{P}_{\boldsymbol{X}}\}_{\boldsymbol{X}\in\mathcal{X}^n}$. In order to construct one process that is conditionally on $\mathbf{X}'_n$, a mean zero Gaussian process with the same conditional covariance structure as that of $\boldsymbol{G}_n(\boldsymbol{q})$, where $\mathbf{X}'_n \overset{d}{=} \mathbf{X}_n$, we need to show that this family of measures is a probability kernel as a function $\mathcal{X}^n \times \mathcal{B}(\mathcal{C}(\mathcal{Q}, \mathbb{R}^K)) \to [0,1]$. This follows by a standard argument: we can take a π-system of sets, generating $\mathcal{B}(\mathcal{C}(\mathcal{Q}, \mathbb{R}^K))$, of the form $B = \{\boldsymbol{f} \in \mathcal{C}(\mathcal{Q}, \mathbb{R}^K) : \boldsymbol{f}(\boldsymbol{q}_1) \in B_1, \dots, \boldsymbol{f}(\boldsymbol{q}_m) \in B_m\}$, where $m \in \{1, 2, \dots\}$, $\boldsymbol{q}_j \in \mathcal{Q}$, and each B_j is a parallelepiped in $\mathbb{R}^K$ with edges parallel to the coordinate axes, and notice that for such sets $\boldsymbol{X} \mapsto \mathbb{P}_{\boldsymbol{X}}(B)$ is a Borel function (since a mean-zero Gaussian vector is a linear transformation of a standard Gaussian vector). The sets $A \in \mathcal{B}(\mathcal{C}(\mathcal{Q}, \mathbb{R}^K))$ such that $\boldsymbol{X} \mapsto \mathbb{P}_{\boldsymbol{X}}(A)$ is a Borel function form a λ-system. It is left to apply the monotone class theorem.

We have shown that there exists a law on $\mathcal{X}^n \times \mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$ which is the joint law of $\bigl(\mathbf{X}'_n, \{\boldsymbol{Z}'_n(\boldsymbol{q})\}_{\boldsymbol{q}\in\mathcal{Q}}\bigr)$, where $\{\boldsymbol{Z}'_n(\boldsymbol{q})\}_{\boldsymbol{q}\in\mathcal{Q}}$ is a conditionally on $\mathbf{X}'_n$ mean zero Gaussian process with the same conditional covariance structure as that of $\boldsymbol{G}'_n(\boldsymbol{q})$ (where $\boldsymbol{G}'_n(\boldsymbol{q})$ is the same function of $\mathbf{X}_n$ as $\boldsymbol{G}_n(\boldsymbol{q})$ except $\mathbf{X}_n$ is replaced by $\mathbf{X}'_n$). Projecting this $\mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$-process onto $\mathcal{Q}_n^\delta$, we obtain a vector $\boldsymbol{Z}_n^{\delta\prime}$ such that

$$\bigl(\mathbf{X}'_n, \boldsymbol{Z}_n^{\delta\prime}\bigr) \overset{d}{=} \bigl(\mathbf{X}_n, \boldsymbol{Z}_n^\delta\bigr).$$

Since $\mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$ is Polish and there exists a uniformly distributed random variable U_3 independent of $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n \cup \{\boldsymbol{Z}_n^\delta\}$, we can apply Theorem 8.17 (transfer) in [14], and obtain that there exists (on the same probability space) a random element $\{\boldsymbol{Z}_n(\boldsymbol{q})\}_{\boldsymbol{q}\in\mathcal{Q}} \in \mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$ such that

$$\bigl(\mathbf{X}_n, \boldsymbol{Z}_n^\delta, \{\boldsymbol{Z}_n(\boldsymbol{q})\}_{\boldsymbol{q}\in\mathcal{Q}}\bigr) = \bigl(\mathbf{X}'_n, \boldsymbol{Z}_n^{\delta\prime}, \{\boldsymbol{Z}'_n(\boldsymbol{q})\}_{\boldsymbol{q}\in\mathcal{Q}}\bigr).$$

In particular, $\{\boldsymbol{Z}_n(\boldsymbol{q})\}_{\boldsymbol{q}\in\mathcal{Q}}$ is the conditionally Gaussian process whose projection on $\mathcal{Q}_n^\delta$ is the vector $\boldsymbol{Z}_n^\delta$ a. s.

Discretization of Z_n Consider the stochastic process X_n defined for $t = (l, \boldsymbol{q}, \tilde{\boldsymbol{q}}) \in T$ with

$$T := \{(l, \boldsymbol{q}, \tilde{\boldsymbol{q}})\colon l \in \{1, 2, \dots, K\}, \boldsymbol{q}, \tilde{\boldsymbol{q}} \in \mathcal{Q}, \|\boldsymbol{q} - \tilde{\boldsymbol{q}}\| \le \delta_n\}$$

as $X_{n,t} := Z_{n,l}(\boldsymbol{q}) - Z_{n,l}(\tilde{\boldsymbol{q}})$. It is a separable (because each $Z_{n,l}(\cdot)$ has a. s. continuous trajectories) mean-zero Gaussian conditionally on $\mathbf{X}_n$ process with the index set T considered a metric space: $\operatorname{dist}\bigl((l, \boldsymbol{q}, \tilde{\boldsymbol{q}}), (l', \boldsymbol{q}', \tilde{\boldsymbol{q}}')\bigr) = \|\boldsymbol{q} - \boldsymbol{q}'\| + \|\tilde{\boldsymbol{q}} - \tilde{\boldsymbol{q}}'\| + \mathbb{1}\{l \ne l'\}$.

We will apply Lemma SA3.7 to this process. Note that on the event $\mathcal{A}$

$$\begin{aligned}
\sigma(X_n)^2 &:= \sup_{t\in T}\mathbb{E}\big[X_{n,t}^2 \,\big|\, \mathbf{X}_n\big] = \sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\max_l \mathbb{E}\big[\big(Z_{n,l}(\boldsymbol{q}) - Z_{n,l}(\tilde{\boldsymbol{q}})\big)^2 \,\big|\, \mathbf{X}_n\big]\\
&= \sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\max_l \mathbb{E}\big[\big(G_{n,l}(\boldsymbol{q}) - G_{n,l}(\tilde{\boldsymbol{q}})\big)^2 \,\big|\, \mathbf{X}_n\big]\\
&= \sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\max_l \frac{1}{n}\sum_{i=1}^n b_l(\boldsymbol{x}_i)^2\mathbb{E}\big[\big(A_n(\boldsymbol{q},\boldsymbol{x}_i,y_i) - A_n(\tilde{\boldsymbol{q}},\boldsymbol{x}_i,y_i)\big)^2 \,\big|\, \boldsymbol{x}_i\big]\\
&\lesssim \sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\max_l \frac{1}{n}\sum_{i=1} b_l(\boldsymbol{x}_i)^2\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\| \le \delta_n \max_l \mathbb{E}_n[b_l(\boldsymbol{x}_i)^2]\\
&\le \delta_n \underbrace{\sup_{\|\boldsymbol{\alpha}\|=1}\mathbb{E}_n\big[\big(\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{b}(\boldsymbol{x}_i)\big)^2\big]}_{\le C_{\mathrm{Gr}}} \lesssim \delta_n.
\end{aligned}$$

Next, we define and bound the semimetric $\rho(t,t')$:

$$\begin{aligned}
\rho(t,t')^2 &:= \mathbb{E}\big[(X_{n,t} - X_{n,t'})^2 \,\big|\, \mathbf{X}_n\big]\\
&= \mathbb{E}\big[\big(\big(Z_{n,l}(\boldsymbol{q}) - Z_{n,l}(\tilde{\boldsymbol{q}})\big) - \big(Z_{n,l'}(\boldsymbol{q}') - Z_{n,l'}(\tilde{\boldsymbol{q}}')\big)\big)^2 \,\big|\, \mathbf{X}_n\big]\\
&\lesssim \mathbb{E}\big[\big(Z_{n,l}(\boldsymbol{q}) - Z_{n,l'}(\boldsymbol{q}')\big)^2 \,\big|\, \mathbf{X}_n\big] + \mathbb{E}\big[\big(Z_{n,l}(\tilde{\boldsymbol{q}}) - Z_{n,l'}(\tilde{\boldsymbol{q}}')\big)^2 \,\big|\, \mathbf{X}_n\big].
\end{aligned}$$

The first term on the right is bounded the following way: if $l \ne l'$,

$$\begin{aligned}
&\mathbb{E}\big[\big(Z_{n,l}(\boldsymbol{q}) - Z_{n,l'}(\boldsymbol{q}')\big)^2 \,\big|\, \mathbf{X}_n\big] = \mathbb{E}\big[\big(G_{n,l}(\boldsymbol{q}) - G_{n,l'}(\boldsymbol{q}')\big)^2 \,\big|\, \mathbf{X}_n\big]\\
&\quad= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\big(A_n(\boldsymbol{q},\boldsymbol{x}_i,y_i)b_l(\boldsymbol{x}_i) - A_n(\boldsymbol{q}',\boldsymbol{x}_i,y_i)b_{l'}(\boldsymbol{x}_i)\big)^2 \,\big|\, \boldsymbol{x}_i\big]\\
&\quad\lesssim \frac{1}{n}\sum_{i=1}^n b_l(\boldsymbol{x}_i)^2\mathbb{E}\big[\big(A_n(\boldsymbol{q},\boldsymbol{x}_i,y_i) - A_n(\boldsymbol{q}',\boldsymbol{x}_i,y_i)\big)^2 \,\big|\, \boldsymbol{x}_i\big]\\
&\qquad+ \frac{1}{n}\sum_{i=1}^n \big(b_l(\boldsymbol{x}_i)^2 + b_{l'}(\boldsymbol{x}_i)^2\big)\mathbb{E}\big[A_n(\boldsymbol{q}',\boldsymbol{x}_i,y_i)^2 \,\big|\, \boldsymbol{x}_i\big]\\
&\quad\lesssim \|\boldsymbol{q}-\boldsymbol{q}'\|\mathbb{E}_n\big[b_l(\boldsymbol{x}_i)^2\big] + \mu_n^{2/\nu}\big(\mathbb{E}_n[b_l(\boldsymbol{x}_i)^2] + \mathbb{E}_n[b_{l'}(\boldsymbol{x}_i)^2]\big) \lesssim \|\boldsymbol{q}-\boldsymbol{q}'\| + \mu_n^{2/\nu}.
\end{aligned}$$

Similarly, if $l = l'$, $\mathbb{E}\big[\big(Z_{n,l}(\boldsymbol{q}) - Z_{n,l'}(\boldsymbol{q}')\big)^2 \,\big|\, \mathbf{X}_n\big] \lesssim \|\boldsymbol{q}-\boldsymbol{q}'\|$.

The term $\mathbb{E}\big[\big(Z_{n,l}(\tilde{\boldsymbol{q}}) - Z_{n,l'}(\tilde{\boldsymbol{q}}')\big)^2 \,\big|\, \mathbf{X}_n\big]$ is bounded the same way, and we conclude

$$\rho(t,t')^2 \lesssim \mu_n^{2/\nu}\big(\|\boldsymbol{q}-\boldsymbol{q}'\| + \|\tilde{\boldsymbol{q}}-\tilde{\boldsymbol{q}}'\| + \mathbb{1}\{l \ne l'\}\big) = \mu_n^{2/\nu}\mathrm{dist}\big((l,\boldsymbol{q},\tilde{\boldsymbol{q}}),(l',\boldsymbol{q}',\tilde{\boldsymbol{q}}')\big).$$

In other words, we have proven that for some positive constant C_{13} we have on $\mathcal{A}$

$$\rho(t,t')^2 \le C_{13}\mu_n^{2/\nu}\mathrm{dist}\big((l,\boldsymbol{q},\tilde{\boldsymbol{q}}),(l',\boldsymbol{q}',\tilde{\boldsymbol{q}}')\big).$$

This means that an $\big(\varepsilon/\mu_n^{1/\nu}\sqrt{C_{13}}\big)^2$-covering of T with respect to $\mathrm{dist}(\cdot)$ induces an ε-covering of T with respect to ρ, and hence

$$N(T,\rho,\varepsilon) \le N\left(T, \mathrm{dist}(\cdot), \left(\frac{\varepsilon}{\mu_n^{1/\nu}\sqrt{C_{13}}}\right)^2\right). \tag{SA6.5}$$

Therefore,

$$\log N(T,\rho,\varepsilon) \lesssim \log(K\mu_n^{1/\nu}/\varepsilon).$$

Applying Lemma SA3.7 gives that on the event $\mathcal{A}$

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in T}|X_{n,t}|\,\Big|\,\mathbf{X}_n\Big] &\lesssim \sigma(X_n) + \int_0^{2\sigma(X_n)} \sqrt{\log N(T,\rho,\varepsilon)}\,\mathrm{d}\varepsilon \\ &\lesssim \sigma(X_n) + \sigma(X_n)\sqrt{\log\bigl(K\mu_n^{1/\nu}/\sigma(X_n)\bigr)} \overset{\text{(a)}}{\lesssim} \left(\delta_n \log\left(K\frac{\mu_n^{1/\nu}}{\delta_n}\right)\right)^{1/2},\end{aligned}$$

where in (a) we used our bound $\sigma(X_n) \lesssim \sqrt{\delta_n}$ above and that $x \mapsto x\log\frac{1}{x}$ is increasing for sufficiently small x. Rewriting and applying Markov's inequality, we obtain that on $\mathcal{A}$

$$\mathbb{P}\left\{\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n} \|\boldsymbol{Z}_n(\boldsymbol{q})-\boldsymbol{Z}_n(\tilde{\boldsymbol{q}})\|_\infty > t_n \,\middle|\, \mathbf{X}_n\right\} \lesssim \frac{1}{t_n}\sqrt{\delta_n \log\frac{K\mu_n^{1/\nu}}{\delta_n}},$$

where the constant in $\lesssim$ does not depend on n. Since $\mathbb{P}\{\mathcal{A}\}\to 1$, this implies

$$\mathbb{P}\left\{\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n} \|\boldsymbol{Z}_n(\boldsymbol{q})-\boldsymbol{Z}_n(\tilde{\boldsymbol{q}})\|_\infty > t_n\right\} \lesssim \frac{1}{t_n}\sqrt{\delta_n \log\frac{K\mu_n^{1/\nu}}{\delta_n}} + o(1). \tag{SA6.6}$$

Choosing δ_n and conclusion Combining the bounds obtained above, for any given positive sequence t_n and any constant $C>0$ of our choice

$$\begin{aligned}\mathbb{P}\left\{\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\boldsymbol{G}_n(\boldsymbol{q})-\boldsymbol{Z}_n(\boldsymbol{q})\|_\infty > t_n\right\} &\le \mathbb{P}\left\{\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\|\boldsymbol{G}_n(\boldsymbol{q})-\boldsymbol{G}_n(\tilde{\boldsymbol{q}})\|_\infty > t_n/3\right\} \\ &\quad + \mathbb{P}\left\{\left\|\boldsymbol{G}_n\circ\pi_n^\delta - \boldsymbol{Z}_n^\delta\right\|_\infty > t_n/3\right\} + \mathbb{P}\left\{\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\|Z_n(\boldsymbol{q})-Z_n(\tilde{\boldsymbol{q}})\|_\infty > t_n/3\right\} \\ &\lesssim \frac{1}{t_n}\sqrt{\delta_n\log\frac{K\zeta_K\mu_n^{1/\nu}}{\sqrt{\delta_n}}} + \frac{\zeta_K(\mu_n n)^{1/\nu}}{t_n\sqrt{n}}\log\frac{K\zeta_K\mu_n^{1/\nu}}{\sqrt{\delta_n}} \\ &\quad + \left(\frac{K}{\delta_n^{d_{\mathcal{Q}}}}\right)^{1-C^2/2} + C^2\frac{\zeta_K^3\mu_n^{3/\nu}}{t_n^3\sqrt{n}\delta_n^{d_{\mathcal{Q}}}}\log^{3/2}\frac{K}{\delta_n^{d_{\mathcal{Q}}}} \\ &\quad + \frac{1}{t_n}\sqrt{\delta_n\log\frac{K\mu_n^{1/\nu}}{\delta_n}} + o(1),\end{aligned}$$

where the constant in $\lesssim$ does not depend on n.

Take, for example, $C=2$ (so that $1-C^2/2$ is negative). Now we approximately (assuming that each $\log(\cdot)$ on the right is $O(\log n)$ and ignoring constant coefficients) optimize this over δ_n. This gives

$$\delta_n := \left(\frac{\zeta_K^3\mu_n^{3/\nu}}{t_n^2\sqrt{n}}\log n\right)^{\frac{2}{1+2d_{\mathcal{Q}}}}.$$

Let $\ell_n \to 0$ be a positive sequence satisfying

$$\left(\frac{\zeta_K^3\mu_n^{3/\nu}}{\sqrt{n}}\right)^{\frac{1}{3+2d_{\mathcal{Q}}}}\sqrt{\log n} + \frac{\zeta_K\mu_n^{1/\nu}}{n^{1/2-1/\nu}}\log n = o(\ell_n r_{\mathtt{YUR}}).$$

Clearly, whenever (SA6.3) holds, such a sequence exists. Putting $t_n := \ell_n r_{\mathrm{yur}}$, we get

$$\mathbb{P}\left\{\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\boldsymbol{G}_n(\boldsymbol{q})-\boldsymbol{Z}_n(\boldsymbol{q})\|_\infty > \ell_n r_{\mathrm{yur}}\right\} = o(1).$$

Fix $\varepsilon > 0$. For n large enough, $\ell_n < \varepsilon$. Then for these n we have

$$\mathbb{P}\left\{\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\boldsymbol{G}_n(\boldsymbol{q})-\boldsymbol{Z}_n(\boldsymbol{q})\|_\infty > \varepsilon r_{\mathrm{yur}}\right\} \le \mathbb{P}\left\{\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\boldsymbol{G}_n(\boldsymbol{q})-\boldsymbol{Z}_n(\boldsymbol{q})\|_\infty > \ell_n r_{\mathrm{yur}}\right\} = o(1).$$

Lemma SA6.3 is proven. □

SA6.2 Main Result

We begin by presenting our main strong approximation result, which is a special case of more technical and lengthy Lemma SA6.3. To simplify exposition, the following notation will be helpful from this point onwards:

$$\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) := -h^{d/2}\frac{\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}, \tag{SA6.7}$$

$$\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) := -h^{d/2}\frac{\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}, \tag{SA6.8}$$

$$\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) := -h^{d/2}\frac{\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})}{\sqrt{\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}, \tag{SA6.9}$$

$$t_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) := h^{-d/2}\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\sqrt{n}\mathbb{E}_n\big[\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\boldsymbol{p}(\boldsymbol{x}_i)\big], \tag{SA6.10}$$

$$T_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) := \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})/n}}, \tag{SA6.11}$$

where $\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})$ is some feasible estimator of variance.

Theorem SA6.4 (Strong approximation)**.**

(a) Suppose Assumptions SA2.1 to SA2.6 hold with $\nu \ge 3$. Furthermore, assume the following conditions hold:

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}|\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})-\mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|}r_{\mathtt{UC}}; \tag{SA6.12}$$

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}|\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})-\mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})-\mathsf{L}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{v}|}r_{\mathtt{BR}} \text{ with } \frac{\log n}{nh^d} \lesssim r_{\mathtt{BR}}; \tag{SA6.13}$$

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}\big|\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})-\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\big| \lesssim_{\mathbb{P}} h^{-2|\boldsymbol{v}|-d}r_{\mathtt{VC}} \text{ with } r_{\mathtt{VC}} = o(1); \tag{SA6.14}$$

$$\begin{aligned}\mathbb{E}[(\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))-\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\tilde{\boldsymbol{q}}));\tilde{\boldsymbol{q}})\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\tilde{\boldsymbol{q}})))^2 \mid \boldsymbol{x}_i]&\\ \lesssim \|\boldsymbol{q}-\tilde{\boldsymbol{q}}\| \text{ for all } \boldsymbol{q},\tilde{\boldsymbol{q}}\in\mathcal{Q}.&\end{aligned} \tag{SA6.15}$$

Let $r_{\mathtt{SA}}$ is any positive sequence of numbers converging to zero[1] *such that*

$$\left(\frac{1}{nh^{3d}}\right)^{\frac{1}{6+4d_{\mathcal{Q}}}}\sqrt{\log n} + \frac{1}{h^{d/2}n^{1/2-1/\nu}}\log n = o(r_{\mathtt{SA}}). \tag{SA6.16}$$

[1] In particular, the left-hand side of (SA6.16) is automatically assumed to be $o(1)$.

Denote
$$\boldsymbol{G}(\boldsymbol{q}) := h^{-d/2}\sqrt{n}\mathbb{E}_n\big[\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))\boldsymbol{p}(\boldsymbol{x}_i)\big].$$

Then (provided the probability space is rich enough) there exists a K-dimensional, conditionally on $\mathbf{X}_n$ mean-zero Gaussian, process $\boldsymbol{Z}(\boldsymbol{q})$ on $\mathcal{Q}$ with a. s. continuous trajectories, satisfying

$$\mathbb{E}[\boldsymbol{G}(\boldsymbol{q})\boldsymbol{G}(\tilde{\boldsymbol{q}})^\mathsf{T} \mid \mathbf{X}_n] = \mathbb{E}[\boldsymbol{Z}(\boldsymbol{q})\boldsymbol{Z}(\tilde{\boldsymbol{q}})^\mathsf{T} \mid \mathbf{X}_n]; \tag{SA6.17}$$
$$\sup_{\boldsymbol{q}}\|\boldsymbol{G}(\boldsymbol{q}) - \boldsymbol{Z}(\boldsymbol{q})\|_\infty = o_\mathbb{P}(r_{\mathtt{SA}}); \tag{SA6.18}$$
$$\sup_{\boldsymbol{x}\in\mathcal{X}}\sup_{\boldsymbol{q}\in\mathcal{Q}}\big|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})^\mathsf{T}\boldsymbol{G}(\boldsymbol{q}) - \bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})^\mathsf{T}\boldsymbol{Z}(\boldsymbol{q})\big| = o_\mathbb{P}(r_{\mathtt{SA}}), \tag{SA6.19}$$
$$\sup_{\boldsymbol{q},\boldsymbol{x}}\big|T_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q}) - \bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})^\mathsf{T}\boldsymbol{Z}(\boldsymbol{q})\big| \lesssim_\mathbb{P} \sqrt{nh^d}(r_{\mathtt{UC}}r_{\mathtt{VC}} + r_{\mathtt{BR}}) + o(r_{\mathtt{SA}}). \tag{SA6.20}$$

(b) If, in addition to the previous conditions, $\overline{\psi}(\boldsymbol{x}_i, y_i)$ is σ^2-sub-Gaussian conditionally on $\boldsymbol{x}_i$, and $r_{\mathtt{SA}}^{\mathrm{sub}} \to 0$ is a positive sequence of numbers satisfying

$$\left(\frac{1}{nh^{3d}}\right)^{\frac{1}{6+4d_\mathcal{Q}}}\sqrt{\log n} + \frac{\log^{3/2} n}{\sqrt{nh^d}} = o(r_{\mathtt{SA}}^{\mathrm{sub}}),$$

then
$$\sup_{\boldsymbol{x}\in\mathcal{X}}\sup_{\boldsymbol{q}\in\mathcal{Q}}\big|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})^\mathsf{T}\boldsymbol{G}(\boldsymbol{q}) - \bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})^\mathsf{T}\boldsymbol{Z}(\boldsymbol{q})\big| = o_\mathbb{P}(r_{\mathtt{SA}}^{\mathrm{sub}}). \tag{SA6.21}$$

Proof. Note that (SA6.18) implies (SA6.19) because $\sup_{\boldsymbol{q},\boldsymbol{x}}\big\|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})\big\|_1 \lesssim_\mathbb{P} 1$ by Lemma SA6.7 below. To prove (SA6.18), apply Lemma SA6.3 with $\boldsymbol{G}(\cdot)$ as in the statement,

$$\begin{aligned}
&A_n(\boldsymbol{q}, y) = A_n(\boldsymbol{q}, \boldsymbol{x}, y) := \psi(y, \eta(\mu_0(\boldsymbol{x}, \boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}, \boldsymbol{q})),\\
&\boldsymbol{b}(\cdot) := h^{-d/2}\boldsymbol{p}(\cdot),\\
&\zeta_K \lesssim K^{1/2} \lesssim h^{-d/2},\\
&\mu_n \lesssim 1.
\end{aligned}$$

Finally, (SA6.20) follows from combining (SA6.19) with Lemma SA6.8. □

Remark SA6.5. This theorem contains Theorem 2 in the main paper; the notation there is simplified for better readability. Specifically, $T(\boldsymbol{x}, q)$ in the main paper corresponds to $T_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})$ here and the approximating process $Z(\boldsymbol{x}, q)$ corresponds to $\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})^\mathsf{T}\boldsymbol{Z}(\boldsymbol{q})$ here. The reason for differing notation is that we require more precision here: given the form of Lemma SA6.3, it is more natural to use K-dimensional processes such as $\boldsymbol{Z}(\boldsymbol{q})$, also making for simpler presentation of precise results in Section SA7 later.

The rest of this subsection will be devoted to proving the lemmas used in the proof of Theorem SA6.4. First, we prove bounds on the asymptotic variance and its approximation. Similar bounds were proven in [4, 6].

Lemma SA6.6 (Asymptotic variance)**.** *Suppose Assumptions Assumptions SA2.1 to SA2.6 hold, and $\frac{\log(1/h)}{nh^d} = o(1)$. Then*

$$h^{-d-2|\boldsymbol{v}|} \lesssim \inf_{\boldsymbol{q}\in\mathcal{Q}}\inf_{\boldsymbol{x}\in\mathcal{X}}\Omega_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q}) \le \sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}\Omega_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q}) \lesssim h^{-d-2|\boldsymbol{v}|}, \tag{SA6.22}$$

$$\sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})-\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\right| \lesssim_{\mathbb{P}} h^{-d-2|\boldsymbol{v}|}\sqrt{\frac{\log(1/h)}{nh^d}}, \tag{SA6.23}$$

$$h^{-d-2|\boldsymbol{v}|} \lesssim_{\mathbb{P}} \inf_{\boldsymbol{q}\in\mathcal{Q}}\inf_{\boldsymbol{x}\in\mathcal{X}}\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) \le \sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) \lesssim_{\mathbb{P}} h^{-d-2|\boldsymbol{v}|}. \tag{SA6.24}$$

Proof. For the lower bound in (SA6.22), we have

$$\begin{aligned}\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) &\ge \lambda_{\min}(\boldsymbol{\Sigma}_{0,\boldsymbol{q}})\cdot\|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|^2 \ge \lambda_{\min}(\boldsymbol{\Sigma}_{0,\boldsymbol{q}})\cdot\|\mathbf{Q}_{0,\boldsymbol{q}}\|^{-2}\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|^2\\ &\gtrsim \lambda_{\min}(\boldsymbol{\Sigma}_{0,\boldsymbol{q}})\cdot h^{-2d}\cdot h^{-2|\boldsymbol{v}|} \gtrsim h^d\cdot h^{-2d}\cdot h^{-2|\boldsymbol{v}|} = h^{-d-2|\boldsymbol{v}|}\end{aligned} \tag{SA6.25}$$

by Assumptions SA2.2, SA2.4 ($\sigma^2_{\boldsymbol{q}}(\boldsymbol{x})$ and $\eta^{(1)}\big(\mu_0(\boldsymbol{x},\boldsymbol{q})\big)^2$ are uniformly over $\boldsymbol{x}$ bounded away from zero) and Lemma SA3.11.

For the upper bound in (SA6.22), we have

$$\begin{aligned}\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) &\le \lambda_{\max}(\boldsymbol{\Sigma}_{0,\boldsymbol{q}})\cdot\left\|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\right\|^2 \le \lambda_{\max}(\boldsymbol{\Sigma}_{0,\boldsymbol{q}})\cdot\left[\lambda_{\max}(\mathbf{Q}_{0,\boldsymbol{q}}^{-1})\right]^2\left\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\right\|^2\\ &\overset{\text{(a)}}{\lesssim} \lambda_{\max}(\boldsymbol{\Sigma}_{0,\boldsymbol{q}})\cdot h^{-2d}\cdot h^{-2|\boldsymbol{v}|} \overset{\text{(b)}}{\lesssim} h^d\cdot h^{-2d}\cdot h^{-2|\boldsymbol{v}|} = h^{-d-2|\boldsymbol{v}|},\end{aligned}$$

where (a) is by Assumption SA2.2 and Lemma SA3.11, (b) is by Assumption SA2.4 ($\sigma^2_{\boldsymbol{q}}(\boldsymbol{x})$ and $\eta^{(1)}\big(\mu_0(\boldsymbol{x},\boldsymbol{q})\big)^2$ are uniformly over $\boldsymbol{x}$ bounded) and Lemma SA3.11.

We will now prove (SA6.23). We start by noticing

$$\begin{aligned}&\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^\mathsf{T}\mathbf{Q}_{0,\boldsymbol{q}}^{-1}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}-\boldsymbol{\Sigma}_{0,\boldsymbol{q}})\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\right|\\ &\quad\lesssim_{\mathbb{P}} \sup_{\boldsymbol{q},\boldsymbol{x}}\|\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}-\boldsymbol{\Sigma}_{0,\boldsymbol{q}}\|\cdot\|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\|^2\cdot\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|^2\\ &\quad\overset{\text{(a)}}{\lesssim_{\mathbb{P}}} h^{-2d}\cdot h^{-2|\boldsymbol{v}|}\sup_{\boldsymbol{q}}\|\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}-\boldsymbol{\Sigma}_{0,\boldsymbol{q}}\| \overset{\text{(b)}}{\lesssim_{\mathbb{P}}} h^{-2d-2|\boldsymbol{v}|}h^d\sqrt{\frac{\log(1/h)}{nh^d}} = h^{-d-2|\boldsymbol{v}|}\sqrt{\frac{\log(1/h)}{nh^d}},\end{aligned}$$

where in (a) we used $\sup_{\boldsymbol{q}}\|\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\| \lesssim_{\mathbb{P}} h^{-d}$ by Lemma SA3.11 and $\sup_{\boldsymbol{x}}\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\| \lesssim h^{-|\boldsymbol{v}|}$ by Assumption SA2.2, in (b) we used $\sup_{\boldsymbol{q}}\|\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}-\boldsymbol{\Sigma}_{0,\boldsymbol{q}}\| \lesssim_{\mathbb{P}} h^d\sqrt{\frac{\log(1/h)}{nh^d}}$ which is proven by the same argument as $\sup_{\boldsymbol{q}}\|\bar{\mathbf{Q}}_{\boldsymbol{q}}-\mathbf{Q}_{0,\boldsymbol{q}}\| \lesssim_{\mathbb{P}} h^d\sqrt{\frac{\log(1/h)}{nh^d}}$ in Lemma SA3.11. Similarly,

$$\begin{aligned}&\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^\mathsf{T}(\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}-\mathbf{Q}_{0,\boldsymbol{q}}^{-1})\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\right|\\ &\quad\le \sup_{\boldsymbol{q},\boldsymbol{x}}\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}-\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\|\cdot\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|\cdot\|\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}\|\cdot\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|^2\\ &\quad\lesssim_{\mathbb{P}} h^{-d}\sqrt{\frac{\log(1/h)}{nh^d}}\cdot h^{-d}\cdot h^d\cdot h^{-2|\boldsymbol{v}|} = h^{-d-2|\boldsymbol{v}|}\sqrt{\frac{\log(1/h)}{nh^d}}.\end{aligned}$$

Finally,

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^\mathsf{T}\mathbf{Q}_{0,\boldsymbol{q}}^{-1}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}}(\mathbf{Q}_{0,\boldsymbol{q}}^{-1}-\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1})\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\right| \lesssim_{\mathbb{P}} h^{-d-2|\boldsymbol{v}|}\sqrt{\frac{\log(1/h)}{nh^d}}.$$

Combining the bounds above gives $\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})-\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\right| \lesssim_{\mathbb{P}} h^{-d-2|\boldsymbol{v}|}\sqrt{\frac{\log(1/h)}{nh^d}}$.

(SA6.24) is an immediate consequence of (SA6.22) and (SA6.23), since $\frac{\log(1/h)}{nh^d} = o(1)$. ☐

Lemma SA6.7 (Closeness of linear terms). *Suppose Assumptions SA2.1 to SA2.6, (SA6.14) hold, and* $\frac{\log(1/h)}{nh^d} = o(1)$. *Then*

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2} - \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2}\right| \lesssim_{\mathbb{P}} h^{d/2+|\boldsymbol{v}|} r_{\mathtt{VC}}, \tag{SA6.26}$$

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2} - \Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2}\right| \lesssim_{\mathbb{P}} h^{d/2+|\boldsymbol{v}|}\sqrt{\frac{\log(1/h)}{nh^d}}, \tag{SA6.27}$$

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim 1, \tag{SA6.28}$$

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}, \tag{SA6.29}$$

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim 1 \quad \textit{w. p. a. 1.} \tag{SA6.30}$$

If, in addition, Assumption SA2.7 holds, then

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim_{\mathbb{P}} r_{\mathtt{Q}} + r_{\mathtt{VC}}, \tag{SA6.31}$$

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim 1 \quad \textit{w. p. a. 1.} \tag{SA6.32}$$

Proof. Equation (SA6.26) follows from the following chain of inequalities:

$$\begin{aligned}
&\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2} - \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2}\right| \\
&\quad \le \sup_{\boldsymbol{q},\boldsymbol{x}}\left[\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\right]^{-1/2}\left[\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{1/2} + \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{1/2}\right]^{-1}\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\right| \\
&\quad \overset{\text{(a)}}{\lesssim_{\mathbb{P}}} (h^{d+2|\boldsymbol{v}|})^{3/2}\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\right| \overset{\text{(b)}}{\lesssim_{\mathbb{P}}} h^{d/2+|\boldsymbol{v}|} r_{\mathtt{VC}},
\end{aligned}$$

where in (a) we used that $\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})$ is close to $\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})$ by (SA6.14) and the lower bound on $\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})$ from Lemma SA6.6, and in (b) we used (SA6.14).

To prove (SA6.27), recall that

$$\begin{aligned}
&\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\right| \lesssim_{\mathbb{P}} h^{-d-2|\boldsymbol{v}|}\sqrt{\frac{\log(1/h)}{nh^d}}, \quad \text{and} \\
&\inf_{\boldsymbol{q},\boldsymbol{x}} \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) \gtrsim_{\mathbb{P}} h^{-d-2|\boldsymbol{v}|}
\end{aligned}$$

by Lemma SA6.6. Therefore,

$$\begin{aligned}
&\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2} - \Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2}\right| \\
&\quad \le \sup_{\boldsymbol{q},\boldsymbol{x}}\left[\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\right]^{-1/2}\left[\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{1/2} + \Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{1/2}\right]^{-1}\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\right| \\
&\quad \lesssim_{\mathbb{P}} (h^{d+2|\boldsymbol{v}|})^{3/2}\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\right| \lesssim_{\mathbb{P}} h^{3d/2+3|\boldsymbol{v}|}h^{-d-2|\boldsymbol{v}|}\sqrt{\frac{\log(1/h)}{nh^d}} \\
&\quad = h^{d/2+|\boldsymbol{v}|}\sqrt{\frac{\log(1/h)}{nh^d}}.
\end{aligned}$$

Equation (SA6.28) follows from

$$h^{d/2}\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|_1 \le h^{d/2}\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_1 \cdot \|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|_1 \lesssim_{\mathbb{P}} h^{-d/2-|\boldsymbol{v}|}$$

and Lemma SA6.6.

We will now prove (SA6.29). We have that

$$\sup_{\boldsymbol{q}}\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_1 \lesssim_{\mathbb{P}} h^{-d}$$

by Lemma SA3.11, and

$$\sup_{\boldsymbol{x}}\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|_1 \lesssim h^{-|\boldsymbol{v}|}$$

by Assumption SA2.2. Combining this with (SA6.27) gives

$$\left\|\frac{h^{d/2}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}} - \frac{h^{d/2}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})}{\sqrt{\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\right\|_1 \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}.$$

It is left to bound

$$\begin{aligned}
&\sup_{\boldsymbol{q},\boldsymbol{x}}\frac{h^{d/2}}{\sqrt{\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\big\|(\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \mathbf{Q}_{0,\boldsymbol{q}}^{-1})\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\big\|_1\\
&\overset{(a)}{\lesssim} h^{d+|\boldsymbol{v}|}\sup_{\boldsymbol{q},\boldsymbol{x}}\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \mathbf{Q}_{0,\boldsymbol{q}}^{-1}\|_1\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|_1\\
&\overset{(b)}{\lesssim_{\mathbb{P}}} h^{d+|\boldsymbol{v}|}\cdot h^{-d}\sqrt{\frac{\log(1/h)}{nh^d}}\cdot h^{-|\boldsymbol{v}|} = \sqrt{\frac{\log(1/h)}{nh^d}},
\end{aligned}$$

where in (a) we used $\inf_{\boldsymbol{q},\boldsymbol{x}}\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) \gtrsim h^{-d-2|\boldsymbol{v}|}$ by Lemma SA6.6, in (b) we used $\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \mathbf{Q}_{0,\boldsymbol{q}}^{-1}\|_1 \lesssim_{\mathbb{P}} h^{-d}\sqrt{\log(1/h)/(nh^d)}$ by Lemma SA3.11 and $\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|_1 \lesssim h^{-|\boldsymbol{v}|}$ by Assumption SA2.2.

Equation (SA6.30) follows from (SA6.29), (SA6.28) and $\log(1/h)/(nh^d) = o(1)$.

We will now prove (SA6.31). By the triangle inequality,

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\left\|\frac{\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})}{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{1/2}} - \frac{\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})}{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{1/2}}\right\|_1 \tag{SA6.33}$$

$$\lesssim \sup_{\boldsymbol{q},\boldsymbol{x}}\left\|\frac{(\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1})\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})}{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{1/2}}\right\|_1 + \sup_{\boldsymbol{q},\boldsymbol{x}}\big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\big(\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2} - \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2}\big)\big\|_1. \tag{SA6.34}$$

To bound the first term in (SA6.34), recall that $\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) \gtrsim_{\mathbb{P}} h^{-d-2|\boldsymbol{v}|}$ by Lemma SA6.6 and (SA6.14). Then

$$\begin{aligned}
\sup_{\boldsymbol{q},\boldsymbol{x}}\left\|\frac{(\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1})\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\right\|_1 &\lesssim_{\mathbb{P}} h^{d/2+|\boldsymbol{v}|}\sup_{\boldsymbol{q},\boldsymbol{x}}\big\|(\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1})\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\big\|_1\\
&\lesssim h^{d/2+|\boldsymbol{v}|}\sup_{\boldsymbol{q},\boldsymbol{x}}\|\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_1\cdot\|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|_1\\
&\lesssim h^{d/2}\|\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_\infty \lesssim_{\mathbb{P}} h^{d/2}\cdot h^{-d}r_{\mathtt{Q}} = h^{-d/2}r_{\mathtt{Q}},
\end{aligned}$$

where in the last inequality we used $\|\widehat{\mathbf{Q}}_{\boldsymbol{q}} - \bar{\mathbf{Q}}_{\boldsymbol{q}}\|_\infty \lesssim_{\mathbb{P}} h^d r_{\mathtt{Q}}$ by assumption, and

$$\|\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_\infty \leq \|\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_\infty \cdot \|\widehat{\mathbf{Q}}_{\boldsymbol{q}} - \bar{\mathbf{Q}}_{\boldsymbol{q}}\|_\infty \cdot \|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_\infty \lesssim_{\mathbb{P}} h^{-d}(h^d r_{\mathtt{Q}})h^{-d} = h^{-d} r_{\mathtt{Q}}$$

also by assumption and Lemma SA3.11.

It is left to bound the second term in (SA6.34):

$$\begin{aligned}
&\sup_{\boldsymbol{q},\boldsymbol{x}} \big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\big(\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2} - \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2}\big)\big\|_1 \\
&\quad \leq \sup_{\boldsymbol{q}} \|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\|_1 \sup_{\boldsymbol{x}} \|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|_1 \sup_{\boldsymbol{q},\boldsymbol{x}} \big|\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2} - \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2}\big| \\
&\quad \overset{\text{(a)}}{\lesssim_{\mathbb{P}}} h^{-d} \cdot h^{-|\boldsymbol{v}|} \cdot \sup_{\boldsymbol{q},\boldsymbol{x}} \big|\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2} - \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{-1/2}\big| \\
&\quad \overset{\text{(b)}}{\lesssim_{\mathbb{P}}} h^{-d-|\boldsymbol{v}|} h^{d/2+|\boldsymbol{v}|} r_{\mathtt{VC}} = h^{-d/2} r_{\mathtt{VC}},
\end{aligned}$$

where (a) is by Lemma SA3.11 and Assumption SA2.2, (b) is by (SA6.26). Equation (SA6.31) is proven.

Equation (SA6.32) follows from (SA6.31), (SA6.30) and $r_{\mathtt{Q}} + r_{\mathtt{VC}} = o(1)$. □

Lemma SA6.8 (Hats off)**.** *Suppose Assumptions SA2.1 to SA2.6, (SA6.12), (SA6.13) and (SA6.14) hold, $\nu \geq 3$, and $\frac{\log^3 n}{nh^{3d}} = o(1)$. Define (and fix for all further arguments) $r_{\mathtt{HO}}$ as an arbitrary positive sequence satisfying*

$$\sqrt{nh^d}(r_{\mathtt{UC}} r_{\mathtt{VC}} + r_{\mathtt{BR}}) = o(r_{\mathtt{HO}}). \tag{SA6.35}$$

Then

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \sup_{\boldsymbol{q}\in\mathcal{Q}} |T_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - t_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})| \lesssim_{\mathbb{P}} \sqrt{nh^d}(r_{\mathtt{UC}} r_{\mathtt{VC}} + r_{\mathtt{BR}}) = o(r_{\mathtt{HO}}). \tag{SA6.36}$$

Proof. First, note that by Lemma SA6.6, (SA6.12) and (SA6.14)

$$\begin{aligned}
&\sup_{\boldsymbol{q},\boldsymbol{x}} \left| \frac{\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})/n}} - \frac{\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})/n}} \right| \\
&\leq \sqrt{n} \sup_{\boldsymbol{q},\boldsymbol{x}} \big|\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})\big| \sup_{\boldsymbol{q},\boldsymbol{x}} \frac{1}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}\Big(\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})} + \sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}\Big)} \\
&\quad \cdot \sup_{\boldsymbol{q},\boldsymbol{x}} \big|\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\big| \\
&\lesssim_{\mathbb{P}} \sqrt{n} \cdot h^{-|\boldsymbol{v}|} r_{\mathtt{UC}} \cdot (h^{d+2|\boldsymbol{v}|})^{3/2} \cdot h^{-2|\boldsymbol{v}|-d} r_{\mathtt{VC}} \\
&= \sqrt{n} h^{d/2} r_{\mathtt{UC}} r_{\mathtt{VC}}.
\end{aligned}$$

By (SA6.13), Assumption SA2.2 and Lemma SA3.11, Eq. (SA3.12) in Lemma SA3.12,

$$\begin{aligned}
&\sup_{\boldsymbol{q},\boldsymbol{x}} \big|\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) + \boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}} \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1} \mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})]\big| \\
&\leq \sup_{\boldsymbol{q},\boldsymbol{x}} |\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - \mathsf{L}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})| \\
&\quad + \sup_{\boldsymbol{q},\boldsymbol{x}} \|\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})\|_1 \big\|\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1} - \mathbf{Q}_{0,\boldsymbol{q}}^{-1}\big\|_\infty \big\|\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})]\big\|_\infty
\end{aligned}$$

$$\lesssim h^{-|\boldsymbol{v}|} r_{\mathtt{BR}} + h^{-d-|\boldsymbol{v}|} \sqrt{\frac{\log(1/h)}{nh^d}} \sqrt{\frac{h^d \log(1/h)}{n}}$$
$$= h^{-|\boldsymbol{v}|} r_{\mathtt{BR}} + h^{-|\boldsymbol{v}|} \frac{\log(1/h)}{nh^d} \overset{\text{(a)}}{\lesssim} h^{-|\boldsymbol{v}|} r_{\mathtt{BR}}, \tag{SA6.37}$$

where in (a) we used $\frac{\log(1/h)}{nh^d} \lesssim \frac{\log n}{nh^d} \lesssim r_{\mathtt{BR}}$. Combining this with $h^{-2|\boldsymbol{v}|-d} \lesssim_{\mathbb{P}} \inf_{\boldsymbol{q},\boldsymbol{x}} |\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})|$ (by Lemma SA6.6)

$$\sup_{\boldsymbol{q},\boldsymbol{x}} \left| \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}} \widehat{\boldsymbol{\beta}}(\boldsymbol{q}) - \mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q})}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})/n}} - t_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) \right| \lesssim_{\mathbb{P}} \sqrt{n} \left(h^{d+2|\boldsymbol{v}|}\right)^{1/2} h^{-|\boldsymbol{v}|} r_{\mathtt{BR}} = \sqrt{nh^d} r_{\mathtt{BR}}. \tag{SA6.38}$$

It is left to apply the triangle inequality. □

SA7 Uniform inference

SA7.1 Plug-in approximation

Recall that the conditional covariance structure of $\boldsymbol{Z}(\cdot)$ is

$$\mathbb{E}\left[\boldsymbol{Z}(\boldsymbol{q})\boldsymbol{Z}(\tilde{\boldsymbol{q}})^{\mathsf{T}} \,\middle|\, \mathbf{X}_n\right] = h^{-d} \bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$$

with

$$\begin{aligned} \bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} &:= \mathbb{E}_n\left[S_{\boldsymbol{q},\tilde{\boldsymbol{q}}}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\tilde{\boldsymbol{q}}))\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\right] \in \mathcal{C}(\mathcal{Q}\times\mathcal{Q},\mathbb{R}^{K\times K}), \\ S_{\boldsymbol{q},\tilde{\boldsymbol{q}}}(\boldsymbol{x}) &:= \mathbb{E}\left[\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\tilde{\boldsymbol{q}}));\tilde{\boldsymbol{q}}) \,\middle|\, \boldsymbol{x}_i = \boldsymbol{x}\right]. \end{aligned} \tag{SA7.1}$$

Theorem SA7.1 (Plug-in approximation)**.** *Suppose that all of the following is true:*

(i) All the conditions of Theorem SA6.4(a) and Assumption SA2.7 hold.

(ii) The function $\boldsymbol{x} \mapsto S_{\boldsymbol{q},\tilde{\boldsymbol{q}}}(\boldsymbol{x})$ defined in (SA7.1) is continuous for all $\boldsymbol{q},\tilde{\boldsymbol{q}}$, and

$$\sup_{\boldsymbol{x},\boldsymbol{q},\boldsymbol{q}_1\neq\boldsymbol{q}_2} \frac{|S_{\boldsymbol{q},\boldsymbol{q}_1}(\boldsymbol{x}) - S_{\boldsymbol{q},\boldsymbol{q}_2}(\boldsymbol{x})|}{\|\boldsymbol{q}_1-\boldsymbol{q}_2\|} \lesssim 1. \tag{SA7.2}$$

(iii) There is an estimator $\widehat{S}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}(\boldsymbol{x})$ which is a known (not depending on $\mathbb{P}_{\mathbf{D}_n}$) measurable function of $(\mathbf{D}_n,\boldsymbol{q},\tilde{\boldsymbol{q}},\boldsymbol{x})$, satisfying the bound

$$\sup_{\boldsymbol{q},\tilde{\boldsymbol{q}},\boldsymbol{x}} \left|\widehat{S}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}(\boldsymbol{x}) - S_{\boldsymbol{q},\tilde{\boldsymbol{q}}}(\boldsymbol{x})\right| \lesssim_{\mathbb{P}} r_{\mathtt{S}} \quad \textit{with } r_{\mathtt{S}} = o(1). \tag{SA7.3}$$

Then, on the same probability space, there is a mean-zero Gaussian conditionally on $\mathbf{D}_n$ process $\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})$ in $\mathcal{C}(\mathcal{Q},\mathbb{R}^K)$, whose distribution is known (does not depend on $\mathbb{P}_{\mathbf{D}_n}$), such that

$$\mathbb{E}\left[\sup_{\boldsymbol{q}\in\mathcal{Q}} \left\|\boldsymbol{Z}^*(\boldsymbol{q}) - \widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\right\|_\infty \,\middle|\, \mathbf{D}_n\right] \lesssim_{\mathbb{P}} (r_{\mathtt{UC}} + r_{\mathtt{S}})^{1/(2d_{\mathcal{Q}}+2)} \sqrt{\log n}, \tag{SA7.4}$$

where $\boldsymbol{Z}^*(\boldsymbol{q})$ is a process in $\mathcal{C}(\mathcal{Q},\mathbb{R}^K)$ satisfying

$$(\mathbf{X}_n, \boldsymbol{Z}(\cdot)) \overset{d}{=} (\mathbf{X}_n, \boldsymbol{Z}^*(\cdot)), \quad \boldsymbol{Z}^*(\cdot) \underset{\mathbf{X}_n}{\perp\!\!\!\perp} \boldsymbol{y}_n.$$

Further,

$$h^{d/2} \sup_{\boldsymbol{q}\in\mathcal{Q}} \sup_{\boldsymbol{x}\in\mathcal{X}} \left| \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}} \widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}} \widehat{\boldsymbol{Z}}^*(\boldsymbol{q}) - \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}} \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}} \boldsymbol{Z}^*(\boldsymbol{q}) \right| \lesssim_{\mathbb{P}} r_{\mathtt{PI}}, \tag{SA7.5}$$

$$\text{where}\quad r_{\texttt{PI}} := \left[(r_{\texttt{UC}} + r_{\texttt{S}})^{1/(2d_{\mathcal{Q}}+2)} + r_{\texttt{Q}} + r_{\texttt{VC}}\right]\sqrt{\log n}. \tag{SA7.6}$$

Equivalently (see Lemma SA3.3), fixing an arbitrary positive sequence $R_{\texttt{PI}}$ *such that* $r_{\texttt{PI}} = o(R_{\texttt{PI}})$*, we have*

$$\mathbb{P}\left\{h^{d/2}\sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q}) - \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\boldsymbol{Z}^*(\boldsymbol{q})\right| > R_{\texttt{PI}}\;\middle|\;\boldsymbol{D}_n\right\} = o_{\mathbb{P}}(1).$$

Remark SA7.2. One-dimensional processes $Z^\star(\boldsymbol{x},\boldsymbol{q})$ and $\widehat{Z}(\boldsymbol{x},\boldsymbol{q})$ in the main paper are defined as follows:

$$Z^\star(\boldsymbol{x},\boldsymbol{q}) := \bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}^*(\boldsymbol{q}) = -h^{d/2}\frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\boldsymbol{Z}^*(\boldsymbol{q}),$$
$$\widehat{Z}(\boldsymbol{x},\boldsymbol{q}) := \hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q}) = -h^{d/2}\frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q}).$$

The rest of Section SA7.1 is devoted to proving Theorem SA7.1.

SA7.1.1 Strategy

First, $\boldsymbol{Z}^*(\cdot)$ exists by Theorem 8.17 (transfer) in [14] with $(\xi,\eta,\zeta) = (\mathbf{X}_n, \boldsymbol{Z}(\cdot), \boldsymbol{y}_n)$, $\tilde{\xi} = \xi$.

The rest of the proof will be broken up in two steps.

First, we will prove the existence of $\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})$. A natural approach is to approximate the covariance structure $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ using plug-in, giving an estimated covariance structure $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$, discretize $\mathcal{Q}$ by a suitable δ_n-covering $\mathcal{Q}^{\delta_n}$ and argue that since the discrete versions $\bar{\boldsymbol{\Sigma}}^{\delta_n}$ and $\widehat{\boldsymbol{\Sigma}}^{\delta_n}$ of $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ are close, the Gaussian vector $\boldsymbol{Z}(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}) \in \mathbb{R}^{K|\mathcal{Q}^{\delta_n}|}$ is close to a "feasible" Gaussian vector with covariance $\widehat{\boldsymbol{\Sigma}}^{\delta_n}$. Then we can embed this vector into a conditionally Gaussian process $\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})$. One technical caveat is that a simple plug-in method can lead to $\widehat{\boldsymbol{\Sigma}}^{\delta_n}$ not being positive-semidefinite, or to $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ not having a suitable Lipschitz property in $\boldsymbol{q}$ and $\tilde{\boldsymbol{q}}$ (necessary for controlling the discretization error between a continuous process and a discrete vector). For this reason, we just project $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ on the space of such structures that the wanted properties hold (program (SA7.9)).

The second step is proving (SA7.4) (given $\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})$ and $\boldsymbol{Z}^*(\boldsymbol{q})$ that are close) by bounding $\sup_{\boldsymbol{q}\in\mathcal{Q}}\boldsymbol{Z}^*(\boldsymbol{q})$ using the maximal inequality in Lemma SA3.7, and applying Lemma SA6.7.

SA7.1.2 Existence of the feasible process

As the first step, we will prove that under Conditions (i) to (iii), the process $\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})$ as described in Theorem SA7.1 exists.

Estimating the covariance structure By Lemma SA3.11, for a large enough constant C_{14} the probability of the event

$$\mathcal{A} := \{\lambda_{\max}(\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}]) \le C_{14}h^d\} \tag{SA7.7}$$

converges to one. Note that $S_{\boldsymbol{q},\tilde{\boldsymbol{q}}}(\boldsymbol{x})$ is bounded uniformly in $\boldsymbol{x}, \boldsymbol{q}, \tilde{\boldsymbol{q}}$ because it is continuous in these arguments and $\mathcal{Q}\times\mathcal{Q}\times\mathcal{X}$ is compact. Combining with Eq. (SA7.2), we see that there is a large enough C_{15} such that on $\mathcal{A}$

$$\sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\|\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\| \le C_{15}h^d$$

and

$$\sup_{\boldsymbol{q},\boldsymbol{q}_1\neq\boldsymbol{q}_2} \frac{|\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\boldsymbol{q}_1} - \bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\boldsymbol{q}_2}|}{\|\boldsymbol{q}_1 - \boldsymbol{q}_2\|} \leq C_{15} h^d.$$

We first approximate $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ by

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} := \mathbb{E}_n\big[\widehat{S}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}(\boldsymbol{x}_i)\eta^{(1)}(\widehat{\mu}(\boldsymbol{x}_i,\boldsymbol{q}))\eta^{(1)}(\widehat{\mu}(\boldsymbol{x}_i,\tilde{\boldsymbol{q}}))\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\big]. \tag{SA7.8}$$

The latter is known (a function of $\mathbf{D}_n$ not depending on $\mathbb{P}_{\mathbf{D}_n}$), but may not be positive semidefinite or Lipschitz in the arguments. In order to recover these properties, we solve the following program:

$$\begin{aligned}
\text{minimize} \quad & \sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \|\mathbf{M}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\| \qquad\qquad (\text{SA7.9})\\
\text{over} \quad & \mathbf{M}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \in \mathcal{C}(\mathcal{Q}\times\mathcal{Q},\mathbb{R}^{K\times K}),\\
& \sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \|\mathbf{M}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\| \leq C_{15} h^d,\\
& \mathbf{M}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} = \mathbf{M}_{\tilde{\boldsymbol{q}},\boldsymbol{q}} = \mathbf{M}^{\mathsf{T}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} = \mathbf{M}^{\mathsf{T}}_{\tilde{\boldsymbol{q}},\boldsymbol{q}}, \quad \boldsymbol{q},\tilde{\boldsymbol{q}} \in \mathcal{Q},\\
& \big(\mathbf{M}_{\boldsymbol{q}_k,\boldsymbol{q}_m}\big)_{k,m=1}^M \in \mathbb{R}^{MK\times MK} \text{ is symmetric positive semidefinite for any } (\boldsymbol{q}_k)_{k=1}^M,\\
& \|\mathbf{M}_{\boldsymbol{q},\boldsymbol{q}_1} - \mathbf{M}_{\boldsymbol{q},\boldsymbol{q}_2}\| \leq C_{15} h^d \|\boldsymbol{q}_1 - \boldsymbol{q}_2\|, \quad \boldsymbol{q},\boldsymbol{q}_1,\boldsymbol{q}_2 \in \mathcal{Q}.
\end{aligned}$$

Lemma SA7.3 (Solution to the program)**.** *There exists an exact minimizer $\widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ of (SA7.9) which is a (known) measurable function of $\mathbf{D}_n$ with values in $\mathcal{C}(\mathcal{Q}\times\mathcal{Q},\mathbb{R}^{K\times K})$.*

Proof. The space $\mathcal{C}(\mathcal{Q}\times\mathcal{Q},\mathbb{R}^{K\times K})$ with the sup-norm

$$\sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \|\mathbf{M}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\|$$

is Polish, and the feasible set is non-empty and compact by the multi-dimensional Arzelà–Ascoli theorem. The function

$$\big(\mathcal{X}^n\times\mathcal{Y},\mathcal{C}(\mathcal{Q}\times\mathcal{Q},\mathbb{R}^{K\times K})\big) \ni \big(\mathbf{D}_n,\mathbf{M}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\big) \mapsto \sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \|\mathbf{M}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\| \in \mathbb{R}$$

is a Carathéodory function, that is, measurable in $\mathbf{D}_n$ and continuous in $\mathbf{M}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$. By the Measurable Maximum Theorem 18.19 in [1] (here minimum), the result follows. □

Now we prove that $\widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ is sufficiently close to $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$.

Lemma SA7.4 (Consistency of $\widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$)**.** *The function $\widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ defined in Lemma SA7.3 satisfies*

$$\sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \big\|\widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\big\| \lesssim_{\mathbb{P}} h^d(r_{\mathtt{UC}} + r_{\mathtt{S}}).$$

Proof. From Eqs. (SA6.12) and (SA7.3) and Lemma SA3.11 we have

$$\sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \big\|\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\big\| \lesssim_{\mathbb{P}} h^d(r_{\mathtt{UC}} + r_{\mathtt{S}}). \tag{SA7.10}$$

But $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ lies in the feasibility set of (SA7.9) on $\mathcal{A}$, which implies that

$$\sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \big\|\widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\big\| \leq \sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \big\|\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\big\| \quad \text{on } \mathcal{A}.$$

Since $\mathbb{P}(\mathcal{A}) = 1 - o(1)$, we have

$$\sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \big\|\widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\big\| \lesssim_{\mathbb{P}} h^d(r_{\mathtt{UC}} + r_{\mathtt{S}}),$$

and by the triangle inequality the result follows. □

Discretization and controlling the deviations of the infeasible process Having estimated the covariance structure, we proceed with the same steps as in Lemma SA6.3. First, we discretize $\mathcal{Q}$. Let $\mathcal{Q}^{\delta_n} := \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_{|\mathcal{Q}^{\delta_n}|}\}$ be an internal δ_n-covering[2] of $\mathcal{Q}$ with respect to the 2-norm $\|\cdot\|$ of cardinality $|\mathcal{Q}^{\delta_n}| \lesssim 1/\delta_n^{d_{\mathcal{Q}}}$, where δ_n is chosen later. As already proven in Lemma SA6.3, there is a bound on the deviations of $\boldsymbol{Z}^*(\cdot)$:

$$\mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n} \|\boldsymbol{Z}^*(\boldsymbol{q}) - \boldsymbol{Z}^*(\tilde{\boldsymbol{q}})\|_\infty \,\Big|\, \mathbf{X}_n\Big] = \mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n} \|\boldsymbol{Z}^*(\boldsymbol{q}) - \boldsymbol{Z}^*(\tilde{\boldsymbol{q}})\|_\infty \,\Big|\, \mathbf{D}_n\Big] \lesssim \sqrt{\delta_n \log\frac{K}{\delta_n}} \quad \text{on } \mathcal{A},$$

where in the equality we used that $\boldsymbol{Z}^*(\cdot)$ is independent of $\boldsymbol{y}_n$ conditionally on $\mathbf{X}_n$.

Closeness of the discrete vectors Consider now the conditionally Gaussian vector $\boldsymbol{Z}^*(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}) \in \mathbb{R}^{K|\mathcal{Q}^{\delta_n}|}$ and denote

$$\bar{\boldsymbol{\Sigma}}^{\delta_n} := h^d \mathbb{V}[\boldsymbol{Z}^*(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}) \,|\, \mathbf{X}_n] = (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}_k,\boldsymbol{q}_m})_{k,m=1}^{|\mathcal{Q}^{\delta_n}|} \in \mathbb{R}^{K|\mathcal{Q}^{\delta_n}|\times K|\mathcal{Q}^{\delta_n}|},$$

so that

$$\boldsymbol{Z}^*(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}) = h^{-d/2}(\bar{\boldsymbol{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n, \quad \text{where } \mathbb{P}_{\boldsymbol{\xi}_n|\mathbf{X}_n} = \mathcal{N}\big(0, \mathbf{I}_{\mathbb{R}^{K|\mathcal{Q}^{\delta_n}|}}\big),$$

and $\boldsymbol{\xi}_n$ is independent of $\boldsymbol{y}_n$ given $\mathbf{X}_n$. (Since $\mathbb{P}_{\boldsymbol{\xi}_n|\mathbf{X}_n}$ does not depend on $\mathbf{X}_n$, this vector is in fact independent of $\mathbf{D}_n$.) Discretizing $\widehat{\boldsymbol{\Sigma}}^+_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ in the same way, we can put

$$\widehat{\boldsymbol{\Sigma}}^{\delta_n} := \big(\widehat{\boldsymbol{\Sigma}}^+_{\boldsymbol{q}_k,\boldsymbol{q}_m}\big)_{k,m=1}^{|\mathcal{Q}^{\delta_n}|} \in \mathbb{R}^{K|\mathcal{Q}^{\delta_n}|\times K|\mathcal{Q}^{\delta_n}|}.$$

Since the functions $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ and $\widehat{\boldsymbol{\Sigma}}^+_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ are close, the two matrices are close as well, which we will make precise in the following lemma.

Lemma SA7.5 (The matrices $\bar{\boldsymbol{\Sigma}}^{\delta_n}$ and $\widehat{\boldsymbol{\Sigma}}^{\delta_n}$ are close)**.** *We have*

$$\big\|\bar{\boldsymbol{\Sigma}}^{\delta_n} - \widehat{\boldsymbol{\Sigma}}^{\delta_n}\big\| \lesssim_{\mathbb{P}} |\mathcal{Q}^{\delta_n}| h^d (r_{\mathtt{UC}} + r_{\mathtt{S}}).$$

Proof. We can bound the Frobenius norm using the the bound on each element of the matrix:

$$\big\|\bar{\boldsymbol{\Sigma}}^{\delta_n} - \widehat{\boldsymbol{\Sigma}}^{\delta_n}\big\| \le \big\|\bar{\boldsymbol{\Sigma}}^{\delta_n} - \widehat{\boldsymbol{\Sigma}}^{\delta_n}\big\|_F \le |\mathcal{Q}^{\delta_n}| \sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \big\|\widehat{\boldsymbol{\Sigma}}^+_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\big\|.$$

It is left to combine with Lemma SA7.4. □

Applying Lemma SA3.5, we then have

$$\begin{aligned} h^{-d/2}\mathbb{E}\big[\big\|(\bar{\boldsymbol{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n - (\widehat{\boldsymbol{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n\big\|_\infty \,\big|\, \mathbf{D}_n\big] &\le 2h^{-d/2}\sqrt{\log(K|\mathcal{Q}^{\delta_n}|)}\|\bar{\boldsymbol{\Sigma}}^{\delta_n} - \widehat{\boldsymbol{\Sigma}}^{\delta_n}\|^{1/2} \\ &\lesssim_{\mathbb{P}} |\mathcal{Q}^{\delta_n}|^{1/2}(r_{\mathtt{UC}} + r_{\mathtt{S}})^{1/2}\sqrt{\log(K|\mathcal{Q}^{\delta_n}|)}. \end{aligned} \tag{SA7.11}$$

[2]This means $\mathcal{Q}_n^\delta \subset \mathcal{Q}$ and balls with radius δ_n centered at these points cover $\mathcal{Q}$.

Embedding the conditionally Gaussian vector $(\widehat{\mathbf{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n$ into a conditionally Gaussian process By the same argument as in Lemma SA6.3, there exists a law on $(\mathcal{X}\times\mathcal{Y})^n\times\mathcal{C}(\mathcal{Q},\mathbb{R}^K)$ which is the joint law of $\mathbf{D}_n'\overset{d}{=}\mathbf{D}_n$ and a conditionally on $\mathbf{D}_n'$ mean-zero Gaussian process $\widehat{\boldsymbol{Z}}_n^{*\prime}(\boldsymbol{q})$ with the conditional covariance structure

$$\mathbb{E}[\widehat{\boldsymbol{Z}}_n^{*\prime}(\boldsymbol{q})\widehat{\boldsymbol{Z}}_n^{*\prime}(\tilde{\boldsymbol{q}})^{\mathsf{T}}\,|\,\mathbf{D}_n']=h^{-d}\widehat{\mathbf{\Sigma}}'^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$$

(where $\widehat{\mathbf{\Sigma}}'^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ is the same function of the data as $\widehat{\mathbf{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ with $\mathbf{D}_n$ replaced by $\mathbf{D}_n'$). Continuity follows by the Kolmogorov-Chentsov theorem:

$$\mathbb{E}\big[\big\|\widehat{\boldsymbol{Z}}_n^{*\prime}(\boldsymbol{q})-\widehat{\boldsymbol{Z}}_n^{*\prime}(\tilde{\boldsymbol{q}})\big\|^a\,\big|\,\mathbf{D}_n'\big]\le C_n\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|^{d_{\mathcal{Q}}+b},\tag{SA7.12}$$

for $a,b>0$ chosen as follows. The vector $\widehat{\boldsymbol{Z}}_n^{*\prime}(\boldsymbol{q})-\widehat{\boldsymbol{Z}}_n^{*\prime}(\tilde{\boldsymbol{q}})$ is conditionally on $\mathbf{D}_n'$ Gaussian with covariance

$$h^{-d}\big(\widehat{\mathbf{\Sigma}}'^{+}_{\boldsymbol{q},\boldsymbol{q}}-2\widehat{\mathbf{\Sigma}}'^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}+\widehat{\mathbf{\Sigma}}'^{+}_{\tilde{\boldsymbol{q}},\tilde{\boldsymbol{q}}}\big).$$

Hence, we have for any $m>0$

$$\begin{aligned}&\mathbb{E}\big[\big\|\widehat{\boldsymbol{Z}}_n^{*\prime}(\boldsymbol{q})-\widehat{\boldsymbol{Z}}_n^{*\prime}(\tilde{\boldsymbol{q}})\big\|^{2m}\,\big|\,\mathbf{D}_n'\big]\\&\quad\le h^{-dm}\big\|\widehat{\mathbf{\Sigma}}'^{+}_{\boldsymbol{q},\boldsymbol{q}}-2\widehat{\mathbf{\Sigma}}'^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}+\widehat{\mathbf{\Sigma}}'^{+}_{\tilde{\boldsymbol{q}},\tilde{\boldsymbol{q}}}\big\|^m\mathbb{E}[\|\boldsymbol{\xi}_K\|^{2m}]\\&\quad\le(2C_{15})^m\mathbb{E}[\|\boldsymbol{\xi}_K\|^{2m}]\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|^m,\end{aligned}$$

where $\boldsymbol{\xi}_K\sim\mathcal{N}(0,\mathbf{I}_K)$. So we can take, for example, $m=d_{\mathcal{Q}}+1$, $a=2m$, $b=1$ in (SA7.12).

Projecting this $\mathcal{C}(\mathcal{Q},\mathbb{R}^K)$-process onto $\mathcal{Q}^{\delta_n}$, we obtain a vector $\widehat{\boldsymbol{Z}}_n^{*\prime}\big(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}\big)\in\mathbb{R}^{K|\mathcal{Q}^{\delta_n}|}$ such that

$$\big(\mathbf{D}_n',\widehat{\boldsymbol{Z}}_n^{*\prime}\big(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}\big)\big)\overset{d}{=}\big(\mathbf{D}_n,h^{-d/2}(\widehat{\mathbf{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n\big).$$

Since $\mathcal{C}(\mathcal{Q},\mathbb{R}^K)$ is Polish, by Theorem 8.17 (transfer) in [14] on our probability space there exists a random element $\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\in\mathcal{C}(\mathcal{Q},\mathbb{R}^K)$ such that

$$\big(\mathbf{D}_n',\widehat{\boldsymbol{Z}}_n^{*\prime}\big(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}\big),\widehat{\boldsymbol{Z}}_n^{*\prime}(\boldsymbol{q})\big)\overset{d}{=}\big(\mathbf{D}_n,h^{-d/2}(\widehat{\mathbf{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n,\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big).$$

In particular, almost surely $h^{-d/2}(\widehat{\mathbf{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n$ is the projection of $\widehat{\boldsymbol{Z}}^*$ on $\mathcal{Q}^{\delta_n}$.

Controlling the deviations of $\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})$ Consider the stochastic process X_n with index set

$$T:=\{(l,\boldsymbol{q},\tilde{\boldsymbol{q}})\colon l\in\{1,2,\dots,K\},\boldsymbol{q},\tilde{\boldsymbol{q}}\in\mathcal{Q},\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n\}$$

whose value at $t=(l,\boldsymbol{q},\tilde{\boldsymbol{q}})$ is $X_{n,t}:=\widehat{Z}_l^*(\boldsymbol{q})-\widehat{Z}_l^*(\tilde{\boldsymbol{q}})$. It is a separable (because each $\widehat{Z}_l^*(\cdot)$ has continuous trajectories) mean-zero Gaussian conditionally on $\mathbf{D}_n$ process with the index set T considered a metric space: $\operatorname{dist}\big((l,\boldsymbol{q},\tilde{\boldsymbol{q}}),(l',\boldsymbol{q}',\tilde{\boldsymbol{q}}')\big)=\|\boldsymbol{q}-\boldsymbol{q}'\|+\|\tilde{\boldsymbol{q}}-\tilde{\boldsymbol{q}}'\|+\mathbb{1}\{l\neq l'\}$.

The quantity $\sigma(X_n)^2$ is defined and bounded as follows:

$$\sigma(X_n)^2:=\sup_{t\in T}\mathbb{E}[X_{n,t}^2\,|\,\mathbf{D}_n]=\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\max_l\mathbb{E}\big[\big(\widehat{Z}_l^*(\boldsymbol{q})-\widehat{Z}_l^*(\tilde{\boldsymbol{q}})\big)^2\,\big|\,\mathbf{D}_n\big]\lesssim\delta_n,$$

because $\widehat{Z}_l^*(\boldsymbol{q})-\widehat{Z}_l^*(\tilde{\boldsymbol{q}})$ is conditionally on $\mathbf{D}_n$ mean-zero normal variable with variance

$$h^{-d}\big((\widehat{\Sigma}^{+}_{\boldsymbol{q},\boldsymbol{q}})_{ll}-2(\widehat{\Sigma}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}})_{ll}+(\widehat{\Sigma}^{+}_{\tilde{\boldsymbol{q}},\tilde{\boldsymbol{q}}})_{ll}\big)$$

$$
\begin{aligned}
&\le h^{-d}\big|(\widehat{\Sigma}^{+}_{\boldsymbol{q},\boldsymbol{q}})_{ll} - (\widehat{\Sigma}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}})_{ll}\big| + h^{-d}\big|(\widehat{\Sigma}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}})_{ll} - (\widehat{\Sigma}^{+}_{\tilde{\boldsymbol{q}},\tilde{\boldsymbol{q}}})_{ll}\big| \\
&\le h^{-d}\big\|\widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\boldsymbol{q}} - \widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\big\| + h^{-d}\big\|\widehat{\boldsymbol{\Sigma}}^{+}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \widehat{\boldsymbol{\Sigma}}^{+}_{\tilde{\boldsymbol{q}},\tilde{\boldsymbol{q}}}\big\| \\
&\le 2C_{15}\|\boldsymbol{q} - \tilde{\boldsymbol{q}}\|.
\end{aligned}
$$

Next, we define and bound the semimetric $\rho(t,t')$:

$$
\begin{aligned}
\rho(t,t')^2 &:= \mathbb{E}[(X_{n,t} - X_{n,t'})^2 \mid \mathbf{D}_n] \\
&= \mathbb{E}\Big[\Big(\big(\widehat{Z}^*_l(\boldsymbol{q}) - \widehat{Z}^*_l(\tilde{\boldsymbol{q}})\big) - \big(\widehat{Z}^*_{l'}(\boldsymbol{q}') - \widehat{Z}^*_{l'}(\tilde{\boldsymbol{q}}')\big)\Big)^2 \,\Big|\, \mathbf{D}_n\Big] \\
&\lesssim \mathbb{E}\Big[\big(\widehat{Z}^*_l(\boldsymbol{q}) - \widehat{Z}^*_{l'}(\boldsymbol{q}')\big)^2 \,\Big|\, \mathbf{D}_n\Big] - \mathbb{E}\Big[\big(\widehat{Z}^*_l(\tilde{\boldsymbol{q}}) - \widehat{Z}^*_{l'}(\tilde{\boldsymbol{q}}')\big)^2 \,\Big|\, \mathbf{D}_n\Big].
\end{aligned}
$$

The first term on the right is bounded the following way: if $l \neq l'$,

$$
\begin{aligned}
&\mathbb{E}\Big[\big(\widehat{Z}^*_l(\boldsymbol{q}) - \widehat{Z}^*_{l'}(\boldsymbol{q}')\big)^2 \,\Big|\, \mathbf{D}_n\Big] \\
&\quad\lesssim \mathbb{E}\Big[\big(\widehat{Z}^*_l(\boldsymbol{q}) - \widehat{Z}^*_l(\boldsymbol{q}')\big)^2 \,\Big|\, \mathbf{D}_n\Big] + \mathbb{E}\big[\widehat{Z}^*_l(\boldsymbol{q}')^2 \,\big|\, \mathbf{D}_n\big] + \mathbb{E}\big[\widehat{Z}^*_{l'}(\boldsymbol{q}')^2 \,\big|\, \mathbf{D}_n\big] \\
&\quad\lesssim \|\boldsymbol{q} - \boldsymbol{q}'\| + 1,
\end{aligned}
$$

and if $l = l'$,

$$
\mathbb{E}\Big[\big(\widehat{Z}^*_l(\boldsymbol{q}) - \widehat{Z}^*_{l'}(\boldsymbol{q}')\big)^2 \,\Big|\, \mathbf{D}_n\Big] \lesssim \|\boldsymbol{q} - \boldsymbol{q}'\|.
$$

The term $\mathbb{E}\Big[\big(\widehat{Z}^*_l(\tilde{\boldsymbol{q}}) - \widehat{Z}^*_{l'}(\tilde{\boldsymbol{q}}')\big)^2 \,\Big|\, \mathbf{D}_n\Big]$ is bounded the same way, and we conclude that

$$
\rho(t,t')^2 \le C_{16}\big(\|\boldsymbol{q} - \boldsymbol{q}'\| + \|\tilde{\boldsymbol{q}} - \tilde{\boldsymbol{q}}'\| + \mathbb{1}\{l \neq l'\}\big) = C_{16}\mathrm{dist}\big((l,\boldsymbol{q},\tilde{\boldsymbol{q}}),(l',\boldsymbol{q}',\tilde{\boldsymbol{q}}')\big).
$$

for some positive constant C_{16}. This means that, for any $\varepsilon > 0$, an $C_{16}^{-1}\varepsilon^2$-covering of T with respect to $\mathrm{dist}(\cdot)$ induces an ε-covering of T with respect to ρ, and hence

$$
N(T,\rho,\varepsilon) \le N\big(T,\mathrm{dist}(\cdot),C_{16}^{-1}\varepsilon^2\big).
$$

But the right-hand side is clearly $O(K\varepsilon^{-2})$ because T is a subset of $\{1,\dots,K\} \times \mathcal{Q} \times \mathcal{Q}$, so for $\varepsilon \in (0,1)$

$$
\log N(T,\rho,\varepsilon) \lesssim \log\frac{K}{\varepsilon}.
$$

Applying Lemma SA3.7 gives

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in T}|X_{n,t}| \,\Big|\, \mathbf{D}_n\Big] = \mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\big\|\widehat{\boldsymbol{Z}}^*(\boldsymbol{q}) - \widehat{\boldsymbol{Z}}^*(\tilde{\boldsymbol{q}})\big\|_\infty \,\Big|\, \mathbf{D}_n\Big] \lesssim \sigma(X_n) + \int_0^{2\sigma(X_n)} \sqrt{\log N(T,\rho,\varepsilon)}\,\mathrm{d}\varepsilon \\
&\lesssim \sigma(X_n) + \int_0^{2\sigma(X_n)} \sqrt{\log(K/\varepsilon)}\,\mathrm{d}\varepsilon \lesssim \sigma(X_n) + \int_0^{2\sigma(X_n)} \big(\sqrt{\log K} + \sqrt{\log(\varepsilon^{-1})}\big)\,\mathrm{d}\varepsilon \\
&\lesssim \sigma(X_n)\sqrt{\log K} + \frac{1}{\sqrt{\log(\sigma(X_n)^{-1})}}\int_0^{2\sigma(X_n)} \log(\varepsilon^{-1})\,\mathrm{d}\varepsilon \\
&\lesssim \sigma(X_n)\sqrt{\log K} + \frac{1 + \log(\sigma(X_n)^{-1})}{\sqrt{\log(\sigma(X_n)^{-1})}}\sigma(X_n) \lesssim \sigma(X_n)\sqrt{\log(K\sigma(X_n)^{-1})} \lesssim \sqrt{\delta_n\log\frac{K}{\delta_n}},
\end{aligned}
$$

where in the last inequality we used the bound $\sigma(X_n) \lesssim \sqrt{\delta_n}$ from above and that $x \mapsto x\log(1/x)$ is increasing for sufficiently small x.

Choosing δ_n and conclusion Combining the bounds obtained above, we can write

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\boldsymbol{Z}^*(\boldsymbol{q})-\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big\|_\infty \,\Big|\, \mathbf{D}_n\Big]\\
&\quad\le \mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\big\|\boldsymbol{Z}^*(\boldsymbol{q})-\boldsymbol{Z}^*(\tilde{\boldsymbol{q}})\big\|_\infty \,\Big|\, \mathbf{D}_n\Big] + h^{-d/2}\mathbb{E}\big[\big\|(\bar{\boldsymbol{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n-(\widehat{\boldsymbol{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n\big\|_\infty \,\big|\, \mathbf{D}_n\big]\\
&\qquad + \mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\big\|\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})-\widehat{\boldsymbol{Z}}^*(\tilde{\boldsymbol{q}})\big\|_\infty \,\Big|\, \mathbf{D}_n\Big]\\
&\quad\lesssim_{\mathbb{P}} \left(\sqrt{\delta_n}+\frac{(r_{\mathtt{UC}}+r_{\mathtt{S}})^{1/2}}{\delta_n^{d_{\mathcal{Q}}/2}}\right)\sqrt{\log\frac{K}{\delta_n}}.
\end{aligned}
$$

Choose the approximately optimal

$$\delta_n := (r_{\mathtt{UC}}+r_{\mathtt{S}})^{1/(d_{\mathcal{Q}}+1)},$$

giving

$$\mathbb{E}\Big[\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\boldsymbol{Z}^*(\boldsymbol{q})-\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big\|_\infty \,\Big|\, \mathbf{D}_n\Big] \lesssim_{\mathbb{P}} (r_{\mathtt{UC}}+r_{\mathtt{S}})^{1/(2d_{\mathcal{Q}}+2)}\sqrt{\log n}.$$

SA7.1.3 Proof of (SA7.5)

Having proven the existence of $\widehat{\boldsymbol{Z}}^*(\cdot)$, we will now prove that under Conditions (i) to (iii), Eq. (SA7.5) holds. We build on the proof of Theorem 6.3 in [6].

Lemma SA7.6 (Bounding the supremum of the Gaussian process)**.** *In the setting of Theorem SA7.1, we have* $\sup_{\boldsymbol{q}\in\mathcal{Q}}\|\boldsymbol{Z}(\boldsymbol{q})\|_\infty \lesssim_{\mathbb{P}} \sqrt{\log K}$.

Proof. Consider the stochastic process X_n defined for $t=(l,\boldsymbol{q})\in T$ with

$$T := \{(l,\boldsymbol{q}) : l\in\{1,2,\dots,K\}, \boldsymbol{q}\in\mathcal{Q}\}$$

as $X_{n,t} := Z_{n,l}(\boldsymbol{q})$. It is a separable mean-zero Gaussian process with the index set T considered a metric space: $\operatorname{dist}\big((l,\boldsymbol{q}),(l',\boldsymbol{q}')\big) = \|\boldsymbol{q}-\boldsymbol{q}'\| + \mathbb{1}\{l\neq l'\}$. Note that

$$
\begin{aligned}
\sigma(X_n)^2 &:= \sup_{t\in T}\mathbb{E}\big[X_{n,t}^2 \,\big|\, \mathbf{X}_n\big] = \sup_{\boldsymbol{q}\in\mathcal{Q}}\max_l \mathbb{E}\big[Z_{n,l}^2(\boldsymbol{q}) \,\big|\, \mathbf{X}_n\big]\\
&= h^{-d}\sup_{\boldsymbol{q}\in\mathcal{Q}}\max_l \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})^2 \,\big|\, \boldsymbol{x}_i\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))^2 p_l(\boldsymbol{x}_i)^2\\
&\lesssim h^{-d}\mathbb{E}_n[p_l(\boldsymbol{x}_i)^2] \le h^{-d}\sup_{\boldsymbol{\alpha}\in\mathcal{S}^{K-1}}\mathbb{E}_n\big[\boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{\alpha}\big] \lesssim 1 \quad \text{w. p. a. 1.}
\end{aligned}
$$

Next, we will bound

$$\rho(t,t')^2 := \mathbb{E}\big[(X_{n,t}-X_{n,t'})^2 \,\big|\, \mathbf{X}_n\big] = \mathbb{E}\big[(Z_{n,l}(\boldsymbol{q})-Z_{n,l'}(\boldsymbol{q}'))^2 \,\big|\, \mathbf{X}_n\big].$$

If $l\neq l'$,

$$
\begin{aligned}
&\mathbb{E}\big[(Z_{n,l}(\boldsymbol{q})-Z_{n,l'}(\boldsymbol{q}'))^2 \,\big|\, \mathbf{X}_n\big]\\
&\quad= \frac{1}{n}\sum_{i=1}^n h^{-d}\mathbb{E}\Big[\Big(\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}));\boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))p_l(\boldsymbol{x}_i)
\end{aligned}
$$

$$
\begin{aligned}
&\quad -\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}')); \boldsymbol{q}')\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}'))p_{l'}(\boldsymbol{x}_i)\Big)^2 \mathbf{X}_n\Big] \\
&= h^{-d}\mathbb{E}_n\big[\mathbb{E}\big[(A_n(\boldsymbol{q}, y_i, \boldsymbol{x}_i) - A_n(\boldsymbol{q}', y_i, \boldsymbol{x}_i))^2 \,\big|\, \mathbf{X}_n\big] p_l(\boldsymbol{x}_i)^2\big] \\
&\quad + h^{-d}\mathbb{E}_n\big[\mathbb{E}\big[A_n(\boldsymbol{q}', y_i, \boldsymbol{x}_i)^2 \,\big|\, \mathbf{X}_n\big]\big(p_l(\boldsymbol{x}_i)^2 + p_{l'}(\boldsymbol{x}_i)^2\big)\big] \\
&\lesssim h^{-d}\|\boldsymbol{q} - \boldsymbol{q}'\|\mathbb{E}_n\big[p_l(\boldsymbol{x}_i)^2\big] + h^{-d}\mathbb{E}_n\big[p_l(\boldsymbol{x}_i)^2 + p_{l'}(\boldsymbol{x}_i)^2\big] \\
&\lesssim \|\boldsymbol{q} - \boldsymbol{q}'\| + 1 \quad \text{w. p. a. } 1,
\end{aligned}
$$

where we denoted $A_n(\boldsymbol{q}, y_i, \boldsymbol{x}_i) := \psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \boldsymbol{q})); \boldsymbol{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \boldsymbol{q}))$ to simplify notation. Similarly, if $l = l'$,

$$
\mathbb{E}\big[(Z_{n,l}(\boldsymbol{q}) - Z_{n,l'}(\boldsymbol{q}'))^2 \,\big|\, \mathbf{X}_n\big] \lesssim \|\boldsymbol{q} - \boldsymbol{q}'\| \quad \text{w. p. a. } 1.
$$

We conclude

$$
\rho(t, t')^2 \lesssim \operatorname{dist}\big((l, \boldsymbol{q}), (l', \boldsymbol{q}')\big).
$$

This means that an ε^2-covering of T with respect to $\operatorname{dist}(\cdot)$ induces an ε-covering of T with respect to ρ, and hence

$$
N(T, \rho, \varepsilon) \le N(T, \operatorname{dist}(\cdot), \varepsilon^2). \tag{SA7.13}
$$

On the other hand, since $\mathcal{Q}$ does not depend on n, clearly for (sufficiently small) $\tilde{\varepsilon} > 0$, $N(T, \operatorname{dist}(\cdot), \tilde{\varepsilon}) \lesssim K(C_1/\tilde{\varepsilon})^{C_2}$, where C_1 and C_2 are both constants. Combining this with (SA7.13) we get

$$
\log N(T, \rho, \varepsilon) \lesssim \log(K/\varepsilon).
$$

Now we apply Lemma SA3.7:

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{t \in T}|X_{n,t}| \,\Big|\, \mathbf{X}_n\Big] &\lesssim \sigma(X_n) + \int_0^{2\sigma(X_n)} \sqrt{\log N(T, \rho, \varepsilon)}\, \mathrm{d}\varepsilon \\
&\lesssim \sigma(X_n) + \sigma(X_n)\sqrt{\log(K/\sigma(X_n))} \lesssim_1 \sqrt{\log K} \quad \text{w. p. a. } 1,
\end{aligned}
$$

where in $\lesssim_1$ we used our bound $\sigma(X_n) \lesssim 1$ above. Rewriting, we obtain

$$
\mathbb{E}\Big[\sup_{\boldsymbol{q} \in \mathcal{Q}}\|\boldsymbol{Z}(\boldsymbol{q})\|_\infty \,\Big|\, \mathbf{X}_n\Big] \lesssim \sqrt{\log K} \quad \text{w. p. a. } 1. \tag{SA7.14}
$$

By Markov's inequality this gives $\sup_{\boldsymbol{q} \in \mathcal{Q}}\|\boldsymbol{Z}(\boldsymbol{q})\|_\infty \lesssim_{\mathbb{P}} \sqrt{\log K}$. □

By the triangle inequality, Eq. (SA7.5) will follow from the two bounds

$$
h^{d/2}\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q}) - \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})}}\boldsymbol{Z}^*(\boldsymbol{q})\right| \lesssim_{\mathbb{P}} (r_{\mathtt{UC}} + r_{\mathtt{S}})^{1/(2d_{\mathcal{Q}}+2)}\sqrt{\log n}, \tag{SA7.15}
$$

$$
h^{d/2}\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})}}\boldsymbol{Z}^*(\boldsymbol{q}) - \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})}}\boldsymbol{Z}^*(\boldsymbol{q})\right| \lesssim_{\mathbb{P}} (r_{\mathtt{Q}} + r_{\mathtt{VC}})\sqrt{\log n}. \tag{SA7.16}
$$

To prove (SA7.15), combine $\sup_{\boldsymbol{q},\boldsymbol{x}}\big\|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})\big\|_1 \lesssim_{\mathbb{P}} 1$ by Lemma SA6.7 with (SA7.4).

To prove (SA7.16), combine Lemma SA7.6 with $\sup_{\boldsymbol{q},\boldsymbol{x}}\big\|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q}) - \bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})\big\|_1 \lesssim_{\mathbb{P}} r_{\mathtt{Q}} + r_{\mathtt{VC}}$ by Lemma SA6.7.

This completes the proof of Theorem SA7.1.

SA7.2 Confidence bands

As a corollary of Theorem SA7.1, we can obtain the following result, whose proof is based on [2].

Theorem SA7.7 (Confidence bands)**.** *Suppose all conditions of Theorem SA7.1 are true, and* $(R_{\mathtt{PI}} + r_{\mathtt{SA}} + r_{\mathtt{HO}})\sqrt{\log n} = o(1)$. *Then* $\mathbb{P}\{\sup_{\boldsymbol{q},\boldsymbol{x}} |T_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})| < k^*(1-\alpha)\} = 1-\alpha+o(1)$, *where* $k^*(\eta)$ *is the conditional* η*-quantile of* $\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big|$ *given the data. Equivalently,* $\mathbb{P}\{\mu_0^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) \in \mathsf{CB}_{1-\alpha}(\boldsymbol{x},\boldsymbol{q},\boldsymbol{v})\} = 1-\alpha+o(1)$ *with*

$$\mathsf{CB}_{1-\alpha}(\boldsymbol{x},\boldsymbol{q},\boldsymbol{v}) := \Big(\widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) - k^*(1-\alpha)\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})/n}, \widehat{\mu}^{(\boldsymbol{v})}(\boldsymbol{x},\boldsymbol{q}) + k^*(1-\alpha)\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})/n}\Big). \tag{SA7.17}$$

To prove this result we first introduce some auxiliary lemmas. To simplify the exposition, let

$$V := \sup_{\boldsymbol{q},\boldsymbol{x}} |T_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})|, \tag{SA7.18}$$

$$V^* := \sup_{\boldsymbol{q},\boldsymbol{x}} \left| h^{d/2} \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}} \widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}} \widehat{\boldsymbol{Z}}^*(\boldsymbol{q}) \right|, \tag{SA7.19}$$

$$\widetilde{V} := \sup_{\boldsymbol{q},\boldsymbol{x}} \left| h^{d/2} \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}} \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}} \boldsymbol{Z}(\boldsymbol{q}) \right|, \tag{SA7.20}$$

$$\widetilde{V}^* := \sup_{\boldsymbol{q},\boldsymbol{x}} \left| h^{d/2} \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}} \bar{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}} \boldsymbol{Z}^*(\boldsymbol{q}) \right|. \tag{SA7.21}$$

Let $\tilde{k}^*(\eta)$ denote the conditional η-quantile of $\widetilde{V}^*$ given $\mathbf{X}_n$.

Lemma SA7.8 (Closeness rates)**.** *Random variables* V, $\widetilde{V}$, V^*, $\widetilde{V}^*$ *satisfy the following:*

(a) $|V - \widetilde{V}| = o_{\mathbb{P}}(r_{\mathtt{SA}} + r_{\mathtt{HO}})$;

(b) $|V^* - \widetilde{V}^*| = o_{\mathbb{P}}(R_{\mathtt{PI}})$, *where* $R_{\mathtt{PI}}$ *is defined in Theorem SA7.1;*

(c) $\widetilde{V}^* \perp\!\!\!\perp_{\mathbf{X}_n} \boldsymbol{y}_n$.

Proof. By Theorem SA6.4, we have

$$\left| V - \sup_{\boldsymbol{q},\boldsymbol{x}} \left|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}} \boldsymbol{Z}(\boldsymbol{q})\right| \right| = o_{\mathbb{P}}(r_{\mathtt{SA}} + r_{\mathtt{HO}}),$$

which is Assertion (a).

By Theorem SA7.1, we have

$$\left| V^* - \sup_{\boldsymbol{q},\boldsymbol{x}} \left|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}} \boldsymbol{Z}^*(\boldsymbol{q})\right| \right| = o_{\mathbb{P}}(R_{\mathtt{PI}}),$$

which is Assertion (b).

Assertion (c) follows from the definition of the process $\boldsymbol{Z}^*(\cdot)$ and the fact that $\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})$ only depends on the data via $\mathbf{X}_n$. □

Lemma SA7.9 (First sequence)**.** *There exists a sequence of positive numbers* $\nu_{n,1} \to 0$ *such that w. p. a. 1*

$$k^*(1-\alpha) \le \tilde{k}^*(1-\alpha+\nu_{n,1}) + R_{\mathtt{PI}} \quad \textit{and} \quad k^*(1-\alpha) \ge \tilde{k}^*(1-\alpha-\nu_{n,1}) - R_{\mathtt{PI}}.$$

Proof. This follows from $|V^*-\widetilde{V}^*| = o_{\mathbb{P}}(R_{\texttt{PI}})$ by Lemma SA7.8, directly applying Lemma SA3.8. $\square$

Lemma SA7.10 (Second sequence)**.** *There exists a constant $C_{\widetilde{V}^*} > 0$ such that with $1-o(1)$ probability*

$$\sup_{u\in\mathbb{R}} \mathbb{P}\big\{\big|\widetilde{V}^* - u\big| \le R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}} \,\big|\, \boldsymbol{D}_n\big\} \le C_{\widetilde{V}^*}(R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}})\sqrt{\log(1/h)} =: \nu_{n,2},$$

Moreover, for the sequence $\nu_{n,2} \to 0$ just defined, the following holds w. p. a. 1:

$$\tilde{k}^*(1-\alpha-\nu_{n,1}) - \tilde{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2}) \ge R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}}, \tag{SA7.22}$$

$$\tilde{k}^*(1-\alpha+\nu_{n,1}+\nu_{n,2}) - \tilde{k}^*(1-\alpha+\nu_{n,1}) \ge R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}}. \tag{SA7.23}$$

Proof. By Lemma SA3.9, using that $\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}^*(\boldsymbol{q})$ is a separable mean-zero Gaussian conditionally on $\mathbf{X}_n$ process on $\mathcal{Q}\times\mathcal{X}$ with $\mathbb{E}\big[\big(\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}^*(\boldsymbol{q})\big)^2 \,\big|\, \mathbf{X}_n\big] = 1$, we have with $1-o(1)$ probability

$$\begin{aligned}
&\sup_{u\in\mathbb{R}} \mathbb{P}\big\{\big|\widetilde{V}^* - u\big| \le R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}} \,\big|\, \boldsymbol{D}_n\big\} \\
&\quad= \sup_{u\in\mathbb{R}} \mathbb{P}\big\{\big|\widetilde{V}^* - u\big| \le R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}} \,\big|\, \mathbf{X}_n\big\} \\
&\quad\le 4(R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}})\bigg(\mathbb{E}\Big[\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}^*(\boldsymbol{q})\big| \,\Big|\, \mathbf{X}_n\Big] + 1\bigg) \\
&\quad\le 4(R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}})\bigg(\sup_{\boldsymbol{q},\boldsymbol{x}}\big\|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\big\|_1 \mathbb{E}\Big[\sup_{\boldsymbol{q}}\big\|\boldsymbol{Z}^*(\boldsymbol{q})\big\|_\infty \,\Big|\, \mathbf{X}_n\Big] + 1\bigg) \\
&\quad\overset{\text{(a)}}{\lesssim} (R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}})\Big(\mathbb{E}\Big[\sup_{\boldsymbol{q}}\big\|\boldsymbol{Z}^*(\boldsymbol{q})\big\|_\infty \,\Big|\, \mathbf{X}_n\Big] + 1\Big) \\
&\quad\overset{\text{(b)}}{\lesssim} (R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}})\sqrt{\log(1/h)},
\end{aligned}$$

where in (a) we used $\big\|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\big\|_1 \lesssim 1$ w. p. a. 1 by Lemma SA6.7; (b) is by Lemma SA7.6.

We will now prove (SA7.22). Note that $\sup_{u\in\mathbb{R}} \mathbb{P}\big\{|\widetilde{V}^*-u| \le R_{\texttt{PI}}+r_{\texttt{SA}}+r_{\texttt{HO}} \,\big|\, \boldsymbol{D}_n\big\} \le \nu_{n,2}$ w. p. a. 1 implies that w. p. a. 1

$$\sup_{u\in\mathbb{R}} \mathbb{P}\Big\{u < \widetilde{V}^* \le u + R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}} \,\Big|\, \boldsymbol{D}_n\Big\} \le \nu_{n,2},$$

that is,

$$\sup_{u\in\mathbb{R}} \big\{\mathbb{P}\big\{\widetilde{V}^* \le u + R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}} \,\big|\, \boldsymbol{D}_n\big\} - \mathbb{P}\big\{\widetilde{V}^* \le u \,\big|\, \boldsymbol{D}_n\big\}\big\} \le \nu_{n,2}.$$

Since this is true for any u, we can in particular replace u with a random variable $\tilde{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2})$. Using $\mathbb{P}\big\{\widetilde{V}^* \le \tilde{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2}) \,\big|\, \boldsymbol{D}_n\big\} \ge 1-\alpha-\nu_{n,1}-\nu_{n,2}$, this gives w. p. a. 1

$$\mathbb{P}\big\{\widetilde{V}^* \le \tilde{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2}) + R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}} \,\big|\, \boldsymbol{D}_n\big\} - (1-\alpha-\nu_{n,1}-\nu_{n,2}) \le \nu_{n,2}$$

or

$$\mathbb{P}\big\{\widetilde{V}^* \le \tilde{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2}) + R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}} \,\big|\, \boldsymbol{D}_n\big\} \le 1-\alpha-\nu_{n,1}.$$

By monotonicity of a (conditional) distribution function, this means that w. p. a. 1

$$\tilde{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2}) + R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}} \le \tilde{k}^*(1-\alpha-\nu_{n,1}).$$

This proves the inequality (SA7.22). The inequality (SA7.23) is proven similarly. $\square$

Concluding the proof of Theorem SA7.7. Note that

$$\begin{aligned}
\mathbb{P}\{V > k^*(1-\alpha)\} &\overset{\text{(a)}}{\leq} \mathbb{P}\Big\{V > \tilde{k}^*(1-\alpha-\nu_{n,1}) - R_{\texttt{PI}}\Big\} \\
&\overset{\text{(b)}}{\leq} \mathbb{P}\Big\{\widetilde{V} > \tilde{k}^*(1-\alpha-\nu_{n,1}) - (R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}})\Big\} + o(1) \\
&\overset{\text{(c)}}{\leq} \mathbb{P}\Big\{\widetilde{V} > \tilde{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2})\Big\} + o(1) \\
&\overset{\text{(d)}}{=} \mathbb{E}\Big[\mathbb{P}\Big\{\widetilde{V} > \tilde{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2}) \,\Big|\, \mathbf{X}_n\Big\}\Big] + o(1) \\
&\overset{\text{(e)}}{\leq} \alpha + \nu_{n,1} + \nu_{n,2} + o(1) = \alpha + o(1),
\end{aligned}$$

where (a) is by Lemma SA7.9, (b) is by Assertion (a) in Lemma SA7.8, (c) is by Lemma SA7.10, (d) is by the law of iterated expectations, (e) is by the definition of a conditional quantile and using that $\widetilde{V}$ has the same conditional distribution as $\widetilde{V}^*$.

Similarly,

$$\begin{aligned}
\mathbb{P}\big\{V > k^*(1-\alpha)\big\} &\overset{\text{(a)}}{\geq} \mathbb{P}\big\{V > \tilde{k}^*(1-\alpha+\nu_{n,1}) + R_{\texttt{PI}}\big\} + o(1) \\
&\overset{\text{(b)}}{\geq} \mathbb{P}\big\{\widetilde{V} > \tilde{k}^*(1-\alpha+\nu_{n,1}) + (R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}})\big\} + o(1) \\
&\overset{\text{(c)}}{\geq} \mathbb{P}\big\{\widetilde{V} > \tilde{k}^*(1-\alpha+\nu_{n,1}+\nu_{n,2})\big\} + o(1) \\
&\overset{\text{(d)}}{=} \mathbb{E}\big[\mathbb{P}\big\{\widetilde{V} > \tilde{k}^*(1-\alpha+\nu_{n,1}+\nu_{n,2}) \,\big|\, \mathbf{X}_n\big\}\big] + o(1) \\
&\overset{\text{(e)}}{=} \alpha - \nu_{n,1} - \nu_{n,2} + o(1) = \alpha + o(1),
\end{aligned}$$

where (a) is by Lemma SA7.9, (b) is by Assertion (a) in Lemma SA7.8, (c) is by Lemma SA7.10, (d) is by the law of iterated expectations. In (e) we used that the distribution function of $\widetilde{V}$ conditional on $\mathbf{X}_n$ is continuous w. p. a. 1, because by the same anti-concentration argument as in the proof of Lemma SA7.10 there is a positive constant C such that on an event $\mathcal{A}_n$ satisfying $\mathbb{P}\{\mathcal{A}_n\} \to 1$ we have for any $\varepsilon > 0$

$$\sup_{u\in\mathbb{R}} \mathbb{P}\big\{|\widetilde{V} - u| \leq \varepsilon \,\big|\, \mathbf{X}_n\big\} \leq C\varepsilon\sqrt{\log(1/h)}.$$

In particular, on $\mathcal{A}_n$ all jumps of the distribution function of $\widetilde{V}$ conditional on $\mathbf{X}_n$ are bounded by $C\varepsilon\sqrt{\log(1/h)}$, which implies that the distribution function is continuous on $\mathcal{A}_n$, since ε is arbitrary. Theorem SA7.7 is proven. □

The following theorem uses a Gaussian anti-concentration result from [11] to convert $o_{\mathbb{P}}(\cdot)$ bounds obtained above to the bound on the Kolmogorov-Smirnov distance (sup-norm distance of distribution functions), similarly to [6].

SA7.3 Kolmogorov-Smirnov distance bound

Theorem SA7.11 (Kolmogorov-Smirnov Distance)**.** *Suppose all conditions of Theorem SA7.1 are true, and* $(R_{\texttt{PI}} + r_{\texttt{SA}} + r_{\texttt{HO}})\sqrt{\log n} = o(1)$*. Then*

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\Big\{\sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}|T_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})| \leq u\Big\} - \mathbb{P}\left\{\sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\frac{h^{d/2}\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}}{\sqrt{\widehat{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\right| \leq u \,\middle|\, \boldsymbol{D}_n\right\}\right| = o_{\mathbb{P}}(1).$$

Proof. We will rely on the following lemma which is proven in Section SA7.3.1 by the same discretization argument as in Theorem SA7.1.

Lemma SA7.12. *On the same probability space, there exists a mean-zero unconditionally Gaussian process $\widetilde{\boldsymbol{Z}}(\cdot)$ in $\mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$ such that*

$$\widetilde{\boldsymbol{Z}}(\cdot) \perp\!\!\!\perp \mathbf{X}_n, \tag{SA7.24}$$

$$\mathbb{E}\Big[\big(\boldsymbol{\ell_v}(\boldsymbol{x}, \boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big)^2\Big] = 1, \tag{SA7.25}$$

$$\mathbb{E}\Big[\sup_{\boldsymbol{q}\in\mathcal{Q}} \big\|\boldsymbol{Z}(\boldsymbol{q}) - \widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big\|_\infty \,\Big|\, \mathbf{X}_n\Big] \lesssim_{\mathbb{P}} \left(\frac{\log n}{nh^d}\right)^{1/(4d_{\mathcal{Q}}+4)} \sqrt{\log n}. \tag{SA7.26}$$

Also, there exists a process $\widetilde{\boldsymbol{Z}}^*(\cdot)$ in $\mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$, such that

$$\widetilde{\boldsymbol{Z}}^*(\cdot) \perp\!\!\!\perp \mathbf{D}_n, \tag{SA7.27}$$

$$\big(\mathbf{X}_n, \boldsymbol{Z}(\cdot), \widetilde{\boldsymbol{Z}}(\cdot)\big) \overset{d}{=} \big(\mathbf{X}_n, \boldsymbol{Z}^*(\cdot), \widetilde{\boldsymbol{Z}}^*(\cdot)\big), \tag{SA7.28}$$

and in particular

$$\mathbb{E}\Big[\sup_{\boldsymbol{q}\in\mathcal{Q}} \big\|\boldsymbol{Z}^*(\boldsymbol{q}) - \widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\big\|_\infty \,\Big|\, \mathbf{X}_n\Big] \lesssim_{\mathbb{P}} \left(\frac{\log n}{nh^d}\right)^{1/(4d_{\mathcal{Q}}+4)} \sqrt{\log n}. \tag{SA7.29}$$

Since $\widetilde{\boldsymbol{Z}}^*(\cdot) \overset{d}{=} \widetilde{\boldsymbol{Z}}(\cdot)$,

$$\mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}} \big|\boldsymbol{\ell_v}(\boldsymbol{x}, \boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big| \le u\Big\} = \mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}} \big|\boldsymbol{\ell_v}(\boldsymbol{x}, \boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\big| \le u\Big\}.$$

This means that, by the triangle inequality, it is sufficient to prove the following bounds:

$$\sup_{u\in\mathbb{R}} \Big|\mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}} \big|T_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})\big| \le u\Big\} - \mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}} \big|\boldsymbol{\ell_v}(\boldsymbol{x}, \boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big| \le u\Big\}\Big| = o(1), \tag{SA7.30}$$

$$\sup_{u\in\mathbb{R}} \Big|\mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}} \big|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})^{\mathsf{T}} \widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big| \le u \,\Big|\, \boldsymbol{D}_n\Big\} - \mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}} \big|\boldsymbol{\ell_v}(\boldsymbol{x}, \boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\big| \le u\Big\}\Big| = o_{\mathbb{P}}(1). \tag{SA7.31}$$

We will now prove (SA7.30). Note that for any random variables ξ and η and any $s > 0$,

$$\sup_{u\in\mathbb{R}} \big|\mathbb{P}\{\xi \le u\} - \mathbb{P}\{\eta \le u\}\big| \le \sup_{u\in\mathbb{R}} \mathbb{P}\{|\eta - u| \le s\} + \mathbb{P}\{|\xi - \eta| > s\}, \tag{SA7.32}$$

which follows from the two bounds

$$\begin{aligned}
&\mathbb{P}\{\xi \le u\} \le \mathbb{P}\{\xi \le u \text{ and } |\xi - \eta| \le s\} + \mathbb{P}\{|\xi - \eta| > s\}\\
&\quad \le \mathbb{P}\{\eta \le u + s\} + \mathbb{P}\{|\xi - \eta| > s\}\\
&\quad \le \mathbb{P}\{\eta \le u\} + \mathbb{P}\{u < \eta \le u + s\} + \mathbb{P}\{|\xi - \eta| > s\};\\
&\mathbb{P}\{\xi > u\} \le \mathbb{P}\{\xi > u \text{ and } |\xi - \eta| \le s\} + \mathbb{P}\{|\xi - \eta| > s\}\\
&\quad \le \mathbb{P}\{\eta > u - s\} + \mathbb{P}\{|\xi - \eta| > s\}\\
&\quad \le \mathbb{P}\{\eta > u\} + \mathbb{P}\{u - s < \eta \le u\} + \mathbb{P}\{|\xi - \eta| > s\}.
\end{aligned}$$

Strategy To obtain (SA7.30), we will apply (SA7.32) with $\xi = \sup_{\boldsymbol{q},\boldsymbol{x}} |T_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})|$, $\eta = \sup_{\boldsymbol{q},\boldsymbol{x}} \big|\boldsymbol{\ell_v}(\boldsymbol{x}, \boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big|$ and $s = r_{\mathtt{SA}} + r_{\mathtt{HO}}$, use that the term $\mathbb{P}\{|\xi - \eta| > s\}$ is $o(1)$ because $\big|T_{\boldsymbol{v}}(\boldsymbol{x}, \boldsymbol{q})\big|$ and $\big|\boldsymbol{\ell_v}(\boldsymbol{x}, \boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big|$ are close, and apply Gaussian anti-concentration Lemma SA3.9 to show that $\sup_{u\in\mathbb{R}} \mathbb{P}\{|\eta - u| \le s\}$ is also $o(1)$. The argument for (SA7.31) will be similar.

Proof of (SA7.30) By $\sup_{\boldsymbol{q},\boldsymbol{x}}\|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim 1$ (from Lemma SA6.7) and (SA7.26), a simple application of Lemma SA3.2 gives

$$h^{d/2}\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\mathbf{Q}_{0,\boldsymbol{q}}^{-1}}{\sqrt{\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\boldsymbol{Z}(\boldsymbol{q}) - \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\mathbf{Q}_{0,\boldsymbol{q}}^{-1}}{\sqrt{\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\right| = o_{\mathbb{P}}(r_{\mathtt{SA}}). \tag{SA7.33}$$

By (SA6.29) in Lemma SA6.7, we have $\sup_{\boldsymbol{q},\boldsymbol{x}}\|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$, which by Lemma SA7.6 gives $\sup_{\boldsymbol{q},\boldsymbol{x}}|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}(\boldsymbol{q}) - \boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}(\boldsymbol{q})| \lesssim_{\mathbb{P}} \frac{\log(1/h)}{\sqrt{nh^d}} = o(r_{\mathtt{SA}})$. Combining it with Theorem SA6.4, Lemma SA6.8 and (SA7.33),

$$\mathbb{P}\Big\{\Big|\sup_{\boldsymbol{q},\boldsymbol{x}}|T_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})| - \sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big|\Big| > r_{\mathtt{SA}} + r_{\mathtt{HO}}\Big\} = o(1).$$

Then we can apply (SA7.32) with $\xi = \sup_{\boldsymbol{q},\boldsymbol{x}}|T_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})|$, $\eta = \sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big|$ and $s = r_{\mathtt{SA}} + r_{\mathtt{HO}}$, and get

$$\begin{aligned}
&\sup_{u\in\mathbb{R}}\Big|\mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}}|T_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})| \le u\Big\} - \mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big| \le u\Big\}\Big| \\
&\quad\le \sup_{u\in\mathbb{R}}\mathbb{P}\Big\{\Big|\sup_{\boldsymbol{q},\boldsymbol{x}}\Big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\Big| - u\Big| \le r_{\mathtt{SA}} + r_{\mathtt{HO}}\Big\} + o(1).
\end{aligned}$$

For (SA7.30), it is left to show that

$$\sup_{u\in\mathbb{R}}\mathbb{P}\Big\{\Big|\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big| - u\Big| \le r_{\mathtt{SA}} + r_{\mathtt{HO}}\Big\} = o(1).$$

We will show a stronger version

$$\sup_{u\in\mathbb{R}}\mathbb{P}\Big\{\Big|\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big| - u\Big| \le R_{\mathtt{PI}} + r_{\mathtt{SA}} + r_{\mathtt{HO}}\Big\} = o(1). \tag{SA7.34}$$

We apply Lemma SA3.9:

$$\begin{aligned}
&\sup_{u\in\mathbb{R}}\mathbb{P}\Big\{\Big|\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big| - u\Big| \le R_{\mathtt{PI}} + r_{\mathtt{SA}} + r_{\mathtt{HO}}\Big\} \\
&\quad\le 4(R_{\mathtt{PI}} + r_{\mathtt{SA}} + r_{\mathtt{HO}})\Big(\mathbb{E}\Big[\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big|\Big] + 1\Big) \\
&\quad\le 4(R_{\mathtt{PI}} + r_{\mathtt{SA}} + r_{\mathtt{HO}})\Big(\mathbb{E}\Big[\sup_{\boldsymbol{q},\boldsymbol{x}}\|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1\big\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big\|_\infty\Big] + 1\Big) \\
&\quad\overset{\text{(a)}}{\lesssim} (R_{\mathtt{PI}} + r_{\mathtt{SA}} + r_{\mathtt{HO}})\Big(\mathbb{E}\Big[\sup_{\boldsymbol{q}}\big\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big\|_\infty\Big] + 1\Big) \\
&\quad\overset{\text{(b)}}{\lesssim} (R_{\mathtt{PI}} + r_{\mathtt{SA}} + r_{\mathtt{HO}})\sqrt{\log(1/h)},
\end{aligned} \tag{SA7.35}$$

where in (a) we used $\|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim 1$, see (SA6.28); (b) is by

$$\begin{aligned}
\mathbb{E}\Big[\sup_{\boldsymbol{q}}\big\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big\|_\infty\Big] &= \mathbb{E}\Big[\sup_{\boldsymbol{q}}\big\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big\|_\infty\,\Big|\,\mathbf{X}_n\Big] \\
&\le \mathbb{E}\Big[\sup_{\boldsymbol{q}}\big\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q}) - \boldsymbol{Z}(\boldsymbol{q})\big\|_\infty\,\Big|\,\mathbf{X}_n\Big] + \mathbb{E}\Big[\sup_{\boldsymbol{q}}\big\|\boldsymbol{Z}(\boldsymbol{q})\big\|_\infty\,\Big|\,\mathbf{X}_n\Big]
\end{aligned}$$

$$\overset{\text{(c)}}{=} o_{\mathbb{P}}(r_{\mathtt{SA}}) + O_{\mathbb{P}}\big(\sqrt{\log(1/h)}\big) \overset{\text{(d)}}{=} O_{\mathbb{P}}\big(\sqrt{\log(1/h)}\big)$$

using (SA7.26) and Lemma SA7.6 for (c), $r_{\mathtt{SA}} \lesssim 1 \lesssim \sqrt{\log(1/h)}$ for (d), and noting that

$$\mathbb{E}\Big[\sup_{\boldsymbol{q}}\big\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big\|_\infty\Big] \lesssim_{\mathbb{P}} \sqrt{\log(1/h)}$$

is equivalent to

$$\mathbb{E}\Big[\sup_{\boldsymbol{q}}\big\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big\|_\infty\Big] \lesssim \sqrt{\log(1/h)}.$$

Since $(R_{\mathtt{PI}} + r_{\mathtt{SA}} + r_{\mathtt{HO}})\sqrt{\log(1/h)} = o(1)$, the right-hand side in (SA7.35) is $o(1)$, proving (SA7.34), which was sufficient for (SA7.30).

Proof of (SA7.31) By $\sup_{\boldsymbol{q},\boldsymbol{x}}\|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim 1$ (from Lemma SA6.7) and (SA7.29), a simple application of Lemma SA3.2 gives

$$h^{d/2}\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\mathbf{Q}_{0,\boldsymbol{q}}^{-1}}{\sqrt{\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\boldsymbol{Z}^*(\boldsymbol{q}) - \frac{\boldsymbol{p}^{(\boldsymbol{v})}(\boldsymbol{x})^{\mathsf{T}}\mathbf{Q}_{0,\boldsymbol{q}}^{-1}}{\sqrt{\Omega_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})}}\widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\right| = o_{\mathbb{P}}(r_{\mathtt{SA}}). \tag{SA7.36}$$

By Theorem SA7.1, we have $\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q}) - \bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\boldsymbol{Z}^*(\boldsymbol{q})\big| = o_{\mathbb{P}}(R_{\mathtt{PI}})$. By (SA6.29) in Lemma SA6.7, we have $\sup_{\boldsymbol{q},\boldsymbol{x}}\|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q}) - \boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})\|_1 \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$, which by Lemma SA7.6 gives $\sup_{\boldsymbol{q},\boldsymbol{x}}|\bar{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}^*(\boldsymbol{q}) - \boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}^*(\boldsymbol{q})| \lesssim_{\mathbb{P}} \frac{\log(1/h)}{\sqrt{nh^d}} = o(r_{\mathtt{SA}})$. By the triangle inequality,

$$\sup_{\boldsymbol{q},\boldsymbol{x}}|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q}) - \boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}^*(\boldsymbol{q})| = o_{\mathbb{P}}(R_{\mathtt{PI}} + r_{\mathtt{SA}}).$$

Combining with (SA7.36) and applying the triangle inequality again, we obtain

$$\mathbb{P}\Big\{\Big|\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big| - \sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\big|\Big| > R_{\mathtt{PI}} + r_{\mathtt{SA}}\Big\} = o(1),$$

implying by Markov's inequality that for any $\varepsilon > 0$,

$$\mathbb{P}\Big\{\mathbb{P}\Big\{\Big|\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big| - \sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\big|\Big| > R_{\mathtt{PI}} + r_{\mathtt{SA}}\,\Big|\,\boldsymbol{D}_n\Big\} > \varepsilon\Big\} = o(1),$$

that is,

$$\mathbb{P}\Big\{\Big|\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big| - \sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\big|\Big| > R_{\mathtt{PI}} + r_{\mathtt{SA}}\,\Big|\,\boldsymbol{D}_n\Big\} = o_{\mathbb{P}}(1).$$

Then we can apply (SA7.32) with $\mathbb{P}\{\cdot\,|\,\boldsymbol{D}_n\}$ instead of $\mathbb{P}\{\cdot\}$, $\xi = \sup_{\boldsymbol{q},\boldsymbol{x}}\big|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big|$, $\eta = \sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\big|$ and $s = R_{\mathtt{PI}} + r_{\mathtt{SA}}$, and get

$$\begin{aligned}
&\sup_{u\in\mathbb{R}}\Big|\mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\hat{\boldsymbol{\ell}}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widehat{\boldsymbol{Z}}^*(\boldsymbol{q})\big| \le u\,\Big|\,\boldsymbol{D}_n\Big\} - \mathbb{P}\Big\{\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\big| \le u\Big\}\Big| \\
&\quad \le \sup_{u\in\mathbb{R}}\mathbb{P}\Big\{\Big|\sup_{\boldsymbol{q},\boldsymbol{x}}\big|\boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})\big| - u\Big| \le R_{\mathtt{PI}} + r_{\mathtt{SA}}\Big\} + o_{\mathbb{P}}(1)
\end{aligned}$$

$$= \sup_{u\in\mathbb{R}} \mathbb{P}\Big\{ \Big| \sup_{\boldsymbol{q},\boldsymbol{x}} \Big| \boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}(\boldsymbol{q}) \Big| - u \Big| \le R_{\mathtt{PI}} + r_{\mathtt{SA}} \Big\} + o_{\mathbb{P}}(1),$$

where we used that $\widetilde{\boldsymbol{Z}}^*(\boldsymbol{q})$ is independent of the data allowing us to remove the conditioning on $\boldsymbol{D}_n$, and again that $\widetilde{\boldsymbol{Z}}(\cdot)$ and $\widetilde{\boldsymbol{Z}}^*(\cdot)$ have the same laws.

It is left to use that

$$\begin{aligned} &\sup_{u\in\mathbb{R}} \mathbb{P}\Big\{ \Big| \sup_{\boldsymbol{q},\boldsymbol{x}} \big| \boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}(\boldsymbol{q}) \big| - u \Big| \le R_{\mathtt{PI}} + r_{\mathtt{SA}} \Big\} \\ &\qquad \le \sup_{u\in\mathbb{R}} \mathbb{P}\Big\{ \Big| \sup_{\boldsymbol{q},\boldsymbol{x}} \big| \boldsymbol{\ell}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}} \widetilde{\boldsymbol{Z}}(\boldsymbol{q}) \big| - u \Big| \le R_{\mathtt{PI}} + r_{\mathtt{SA}} + r_{\mathtt{HO}} \Big\} \overset{\text{(a)}}{=} o(1), \end{aligned}$$

where (a) is by (SA7.34).

Theorem SA7.11 is proven. □

SA7.3.1 Proof of Lemma SA7.12: construction of $\widetilde{\boldsymbol{Z}}(\cdot)$ and $\widetilde{\boldsymbol{Z}}^*(\cdot)$

Recall that the conditional covariance structure of $\boldsymbol{Z}(\cdot)$ is

$$\mathbb{E}\big[\boldsymbol{Z}(\boldsymbol{q})\boldsymbol{Z}(\tilde{\boldsymbol{q}})^{\mathsf{T}} \,\big|\, \mathbf{X}_n\big] = h^{-d} \bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$$

with $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ defined in (SA7.1). The matrix $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ is approximated by the non-random matrix $\boldsymbol{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ defined by

$$\boldsymbol{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} := \mathbb{E}\big[S_{\boldsymbol{q},\tilde{\boldsymbol{q}}}(\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\boldsymbol{q}))\eta^{(1)}(\mu_0(\boldsymbol{x}_i,\tilde{\boldsymbol{q}}))\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\big].$$

By the same argument as in Lemma SA3.11, we have

$$\sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \|\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \boldsymbol{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\| \lesssim_{\mathbb{P}} h^d \bigg(\frac{\log(1/h)}{nh^d}\bigg)^{1/2} =: h^d r_{\boldsymbol{\Sigma}}. \tag{SA7.37}$$

Discretization and controlling the deviations of $\boldsymbol{Z}(\cdot)$ Then we proceed with the same discretization as in Theorem SA7.1. Let $\mathcal{Q}^{\delta_n} := \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_{|\mathcal{Q}^{\delta_n}|}\}$ be an internal δ_n-covering of $\mathcal{Q}$ with respect to the 2-norm $\|\cdot\|$ of cardinality $|\mathcal{Q}^{\delta_n}| \lesssim 1/\delta_n^{d_{\mathcal{Q}}}$, where δ_n is chosen later. As already proven in Lemma SA6.3, there is a bound on the deviations of $\boldsymbol{Z}(\cdot)$:

$$\mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n} \|\boldsymbol{Z}(\boldsymbol{q}) - \boldsymbol{Z}(\tilde{\boldsymbol{q}})\|_\infty \,\Big|\, \mathbf{X}_n\Big] \lesssim \sqrt{\delta_n \log\frac{K}{\delta_n}} \quad \text{on } \mathcal{A},$$

where $\mathcal{A}$ is the event from (SA7.7).

Closeness of the discrete vectors Consider now the conditionally Gaussian vector $\boldsymbol{Z}(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}) \in \mathbb{R}^{K|\mathcal{Q}^{\delta_n}|}$ and denote

$$\bar{\boldsymbol{\Sigma}}^{\delta_n} := h^d \mathbb{V}[\boldsymbol{Z}(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}) \,|\, \mathbf{X}_n] = (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q}_k,\boldsymbol{q}_m})_{k,m=1}^{|\mathcal{Q}^{\delta_n}|} \in \mathbb{R}^{K|\mathcal{Q}^{\delta_n}|\times K|\mathcal{Q}^{\delta_n}|},$$

so that

$$\boldsymbol{Z}(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}) = h^{-d/2}(\bar{\boldsymbol{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n, \quad \text{where } \mathbb{P}_{\boldsymbol{\xi}_n|\mathbf{X}_n} = \mathcal{N}\big(0, \mathbf{I}_{\mathbb{R}^{K|\mathcal{Q}^{\delta_n}|}}\big).$$

(Since $\mathbb{P}_{\boldsymbol{\xi}_n|\mathbf{X}_n}$ does not depend on $\mathbf{X}_n$, the vector $\boldsymbol{\xi}_n$ is independent of $\mathbf{X}_n$.) Discretizing $\boldsymbol{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ in the same way, we can put

$$\boldsymbol{\Sigma}^{\delta_n} := \big(\boldsymbol{\Sigma}_{\boldsymbol{q}_k,\boldsymbol{q}_m}\big)_{k,m=1}^{|\mathcal{Q}^{\delta_n}|} \in \mathbb{R}^{K|\mathcal{Q}^{\delta_n}|\times K|\mathcal{Q}^{\delta_n}|}.$$

Since the functions $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ are close, the two matrices are close as well, which we will make precise in the following lemma.

Lemma SA7.13 (The matrices $\bar{\mathbf{\Sigma}}^{\delta_n}$ and $\mathbf{\Sigma}^{\delta_n}$ are close)**.** *We have*

$$\big\|\bar{\mathbf{\Sigma}}^{\delta_n} - \mathbf{\Sigma}^{\delta_n}\big\| \lesssim_{\mathbb{P}} |\mathcal{Q}^{\delta_n}| h^d r_{\mathbf{\Sigma}}.$$

Proof. We can bound the Frobenius norm using the the bound on each element of the matrix:

$$\big\|\bar{\mathbf{\Sigma}}^{\delta_n} - \mathbf{\Sigma}^{\delta_n}\big\| \le \big\|\bar{\mathbf{\Sigma}}^{\delta_n} - \mathbf{\Sigma}^{\delta_n}\big\|_F \le |\mathcal{Q}^{\delta_n}| \sup_{\boldsymbol{q},\tilde{\boldsymbol{q}}} \big\|\bar{\mathbf{\Sigma}}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} - \mathbf{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}\big\|.$$

It is left to combine with Eq. (SA7.37). $\square$

Applying Lemma SA3.5, we then have

$$\begin{aligned} h^{-d/2}\mathbb{E}\big[\big\|(\bar{\mathbf{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n - (\mathbf{\Sigma}^{\delta_n})^{1/2}\boldsymbol{\xi}_n\big\|_\infty \,\big|\, \mathbf{X}_n\big] &\le 2h^{-d/2}\sqrt{\log(K|\mathcal{Q}^{\delta_n}|)}\|\bar{\mathbf{\Sigma}}^{\delta_n} - \mathbf{\Sigma}^{\delta_n}\|^{1/2} \\ &\lesssim_{\mathbb{P}} |\mathcal{Q}^{\delta_n}|^{1/2} {r_{\mathbf{\Sigma}}}^{1/2}\sqrt{\log(K|\mathcal{Q}^{\delta_n}|)}. \end{aligned} \tag{SA7.38}$$

Embedding the conditionally Gaussian vector into a conditionally Gaussian process By the same argument as in Lemma SA6.3, there exists a law on $\mathcal{X}^n \times \mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$ which is the joint law of $\mathbf{X}'_n \overset{d}{=} \mathbf{X}_n$ and a conditionally on $\mathbf{X}'_n$ mean-zero Gaussian process $\widetilde{\boldsymbol{Z}}'_n(\boldsymbol{q})$ with the conditional covariance structure

$$\mathbb{E}\big[\widetilde{\boldsymbol{Z}}'_n(\boldsymbol{q})\widetilde{\boldsymbol{Z}}'_n(\tilde{\boldsymbol{q}})^{\mathsf{T}} \,\big|\, \mathbf{X}'_n\big] = h^{-d}\mathbf{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}.$$

Continuity follows by the Kolmogorov-Chentsov theorem:

$$\mathbb{E}\big[\big\|\widetilde{\boldsymbol{Z}}'_n(\boldsymbol{q}) - \widetilde{\boldsymbol{Z}}'_n(\tilde{\boldsymbol{q}})\big\|^a \,\big|\, \mathbf{X}'_n\big] \le C_n \|\boldsymbol{q} - \tilde{\boldsymbol{q}}\|^{d_{\mathcal{Q}}+b}, \tag{SA7.39}$$

for $a, b > 0$ chosen as follows. The vector $\widetilde{\boldsymbol{Z}}'_n(\boldsymbol{q}) - \widetilde{\boldsymbol{Z}}'_n(\tilde{\boldsymbol{q}})$ is conditionally on $\mathbf{X}'_n$ Gaussian with covariance

$$h^{-d}\big(\mathbf{\Sigma}_{\boldsymbol{q},\boldsymbol{q}} - 2\mathbf{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} + \mathbf{\Sigma}_{\tilde{\boldsymbol{q}},\tilde{\boldsymbol{q}}}\big).$$

Hence, we have for any $m > 0$

$$\begin{aligned} &\mathbb{E}\big[\big\|\widetilde{\boldsymbol{Z}}'_n(\boldsymbol{q}) - \widetilde{\boldsymbol{Z}}'_n(\tilde{\boldsymbol{q}})\big\|^{2m} \,\big|\, \mathbf{X}'_n\big] \\ &\quad \le h^{-dm}\big\|\mathbf{\Sigma}_{\boldsymbol{q},\boldsymbol{q}} - 2\mathbf{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}} + \mathbf{\Sigma}_{\tilde{\boldsymbol{q}},\tilde{\boldsymbol{q}}}\big\|^m \mathbb{E}[\|\boldsymbol{\xi}_K\|^{2m}] \\ &\quad \le C_n \|\boldsymbol{q} - \tilde{\boldsymbol{q}}\|^m, \end{aligned}$$

where $\boldsymbol{\xi}_K \sim \mathcal{N}(0, \mathbf{I}_K)$, and we used that $\mathbf{\Sigma}_{\boldsymbol{q},\tilde{\boldsymbol{q}}}$ is Lipschitz in $\tilde{\boldsymbol{q}}$ a. s. (with the constant allowed to depend on n). So we can take, for example, $m = d_{\mathcal{Q}} + 1$, $a = 2m$, $b = 1$ in (SA7.39).

Projecting this $\mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$-process onto $\mathcal{Q}^{\delta_n}$, we obtain a vector $\widetilde{\boldsymbol{Z}}'_n\big(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}\big) \in \mathbb{R}^{K|\mathcal{Q}^{\delta_n}|}$ such that

$$\big(\mathbf{X}'_n, \widetilde{\boldsymbol{Z}}'_n\big(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}\big)\big) \overset{d}{=} \big(\mathbf{X}_n, h^{-d/2}(\mathbf{\Sigma}^{\delta_n})^{1/2}\boldsymbol{\xi}_n\big).$$

Since $\mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$ is Polish, by Theorem 8.17 (transfer) in [14] on our probability space there exists a random element $\widetilde{\boldsymbol{Z}}(\boldsymbol{q}) \in \mathcal{C}(\mathcal{Q}, \mathbb{R}^K)$ such that

$$\big(\mathbf{X}'_n, \widetilde{\boldsymbol{Z}}'_n\big(\boldsymbol{q}|_{\mathcal{Q}^{\delta_n}}\big), \widetilde{\boldsymbol{Z}}'_n(\boldsymbol{q})\big) \overset{d}{=} \big(\mathbf{X}_n, h^{-d/2}(\mathbf{\Sigma}^{\delta_n})^{1/2}\boldsymbol{\xi}_n, \widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big),$$

In particular, almost surely $h^{-d/2}(\mathbf{\Sigma}^{\delta_n})^{1/2}\boldsymbol{\xi}_n$ is the projection of $\widetilde{\boldsymbol{Z}}(\cdot)$ on $\mathcal{Q}^{\delta_n}$. Note that $\mathbb{P}_{\widetilde{\boldsymbol{Z}}(\boldsymbol{q})|\mathbf{X}_n}$ does not depend on $\mathbf{X}_n$, which means $\widetilde{\boldsymbol{Z}}(\boldsymbol{q})$ is independent of $\mathbf{X}_n$. In addition, note that (SA7.25) holds.

Controlling the deviations of $\widetilde{\boldsymbol{Z}}(\boldsymbol{q})$ It is proven by the same argument as in Lemma SA6.3 and Theorem SA7.1 that

$$\mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q})-\widetilde{\boldsymbol{Z}}(\tilde{\boldsymbol{q}})\|_\infty \,\Big|\, \mathbf{X}_n\Big]=\mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q})-\widetilde{\boldsymbol{Z}}(\tilde{\boldsymbol{q}})\|_\infty\Big]\lesssim\sqrt{\delta_n\log\frac{K}{\delta_n}}.$$

Choosing δ_n and conclusion Combining the bounds obtained above, we can write

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\boldsymbol{Z}(\boldsymbol{q})-\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big\|_\infty \,\Big|\, \mathbf{X}_n\Big]\\
&\quad\le \mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\big\|\boldsymbol{Z}(\boldsymbol{q})-\boldsymbol{Z}(\tilde{\boldsymbol{q}})\big\|_\infty \,\Big|\, \mathbf{X}_n\Big]+h^{-d/2}\mathbb{E}\big[\big\|(\bar{\boldsymbol{\Sigma}}^{\delta_n})^{1/2}\boldsymbol{\xi}_n-(\boldsymbol{\Sigma}^{\delta_n})^{1/2}\boldsymbol{\xi}_n\big\|_\infty \,\big|\, \mathbf{X}_n\big]\\
&\qquad+\mathbb{E}\Big[\sup_{\|\boldsymbol{q}-\tilde{\boldsymbol{q}}\|\le\delta_n}\big\|\widetilde{\boldsymbol{Z}}(\boldsymbol{q})-\widetilde{\boldsymbol{Z}}(\tilde{\boldsymbol{q}})\big\|_\infty \,\Big|\, \mathbf{X}_n\Big]\\
&\quad\lesssim_{\mathbb{P}}\left(\sqrt{\delta_n}+\frac{r_{\boldsymbol{\Sigma}}{}^{1/2}}{\delta_n^{d_{\mathcal{Q}}/2}}\right)\sqrt{\log\frac{K}{\delta_n}}.
\end{aligned}$$

Choose the approximately optimal

$$\delta_n:=(r_{\boldsymbol{\Sigma}})^{1/(d_{\mathcal{Q}}+1)},$$

giving

$$\mathbb{E}\Big[\sup_{\boldsymbol{q}\in\mathcal{Q}}\big\|\boldsymbol{Z}(\boldsymbol{q})-\widetilde{\boldsymbol{Z}}(\boldsymbol{q})\big\|_\infty \,\Big|\, \mathbf{X}_n\Big]\lesssim_{\mathbb{P}}(r_{\boldsymbol{\Sigma}})^{1/(2d_{\mathcal{Q}}+2)}\sqrt{\log n}.$$

Constructing $\widetilde{\boldsymbol{Z}}^*(\cdot)$ By Theorem 8.17 (transfer) in [14], on our probability space there is a random element $\widetilde{\boldsymbol{Z}}^*(\cdot)$ with values in $\mathcal{C}(\mathcal{Q},\mathbb{R}^K)$ such that

$$\big(\mathbf{X}_n,\boldsymbol{Z}(\cdot),\widetilde{\boldsymbol{Z}}(\cdot)\big)\overset{d}{=}\big(\mathbf{X}_n,\boldsymbol{Z}^*(\cdot),\widetilde{\boldsymbol{Z}}^*(\cdot)\big),\tag{SA7.40}$$

and $\widetilde{\boldsymbol{Z}}^*(\cdot)$ is conditionally on $(\mathbf{X}_n,\boldsymbol{Z}^*(\cdot))$ independent of $\boldsymbol{y}_n$. Since also $\boldsymbol{Z}^*(\cdot)$ conditionally on $\mathbf{X}_n$ independent of $\boldsymbol{y}_n$, we have by Theorem 8.12 (chain rule) in [14] that $(\boldsymbol{Z}^*(\cdot),\widetilde{\boldsymbol{Z}}^*(\cdot))$ is conditionally on $\mathbf{X}_n$ independent of $\boldsymbol{y}_n$; in particular, $\widetilde{\boldsymbol{Z}}^*(\cdot)$ is conditionally on $\mathbf{X}_n$ independent of $\boldsymbol{y}_n$. But by (SA7.40), $\widetilde{\boldsymbol{Z}}^*(\cdot)$ is independent of $\mathbf{X}_n$. Again by the chain rule, $\widetilde{\boldsymbol{Z}}^*(\cdot)$ is independent of $(\mathbf{X}_n,\boldsymbol{y}_n)=\mathbf{D}_n$.

SA8 Examples

This section discusses in detail the four motivating examples introduced in the paper.

SA8.1 Quantile regression

The next result verifies that the quantile regression case under the same conditions on $f_{Y|X}$ as Condition S.2 in [2] is a special case of our setting.

Proposition SA8.1 (Verification of Assumption SA2.4 for quantile regression)**.** *Suppose Assumptions SA2.1 to SA2.3 hold with $\mathcal{Q}=[\varepsilon_0,1-\varepsilon_0]$ for some $\varepsilon_0\in(0,1/2)$, the loss is given by $\rho(y,\eta;q)=(q-\mathbb{1}\{y<\eta\})(y-\eta)$, and $\mathbb{E}[|y_1|]<\infty$. Assume further that $\eta(\cdot)\colon\mathbb{R}\to\mathcal{E}$ is strictly monotonic and twice continuously differentiable with $\mathcal{E}$ an open connected subset of $\mathbb{R}$ containing the conditional q-quantile of $y_1|\boldsymbol{x}_1=\boldsymbol{x}$, given by $\eta\big(\mu_0(\boldsymbol{x},q)\big)$ for all $(\boldsymbol{x},q)$; $y\mapsto F_{Y|X}(y|\boldsymbol{x})$ is* *twice continuously differentiable with first derivative $f_{Y|X}(y|\boldsymbol{x})$ (in particular, $\mathfrak{M}$ is the Lebesgue measure); $f_{Y|X}\big(\eta\big(\mu_0(\boldsymbol{x},q)\big)|\boldsymbol{x}\big)$ is bounded away from zero uniformly over $q \in \mathcal{Q}$, $\boldsymbol{x} \in \mathcal{X}$, and the derivative of $y \mapsto f_{Y|X}(y|\boldsymbol{x})$ is continuous and bounded in absolute value from above uniformly over $y \in \mathcal{Y}_{\boldsymbol{x}}, \boldsymbol{x} \in \mathcal{X}$. Then Assumptions SA2.4, SA2.5 and SA2.8 and Eq. (SA6.15) are also true.*

Remark. Taking $\eta\big(\mu_0(\boldsymbol{x},q)\big)$ to be the conditional q-quantile does not violate (SA1.1) by Lemma SA3.10.

Remark. In the setting of Proposition SA8.1, it is not necessary to assume that $\mu_0(\boldsymbol{x},q)$ is Lipschitz in parameter (as we do in Assumption SA2.3(iv)). Since

$$\frac{\partial}{\partial q}\mu_0(\boldsymbol{x},q) = \frac{1}{f_{Y|X}(\mu_0(\boldsymbol{x},q)|\boldsymbol{x})},$$

the Lipschitz property follows from $f_{Y|X}(\mu_0(\boldsymbol{x},q)|\boldsymbol{x})$ being bounded from below uniformly over $\boldsymbol{x} \in \mathcal{X}$, $q \in \mathcal{Q}$.

Proof. We will verify the assumptions one by one.

Verifying Assumption SA2.4(i) It is easy to see that the a. e. derivative $\eta \mapsto \psi(y,\eta;q) \equiv \mathbb{1}\{y-\eta<0\}-q$ of $\eta \mapsto \rho(y,\eta;q)$ is Lebesgue integrable and satisfies

$$\int_a^b \psi(y,\eta;q)\,\mathrm{d}\eta = \rho(y,b;q) - \rho(y,a;q)$$

for any $[a,b] \subset \mathcal{E}$.

Verifying Assumption SA2.4(ii) Since $\eta(\mu_0(\cdot,q))$ is the conditional q-quantile, we have

$$\mathbb{E}[\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,q));q)\,|\,\boldsymbol{x}_i] = \mathbb{E}[\mathbb{1}\{y_i<\eta(\mu_0(\boldsymbol{x}_i,q))\}-q\,|\,\boldsymbol{x}_i] = q-q=0$$

and

$$\begin{aligned}\sigma_q^2(\boldsymbol{x}) &= \mathbb{E}[\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,q));q)^2\,|\,\boldsymbol{x}_i=\boldsymbol{x}]\\ &= \mathbb{E}[(\mathbb{1}\{y_i<\eta(\mu_0(\boldsymbol{x}_i,q))\}-q)^2\,|\,\boldsymbol{x}_i=\boldsymbol{x}] = q-2q^2+q^2 = q(1-q)\end{aligned}$$

is constant in $\boldsymbol{x}$ (in particular continuous in $\boldsymbol{x}$) and bounded away from zero since both q and $1-q$ are bounded away from zero. Since $q(1-q)$ is smooth, $\sigma_q^2(\boldsymbol{x})$ is Lipshitz in q. The family $\{\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i,q));q)\colon q\in\mathcal{Q}\}$ has a positive measurable envelope 1 which has uniformly bounded conditional moments of any order.

Verifying Assumption SA2.4(iii) Clearly, $\rho(y,\eta;q)$ is convex in η and the a. e. derivative in η is $\psi(y,\eta;q)\equiv\mathbb{1}\{y-\eta<0\}-q$ is piecewise constant with only one jump (therefore piecewise Hölder with $\alpha=1$). The link function $\eta(\cdot)$ is strictly monotonic and twice continuously differentiable by assumption.

Verifying Assumption SA2.4(iv) The conditional expectation

$$\Psi(x,\eta;q) = \mathbb{E}[\mathbb{1}\{y<\eta\}-q\,|\,\boldsymbol{x}_i=\boldsymbol{x}] = \int_{-\infty}^{\eta} f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y - q$$

is twice continuously differentiable in η (the integral on the right is a Riemann integral, possibly improper) and its second derivative $f'_{Y|X}(\eta|\boldsymbol{x})$ is continuous and bounded in absolute value. By the mean value theorem, this means that $f_{Y|X}(\eta(\mu_0(\boldsymbol{x},q))|\boldsymbol{x})$ being bounded away from zero implies $f_{Y|X}(\eta(\zeta)|\boldsymbol{x})$ is bounded away from zero for ζ sufficiently close to $\mu_0(\boldsymbol{x},q)$. The bound on $|\Psi_1(\boldsymbol{x},\eta(\zeta);q)|$ from above in such a neighborhood (and in fact everywhere) is automatic since $f_{Y|X}(y|\boldsymbol{x})$ is bounded from above (uniformly over $y\in\mathcal{Y}_{\boldsymbol{x}}$, $\boldsymbol{x}\in\mathcal{X}$).

Verifying Assumption SA2.5 This verification proceeds similarly to Lemmas 25–28 in [2].

The class of functions

$$\mathcal{W}_1 := \big\{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\} \colon \boldsymbol{\beta} \in \mathbb{R}^K\big\}$$

is VC-subgraph with index $O(K)$ by Lemmas 2.6.15 and 2.6.18 in [18] (since $\eta(\cdot)$ is monotone). The class of functions

$$\mathcal{G}_2 = \big\{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y < \eta(\mu_0(\boldsymbol{x}, q))\} \colon q \in \mathcal{Q}\big\}$$

is VC-subgraph with index 2 since $\eta(\mu_0(\boldsymbol{x}, q))$ is increasing in q for any $\boldsymbol{x} \in \mathcal{X}$, giving that the class of sets $\{(\boldsymbol{x}, y) : y < \eta(\mu_0(\boldsymbol{x}, q))\}$ with $q \in \mathcal{Q}$ is linearly ordered by inclusion. The VC property of $\mathcal{W}_1$ with envelope 1 implies that it satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$. The VC property of $\mathcal{G}_2$ with envelope 1 implies that it satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim 1$. By Lemma SA3.4, for any fixed $r > 0$ the class

$$\mathcal{G}_1 = \big\{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\} - \mathbb{1}\{y < \eta(\mu_0(\boldsymbol{x}, q))\} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, q \in \mathcal{Q}\big\}$$

with envelope 2 satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$ because it is a subclass of $\mathcal{W}_1 - \mathcal{G}_2$.

For a fixed vector space $\mathcal{B}$ of dimension $\dim \mathcal{B}$,

$$\mathcal{W}_{\mathcal{B}} := \{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\} : \boldsymbol{\beta} \in \mathcal{B}\}$$

is VC with index $O(\dim \mathcal{B})$ by Lemmas 2.6.15 and 2.6.18 in [18] (since $\eta(\cdot)$ is monotone). Therefore, for any fixed $c > 0$, $\delta \in \Delta$, the class

$$\mathcal{W}_{2,\delta} := \big\{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\}\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\} : \boldsymbol{\beta} \in \mathbb{R}^K\big\}$$

is VC-subgraph with index $O(\log^d n)$ by Lemma 2.6.18 in [18] because it is contained in the product of $\mathcal{W}_{\mathcal{B}_{2,\delta}}$ for some vector space $\mathcal{B}_{2,\delta}$ of dimension $\dim \mathcal{B}_{2,\delta} \lesssim \log^d n$ and a fixed function $\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\}$. This means $\mathcal{W}_{2,\delta}$ with envelope 1 satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$. Then the union of $O(h^{-d})$ such classes

$$\mathcal{W}_2 := \big\{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\}\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\} : \boldsymbol{\beta} \in \mathbb{R}^K, \delta \in \Delta\big\}$$

with envelope 1 satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$, see (SA5.34). By Lemma SA3.4, the same is true of

$$\begin{aligned}\mathcal{G}_3 = \big\{&(\boldsymbol{x}, y) \mapsto \\ &\big[\mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\} - \mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q))\}\big]\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)\} : \\ &\boldsymbol{\beta} \in \mathbb{R}^K, \delta \in \Delta, q \in \mathcal{Q}\big\}\end{aligned}$$

with envelope 2 because it is a subclass of $\mathcal{W}_2 - \mathcal{W}_2$.

The class of functions

$$\mathcal{G}_4 = \{\boldsymbol{x} \mapsto f_{Y|X}(\eta(\mu_0(\boldsymbol{x}, q))|\boldsymbol{x}) : q \in \mathcal{Q}\}$$

has a bounded envelope by the assumptions of the lemma. Moreover, $\mathcal{G}_4$ has the following property: for any $q_1, q_2 \in \mathcal{Q}$ we have for some $\xi_{\boldsymbol{x}, q_1, q_2}$ between $\eta(\mu_0(\boldsymbol{x}, q_1))$ and $\eta(\mu_0(\boldsymbol{x}, q_2))$

$$\big|f_{Y|X}(\eta(\mu_0(\boldsymbol{x}, q_1))|\boldsymbol{x}) - f_{Y|X}(\eta(\mu_0(\boldsymbol{x}, q_2))|\boldsymbol{x})\big|$$

$$
\begin{aligned}
&= \left|f'_{Y|X}(\xi_{\boldsymbol{x},q_1,q_2})\right| \cdot |\eta(\mu_0(\boldsymbol{x},q_1)) - \eta(\mu_0(\boldsymbol{x},q_2))| \\
&\lesssim \left|\eta(\mu_0(\boldsymbol{x},q_1)) - \eta(\mu_0(\boldsymbol{x},q_2))\right| && \text{since } f'_{Y|X} \text{ is uniformly bounded} \\
&\lesssim \left|\mu_0(\boldsymbol{x},q_1) - \mu_0(\boldsymbol{x},q_2)\right| && \text{since } \eta(\cdot) \text{ is Lipschitz} \\
&\lesssim \left|q_1 - q_2\right| && \text{since } \mu_0(\boldsymbol{x},q) \text{ is Lipschitz in } q.
\end{aligned}
$$

with constants in $\lesssim$ not depending on $\boldsymbol{x}, q_1, q_2$ or n (this is also proven in Lemma 20 in [2]). Since $\mathcal{Q}$ is a fixed one-dimensional segment, this implies that $\mathcal{G}_4$ satisfies the uniform entropy bound (SA1.3) with $A, V \lesssim 1$.

For a fixed l, the class of functions

$$\{(\boldsymbol{x},y) \mapsto p_l(\boldsymbol{x})\mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\} : \boldsymbol{\beta} \in \mathbb{R}^K\}$$

is VC-subgraph with index $O(1)$ because it is contained in the product of $\mathcal{W}_{\mathcal{B}_{5,l}}$ for some vector space $\mathcal{B}_{5,l}$ of dimension $\dim \mathcal{B}_{5,l} \lesssim 1$, and a fixed function $p_l(\boldsymbol{x})$. Then this class with envelope $O(1)$ satisfies the uniform entropy bound (SA1.3) with $A, V \lesssim 1$. Since, as we have shown above, the same is true of $\mathcal{G}_2$, by Lemma SA3.4, it is also true of

$$
\begin{aligned}
\mathcal{G}_5 = \big\{&(\boldsymbol{x},y) \mapsto \\
&p_l(\boldsymbol{x})\big[\mathbb{1}\{y < \eta(\mu_0(\boldsymbol{x},q))\} - \mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q))\}\big] : q \in \mathcal{Q}\big\}.
\end{aligned}
$$

Verifying Assumption SA2.8 The functions in the class have the form

$$
\begin{aligned}
&\rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})); q)\mathbb{1}(\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)) \\
&\quad - \rho(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})); q)\mathbb{1}(\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)) \\
&\quad - \big[\eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})\big) - \eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})\big)\big]\psi(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q)); q) \\
&\quad \times \mathbb{1}(\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)) \\
&\quad =: T_1 + T_2 + T_3 + T_4,
\end{aligned}
$$

where

$$
\begin{aligned}
T_1 &:= y\big[\mathbb{1}\big\{y < \eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})\big)\big\} - \mathbb{1}\big\{y < \eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})\big)\big\}\big] \\
&\quad \times \mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\}, \\
T_2 &:= \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta}))\mathbb{1}\big\{y < \eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})\big)\big\}\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\}, \\
T_3 &:= -\eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})\big)\mathbb{1}\big\{y < \eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})\big)\big\}\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\}, \\
T_4 &:= -\mathbb{1}\big\{y < \eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q)\big)\big\}\big[\eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})\big) - \eta\big(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})\big)\big] \\
&\quad \times \mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\}.
\end{aligned}
$$

Note that for T_1 to be nonzero, y has to lie in a fixed interval (not depending on n), say $[-\tilde{R}, \tilde{R}]$. The class of functions

$$\{(\boldsymbol{x},y) \mapsto y\mathbb{1}\{|y| \le \tilde{R}\}\mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\} : \boldsymbol{\beta} \in \mathcal{B}\},$$

where $\mathcal{B}$ is any linear subspace of $\mathbb{R}^K$, is VC-subgraph with index $O(\dim \mathcal{B})$ by Lemmas 2.6.15 and 2.6.18 in [18] (since $\eta(\cdot)$ is monotone and $y \mapsto y\mathbb{1}\{|y| \le \tilde{R}\}$ is one fixed function). The class $\{(\boldsymbol{x},y) \mapsto T_1, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, \|\boldsymbol{v}\|_\infty \le \varepsilon_n, q \in \mathcal{Q}\}$ with δ fixed is a subclass of the difference of two such classes, so by Lemma SA3.4 this class with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$. Therefore, the union of $O(h^{-d})$ such classes

$$\{(\boldsymbol{x}, y) \mapsto T_1, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, \|\boldsymbol{v}\|_\infty \le \varepsilon_n, \delta \in \Delta, q \in \mathcal{Q}\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$, see (SA5.34).

Similarly, the classes

$$\begin{aligned}
&\{(\boldsymbol{x}, y) \mapsto T_2, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, \|\boldsymbol{v}\|_\infty \le \varepsilon_n, \delta \in \Delta, q \in \mathcal{Q}\},\\
&\{(\boldsymbol{x}, y) \mapsto T_3, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, \|\boldsymbol{v}\|_\infty \le \varepsilon_n, \delta \in \Delta, q \in \mathcal{Q}\},\\
&\{(\boldsymbol{x}, y) \mapsto T_4, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, \|\boldsymbol{v}\|_\infty \le \varepsilon_n, \delta \in \Delta, q \in \mathcal{Q}\}
\end{aligned}$$

with large enough constant envelopes also satisfy the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$. We used that $\varepsilon_n \to 0$ giving that $\boldsymbol{v}$ is bounded in ∞-norm (like $\boldsymbol{\beta}$). Applying Lemma SA3.4 one more time, we have that there exist some constants $C_{17} \ge e$, $C_{18} \ge 1$ and $C_{19} > 0$ such that

$$\sup_{\mathbb{Q}} N(\mathcal{G}, \|\cdot\|_{\mathbb{Q},2}, \varepsilon C_{19}) \le \left(\frac{C_{17}}{\varepsilon}\right)^{C_{18} \log^d n} \tag{SA8.1}$$

for all $0 < \varepsilon \le 1$, where the supremum is taken over all finite discrete probability measures $\mathbb{Q}$ and $\mathcal{G}$ is the class defined in Assumption SA2.8. Note that the integral representation of $\mathcal{G}$ makes it clear that this class not only has a large enough constant envelope, but is also bounded by $C_{20}\varepsilon_n$, where C_{20} is a large enough constant.

For large enough n we can replace ε with $C_{20}\varepsilon\varepsilon_n/C_{19}$ in (SA8.1), giving

$$\sup_{\mathbb{Q}} N(\mathcal{G}, \|\cdot\|_{\mathbb{Q},2}, C_{20}\varepsilon\varepsilon_n) \le \left(\frac{C_{17}C_{19}}{C_{20}\varepsilon\varepsilon_n}\right)^{C_{18} \log^d n}$$

for all $0 < \varepsilon \le 1$. For large enough n, $C_{17}C_{19}/(C_{20}\varepsilon_n) \ge e$. The verification is complete.

Verifying (SA6.15) In this case, $\psi(y, \eta; q) = \mathbb{1}\{y < \eta\} - q$ and $\eta(\mu_0(\boldsymbol{x}_i, q))$ is the q-quantile of y_i conditional on $\boldsymbol{x}_i$. Without loss of generality, we will assume that $\eta(\cdot)$ is strictly increasing and $q \le \tilde{q}$ (the other cases are symmetric).

$$\begin{aligned}
&\mathbb{E}\big[\big|\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, q)); q)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q)) - \psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \tilde{q})); \tilde{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\big|^2 \,\big|\, \boldsymbol{x}_i\big]\\
&\quad = \mathbb{E}\big[\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, q)); q)^2 \,\big|\, \boldsymbol{x}_i\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q))^2\\
&\qquad + \mathbb{E}\big[\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \tilde{q})); \tilde{q})^2 \,\big|\, \boldsymbol{x}_i\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))^2\\
&\qquad - 2\mathbb{E}\big[\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, q)); q)\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \tilde{q})); \tilde{q}) \,\big|\, \boldsymbol{x}_i\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q))\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\\
&\quad = (q - q^2)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q))^2 + (\tilde{q} - \tilde{q}^2)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))^2\\
&\qquad - 2(q - q\tilde{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q))\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\\
&\quad = q\big|\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q)) - \eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\big|^2 + (\tilde{q} - q)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))^2\\
&\qquad - \big|q\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q)) - \tilde{q}\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\big|^2\\
&\quad \le q\big|\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q)) - \eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\big|^2 + (\tilde{q} - q)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))^2\\
&\quad \overset{\text{(a)}}{\lesssim} q(\tilde{q} - q)^2 + \tilde{q} - q \lesssim \tilde{q} - q,
\end{aligned}$$

where in (a) we used that $\eta^{(1)}(\cdot)$ on a fixed compact is Lipschitz and $\mu_0(\boldsymbol{x}, q)$ is Lipschitz in q uniformly over $\boldsymbol{x}$, as well as boundedness of $\mu_0(\boldsymbol{x}, q)$ uniformly over q and $\boldsymbol{x}$.

This concludes the proof of Proposition SA8.1. $\square$

Proposition SA8.2 (Verification of the conditions of Lemma SA4.3)**.** *Suppose all conditions of Proposition SA8.1 hold. In addition, suppose there is a positive constant C_{21} such that we have $\inf f_{Y|X}(y|\boldsymbol{x}) > C_{21}$, where the infimum is over $\boldsymbol{x} \in \mathcal{X}$, $q \in \mathcal{Q}$, $\|\boldsymbol{\beta}\|_\infty \leq R$ for R described in Lemma SA4.3, y between $\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})$ and $\eta(\mu_0(\boldsymbol{x}, q))$. Then conditions in Conditions (v) and (vi) of Lemma SA4.3 also hold.*

Proof. We only need to verify Lemma SA4.3(vi) since Lemma SA4.3(v) is directly assumed in this lemma ($\Psi_1(\boldsymbol{x}, \eta; q) = f_{Y|X}(\eta|\boldsymbol{x})$ in this case).

In this verification, we will use $\theta_1 := \boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}$, $\theta_2 := \boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q)$ to simplify notations. Rewrite

$$\begin{aligned}
&\rho(y, \eta(\theta_1); q) - \rho(y, \eta(\theta_2); q)\\
&\quad = y[\mathbb{1}\{y < \eta(\theta_2)\} - \mathbb{1}\{y < \eta(\theta_1)\}] + q[\eta(\theta_1) - \eta(\theta_2)]\\
&\qquad + \eta(\theta_1)\mathbb{1}\{y < \eta(\theta_1)\} - \eta(\theta_2)\mathbb{1}\{y < \eta(\theta_2)\}.
\end{aligned}$$

By the same argument as in the proof of Proposition SA8.1, the class

$$\{(\boldsymbol{x}, y) \mapsto y[\mathbb{1}\{y < \eta(\theta_2)\} - \mathbb{1}\{y < \eta(\theta_1)\}] : \|\boldsymbol{\beta}\|_\infty \leq R, q \in \mathcal{Q}\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$.

The class $\{(\boldsymbol{x}, y) \mapsto q : q \in \mathcal{Q}\}$ is of course VC with a constant index (as a subclass of the class of constant functions), and the class $\{(\boldsymbol{x}, y) \mapsto \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}) : \boldsymbol{\beta} \in \mathbb{R}^K\}$ is VC with index $O(K)$ because the space of functions $\boldsymbol{p}(\boldsymbol{x})\boldsymbol{\beta}$ is a linear space with $O(K)$ dimension, and $\eta(\cdot)$ is monotone. Applying Lemma SA3.4, we see that the class $\{(\boldsymbol{x}, y) \mapsto q[\eta(\theta_1) - \eta(\theta_2)] : \|\boldsymbol{\beta}\|_\infty \leq R, q \in \mathcal{Q}\}$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$.

The class $\{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y < \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\} : \boldsymbol{\beta} \in \mathbb{R}^K\}$ is VC with index $O(K)$ because the space of functions $\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}$ is a linear space with $O(K)$ dimension, and $\eta(\cdot)$ is monotone. Again applying Lemma SA3.4, we see that the class $\{(\boldsymbol{x}, y) \mapsto \eta(\theta_1)\mathbb{1}\{y < \eta(\theta_1)\} : \|\boldsymbol{\beta}\|_\infty \leq R\}$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$. The same is true of its subclass $\{(\boldsymbol{x}, y) \mapsto \eta(\theta_2)\mathbb{1}\{y < \eta(\theta_2)\} : q \in \mathcal{Q}\}$.

It is left to apply Lemma SA3.4 again, concluding that the class described in Lemma SA4.3(vi) with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$. $\square$

SA8.2 Distribution regression

Proposition SA8.3 (Verification of assumptions for distribution regression)**.** *Let $\mathcal{Q} = [-A, A]$ for some $A > 0$. Suppose Assumption SA2.3 holds with the loss function $\rho(y, \eta; q) = (\mathbb{1}\{y \leq q\} - \eta)^2$, the link function $\eta(\cdot)\colon \mathbb{R} \to (0, 1)$ is strictly monotonic and twice continuously differentiable, $q \mapsto F_{Y|X}(q|\boldsymbol{x})$ is continuously differentiable with derivative $f_{Y|X}(q|\boldsymbol{x})$ (in particular, $\mathfrak{M}$ is Lebesgue measure), $\boldsymbol{x} \mapsto F_{Y|X}(q|\boldsymbol{x})$ is a continuous function, and $F_{Y|X}(q|\boldsymbol{x}) = \eta(\mu_0(\boldsymbol{x}, q))$ lies in a compact subset of $(0, 1)$ for all $q \in \mathcal{Q}, \boldsymbol{x} \in \mathcal{X}$ (this subset does not depend on q or $\boldsymbol{x}$). Then Assumptions SA2.4, SA2.5 and SA2.8, Eq. (SA6.15), and Conditions (v) and (vi) of Lemma SA4.3 are also true.*

Remark. Taking $\eta(\mu_0(\boldsymbol{x}, q))$ to be the conditional distribution function does not violate Eq. (SA1.1) by the following standard argument. For any Borel function $\mu(\cdot)\colon \mathcal{X} \to \mathbb{R}$ one has

$$
\begin{aligned}
&\mathbb{E}\big[\big(\mathbb{1}\{y_1 \le q\} - \eta(\mu(\boldsymbol{x}_1))\big)^2\big] \\
&\quad = \mathbb{E}\big[\big(\mathbb{1}\{y_1 \le q\} - F_{Y|X}(q|\boldsymbol{x}_1) + F_{Y|X}(q|\boldsymbol{x}_1) - \eta(\mu(\boldsymbol{x}_1))\big)^2\big] \\
&\quad = \mathbb{E}\big[\big(\mathbb{1}\{y_1 \le q\} - F_{Y|X}(q|\boldsymbol{x}_1)\big)^2\big] \\
&\qquad + 2\mathbb{E}\big[\big(\mathbb{1}\{y_1 \le q\} - F_{Y|X}(q|\boldsymbol{x}_1)\big)\big(F_{Y|X}(q|\boldsymbol{x}_1) - \eta(\mu(\boldsymbol{x}_1))\big)\big] \\
&\qquad + \mathbb{E}\big[\big(F_{Y|X}(q|\boldsymbol{x}_1) - \eta(\mu(\boldsymbol{x}_1))\big)^2\big].
\end{aligned}
$$

Since the cross term is zero (proven by conditioning on $\boldsymbol{x}_1$), this means

$$
\begin{aligned}
&\mathbb{E}\big[\big(\mathbb{1}\{y_1 \le q\} - \eta(\mu(\boldsymbol{x}_1))\big)^2\big] \\
&\quad = \mathbb{E}\big[\big(\mathbb{1}\{y_1 \le q\} - F_{Y|X}(q|\boldsymbol{x}_1)\big)^2\big] + \mathbb{E}\big[\big(F_{Y|X}(q|\boldsymbol{x}_1) - \eta(\mu(\boldsymbol{x}_1))\big)^2\big] \\
&\quad \ge \mathbb{E}\big[\big(\mathbb{1}\{y_1 \le q\} - F_{Y|X}(q|\boldsymbol{x}_1)\big)^2\big].
\end{aligned}
$$

Pointwise in $q \in \mathcal{Q}$, equality holds if and only if $\eta(\mu(\boldsymbol{x}_1)) = F_{Y|X}(q|\boldsymbol{x}_1)$ almost surely.

Remark SA8.4. In this case,

$$
\mathbb{E}\big[\boldsymbol{Z}(q)\boldsymbol{Z}(\tilde{q})^{\mathsf{T}} \,\big|\, \mathbf{X}_n\big] = h^{-d}\mathbb{E}_n\big[S_{q,\tilde{q}}(\boldsymbol{x})\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q))\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\big]
$$

with

$$
S_{q,\tilde{q}}(\boldsymbol{x}) = 4F_{Y|X}(q \wedge \tilde{q}|\boldsymbol{x}_i)\big(1 - F_{Y|X}(q \vee \tilde{q}|\boldsymbol{x}_i)\big).
$$

This covariance structure is not known, but it can be estimated by

$$
h^{-d}\mathbb{E}_n\big[\widehat{S}_{q,\tilde{q}}(\boldsymbol{x})\eta^{(1)}(\widehat{\mu}(\boldsymbol{x}_i, q))\eta^{(1)}(\widehat{\mu}(\boldsymbol{x}_i, \tilde{q}))\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\big]. \tag{SA8.2}
$$

with

$$
\widehat{S}_{q,\tilde{q}}(\boldsymbol{x}) = 4\eta(\widehat{\mu}(\boldsymbol{x}_i, q \wedge \tilde{q}))\big(1 - \eta(\widehat{\mu}(\boldsymbol{x}_i, q \vee \tilde{q}))\big).
$$

Proof. We will verify the assumptions one by one.

Verifying Assumption SA2.4(i) It is easy to see that the a. e. derivative $\eta \mapsto \psi(y, \eta; q) \equiv 2(\eta - \mathbb{1}\{y \le q\})$ of $\eta \mapsto \rho(y, \eta; q)$ is Lebesgue integrable and satisfies

$$
\int_a^b \psi(y, \eta; q)\,\mathrm{d}\eta = \rho(y, b; q) - \rho(y, a; q)
$$

for any $[a, b] \subset (0, 1)$.

Verifying Assumption SA2.4(ii) The first-order optimality condition

$$
\mathbb{E}\big[2(\eta(\mu_0(\boldsymbol{x}_i, q)) - \mathbb{1}\{y_i \le q\}) \,\big|\, \boldsymbol{x}_i\big] = 2\big(F_{Y|X}(q|\boldsymbol{x}_i) - \mathbb{E}[\mathbb{1}\{y_i \le q\} \,|\, \boldsymbol{x}_i]\big) = 0
$$

indeed holds. The conditional variance

$$
\sigma_q^2(\boldsymbol{x}) := 4\mathbb{E}[(\eta(\mu_0(\boldsymbol{x}_i, q)) - \mathbb{1}\{y_i \le q\})^2 \,|\, \boldsymbol{x}_i = \boldsymbol{x}] = 4F_{Y|X}(q|\boldsymbol{x})(1 - F_{Y|X}(q|\boldsymbol{x}))
$$

is continuous and bounded away from zero by the assumptions ($F_{Y|X}(q|\boldsymbol{x})$ cannot achieve 0 or 1).

The family $\{2(F_{Y|X}(q|\boldsymbol{x}) - \mathbb{1}\{y \le q\}) : q \in \mathcal{Q}\}$ is bounded in absolute value by $\overline{\psi}(\boldsymbol{x}, y) \equiv 2$.

Verifying Assumption SA2.4(iii) The loss function $\rho(y, \eta; q)$ is infinitely smooth with respect to η. Its first derivative is $\psi(y, \eta; q) = 2(\eta - \mathbb{1}\{y \leq q\})$. Since the derivative of $\eta(\cdot)$ is bounded on a compact interval, the function $\psi(y, \eta(\theta); q)$ is Lipschitz in θ on a compact interval.

Verifying Assumption SA2.4(iv) The conditional expectation

$$\Psi(\boldsymbol{x}, \eta; q) := \mathbb{E}[2(\eta - \mathbb{1}\{y_i \leq q\}) \,|\, \boldsymbol{x}_i = \boldsymbol{x}] = 2\eta - 2F_{Y|X}(q|\boldsymbol{x})$$

is linear, and in particular infinitely smooth, in η. Its first partial derivative

$$\Psi_1(\boldsymbol{x}, \eta; q) = \frac{\partial}{\partial \eta}\Psi(\boldsymbol{x}, \eta; q) = 2$$

is a nonzero constant, so it is bounded and bounded away from zero everywhere. The second partial derivative is zero, and so it is also bounded.

Verifying Assumption SA2.5 The class of functions

$$\mathcal{G}_{11} := \{(\boldsymbol{x}, y) \mapsto F_{Y|X}(q|\boldsymbol{x}) : q \in \mathcal{Q}\}$$

with a constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim 1$ because it is VC-subgraph with index 1 since the subgraphs are linearly ordered by inclusion (by monotonicity in q).

The class

$$\mathcal{G}_{12} := \{(\boldsymbol{x}, y) \mapsto \boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^K\}$$

is VC with index not exceeding $K + 2$ by Lemma 2.6.15 in [18]. Since in a fixed bounded interval $\eta(\cdot)$ is Lipschitz, the class

$$\mathcal{G}_{13} := \{(\boldsymbol{x}, y) \mapsto \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \leq r\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$. By Lemma SA3.4, the class

$$\mathcal{G}_1 := \big\{(\boldsymbol{x}, y) \mapsto 2\big(\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}) - F_{Y|X}(q|\boldsymbol{x})\big) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \leq r, q \in \mathcal{Q}\big\},$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$.

The class of sets $\{(\boldsymbol{x}, y) : y \leq q\}$ with $q \in \mathcal{Q}$ is linearly ordered by inclusion, so it is VC with a constant index and so is the class of functions

$$\mathcal{G}_{21} := \{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y \leq q\} : q \in \mathcal{Q}\},$$

giving by Lemma SA3.4 that

$$\mathcal{G}_2 := \{(\boldsymbol{x}, y) \mapsto 2(F_{Y|X}(q|\boldsymbol{x}) - \mathbb{1}\{y \leq q\}) : q \in \mathcal{Q}\},$$

which is a subclass of $2(\mathcal{G}_{11} - \mathcal{G}_{21})$, satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim 1$.

For a fixed vector space $\mathcal{B}$ of dimension $\dim \mathcal{B}$,

$$\mathcal{W}_{\mathcal{B}} := \{(\boldsymbol{x}, y) \mapsto \boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathcal{B}\}$$

is VC-subgraph with index $O(\dim \mathcal{B})$ by Lemmas 2.6.15 and 2.6.18 in [18]. Therefore, again using that $\eta(\cdot)$ is Lipschitz in a compact interval, by Lemma SA3.4 we have that for any fixed $c > 0$, $\delta \in \Delta$, the class

$$\mathcal{W}_{3,\delta} := \{(\boldsymbol{x}, y) \mapsto \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, q \in \mathcal{Q}\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$. Then the union of $O(h^{-d})$ such classes

$$\mathcal{W}_3 := \{(\boldsymbol{x}, y) \mapsto \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, q \in \mathcal{Q}, \delta \in \Delta\}$$

satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$ (see (SA5.34)). The same is true of

$$\begin{aligned}\mathcal{G}_3 := \big\{(\boldsymbol{x}, y) \mapsto 2\big(\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}) - \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(q))\big)\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\} :& \\ \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, q \in \mathcal{Q}, \delta \in \Delta\big\}&\end{aligned}$$

because it is a subclass of $2\mathcal{W}_3 - 2\mathcal{W}_3$.

The class

$$\mathcal{G}_4 := \{\boldsymbol{x} \mapsto 2\}$$

consists of just one bounded function, so clearly it satisfies the uniform entropy bound (SA1.3) with envelope 2, $A \lesssim 1$, $V \lesssim 1$.

Finally, for any fixed $l \in \{1, \dots, K\}$ the class

$$\mathcal{G}_{51} := \{(\boldsymbol{x}, y) \mapsto p_l(\boldsymbol{x})\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(q)) : q \in \mathcal{Q}\}$$

satisfies the uniform entropy bound (SA1.3) with a large enough constant envelope, $A \lesssim 1$ and $V \lesssim 1$ because $\eta(\cdot)$ is Lipschitz and $\mathcal{G}_{51}$ is contained in a fixed function multiplied by $\eta(\mathcal{W}_{\mathcal{B}})$ for a linear space $\mathcal{B}$ of a constant dimension, and by Lemma SA3.4

$$\mathcal{G}_5 := \{(\boldsymbol{x}, y) \mapsto 2p_l(\boldsymbol{x})(F_{Y|X}(q|\boldsymbol{x}) - \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(q))) : q \in \mathcal{Q}\}$$

with a constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim 1$.

Verifying Assumption SA2.8 Assumption SA2.8 holds because the class of functions

$$\begin{aligned}&\big\{(\boldsymbol{x}, y) \mapsto \big[\eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})\big) - \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v}))\big] \\ &\quad \times \big[\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})) + \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})) - 2\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(q))\big] \\ &\quad \times \mathbb{1}(\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)) : \\ &\|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, \|\boldsymbol{v}\|_\infty \le \varepsilon_n, \delta \in \Delta, q \in \mathcal{Q}\big\}\end{aligned}$$

is contained in the product of two classes

$$\begin{aligned}\mathcal{V}_1 := \big\{(\boldsymbol{x}, y) \mapsto \big[\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}) - \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta} - \boldsymbol{v}))\big]\mathbb{1}(\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)) :& \\ \|\boldsymbol{\beta}\|_\infty \le \tilde{r}, \|\boldsymbol{v}\|_\infty \le \varepsilon_n, \delta \in \Delta, q \in \mathcal{Q}\big\},&\end{aligned}$$

and

$$\mathcal{V}_2 := \big\{(\boldsymbol{x}, y) \mapsto \big[\eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta})\big) + \eta\big(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}_0(q) + \boldsymbol{\beta} - \boldsymbol{v})\big) - 2\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}_0(q))\big]$$

$$\times \mathbb{1}(\boldsymbol{x} \in \mathcal{N}_{[c \log n]}(\delta)) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \leq r, \|\boldsymbol{v}\|_\infty \leq \varepsilon_n, \delta \in \Delta, q \in \mathcal{Q}\big\}$$

for some fixed $\tilde{r} > 0$. Class $\mathcal{V}_1$ with envelope ε_n multiplied by a large enough constant (since η is Lipschitz) satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1/\varepsilon_n$ and $V \lesssim \log^d n$ (this can be shown by further breaking down $\mathcal{V}_1$ into classes $\{\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\}$ and $\{\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}(\boldsymbol{\beta} - \boldsymbol{v}))\}$ with constant envelopes, using Lemma SA3.4 and then replacing ε in the uniform entropy bound by $\varepsilon \cdot \varepsilon_n$). Class $\mathcal{V}_2$ with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim \log^d n$ because it is true for each of the three additive terms it can be broken down into. We omit the details since they are the same as in the verification of Assumption SA2.5.

Verifying Condition (v) in Lemma SA4.3 This condition holds trivially because $\Psi_1(\cdot, \cdot; q)$ is a positive constant.

Verifying Condition (vi) in Lemma SA4.3 The class of functions described in this condition is

$$\big\{(\boldsymbol{x}, y) \mapsto \big(\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q)) - \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})\big) \cdot \big(2\mathbb{1}\{y \leq q\} - \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}) - \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0(q))\big) : \\ \|\boldsymbol{\beta}\|_\infty \leq R, q \in \mathcal{Q}\big\}.$$

The assertion follows by Lemma SA3.4 since

$$\{(\boldsymbol{x}, y) \mapsto \mathbb{1}\{y \leq q\} : q \in \mathcal{Q}\}$$

is a VC-subgraph class with a constant index and

$$\{(\boldsymbol{x}, y) \mapsto \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}) : \|\boldsymbol{\beta}\|_\infty \leq R\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$ because $\eta(\cdot)$ is Lipschitz.

Verifying (SA6.15) Without loss of generality, assume $q \leq \tilde{q}$ (the other case is symmetric).

$$\begin{aligned}
&\mathbb{E}\big[|\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, q)); q)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q)) - \psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \tilde{q})); \tilde{q})\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))|^2 \,\big|\, \boldsymbol{x}_i\big] \\
&\quad = \mathbb{E}\big[\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, q)); q)^2 \,\big|\, \boldsymbol{x}_i\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q))^2 \\
&\qquad + \mathbb{E}\big[\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \tilde{q})); \tilde{q})^2 \,\big|\, \boldsymbol{x}_i\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))^2 \\
&\qquad - 2\mathbb{E}\big[\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, q)); q)\psi(y_i, \eta(\mu_0(\boldsymbol{x}_i, \tilde{q})); \tilde{q}) \,\big|\, \boldsymbol{x}_i\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q))\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q})) \\
&\quad = 4\big[F_{Y|X}(q|\boldsymbol{x}_i) - F_{Y|X}(q|\boldsymbol{x}_i)^2\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q))^2 \\
&\qquad + 4\big[F_{Y|X}(\tilde{q}|\boldsymbol{x}_i) - F_{Y|X}(\tilde{q}|\boldsymbol{x}_i)^2\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))^2 \\
&\qquad - 8\big[F_{Y|X}(q|\boldsymbol{x}_i) - F_{Y|X}(q|\boldsymbol{x}_i)F_{Y|X}(\tilde{q}|\boldsymbol{x}_i)\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q))\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q})) \\
&\quad = 4F_{Y|X}(q|\boldsymbol{x}_i)\big|\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q)) - \eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\big|^2 \\
&\qquad + 4\big[F_{Y|X}(\tilde{q}|\boldsymbol{x}_i) - F_{Y|X}(q|\boldsymbol{x}_i)\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))^2 \\
&\qquad - 4\big|F_{Y|X}(q|\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q)) - F_{Y|X}(\tilde{q}|\boldsymbol{x}_i)\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\big|^2 \\
&\quad \leq 4F_{Y|X}(q|\boldsymbol{x}_i)\big|\eta^{(1)}(\mu_0(\boldsymbol{x}_i, q)) - \eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))\big|^2 \\
&\qquad + 4\big[F_{Y|X}(\tilde{q}|\boldsymbol{x}_i) - F_{Y|X}(q|\boldsymbol{x}_i)\big]\eta^{(1)}(\mu_0(\boldsymbol{x}_i, \tilde{q}))^2
\end{aligned}$$

$$\overset{\text{(a)}}{\lesssim} (\tilde{q} - q)^2 + (\tilde{q} - q) \lesssim \tilde{q} - q,$$

where in (a) we used that $\eta^{(1)}(\cdot)$ and $\eta(\cdot)$ on a fixed compact are Lipschitz and $\mu_0(\boldsymbol{x}, q)$ is Lipschitz in q uniformly over $\boldsymbol{x}$ (therefore, $\eta(\mu_0(\boldsymbol{x}, q)) = F_{Y|X}(q|\boldsymbol{x})$ is also Lipschitz in q uniformly over $\boldsymbol{x}$), as well as boundedness of $\mu_0(\boldsymbol{x}, q)$ uniformly over q and $\boldsymbol{x}$.

Proposition SA8.3 is proven. □

SA8.3 L_p regression

Proposition SA8.5 (Verification of Assumption SA2.4 for L_p regression)**.** *Suppose Assumptions SA2.1 to SA2.3 hold with $\mathcal{Q}$ a singleton, loss function $\rho(y, \eta) = |y - \eta|^p$, $p \in (1, 2]$, $\mu_0(\cdot)$ as defined in (SA3.1), $\mathfrak{M}$ the Lebesgue measure. Assume the real inverse link function $\eta(\cdot)\colon \mathbb{R} \to \mathcal{E}$ is strictly monotonic and twice continuously differentiable with $\mathcal{E}$ an open connected subset of $\mathbb{R}$. Denoting by a_l and a_r the left and the right ends of $\mathcal{E}$ respectively (possibly $\pm\infty$), assume that $\int_{\mathbb{R}} \psi(y, a_l) f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y < 0$ if a_l is finite, and $\int_{\mathbb{R}} \psi(y, a_r) f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y > 0$ if a_r is finite. Also assume that $\mathbb{E}[|y_1|^{\nu(p-1)}] < \infty$ for some $\nu > 2$, and that $\boldsymbol{x} \mapsto f_{Y|X}(y|\boldsymbol{x})$ is continuous for any $y \in \mathcal{Y}$. In addition, assume that $\eta \mapsto \int_{\mathbb{R}} |\eta(\zeta) - y|^{p-1} \operatorname{sign}(\eta(\zeta) - y) f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y$ is twice continuously differentiable with derivatives $\frac{\mathrm{d}^j}{\mathrm{d}\eta^j} \int_{\mathbb{R}} |\eta(\zeta) - y|^{p-1} \operatorname{sign}(\eta(\zeta) - y) f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y = \int_{\mathbb{R}} |\eta(\zeta) - y|^{p-1} \operatorname{sign}(\eta(\zeta) - y) \frac{\partial^j}{\partial y^j} f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y$ for $j \in \{1, 2\}$. Moreover, the function $\int_{\mathbb{R}} |\eta(\zeta) - y|^{p-1} \operatorname{sign}(\eta(\zeta) - y) \frac{\partial}{\partial y} f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y$ is bounded and bounded away from zero uniformly over $\boldsymbol{x} \in \mathcal{X}$, $\zeta \in B(\boldsymbol{x})$ with $B(\boldsymbol{x})$ defined in (SA2.1), and the function $\int_{\mathbb{R}} |\eta(\zeta) - y|^{p-1} \operatorname{sign}(\eta(\zeta) - y) \frac{\partial^2}{\partial y^2} f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y$ is bounded in absolute value uniformly over $\boldsymbol{x} \in \mathcal{X}$, $\zeta \in B(\boldsymbol{x})$. Then Assumptions SA2.4, SA2.5 and SA2.8, Eq. (SA6.15) and Condition (vi) of Lemma SA4.3 are also true.*

Proof. Since $\mathcal{Q}$ is a singleton, we will omit the index q in notations.

Verifying the assumptions of Lemma SA3.10 The fact that

$$\zeta \mapsto \int_{\mathbb{R}} \psi(y, \zeta) f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y$$

is continuous is proven below in the verification of Assumption SA2.4(iii). To ensure that it crosses zero if a_l or a_r is not finite, we show

$$\int_{\mathbb{R}} \psi(y, \zeta) f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y \to -\infty \text{ as } \zeta \to -\infty, \tag{SA8.3}$$

$$\int_{\mathbb{R}} \psi(y, \zeta) f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y \to +\infty \text{ as } \zeta \to +\infty. \tag{SA8.4}$$

To prove (SA8.3), recall that $\psi(y, \zeta) = p|y - \zeta|^{p-1} \operatorname{sign}(\zeta - y)$ and therefore

$$\begin{aligned}
\int_{\mathbb{R}} \psi(y, \zeta) f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y &= -p \int_{\zeta}^{+\infty} (y - \zeta)^{p-1} f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y + p \int_{-\infty}^{\zeta} (\zeta - y)^{p-1} f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y \\
&= -p(-\zeta)^{p-1} \underbrace{\int_{\mathbb{R}} \left(1 + \frac{y}{-\zeta}\right)^{p-1} \mathbb{1}\{y > \zeta\} f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y}_{\to 1} \\
&\quad + p \underbrace{\int_{\mathbb{R}} (\zeta - y)^{p-1} \mathbb{1}\{y \leq \zeta\} f_{Y|X}(y|\boldsymbol{x})\,\mathrm{d}y}_{\to 0} \to -\infty,
\end{aligned}$$

where we used dominated convergence because for $-\zeta > 1$ we have $1 + y/(-\zeta) \le 1 + |y|$ in the first integral and $\zeta - y \le -y = |y|$ in the second integral. Equation (SA8.4) is proven similarly.

Verifying Assumption SA2.4(i) The function $\rho(y, \eta)$ is continuously differentiable with respect to $\eta \in \mathbb{R}$, and its first derivative is the continuous function $\psi(y, \eta) = p|y-\eta|^{p-1} \operatorname{sign}(\eta - y)$, therefore $\rho(y, \eta)$ for any fixed y is absolutely continuous with respect to η on bounded intervals.

Verifying Assumption SA2.4(ii) The first-order optimality condition is true because $\mu_0(\cdot)$ is defined this way in (SA3.1).

The function

$$\sigma^2(\boldsymbol{x}) := \mathbb{E}\big[\psi(y_i, \mu_0(\boldsymbol{x}))^2 \,\big|\, \boldsymbol{x}_i = \boldsymbol{x}\big] = p^2 \int_{\mathbb{R}} |y - \mu_0(\boldsymbol{x})|^{2p-2} f_{Y|X}(y|\boldsymbol{x})\, \mathrm{d}y$$

is continuous because $\boldsymbol{x} \mapsto |y - \mu_0(\boldsymbol{x})|^{2p-2} f_{Y|X}(y|\boldsymbol{x})$ is continuous and dominated by $(|y|^{2p-2} + C)C'$ for large enough constants C and C'. As a continuous function on a compact set, $\sigma^2(\boldsymbol{x})$ is bounded away from zero because it is non-zero since y_1 has a conditional density.

The family of functions $\{p|y - \eta(\mu_0(\boldsymbol{x}))|^{p-1} \operatorname{sign}(\eta(\mu_0(\boldsymbol{x})) - y)\}$ only contains one element. Note that $\boldsymbol{x} \mapsto |y - \eta(\mu_0(\boldsymbol{x}))|^{\nu(p-1)} f_{Y|X}(y|\boldsymbol{x})$ is continuous and dominated by $(|y|^{\nu(p-1)} + C)C'$ for large enough constants C and C'. Therefore, the function

$$\boldsymbol{x} \mapsto \int_{\mathbb{R}} |y - \eta(\mu_0(\boldsymbol{x}))|^{\nu(p-1)} f_{Y|X}(y|\boldsymbol{x})\, \mathrm{d}y$$

is also continuous. As a continuous function on a compact set, it is bounded.

Verifying Assumption SA2.4(iii) The function $x \mapsto |x|^\alpha \operatorname{sign}(x)$ for $\alpha \in (0, 1]$ is α-Hölder for $x \in \mathbb{R}$ (with constant 2). Therefore, putting $\alpha := p - 1$, for any pair of reals ζ_1 and ζ_2 in a fixed bounded interval we have

$$\begin{aligned}
&\sup_{\boldsymbol{x}} \sup_{\lambda \in [0,1]} \sup_{y} |\psi(y, \eta(\zeta_1 + \lambda(\zeta_2 - \zeta_1))) - \psi(y, \eta(\zeta_2))| \\
&\qquad \le 2p|\eta(\zeta_1 + \lambda(\zeta_2 - \zeta_1)) - \eta(\zeta_2)|^{p-1} \overset{\text{(a)}}{\lesssim} |\zeta_1 - \zeta_2|^{p-1},
\end{aligned}$$

where in (a) we used that the link function $\eta(\cdot)$ in a fixed bounded interval is Lipschitz.

Verifying Assumption SA2.4(iv) The conditions of Assumption SA2.4(iv) are directly assumed in the statement of Proposition SA8.5.

Verifying Assumption SA2.5 Since $\mathcal{G}_2$, $\mathcal{G}_4$, $\mathcal{G}_5$ are just singletons (and the existence of corresponding envelopes holds trivially), it is enough to consider $\mathcal{G}_1$ and $\mathcal{G}_3$.

Assume that $\boldsymbol{\beta}$ and $\widetilde{\boldsymbol{\beta}}$ are such that $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_\infty \le r$ and $\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_\infty \le r$. Note that

$$\begin{aligned}
&\big|\big[\psi(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})) - \psi(y, \eta(\mu_0(\boldsymbol{x})))\big] - \big[\psi(y, \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\widetilde{\boldsymbol{\beta}})) - \psi(y, \eta(\mu_0(\boldsymbol{x})))\big]\big| \\
&\quad \le 2p\big|\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}) - \eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\widetilde{\boldsymbol{\beta}})\big|^{p-1} \lesssim \|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}\|_\infty^{p-1}.
\end{aligned}$$

The result for $\mathcal{G}_1$ follows.

Similarly,

$$\begin{aligned}
&\big|\big[\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{\beta}))-\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{\beta}_0))\big]-\big[\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\widetilde{\boldsymbol{\beta}}))-\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{\beta}_0))\big]\big|\\
&\qquad\le 2p\big|\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{\beta})-\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\widetilde{\boldsymbol{\beta}})\big|^{p-1}\lesssim\|\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}\|_\infty^{p-1}.
\end{aligned}\tag{SA8.5}$$

For a fixed cell $\delta\in\Delta$ the class of functions of the form

$$\big[\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{\beta});q)-\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{\beta}_0);q)\big]\mathbb{1}\{\boldsymbol{x}\in\mathcal{N}_{[c\log n]}(\delta)\}$$

can be parametrized by $\boldsymbol{\beta}$ lying in a fixed vector space $\mathcal{B}_\delta$ of dimension $O(\log^d n)$. The result now follows from the bound (SA8.5) (by using (SA5.34) similarly to the proof of Proposition SA8.6).

Verifying the addition to Assumption SA2.8 Fix $\delta\in\Delta$. Let $\boldsymbol{\beta}$ and $\widetilde{\boldsymbol{\beta}}$ be such that $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_\infty\le r$ and $\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_\infty\le r$; let $\boldsymbol{v}$ and $\widetilde{\boldsymbol{v}}$ be such that $\|\boldsymbol{v}\|_\infty\le\varepsilon_n$ and $\|\widetilde{\boldsymbol{v}}\|_\infty\le\varepsilon_n$. To declutter notation, put

$$\begin{aligned}
g(t)&:=\big[\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}(\boldsymbol{\beta}_0+\boldsymbol{\beta})+t))-\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{\beta}_0))\big]\\
&\quad\times\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}(\boldsymbol{\beta}_0+\boldsymbol{\beta})+t),\\
\tilde g(t)&:=\big[\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}(\boldsymbol{\beta}_0+\widetilde{\boldsymbol{\beta}})+t))-\psi(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{\beta}_0))\big]\\
&\quad\times\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}(\boldsymbol{\beta}_0+\widetilde{\boldsymbol{\beta}})+t).
\end{aligned}$$

Note that

$$\int_{-\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{v}}^0 g(t)\,\mathrm{d}t-\int_{-\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\widetilde{\boldsymbol{v}}}^0\tilde g(t)\,\mathrm{d}t=\int_{-\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{v}}^0\big(g(t)-\tilde g(t)\big)\,\mathrm{d}t+\int_{-\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{v}}^{-\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\widetilde{\boldsymbol{v}}}\tilde g(t)\,\mathrm{d}t.$$

Since $\psi(y,\cdot)$ is $(p-1)$-Hölder continuous in the second argument and functions $\eta(\cdot)$, $\eta^{(1)}(\cdot)$ in a fixed bounded interval are Lipschitz, we get that uniformly over t and $\boldsymbol{x}$ in these integrals

$$\big|g(t)-\tilde g(t)\big|\lesssim\big\|\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}\big\|_\infty^{p-1}+\big\|\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}\big\|_\infty,$$

and $\big|\tilde g(t)\big|$ is bounded. This gives

$$\left|\int_{-\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{v}}^0 g(t)\,\mathrm{d}t-\int_{-\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\widetilde{\boldsymbol{v}}}^0\tilde g(t)\,\mathrm{d}t\right|\lesssim\varepsilon_n\big(\big\|\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}\big\|_\infty^{p-1}+\big\|\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}\big\|_\infty\big)+\big\|\boldsymbol{v}-\widetilde{\boldsymbol{v}}\big\|_\infty$$

It means that taking an ε-net (for ε smaller than 1) in the space of $\boldsymbol{\beta}$ and an $\varepsilon_n\varepsilon$-net in the space of $\boldsymbol{v}$ induces an $C_{22}\varepsilon_n\varepsilon$-net in the space of functions

$$\left\{(\boldsymbol{x},y)\mapsto\int_{-\boldsymbol{p}(\boldsymbol{x})^{\mathsf T}\boldsymbol{v}}^0 g(t)\,\mathrm{d}t\cdot\mathbb{1}\{\boldsymbol{x}\in\mathcal{N}_{[c\log n]}(\delta)\}:\|\boldsymbol{\beta}-\boldsymbol{\beta}_0(q)\|_\infty\le r,\|\boldsymbol{v}\|_\infty\le\varepsilon_n\right\}$$

in terms of the sup-norm, where C_{22} is some constant. Possibly increasing C_{22}, we can conclude that this class with envelope C_{22} satisfies the uniform entropy bound (SA1.3) with $A\lesssim 1$ and $V\lesssim\log^d n$ (where we used that $\boldsymbol{\beta}$ and $\boldsymbol{v}$ can be assumed to lie in a vector space of dimension $O(\log^d n)$). By (SA5.34), the same can be said about the union of $O(h^{-d})$ such classes (corresponding to different δ). The verification is concluded.

Verifying (SA6.15) This is obvious because $\mathcal{Q}$ is a singleton.

Verifying Lemma SA4.3(vi) It follows from the proof of Lemma SA4.2 that

$$\begin{aligned}\big|\rho(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}))-\rho(y,\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\widetilde{\boldsymbol{\beta}}))\big| &\lesssim \big(1+\overline{\psi}(\boldsymbol{x},y)\big)\big|\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}})\big|\\ &\lesssim \big(1+\overline{\psi}(\boldsymbol{x},y)\big)\big\|\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}\big\|_{\infty}.\end{aligned}$$

The required uniform entropy bound follows immediately from this.

Proposition SA8.5 is proven. □

SA8.4 Logistic regression

Proposition SA8.6 (Verification of Assumption SA2.4 and others for logistic regression)**.** *Suppose Assumptions SA2.1 to SA2.3 hold with $\mathcal{Q}$ a singleton, $\mathcal{Y}=\{0,1\}$, $\eta(\theta)=1/(1+e^{-\theta})$, $\mathfrak{M}$ is the counting measure on $\{0,1\}$, and $\rho(y,\eta)=-y\log(\eta)-(1-y)\log(1-\eta)$. Assume also $\pi(\boldsymbol{x}):=\mathbb{P}\{y_1=1\,|\,\boldsymbol{x}_1=\boldsymbol{x}\}$ is continuous and $\pi(\boldsymbol{x})$ lies in the interval $(0,1)$ for any $\boldsymbol{x}\in\mathcal{X}$. Then Assumptions SA2.4, SA2.5 and SA2.8, (SA6.15), Conditions (v) and (vi) of Lemma SA4.3 are true.*

We will prove Proposition SA8.6 now. Since $\mathcal{Q}$ is a singleton, we will omit the index q in notations.

Verifying Assumption SA2.4(i) The function $\rho(y,\eta)$ is infinitely smooth with respect to $\eta\in(0,1)$, and its first derivative is $\psi(y,\eta)=(1-y)/(1-\eta)-y/\eta$, therefore $\rho(y,\eta)$ for any fixed y is absolutely continuous with respect to η on bounded intervals.

Verifying Assumption SA2.4(ii) We have $\mathbb{E}[\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i)))\,|\,\boldsymbol{x}_i]=0$ since

$$\eta(\mu_0(\boldsymbol{x}_i))=\mathbb{E}[y_i\,|\,\boldsymbol{x}_i]=\mathbb{P}\{y_i=1\,|\,\boldsymbol{x}_i\}=\pi(\boldsymbol{x}_i).$$

Next,

$$\begin{aligned}\sigma^2(\boldsymbol{x})=\mathbb{E}\big[\psi(y_i,\eta(\mu_0(\boldsymbol{x}_i)))^2\,\big|\,\boldsymbol{x}_i=\boldsymbol{x}\big]&=\frac{\mathbb{E}\big[(y_i-\pi(\boldsymbol{x}_i))^2\,\big|\,\boldsymbol{x}_i=\boldsymbol{x}\big]}{\eta^{(1)}(\mu_0(\boldsymbol{x}))^2}\\ &=\frac{\pi(\boldsymbol{x})(1-\pi(\boldsymbol{x}))}{\pi(\boldsymbol{x})^2(1-\pi(\boldsymbol{x}))^2}=\frac{1}{\pi(\boldsymbol{x})(1-\pi(\boldsymbol{x}))}\end{aligned}$$

is continuous and bounded away from zero (is not less than 4).

Since $\boldsymbol{x}_i$ lies in a compact set, $\psi(y,\eta(\mu_0(\boldsymbol{x}_i)))$ is bounded, so it has moments of any order.

Verifying Assumption SA2.4(iii) The function $\rho(y,\eta)$ is infinitely smooth with respect to $\eta\in(0,1)$, and its first derivative is $\psi(y,\eta)=(1-y)/(1-\eta)-y/\eta$. Using the famous expression for the derivative of the logistic function $\eta^{(1)}(\theta)=\eta(\theta)(1-\eta(\theta))$, we get

$$\begin{aligned}&\frac{\partial}{\partial\theta}\rho(y,\eta(\theta))=\psi(y,\eta(\theta))\eta^{(1)}(\theta)=(1-y)\eta(\theta)-y(1-\eta(\theta))=\eta(\theta)-y,\\ &\frac{\partial^2}{\partial\theta^2}\rho(y,\eta(\theta))=\eta^{(1)}(\theta)=\eta(\theta)(1-\eta(\theta)).\end{aligned}$$

Since the logistic link maps to $(0,1)$, the second derivative is positive (and does not depend on y). Therefore, $\rho(y,\eta(\theta))$ is convex with respect to θ for any y.

Uniformly over ζ_1 and ζ_2 in a fixed bounded interval, we have

$$\sup_y \bigl|\psi(y,\eta(\zeta_1)) - \psi(y,\eta(\zeta_2))\bigr| \lesssim \bigl|\eta(\zeta_1) - \eta(\zeta_2)\bigr| \overset{\text{(a)}}{\leq} |\zeta_1 - \zeta_2|,$$

where in (a) we used that the derivative of $\eta(\cdot)$ does not exceed 1. So in this case the Hölder parameter $\alpha = 1$.

The logistic function $\eta(\cdot)$ is strictly monotonic and infinitely smooth on $\mathbb{R}$.

Verifying Assumption SA2.4(iv) In this case

$$\Psi(\boldsymbol{x};\eta) = \frac{\eta - \mathbb{E}[y_i \mid \boldsymbol{x}_i = \boldsymbol{x}]}{\eta(1-\eta)} = \frac{\eta - \pi(\boldsymbol{x})}{\eta(1-\eta)}.$$

This function is infinitely smooth in η for $\eta \in (0,1)$. Its first derivative

$$\Psi_1(\boldsymbol{x},\eta) = \frac{\partial}{\partial\eta}\Psi(\boldsymbol{x};\eta) = \frac{\eta^2 - 2\eta\pi(\boldsymbol{x}) + \pi(\boldsymbol{x})}{\eta^2(1-\eta)^2}.$$

Therefore,

$$\Psi_1(\boldsymbol{x},\eta(\zeta))\eta^{(1)}(\zeta)^2 = \eta(\zeta)^2 - 2\eta(\zeta)\pi(\boldsymbol{x}) + \pi(\boldsymbol{x}). \tag{SA8.6}$$

If $|\zeta - \mu_0(\boldsymbol{x})| \leq r$, then $\bigl|\eta(\zeta) - \eta(\mu_0(\boldsymbol{x}))\bigr| = \bigl|\eta(\zeta) - \pi(\boldsymbol{x})\bigr| \leq r$ (since the derivative of $\eta(\cdot)$ does not exceed 1). Since $0 < \min_{\boldsymbol{x}\in\mathcal{X}} \pi(\boldsymbol{x})(1-\pi(\boldsymbol{x})) \leq \max_{\boldsymbol{x}\in\mathcal{X}} \pi(\boldsymbol{x})(1-\pi(\boldsymbol{x})) < 1$, for small enough r the right-hand side of (SA8.6) for such ζ is also bounded away from zero and one.

Finally,

$$\Psi_2(\boldsymbol{x},\eta) = \frac{\partial}{\partial\eta}\Psi_1(\boldsymbol{x},\eta) = \frac{2\bigl(\eta^3 - 3\pi(\boldsymbol{x})\eta^2 + 3\pi(\boldsymbol{x})\eta - \pi(\boldsymbol{x})\bigr)}{\eta^3(1-\eta)^3}.$$

Again since $0 < \min_{\boldsymbol{x}\in\mathcal{X}} \pi(\boldsymbol{x})(1-\pi(\boldsymbol{x})) \leq \max_{\boldsymbol{x}\in\mathcal{X}} \pi(\boldsymbol{x})(1-\pi(\boldsymbol{x})) < 1$, for such ζ that $|\zeta - \mu_0(\boldsymbol{x})| \leq r$ and r small enough, the product $\eta(\zeta)(1-\eta(\zeta))$ is bounded away from zero. So for such ζ, $|\Psi_2(\boldsymbol{x},\eta(\zeta))|$ is uniformly bounded.

Lemma SA8.7 (Class G_1)**.** *The class*

$$\mathcal{G}_1 = \left\{\mathcal{X}\times\mathbb{R} \ni (\boldsymbol{x},y) \mapsto \frac{\eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}) - y}{\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})} - \frac{\pi(\boldsymbol{x}) - y}{\pi(\boldsymbol{x})\bigl(1-\pi(\boldsymbol{x})\bigr)} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq r\right\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$.

Proof of Lemma SA8.7. The class

$$\mathcal{G}_{11} = \bigl\{(\boldsymbol{x},y) \mapsto \boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^K\bigr\}$$

is VC with index not exceeding $K+2$ by Lemma 2.6.15 in [18]. Since in a fixed bounded interval $\eta(\cdot)$ is Lipschitz and $(\boldsymbol{x},y) \mapsto y$ is one fixed function, the class

$$\mathcal{G}_{12} = \bigl\{(\boldsymbol{x},y) \mapsto \eta(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta}) - y : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_\infty \leq r\bigr\}$$

with a large enough constant envelope (recall that $\mathcal{Y}$ is a bounded set) satisfies the uniform entropy bound (SA1.3) with $A \lesssim 1$ and $V \lesssim K$. Since $1/\eta^{(1)}(\cdot)$ in a fixed bounded interval is Lipschitz, the same is true of

$$\mathcal{G}_{13} = \bigl\{(\boldsymbol{x},y) \mapsto \eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})^{-1} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_\infty \leq r\bigr\},$$

where we used again that under these constraints $\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{\beta})$ is bounded away from zero. This implies by Lemma SA3.4 that it is true of $\mathcal{G}_{12}\cdot\mathcal{G}_{13} - \psi(y,\eta(\boldsymbol{x}))$ (since $\psi(y,\eta(\boldsymbol{x}))$ is one fixed function), which is what we need. □

Lemma SA8.8 (Class $\mathcal{G}_3$)**.** *The class*

$$\mathcal{G}_3 := \Big\{\mathcal{X} \times \mathbb{R} \ni (\boldsymbol{x}, y) \mapsto \left[\frac{\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}) - y}{\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})} - \frac{\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0) - y}{\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0)}\right]\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r, \delta \in \Delta\Big\}$$

with a large enough constant envelope satisfies the uniform entropy bound (SA1.3) *with $A \lesssim 1$ and $V \lesssim \log^d n$.*

Proof of Lemma SA8.8. For a fixed vector space $\mathcal{B}$ of dimension $\dim \mathcal{B}$,

$$\mathcal{W}_{\mathcal{B}} := \left\{(\boldsymbol{x}, y) \mapsto \frac{\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}) - y}{\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})} : \boldsymbol{\beta} \in \mathcal{B}, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r\right\}$$

with a large enough constant envelope (recall that $\mathcal{Y}$ is a bounded set) satisfies the uniform entropy bound with $A \lesssim 1$ and $V \lesssim \dim \mathcal{B}$ by the same argument as in the proof of Lemma SA8.7. Therefore, for any fixed $c > 0$, $\delta \in \Delta$, the class

$$\mathcal{W}_{3,\delta} := \left\{(\boldsymbol{x}, y) \mapsto \frac{\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}) - y}{\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})}\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\} : \boldsymbol{\beta} \in \mathbb{R}^K, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r\right\}$$

with a large enough constant envelope also satisfies the uniform entropy bound with $A \lesssim 1$ and $V \lesssim \log^d n$ because it is contained in the product of $\mathcal{W}_{\mathcal{B}_{3,\delta}}$ for some vector space $\mathcal{B}_{3,\delta}$ of dimension $\dim \mathcal{B}_{3,\delta} \lesssim \log^d n$ and a fixed function $\mathbb{1}\{\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)\}$. Subtracting a fixed bounded function does not change this fact, so the same is true of

$$\mathcal{G}_{3,\delta} := \Big\{(\boldsymbol{x}, y) \mapsto \left[\frac{\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}) - y}{\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta})} - \frac{\eta(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0) - y}{\eta^{(1)}(\boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0)}\right]\mathbb{1}(\boldsymbol{x} \in \mathcal{N}_{[c\log n]}(\delta)) : \boldsymbol{\beta} \in \mathbb{R}^K, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(q)\|_\infty \le r\Big\}$$

Since there are $O(h^{-d})$ such classes and $\log(1/h) \lesssim \log n$, using the chain (SA5.34) we obtain that $\mathcal{G}_3$ satisfies the uniform entropy bound (SA1.3) with with $A \lesssim 1$ and $V \lesssim \log^d n$. $\square$

Verifying Assumption SA2.5 Classes $\mathcal{G}_2$, $\mathcal{G}_4$, $\mathcal{G}_5$ are just singletons (and the existence of corresponding envelopes holds trivially). Classes $\mathcal{G}_1$ and $\mathcal{G}_3$ are tackled in Lemma SA8.7 and Lemma SA8.8.

Verifying Assumption SA2.8 This is verified (in a more general setting) in Section SA2.1.2.

Verifying (SA6.15) This is obvious since $\mathcal{Q}$ is a singleton.

Verifying the condition in Lemma SA4.3(v) Recall that in this case

$$\Psi_1(\boldsymbol{x}, \eta) = \frac{\eta^2 - 2\eta\pi(\boldsymbol{x}) + \pi(\boldsymbol{x})}{\eta^2(1-\eta)^2}.$$

The numerator is always positive since $0 < \pi(\boldsymbol{x}) < 1$, and the denominator is also positive since $\eta \in (0,1)$. Since $\Psi_1(\boldsymbol{x}, \eta)$ is continuous in both arguments and the image of a compact set under a continuous mapping is compact, we see that for any fixed compact subset of $(0,1)$, $\Psi_1(\boldsymbol{x}, \eta)$ is bounded away from zero uniformly over $\boldsymbol{x} \in \mathcal{X}$ and η lying in this compact subset.

Verifying the condition in Lemma SA4.3(vi) In this verification, we will use $\theta_1 := \boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}$, $\tilde{\theta}_1 := \boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\widetilde{\boldsymbol{\beta}}$ and $\theta_2 := \boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\boldsymbol{\beta}_0$ to simplify notations. Note that for θ lying in a fixed compact, both functions $\log \eta(\theta)$ and $\log(1-\eta(\theta))$ are Lipschitz in θ, so if $\|\boldsymbol{\beta}\|_\infty \le R$ and $\|\widetilde{\boldsymbol{\beta}}\|_\infty \le R$, we have

$$|\rho(y, \eta(\theta_1)) - \rho(y, \eta(\theta_2)) - \rho(y, \eta(\tilde{\theta}_1)) + \rho(y, \eta(\theta_2))| \lesssim |\theta_1 - \tilde{\theta}_1| \lesssim |\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}|,$$

where the constants in $\lesssim$ are allowed to depend on R but not on $\boldsymbol{\beta}, \widetilde{\boldsymbol{\beta}}$, $\boldsymbol{x}$ or y (we used that y and $1-y$ are bounded by 1). The result follows.

This concludes the proof of Proposition SA8.6.

SA8.5 Examples of basis functions

Fix $0 \le s < m$. Take any segment $[a,b]$ and partition it into J sub-segments by taking $J-1$ knots $a < \tau_1 < \dots \tau_{J-1} < b$. Consider a non-increasing tuple of real numbers $(t_k)_{k=1}^{2m+(m-s)(J-1)}$ defined as follows:

$$\begin{aligned}
&t_1 = \dots = t_m = a,\\
&t_{m+1} = \dots = t_{2m-s} = \tau_1,\\
&\dots\\
&t_{m+(m-s)(J-2)+1} = \dots = t_{m+(m-s)(J-1)} = \tau_{J-1},\\
&t_{m+(m-s)(J-1)+1} = \dots = t_{2m+(m-s)(J-1)} = b,
\end{aligned}$$

and put for $l \in \{1, 2, \dots, m+(m-s)(J-1)\}$

$$\begin{gathered}
p_l(x) = (-1)^m (t_{l+m} - t_l)[t_l, \dots, t_{l+m}](x-t)_+^{m-1}, \quad x \in [a,b),\\
p_{m+(m-s)(J-1)}(b) = \lim_{x \uparrow b} p_{m+(m-s)(J-1)}(x),
\end{gathered}$$

where $[t_l, \dots, t_{l+m}]g(t,x)$ denotes the divided difference [16, Definition 2.49] of $t \mapsto g(t,x)$ over points $t_l, \dots, t_{l+m}$, and $\alpha_+ := \alpha \vee 0$. We will call the basis $\{p_l(\cdot)\}_{l=1}^{m+(m-s)(J-1)}$ the (normalized) *B-spline basis* of order m with knots $\tau_1 < \dots < \tau_{J-1}$, each of multiplicity $m-s$. The functions $\{p_l(\cdot)\}_{l=1}^{m+(m-s)(J-1)}$ form a basis in the function space

$$\big\{f(\cdot) \in \mathcal{C}^{s-1}[a,b] : f(\cdot) \text{ is a polynomial of degree } m-1 \text{ on each } [\tau_j, \tau_{j+1}],\ j \in \{0, \dots, J-1\}\big\},$$

where we put $\tau_0 = a$, $\tau_J = b$ for simplicity ([16, Corollary 4.10])[3]; moreover, each $p_l(\cdot)$ is strictly positive on (t_l, t_{l+m}), zero on the complement of $[t_l, t_{l+m}]$, and does not exceed 1 ([16, Theorem 4.9]).

[3] The symbol $\mathcal{C}^{-1}[a,b]$ here corresponds to the family of all functions $[a,b] \to \mathbb{R}$ (with no smoothness restrictions).

This definition can be extended to dimension d by considering $\mathcal{X} = \bigotimes_{\ell=1}^{d}[a_\ell, b_\ell]$ and the corresponding tensor products

$$p_{l_1,\dots,l_d}(\boldsymbol{x}) = p_{l_1}(x_1)\dots p_{l_d}(x_d)$$

for all tuples $\boldsymbol{l} = (l_1,\dots,l_d)$ such that $l_\ell \in \{1,2,\dots,m+(m-s)(J_\ell-1)\}$. The d partitions of $[a_\ell, b_\ell]$ induce a tensor-product partition Δ of $\mathcal{X}$ (with $\kappa = J_1\dots J_d$ cells), and the basis $\boldsymbol{p}(\boldsymbol{x})$ (arranged in a lexicographic order of $\boldsymbol{l}$) is called the *tensor-product B-spline basis* of order m associated with the tensor-product partition Δ ([16, Definition 12.3]).

We refer to [16] for details.

Proposition SA8.9 (Verifying Assumptions SA2.2 and SA2.6)**.** *Suppose Assumptions SA2.1 and SA2.3 hold with the tensor-product partition Δ as described above. Let $\boldsymbol{p}(\boldsymbol{x})$ be a tensor-product B-spline basis of order m associated with Δ. Then $\boldsymbol{p}(\boldsymbol{x})$ satisfies Assumptions SA2.2 and SA2.6.*

Proof. It follows from the more general argument for Lemma SA-6.1 in [6]. □

Taking $s = m-1$ recovers standard tensor-product B-splines with simple (multiplicity 1) partition knots, whereas $s = 0$ corresponds to piecewise polynomials, where, in particular, each basis function is only supported on one cell (making Lemmas SA4.4 and SA4.5 applicable). Additional examples are provided in [6, Section SA-6].

SA9 Other parameters of interest

This section formalizes the discussion in Section 9 of the paper. The following theorem is now a simple corollary of the previous results presented in this supplemental appendix.

Theorem SA9.1 (Other parameters of interest)**.**

(a) Suppose all the conditions of Theorem SA6.4(a) hold with $\boldsymbol{v} = \boldsymbol{0}$ and $r_{\mathtt{UC}} = o(1)$. Then

$$\begin{aligned}\sup_{\boldsymbol{q}\in\mathcal{Q}}\sup_{\boldsymbol{x}\in\mathcal{X}}&\left|\frac{\eta\big(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q})\big)-\eta\big(\mu_0(\boldsymbol{x},\boldsymbol{q})\big)}{\big|\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\big|\sqrt{\bar{\Omega}_{\boldsymbol{0}}(\boldsymbol{x},\boldsymbol{q})/n}}-\operatorname{sign}\big\{\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\big\}\bar{\boldsymbol{\ell}}_{\boldsymbol{0}}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}(\boldsymbol{q})\right|\\&=o_{\mathbb{P}}(r_{\mathtt{SA}})+O_{\mathbb{P}}\Big(\sqrt{nh^d}\big(r_{\mathtt{UC}}^2+r_{\mathtt{BR}}\big)\Big)\end{aligned}$$

with $\boldsymbol{Z}(\boldsymbol{q})$ defined in Theorem SA6.4.

(b) Fix any $k \in \{1,\dots,d\}$. Suppose all the conditions of Theorem SA6.4(a) hold with $\boldsymbol{v} = \boldsymbol{e}_k$, where $\boldsymbol{e}_k = (0,\dots,0,1,0,\dots,0)^{\mathsf{T}} \in \mathbb{R}^d$ with 1 at the kth place. Then

$$\begin{aligned}\sup_{\boldsymbol{q},\boldsymbol{x}}&\left|\frac{\eta^{(1)}\big(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q})\big)\widehat{\mu}^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})-\eta^{(1)}\big(\mu_0(\boldsymbol{x},\boldsymbol{q})\big)\mu_0^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})}{\big|\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\big|\sqrt{\bar{\Omega}_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})}}\right.\\&\qquad\left.-\operatorname{sign}\big\{\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\big\}\bar{\boldsymbol{\ell}}_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})^{\mathsf{T}}\boldsymbol{Z}(\boldsymbol{q})\right|\\&=o_{\mathbb{P}}(r_{\mathtt{SA}})+O_{\mathbb{P}}\Big(\sqrt{nh^d}\big(r_{\mathtt{UC}}^2+r_{\mathtt{BR}}+hr_{\mathtt{UC}}\big)\Big).\end{aligned}$$

(c) If $\overline{\psi}(\boldsymbol{x}_i, y_i)$ is σ^2-sub-Gaussian conditionally on $\boldsymbol{x}_i$, then in Assertions (a) and (b) $r_{\mathtt{SA}}$ can be replaced with $r_{\mathtt{SA}}^{\mathrm{sub}}$.

Proof. (a) We have uniformly over $\boldsymbol{q}$ and $\boldsymbol{x}$

$$
\begin{aligned}
&\eta(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q})) - \eta(\mu_0(\boldsymbol{x},\boldsymbol{q}))\\
&\overset{\text{(a)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}) - \mu_0(\boldsymbol{x},\boldsymbol{q})) + \frac{\eta^{(2)}(\xi)}{2}(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}) - \mu_0(\boldsymbol{x},\boldsymbol{q}))^2\\
&\overset{\text{(b)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}) - \mu_0(\boldsymbol{x},\boldsymbol{q})) + O_{\mathbb{P}}(r_{\mathtt{UC}}^2)\\
&\overset{\text{(c)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))t_{\mathbf{0}}(\boldsymbol{x},\boldsymbol{q})\sqrt{\bar{\Omega}_{\mathbf{0}}(\boldsymbol{x},\boldsymbol{q})/n} + O_{\mathbb{P}}(r_{\mathtt{UC}}^2 + r_{\mathtt{BR}})\\
&\overset{\text{(d)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\sqrt{\bar{\Omega}_{\mathbf{0}}(\boldsymbol{x},\boldsymbol{q})/n}\Big[t_{\mathbf{0}}(\boldsymbol{x},\boldsymbol{q}) + O_{\mathbb{P}}\big(\sqrt{nh^d}(r_{\mathtt{UC}}^2 + r_{\mathtt{BR}})\big)\Big].
\end{aligned}
$$

Here, (a) is by Taylor expansion, with some $\xi = \xi_{\boldsymbol{q},\boldsymbol{x}}$ between $\widehat{\mu}(\boldsymbol{x},\boldsymbol{q})$ and $\mu_0(\boldsymbol{x},\boldsymbol{q})$. In (b), we used $r_{\mathtt{UC}} = o(1)$ (giving that $\eta^{(2)}(\xi)$ does not exceed a fixed constant not depending on $\boldsymbol{q}$ or $\boldsymbol{x}$) and (SA6.12). (c) is by (SA6.37) and since $\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))$ is uniformly bounded. (d) is by $h^{-2|\boldsymbol{v}|-d} \lesssim_{\mathbb{P}} \inf_{\boldsymbol{q},\boldsymbol{x}} |\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})|$ (by Lemma SA6.6) and since $|\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))|$ is bounded away from zero by Assumption SA2.4(iv).

Rewriting, we obtain

$$
\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\frac{\eta(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q})) - \eta(\mu_0(\boldsymbol{x},\boldsymbol{q}))}{\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\sqrt{\bar{\Omega}_{\mathbf{0}}(\boldsymbol{x},\boldsymbol{q})/n}} - t_{\mathbf{0}}(\boldsymbol{x},\boldsymbol{q})\right| \lesssim_{\mathbb{P}} \sqrt{nh^d}\big(r_{\mathtt{UC}}^2 + r_{\mathtt{BR}}\big).
$$

It is left to combine this with (SA6.19) and use the triangle inequality.

(b) We have uniformly over $\boldsymbol{q}$ and $\boldsymbol{x}$

$$
\begin{aligned}
&\eta^{(1)}(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}))\widehat{\mu}^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q}) - \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\mu_0^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})\\
&\overset{\text{(a)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\big(\widehat{\mu}^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})\big)\\
&\quad + \big(\eta^{(1)}(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q})) - \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\big)\widehat{\mu}^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})\\
&\overset{\text{(b)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\big(\widehat{\mu}^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})\big) + \eta^{(2)}(\zeta)(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}) - \mu_0(\boldsymbol{x},\boldsymbol{q}))\widehat{\mu}^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})\\
&\overset{\text{(c)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))t_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})\sqrt{\bar{\Omega}_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})/n}\\
&\quad + \eta^{(2)}(\zeta)\big(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}) - \mu_0(\boldsymbol{x},\boldsymbol{q})\big)\widehat{\mu}^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q}) + O_{\mathbb{P}}(h^{-1}r_{\mathtt{BR}})\\
&\overset{\text{(d)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))t_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})\sqrt{\bar{\Omega}_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})/n}\\
&\quad + \eta^{(2)}(\zeta)\big(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}) - \mu_0(\boldsymbol{x},\boldsymbol{q})\big)\big(\widehat{\mu}^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q}) - \mu_0^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})\big)\\
&\quad + \eta^{(2)}(\zeta)\big(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}) - \mu_0(\boldsymbol{x},\boldsymbol{q})\big)\mu_0^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q}) + O_{\mathbb{P}}(h^{-1}r_{\mathtt{BR}})\\
&\overset{\text{(e)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))t_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})\sqrt{\bar{\Omega}_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})/n} + O_{\mathbb{P}}\big(h^{-1}(r_{\mathtt{UC}}^2 + r_{\mathtt{BR}}) + r_{\mathtt{UC}}\big)\\
&\overset{\text{(f)}}{=} \eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\sqrt{\bar{\Omega}_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})/n}\Big[t_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q}) + O_{\mathbb{P}}\Big(h\sqrt{nh^d}\big(h^{-1}(r_{\mathtt{UC}}^2 + r_{\mathtt{BR}}) + r_{\mathtt{UC}}\big)\Big)\Big].
\end{aligned}
$$

Here, (a) is just rewriting. In (b) we used the mean-value theorem, with some $\zeta = \zeta_{\boldsymbol{q},\boldsymbol{x}}$ between $\widehat{\mu}(\boldsymbol{x},\boldsymbol{q})$ and $\mu_0(\boldsymbol{x},\boldsymbol{q})$. In (c) we used (SA6.37) with $\boldsymbol{v} = \boldsymbol{e}_k$ and that $\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))$ is bounded uniformly over $\boldsymbol{q},\boldsymbol{x}$. (d) is just rewriting. In (e) we used $r_{\mathtt{UC}} = o(1)$ (giving that $\eta^{(2)}(\zeta)$ does not exceed a fixed constant not depending on $\boldsymbol{q}$ or $\boldsymbol{x}$), (SA6.12) with $\boldsymbol{v} = \mathbf{0}$ and $\boldsymbol{v} = \boldsymbol{e}_k$, and uniform boundedness of first partial derivatives of $\mu_0(\cdot,\boldsymbol{q})$. (f) is by $h^{-2|\boldsymbol{v}|-d} \lesssim_{\mathbb{P}} \inf_{\boldsymbol{q},\boldsymbol{x}} |\bar{\Omega}_{\boldsymbol{v}}(\boldsymbol{x},\boldsymbol{q})|$ (by Lemma SA6.6) and since $\big|\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\big|$ is bounded away from zero by Assumption SA2.4(iv).

Rewriting, we obtain

$$\sup_{\boldsymbol{q},\boldsymbol{x}}\left|\frac{\eta^{(1)}(\widehat{\mu}(\boldsymbol{x},\boldsymbol{q}))\widehat{\mu}^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})-\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\mu_0^{(\boldsymbol{e}_k)}(\boldsymbol{x},\boldsymbol{q})}{\eta^{(1)}(\mu_0(\boldsymbol{x},\boldsymbol{q}))\sqrt{\bar{\Omega}_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})}}-t_{\boldsymbol{e}_k}(\boldsymbol{x},\boldsymbol{q})\right|$$
$$\lesssim_{\mathbb{P}}\sqrt{nh^d}\big(r_{\mathtt{UC}}^2+r_{\mathtt{BR}}+hr_{\mathtt{UC}}\big).$$

It is left to combine this with (SA6.19) and use the triangle inequality.

(c) The argument is the same as for Parts (i) and (ii) with $r_{\mathtt{SA}}$ replaced by $r_{\mathtt{SA}}^{\mathrm{sub}}$. □

SA10 Simulation Evidence

We conducted a small simulation experiment to demonstrate the finite sample properties of the partitioning-based M-estimation methodology. We studied pointwise and uniform (over $q\in\mathcal{Q}$, $\boldsymbol{x}\in\mathcal{X}$) estimation and inference for the conditional distribution function (Example 2 in the paper, and discussed in Section SA8.2 above).

We considered four data generating processes. Model 1 through 3 were chosen to satisfy $y_i=m(\boldsymbol{x}_i)+\epsilon_i$, where $\boldsymbol{x}_i\sim\text{Uniform}(\mathcal{X})$ with $\mathcal{X}=[0,1]^d$, $\epsilon_i\sim\text{Normal}(0,1)$, $\boldsymbol{x}_i\perp\!\!\!\perp\epsilon_i$, $m(\boldsymbol{x})$ is defined in Table 4. For Model 4, we took the treatment effect model from [12] (with a slight change so that the support of $\boldsymbol{x}_i$ is also $[0,1]$):

$$y_i=T_i\left[\mathbf{1}_{V_i\le 1-x_i}\frac{V_i^2}{1-x_i}+\mathbf{1}_{V_i>1-x_i}V_i\right]+(1-T_i)\left[\mathbf{1}_{U_i\le x_i}\frac{U_i^2}{x_i}+\mathbf{1}_{U_i>x_i}U_i\right],$$

where $T_i=\mathbf{1}_{U_i^T<x_i}$, and $x_i,U_i^T,U_i,V_i\sim\text{Uniform}[0,1]$ are independent. The estimand is $\mu_0(\boldsymbol{x},q)$, where

$$\eta(\mu_0(\boldsymbol{x},q))=F(q|\boldsymbol{x})=\mathbb{P}\{y_i\le q\,|\,\boldsymbol{x}_i=\boldsymbol{x}\}$$

with $q\in\mathcal{Q}=[-0.2,0.2]$ for Models 1 through 3 and $\mathcal{Q}=[0.2,0.8]$ for Model 4; the link function $\eta(\cdot)$ is the complementary log-log link: $\eta(t)=1-e^{-e^t}$.

Table 4: Definition of the $m(\boldsymbol{x})$ function.

	$d=1$	$d=2$
Model 1	$\sin(2\pi x)/2$	$(\sin(2\pi x_1)+\sin(2\pi x_2))/4$
Model 2	$\sin(5x)\sin(10x)/2$	$\sin(5x_1)\sin(10x_2)/2$
Model 3	$(1-(4x-2)^2)^2/4$	$(1-(4x_1-2)^2)^2\sin(5x_2)/4$

We use the tensor-product B-spline basis $\boldsymbol{p}(\boldsymbol{x})$ on $[0,1]^d$ as defined in Section SA8.5 with equally spaced knots. (Recall that the number of knots on each segment is $J-1$, and thus $\kappa=J^d$.) The estimator $\widehat{\mu}(\boldsymbol{x},q)=\boldsymbol{p}(\boldsymbol{x})^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(q)$ is constructed by solving the optimization problem

$$\widehat{\boldsymbol{\beta}}(q)\in\operatorname*{arg\,min}_{\boldsymbol{b}\in\mathbb{R}^K}\sum_{i=1}^n\rho(y_i,\eta(\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\boldsymbol{b});q)$$

for each q of interest separately (with $\rho(\cdot)$ defined as in Proposition SA8.3), using the L-BFGS-B algorithm initialized at $\mathbf{0}$. The matrices $\bar{\mathbf{Q}}_q$ and $\bar{\boldsymbol{\Sigma}}_q$ were estimated using, respectively,

$$\widehat{\mathbf{Q}}_{\boldsymbol{q}}=2\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^{\mathsf{T}}\eta^{(1)}(\widehat{\mu}(\boldsymbol{x}_i,q))^2]$$

and

$$\widehat{\mathbf{\Sigma}}_q = 4\mathbb{E}_n[\boldsymbol{p}(\boldsymbol{x}_i)\boldsymbol{p}(\boldsymbol{x}_i)^\mathsf{T}\widehat{F}_{Y|X}(q|\boldsymbol{x})(1-\widehat{F}_{Y|X}(q|\boldsymbol{x}))\eta^{(1)}(\widehat{\mu}(\boldsymbol{x}_i,\boldsymbol{q}))^2],$$

with $\widehat{F}_{Y|X}(q|\boldsymbol{x}) = \eta(\widehat{\mu}(\boldsymbol{x},q))$. Thus, the variance estimator is $\widehat{\Omega}(\boldsymbol{x},q) = \boldsymbol{p}(\boldsymbol{x})^\mathsf{T}\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\widehat{\mathbf{\Sigma}}_{\boldsymbol{q}}\widehat{\mathbf{Q}}_{\boldsymbol{q}}^{-1}\boldsymbol{p}(\boldsymbol{x})$.

The simulation experiment considered two sample sizes ($n = 5\,000$ and $n = 50\,000$), two dimensions ($d = 1$ and $d = 2$), and $1\,000$ replications. The pointwise estimation and inference results are presented in Tables 5 and 7 ($n = 5\,000$) and Tables 9 and 11 ($n = 50\,000$). For each model, we consider two choices of J and three evaluation points for $(q,\boldsymbol{x})$. The root mean squared error (RMSE) for point estimators, the coverage rates and the average widths of pointwise 95% nominal confidence intervals (CIs) are reported. The uniform estimation and inference results are presented in Tables 6 and 8 ($n = 5\,000$) and Tables 10 and 12 ($n = 50\,000$). We consider two choices of J, and use a discrete grid of 8 equally spaced points in place of the continuous segment $\mathcal{Q}$ and 10 (for $d = 1$) or $5 \times 5 = 25$ (for $d = 2$) points in place of $\mathcal{X}$. The uniform 95% nominal confidence bands (CBs) are constructed using the plug-in approach as in Section SA7.1, and the covariance structure of the discrete analogue of $\widehat{\boldsymbol{Z}}(q)$ is obtained as in Remark SA8.4. The maximum estimation error, the uniform coverage rates and the average widths of confidence bands are reported.

Table 5: Pointwise simulation results for points $\{(q^{(k)}, x^{(k)})\}_{k=1}^3$, averaged across $1\,000$ replications with $n = 5\,000$, $d = 1$, where $(q^{(1)}, q^{(2)}, q^{(3)}) = (-0.2, 0, 0.2)$ for Model 1 through 3, $(q^{(1)}, q^{(2)}, q^{(3)}) = (0.45, 0.6, 0.75)$ for Model 4; $(x^{(1)}, x^{(2)}, x^{(3)}) = (0.3, 0.1, 0.2)$.

		RMSE			Coverage			CI width		
Model	J	Point 1	Point 2	Point 3	Point 1	Point 2	Point 3	Point 1	Point 2	Point 3
1	4	0.088	0.074	0.074	0.952	0.953	0.944	0.336	0.289	0.287
	6	0.110	0.077	0.088	0.946	0.951	0.944	0.420	0.303	0.339
2	4	0.069	0.071	0.070	0.953	0.947	0.946	0.270	0.277	0.274
	6	0.088	0.076	0.085	0.950	0.942	0.948	0.343	0.291	0.321
3	4	0.073	0.099	0.068	0.953	0.943	0.945	0.279	0.370	0.258
	6	0.084	0.110	0.072	0.952	0.956	0.954	0.337	0.435	0.292
4	4	0.065	0.058	0.060	0.952	0.961	0.942	0.252	0.237	0.225
	6	0.078	0.063	0.068	0.963	0.953	0.952	0.308	0.246	0.263

Table 6: Uniform simulation results, averaged across 1 000 replications with $n = 5\,000$, $d = 1$.

Model	J	$\sup_{q,\boldsymbol{x}} \lvert\widehat{\mu}(\boldsymbol{x}, q) - \mu_0(\boldsymbol{x}, q)\rvert$	Uniform coverage	Av. CB width
1	4	0.151	0.950	0.399
	6	0.181	0.948	0.475
2	4	0.146	0.939	0.381
	6	0.170	0.940	0.453
3	4	0.187	0.922	0.441
	6	0.223	0.952	0.550
4	4	0.173	0.953	0.415
	6	0.201	0.947	0.494

Table 7: Pointwise simulation results for points $\{(q^{(k)}, \boldsymbol{x}^{(k)})\}_{k=1}^{3}$, averaged across 1 000 replications with $n = 5\,000$, $d = 2$, where $(q^{(1)}, q^{(2)}, q^{(3)}) = (-0.2, 0, 0.2)$, $\boldsymbol{x}^{(1)} = (0.3, 0.1)$, $\boldsymbol{x}^{(2)} = (0.1, 0.4)$ and $\boldsymbol{x}^{(3)} = (0.2, 0.2)$.

		RMSE			Coverage			CI width		
Model	J	Point 1	Point 2	Point 3	Point 1	Point 2	Point 3	Point 1	Point 2	Point 3
1	4	0.272	0.194	0.243	0.957	0.956	0.954	1.052	0.768	0.949
	6	0.455	0.256	0.366	0.944	0.957	0.935	1.417	1.027	1.328
2	4	0.273	0.182	0.237	0.957	0.934	0.941	1.059	0.662	0.904
	6	0.425	0.231	0.318	0.950	0.939	0.955	1.422	0.855	1.248
3	4	0.244	0.260	0.219	0.926	0.936	0.950	0.896	0.944	0.839
	6	0.314	0.784	0.290	0.940	0.932	0.950	1.154	1.558	1.129

Table 8: Uniform simulation results, averaged across 1 000 replications with $n = 5\,000$, $d = 2$.

Model	J	$\sup_{q,\boldsymbol{x}} \lvert\widehat{\mu}(\boldsymbol{x}, q) - \mu_0(\boldsymbol{x}, q)\rvert$	Uniform coverage	Av. CB width
1	4	0.487	0.940	1.163
	6	0.730	0.950	1.593
2	4	0.491	0.935	1.173
	6	0.732	0.944	1.608
3	4	0.478	0.950	1.172
	6	0.806	0.921	1.656

Table 9: Pointwise simulation results for points $\{(q^{(k)}, x^{(k)})\}_{k=1}^3$, averaged across 1 000 replications with $n = 50\,000$, $d = 1$, where $(q^{(1)}, q^{(2)}, q^{(3)}) = (-0.2, 0, 0.2)$ for Model 1 through 3, $(q^{(1)}, q^{(2)}, q^{(3)}) = (0.45, 0.6, 0.75)$ for Model 4; $(x^{(1)}, x^{(2)}, x^{(3)}) = (0.3, 0.1, 0.2)$.

		RMSE			Coverage			CI width		
Model	J	Point 1	Point 2	Point 3	Point 1	Point 2	Point 3	Point 1	Point 2	Point 3
1	4	0.027	0.024	0.023	0.947	0.950	0.956	0.106	0.091	0.091
	6	0.034	0.025	0.027	0.949	0.950	0.957	0.132	0.096	0.107
2	4	0.026	0.023	0.022	0.906	0.940	0.943	0.085	0.088	0.086
	6	0.028	0.023	0.026	0.942	0.963	0.947	0.108	0.092	0.101
3	4	0.023	0.035	0.021	0.946	0.904	0.960	0.088	0.117	0.081
	6	0.028	0.035	0.024	0.946	0.942	0.959	0.107	0.136	0.092
4	4	0.020	0.019	0.019	0.955	0.947	0.943	0.080	0.075	0.071
	6	0.024	0.019	0.022	0.954	0.951	0.950	0.097	0.078	0.083

Table 10: Uniform simulation results, averaged across 1 000 replications with $n = 50\,000$, $d = 1$.

Model	J	$\sup_{q,x} \lvert\widehat{\mu}(x,q) - \mu_0(x,q)\rvert$	Uniform coverage	Av. CB width
1	4	0.047	0.943	0.126
	6	0.057	0.953	0.150
2	4	0.054	0.697	0.120
	6	0.053	0.944	0.143
3	4	0.066	0.871	0.139
	6	0.070	0.941	0.173
4	4	0.055	0.938	0.131
	6	0.063	0.943	0.156

Table 11: Pointwise simulation results for points $\{(q^{(k)}, \boldsymbol{x}^{(k)})\}_{k=1}^{3}$, averaged across 1 000 replications with $n = 50\,000$, $d = 2$, where $(q^{(1)}, q^{(2)}, q^{(3)}) = (-0.2, 0, 0.2)$, $\boldsymbol{x}^{(1)} = (0.3, 0.1)$, $\boldsymbol{x}^{(2)} = (0.1, 0.4)$ and $\boldsymbol{x}^{(3)} = (0.2, 0.2)$.

		RMSE			Coverage			CI width		
Model	J	Point 1	Point 2	Point 3	Point 1	Point 2	Point 3	Point 1	Point 2	Point 3
1	4	0.085	0.064	0.075	0.947	0.940	0.952	0.326	0.239	0.295
	6	0.108	0.084	0.102	0.952	0.941	0.951	0.423	0.313	0.407
2	4	0.083	0.053	0.074	0.952	0.945	0.943	0.328	0.206	0.281
	6	0.107	0.067	0.095	0.951	0.949	0.959	0.424	0.263	0.385
3	4	0.072	0.076	0.068	0.953	0.941	0.944	0.278	0.291	0.262
	6	0.090	0.107	0.088	0.952	0.948	0.952	0.352	0.419	0.350

Table 12: Uniform simulation results, averaged across 1 000 replications with $n = 50\,000$, $d = 2$.

Model	J	$\sup_{q,\boldsymbol{x}} \lvert\widehat{\mu}(\boldsymbol{x}, q) - \mu_0(\boldsymbol{x}, q)\rvert$	Uniform coverage	Av. CB width
1	4	0.146	0.953	0.363
	6	0.193	0.960	0.484
2	4	0.146	0.958	0.366
	6	0.198	0.944	0.489
3	4	0.145	0.930	0.365
	6	0.198	0.950	0.494

References

[1] Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Berlin, Heidelberg: Springer, 3rd edition, DOI: http://dx.doi.org/10.1007/3-540-29587-9.

[2] Belloni, A., Chernozhukov, V., Chetverikov, D., and Fernandez-Val, I. (2019). "Conditional Quantile Processes based on Series or Many Regressors," *Journal of Econometrics*, *213*(1), 4–29.

[3] Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). "Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, *186*(2), 345–366.

[4] Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2024a). "On Binscatter," *American Economic Review*, *114*(5), 1488–1514.

[5] Cattaneo, M. D. and Farrell, M. H. (2013). "Optimal Convergence Rates, Bahadur Representation, and Asymptotic Normality of Partitioning Estimators," *Journal of Econometrics*, *174*(2), 127–143.

[6] Cattaneo, M. D., Farrell, M. H., and Feng, Y. (2020). "Large Sample Properties of Partitioning-Based Series Estimators," *Annals of Statistics*, *48*(3), 1718–1741.

[7] Cattaneo, M. D., Feng, Y., and Underwood, W. G. (2024b). "Uniform Inference for Kernel Density Estimators with Dyadic Data," *Journal of the American Statistical Association*.

[8] Cattaneo, M. D., Masini, R. P., and Underwood, W. G. (2024c). "Yurinskii's Coupling for Martingales," *arXiv preprint arXiv:2210.00362*.

[9] Chen, X. and Christensen, T. M. (2015). "Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions," *Journal of Econometrics*, *188*(2), 447–465.

[10] Chen, X. and Kato, K. (2020). "Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications," *Probability Theory and Related Fields*, *176*, 1097–1163.

[11] Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). "Anti-Concentration and Honest Adaptive Confidence Bands," *Annals of Statistics*, *42*(5), 1787–1818.

[12] Donald, S. G. and Hsu, Y.-C. (2014). "Estimation and inference for distribution functions and quantile functions in treatment effect models," *Journal of Econometrics*, *178*, 383–397.

[13] Huang, J. Z. (2003). "Local Asymptotics for Polynomial Spline Regression," *Annals of Statistics*, *31*(5), 1600–1635.

[14] Kallenberg, O. (2021). *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling: Springer Cham, 3rd edition, XII, 946.

[15] Monrad, D. and Philipp, W. (1991). "Nearby variables with nearby conditional laws and a strong approximation theorem for Hilbert space valued martingales," *Probability Theory and Related Fields*, *88*(3), 381–404.

[16] Schumaker, L. L. (2007). *Spline Functions: Basic Theory*, Cambridge Mathematical Library, Cambridge: Cambridge University Press, 3rd edition, URL: https://www.cambridge.org/core/books/spline-functions-basic-theory/843475201223F90091FFBDDCBF210BFB, DOI: http://dx.doi.org/10.1017/CBO9780511618994.

[17] Shang, Z. and Cheng, G. (2013). "Local and global asymptotic inference in smoothing spline models," *Annals of Statistics*, *41*(5), 2608–2638.

[18] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, Springer Series in Statistics: Springer New York.

[19] Zhou, S., Shen, X., and Wolfe, D. (1998). "Local Asymptotics for Regression Splines and Confidence Regions," *Annals of Statistics*, *26*(5), 1760–1782.