On Frequentist Coverage Errors of Bayesian Credible Sets in High Dimensions

By

Keisuke Yano and Kengo Kato

April 2018

CENTER FOR RESEARCH AND EDUCATION FOR POLICY EVALUATION
DISCUSSION PAPER NO. 27

CENTER FOR RESEARCH AND EDUCATION FOR POLICY EVALUATION (CREPE)
THE UNIVERSITY OF TOKYO
http://www.crepe.e.u-tokyo.ac.jp/

ON FREQUENTIST COVERAGE ERRORS OF BAYESIAN CREDIBLE SETS IN HIGH DIMENSIONS

KEISUKE YANO AND KENGO KATO

Abstract. In this paper, we study frequentist coverage errors of Bayesian credible sets for an approximately linear regression model with (moderately) high dimensional regressors, where the dimension of the regressors may increase with but is smaller than the sample size. Specifically, we consider Bayesian inference on the slope vector by fitting a Gaussian distribution on the error term and putting priors on the slope vector together with the error variance. The Gaussian specification on the error distribution may be incorrect, so that we work with quasi-likelihoods. Under this setup, we derive finite sample bounds on frequentist coverage errors of Bayesian credible rectangles. Derivation of those bounds builds on a novel Berry–Esseen type bound on quasi-posterior distributions and recent results on high-dimensional CLT on hyper-rectangles. We use this general result to quantify coverage errors of Castillo–Nickl and L^∞-credible bands for Gaussian white noise models, linear inverse problems, and (possibly non-Gaussian) nonparametric regression models. In particular, we show that Bayesian credible bands for those nonparametric models have coverage errors decaying polynomially fast in the sample size, implying advantages of Bayesian credible bands over confidence bands based on extreme value theory.

1. Introduction

Bayesian inference for high or nonparametric statistical models is an active research area in the recent statistics literature. Posterior distributions provide not only point estimates but also credible sets. In a classical regular statistical model with a fixed finite dimensional parameter space, it is well known that the Bernstein–von Mises (BvM) theorem holds under mild conditions and the posterior distribution can be approximated (under the total variation distance) by a normal distribution centered at an efficient estimator (e.g. MLE) and with covariance matrix identical to the inverse of the Fisher information matrix as the sample size increases. The BvM theorem implies that a Bayesian credible set is typically a valid confidence set in the frequentist sense, namely, the coverage probability of a $(1-\alpha)$-Bayesian credible set evaluated under the true parameter value is approaching $(1-\alpha)$ as the sample size increases. There exists a literature on the frequentist behavior of Bayesian credible sets in a nonparametric statistical model. Freedman [22] gave the negative result for the BvM theorem in infinite Gaussian sequence models with Gaussian priors; Johnstone [31], Leahu [33], and Bontemps [7] discovered the conditions under which the BvM theorem holds when using Gaussian priors. Recently, Castillo and Nickl [9] and [10] established the BvM theorem using weaker topologies than ℓ_2.

Date: March 9, 2018.

Key words and phrases. Castillo-Nickl band, credible rectangle, sieve prior.

K. Yano is supported by the Grant-in-Aid for Research Activity start-up (17H06570) from the JSPS. K. Kato is supported by the Grant-in-Aid for Scientific Research (C) (15K03392) from the JSPS.

This paper aims at studying frequentist coverage errors of Bayesian credible rectangles in an approximately linear regression model with an increasing number of regressors. We provide finite sample bounds on frequentist coverage errors of (quasi-)Bayesian credible rectangles based on sieve priors, where the model includes a unknown bias term, the error variance is also unknown, and the distribution of the error term may not be Gaussian. Sieve priors are prior distributions on the slope vectors where the dimension of the regressors may increase. We allow sieve priors to be non-Gaussian or not to be an independent product. We go through "quasi-Bayesian" approach because we fit a Gaussian distribution on the error term but we do not specify Gaussian distributions as the distribution of the error term. The resulting posterior distribution is called "quasi-posterior."

An important application of our results is finite sample quantification of Bayesian nonparametric credible bands based on sieve priors. We presents finite sample bounds of Castillo–Nickl and L^∞-credible bands [1] in Gaussian white noise models, in linear inverse problems, and in (possibly non-Gaussian) nonparametric regression models. There is a literature on nonparametric credible bands. Studies of frequentist approaches go back to Bickel, Rosenblatt, and Smirnov [6, 42]. More recent studies are available in [13, 18, 26]. Studies of Bayesian approaches are relatively few. In particular, there is a limited literature on quantification of frequentist coverage errors of nonparametric credible bands. Castillo and Nickl [10] showed that for Gaussian white noise models Castillo–Nickl credible bands based on product priors have asymptotically optimal frequentist coverage. See also [39] for adaptive Castillo–Nickl credible bands based on a spike and slab prior. Yoo and Ghosal [49] showed that Castillo–Nickl credible bands based on Gaussian series priors have the asymptotically optimal frequentist coverage in (possibly sub-Gaussian) nonparametric regression models. All there work built upon asymptotic results. Recently, Yang et al. [48] obtained a non-asymptotic result using Gaussian process priors. Yet, quantification of frequentist coverage errors of nonparametric credible bands based on general priors had been limited.

Our results have an implication supporting Bayesian approach to constructing nonparametric confidence bands. In the literature on nonparametric confidence bands, it is known that confidence bands based on extreme value theory are not satisfactory due to the slow convergence of Gaussian maxima. Hall [28] showed that bootstrap confidence bands have coverage errors decalying polynomially fast in the sample size. A more general discussion on the fast convergence of coverage errors of bootstrap confidence bands is available in [13]. An alternative to bootstrap approach is Bayesian approach. Castillo and Nickl [10] vaguely discussed the advantage of Bayesian credible bands over confidence bands based on extreme value theory. Our result shows that Bayesian credible bands have also coverage errors comparable to bootstrap confidence bands, yielding the advantage over confidence bands based on extreme value theory. See Remark 3.2.

The main ingredients of the derivation are (i) a novel Berry–Esseen type bound for the BvM theorem in case with sieve priors, that is, a finite sample bound on the total variation distance between the quasi-posterior distribution based on sieve priors and the Gaussian distribution and (ii) recent results on high dimensional CLT on hyper-rectangles. Our Berry–Esseen type bound

[1] In the present paper, a nonparametric credible band whose L^∞-diameter is bounded both below and above is called an L^∞-credible band. A nonparametric credible band whose L^∞-diameter is bounded only above is called a Castillo–Nickl-credible band.

opens the door to non-asymptotic treatment of posteriors going beyond conjugate analyses. The high dimensional CLT on hyper-rectangles is developed by Chernozhukov et al. [12, 16] and is used to approximate in hyper-rectangle regions the sampling distribution by the Gaussian distribution. This enables us to get rid of any assumption on the error distribution except assumptions on the moment of the error distribution.

1.1. **Literature review and contributions.** Closely related are [48, 49]. For nonparametric regression model, Yang et al. [48] obtained finite sample bounds on frequentist coverage errors of Bayesian credible bands based on Gaussian process priors. They worked with (a) Gaussian process priors, (b) the assumption that the error terms follows a sub-Gaussian distribution, and (c) the assumption that the error variance is known. The present paper differs from [48] in that

- we work with possibly non-Gaussian priors;
- we allow a more flexible assumption on the distribution of the error terms;
- we allow the error variance to be unknown.

To deal with non-Gaussian priors, we develop novel Berry–Esseen type bounds on quasi-posterior distributions. To weaken the moment assumption on error terms, we introduce a recently-developed probabilistic tool "high-dimensional CLT on hyper rectangles." In the case with the unknown error variance, the quasi-posterior contraction for a prior on a variance affects on the coverage error and thus a more careful treatment is conducted. Yoo and Ghosal [49] worked with (a) Gaussian series priors, (b) the assumption that the error terms follows a sub-Gaussian distribution, and (d) the requirement that the nominal coverage level $(1-\alpha)$ tends to 1 as the sample size grows. The requirement (d) indicates that their result is available only in the large sample asymptotics.

The present paper works with the BvM theorem in nonparametric statistics. There exists a vast literature on the BvM theorem. The BvM theorem in Gaussian white noise model is studied by [9, 10, 22, 31, 33, 39]. The BvM theorem in linear regression with high dimensional regressors is studies by [7, 24]. The BvM theorem in nonparametric regression with Gaussian process prior is studies by [48, 49]. For the BvM theorem in the other nonparametric models, see also [8, 11, 23, 25, 40]. The present paper goes through quasi-Bayesian approach because we do not use the Gaussian assumption on error terms. The BvM theorem for quasi-posterior distribution was developed by [3, 17, 21, 32, 34].

Importantly, our Berry–Esseen type bound produces a substantially better result in a critical dimension under which the BvM theorem holds. Ghosal [24], Bontemps [7], and Spokoiny [43] investigated critical dimensions when using sieve priors. The result in Bontemps [7] did not cover the case with an unknown error variance; the results in [24, 43] covered the case with the unknown error variance. Our result is consistent to the result in [7] in the case with the known error variance. Our result substantially improves the results in [24, 43] in the case with the unknown error variance. The results in [24, 43] typically indicated that the critical dimension is $p^3 = o(n)$ in the case with the unknown error variance, while our result indicates that the critical dimension is $p^2(\log n)^3 = o(n)$, where p is the number of the regressor and n is the sample size. For the detailed comparison, see Remark 2.2. Our obtained critical dimension is sufficient for the use in nonparametric regression model with the unknown error variance; see Section 3.3.

1.2. **Organization and notation.** Let $\|\cdot\|$ denote the Euclidean norm. Let $\|\cdot\|_\infty$ denote the max or supremum norm for vectors and functions. For a set S of vectors or functions , let $\|S\|_\infty$ denote the supremum of the max or supremum norms of differences between any two elements in S. Let $\mathcal{N}(\mu, \Sigma)$ denote the Gaussian distribution with mean vector μ and covariance matrix Σ. For any $x \in \mathbb{R}$, let $x_+ = \max\{x, 0\}$. For two sequences $\{a_n\}$ and $\{b_n\}$ depending on n, the notation $a_n \lesssim b_n$ signifies that $a_n \leqslant cb_n$ for some universal constant $c > 0$. The notation $a_n \sim b_n$ signifies that $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For any positive semi-definite matrices A and B, the notation $A \preceq B$ signifies that $B - A$ is positive semi-definite. Throughout the paper, constants $c_1, c_2, \ldots, c$, and $\widetilde{c}_1, \widetilde{c}_2$ do not depend on sample size n and dimension p. The values of $c_1, c_2, \ldots, c$, and $\widetilde{c}_1, \widetilde{c}_2$ are different in each theorem, proposition, and proof.

2. Bayesian credible rectangles in high dimensions

Consider an approximately linear regression model

$$Y = X\beta_0 + r + \varepsilon, \tag{1}$$

where $Y = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$ is a vector of outcome variables, X is an $n \times p$ design matrix, $\beta_0 \in \mathbb{R}^p$ is an unknown coefficient vector, $r = (r_1, \ldots, r_n)^\top \in \mathbb{R}^n$ is a deterministic (i.e., non-random) bias term, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top \in \mathbb{R}^n$ is a vector of i.i.d. error terms with mean zero and variance $0 < \sigma_0^2 < \infty$. We are primarily interested in the situation where the number of regressors p increases with the sample size n, i.e., $p = p_n \to \infty$ as $n \to \infty$, but we often suppress the dependence on n for the sake of notational simplicity. In addition, we allow the error variance σ_0^2 to depend on n, i.e., $\sigma_0^2 = \sigma_{0,n}^2$, which allows us to include Gaussian white noise models in the subsequent analysis as a special case. In the general setting, the error variance σ_0^2 is also unknown. In the present paper, we work with the dense model with moderately high-dimensional regressors where β_0 need not be sparse and $p = p_n$ may increase with the sample size n but $p < n$. To be precise, we will maintain the assumption that the design matrix X is of full column rank, i.e., $\operatorname{rank} X = p$. The approximately linear model (1) is flexible enough to cover various nonparametric models such as Gaussian white noise models, linear inverse problems, and nonparametric regression models, via series expansions of functions of interest in those nonparametric models; see Section 3.

We consider Bayesian inference on the slope vector β_0. To this end, we fit a Gaussian distribution on the error ε, but we allow the Gaussian specification on the error distribution to be incorrect. Namely, we work with the *quasi*-likelihood of the form

$$(\beta, \sigma^2) \mapsto (2\pi\sigma^2)^{-n/2} \mathrm{e}^{-\|Y - X\beta\|^2/(2\sigma^2)}.$$

We assume independent priors on β and σ^2, i.e.,

$$\beta \sim \Pi_\beta, \ \sigma^2 \sim \Pi_{\sigma^2}, \ \beta \perp\!\!\!\perp \sigma^2, \tag{2}$$

where we assume that Π_β is absolutely continuous with density π, i.e., $\Pi_\beta(d\beta) = \pi(\beta)d\beta$, and Π_{σ^2} is supported in $(0, \infty)$. Then the resulting quasi-posterior distribution for (β, σ^2) is

$$\Pi(d(\beta, \sigma^2) \mid Y) \propto (2\pi\sigma^2)^{-n/2} \mathrm{e}^{-\|Y - X\beta\|^2/(2\sigma^2)} \pi(\beta) d\beta \Pi_{\sigma^2}(d\sigma^2),$$

and the marginal quasi-posterior distribution for β is $\Pi_\beta(d\beta \mid Y) = \pi(\beta \mid Y)d\beta$, where

$$\pi(\beta \mid Y) = \pi(\beta) \int \frac{\mathrm{e}^{-\|Y-X\beta\|^2/(2\sigma^2)}}{\int \mathrm{e}^{-\|Y-X\widetilde{\beta}\|^2/(2\sigma^2)}\pi(\widetilde{\beta})d\widetilde{\beta}} \Pi_{\sigma^2}(d\sigma^2 \mid Y),$$

and $\Pi_{\sigma^2}(d\sigma^2 \mid Y)$ denotes the quasi-posterior distribution for σ^2. We will assume that Π_{σ^2} may be data-dependent, e.g., $\Pi_{\sigma^2} = \delta_{\widehat{\sigma}^2}$ for some estimator $\widehat{\sigma}^2$ of σ^2 (in that case, $\Pi_{\sigma^2}(\cdot \mid Y) = \delta_{\widehat{\sigma}^2}$), but Π_β is data-independent.

We will derive finite sample bounds on frequentist coverage errors of Bayesian credible rectangles for the approximately linear model (1) under a prior of the form (2). For given $c = (c_1, \dots, c_p)^\top \in \mathbb{R}^p$, $R > 0$, and positive sequence $\{w_j\}_{j=1}^p$, let $I(c, R)$ denote the hyper-rectangle of the form

$$I(c, R) := \left\{ \beta = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p : \frac{|\beta_j - c_j|}{w_j} \leqslant R,\ 1 \leqslant \forall j \leqslant p \right\}.$$

Let $\widehat{\beta}$ denote the OLS estimator for β_0, i.e., $\widehat{\beta} = \widehat{\beta}(Y) = (X^\top X)^{-1}X^\top Y$. For given $\alpha \in (0,1)$, we consider a $(1-\alpha)$-credible rectangle of the form $I(\widehat{\beta}, \widehat{R}_\alpha)$, where the radius $\widehat{R}_\alpha$ is chosen in such a way that the posterior probability of the set $I(\widehat{\beta}, \widehat{R}_\alpha)$ is $1-\alpha$, i.e., $\Pi_\beta\{I(\widehat{\beta}, \widehat{R}_\alpha) \mid Y\} = 1 - \alpha$.

We make the following conditions on Π_β and Π_{σ^2}. For $R > 0$, let

$$B(R) := \{\beta \in \mathbb{R}^p : \|X(\beta - \beta_0)\| \leqslant R\sigma_0\} \quad \text{and} \quad \phi_{\Pi_\beta}(R) := 1 - \inf_{\beta, \widetilde{\beta} \in B(R)} \{\pi(\widetilde{\beta})/\pi(\beta)\}. \tag{3}$$

Condition 2.1. *There exists a positive constant C_1 such that*

$$\pi(\beta_0) \geqslant \sigma_0^{-p}\sqrt{\det(X^\top X)}e^{-C_1 p \log n}.$$

Condition 2.2. *There exist non-negative constants $\delta_1, \delta_2, \delta_3 \in [0,1)$ such that with probability at least $1-\delta_3$, $\Pi_{\sigma^2}\left(\left\{\sigma^2 : \left|\sigma^2/\sigma_0^2 - 1\right| > \delta_1\right\} \mid Y\right) \leqslant \delta_2$.*

Condition 2.3. *The inequality $\phi_{\Pi_\beta}(1/\sqrt{n}) \leqslant 1/2$ holds.*

Condition 2.1 imposes Π_β to put a sufficient mass around β_0. Condition 2.2 is a marginal posterior contraction of Π_{σ^2}. Condition 2.3 is a preliminary flatness condition on Π_β. The detailed discussion on these conditions are provided after the main theorem.

We also make the assumptions on the model.

Assumption 2.1. *There exists a positive constant C_2 such that $\|X(X^\top X)^{-1}X^\top r\| \leqslant C_2\sigma_0\sqrt{p \log n}$.*

Assumption 2.2. *There exists a positive constant C_3 for which either one of the following conditions holds.*

(a) $\mathbb{E}[|\varepsilon_1/(\sigma_0 C_3)|^q] \leqslant 2$ *for some integer $q \geqslant 2$;*
(b) $\mathbb{E}[\exp\{\varepsilon_1^2/(\sigma_0 C_3)^2\}] \leqslant 2$.

Assumption 2.1 controls the norm of the bias term. Assumption 2.2 is a moment assumption on the error terms. These assumptions are substantially weak and satisfied in all applications we present.

The following theorem, which is the main result of this section, provides bounds on frequentist coverage errors of the Bayesian credible rectangle $I(\widehat{\beta}, \widehat{R}_\alpha)$ together with bounds on the max-diameter of $I(\widehat{\beta}, \widehat{R}_\alpha)$. In what follows, let $\overline{\lambda}$ and $\underline{\lambda}$ denote the maximum and minimum eigenvalues of $(X^\top X)^{-1}$, respectively, and let $\overline{w} := \max\{w_1, \ldots, w_p\}$ and $\underline{w} := \min\{w_1, \ldots, w_p\}$.

Theorem 2.1 (Coverage errors of credible rectangles)**.** *Suppose that Conditions 2.1–2.3, Assumption 2.1, and either of Assumption 2.2 (a) or (b) hold. Then, there exist positive constants $c_1, \ldots, c_4$ depending only on C_1, C_2, C_3 and q appearing in Condition 2.1 and Assumptions 2.1 and 2.2 (a) such that the following hold. For every $n \geqslant 2$, we have that*

$$\begin{aligned}&\left|\mathbb{P}\left\{\beta_0 \in I(\widehat{\beta}, \widehat{R}_\alpha)\right\} - (1-\alpha)\right| \\ &\leqslant \phi_{\Pi_\beta}\left(c_1\sqrt{p\log n}\right) + c_1\left(\delta_1 p\log n + \delta_2 + \delta_3 + \frac{\tau}{\sigma_0\underline{\lambda}^{1/2}}\sqrt{\log p} + \omega_n\right)\end{aligned} \tag{4}$$

where $\tau := \|(X^\top X)^{-1}X^\top r\|_\infty$ and

$$\omega_n = \begin{cases} p^{1-q/2}(\log n)^{-q/2} + \left(\frac{\overline{\lambda}}{\underline{\lambda}}\frac{p\log^7(pn)}{n}\right)^{1/6} + \left(\frac{\overline{\lambda}}{\underline{\lambda}}\frac{p\log^3(pn)}{n^{1-2/q}}\right)^{1/3} & \text{under Assumption 2.2 (a)}, \\ \mathrm{e}^{-c_2 p\log n} + \left(\frac{\overline{\lambda}}{\underline{\lambda}}\frac{p\log^7(pn)}{n}\right)^{1/6} & \text{under Assumption 2.2 (b)}, \\ \mathrm{e}^{-c_2 p\log n} & \text{if } \varepsilon_i\text{'s are Gaussian}; \end{cases}$$

In addition, provided that the right hand side on (4) is smaller than $\alpha/2$, for sufficiently large p depending only on α, the max-diameter of $I(\widehat{\beta}, \widehat{R}_\alpha)$ is bounded as

$$c_3\sigma_0\underline{\lambda}^{1/2}\overline{w}\sqrt{\log p} \leqslant |I(\widehat{\beta}, \widehat{R}_\alpha)|_\infty \leqslant c_4\sigma_0\overline{\lambda}^{1/2}\underline{w}\sqrt{\log p}$$

with probability at least

$$\begin{cases} 1 - c_1 p^{1-q/2}(\log n)^{-q/2} - \delta_3 & \text{under Assumption 2.2 (a)}, \\ 1 - c_1 e^{-c_2 p\log n} - \delta_3 & \text{under Assumption 2.2 (b)}. \end{cases}$$

Theorem 2.1 indicates that the frequentist coverage error of the Bayesian credible rectangle is controlled by the flatness function ϕ_{Π_β} up to terms depending on the prior Π_β on β. The discussions below provide a typical bound on ϕ_{Π_β}.

2.1. **Discussions on conditions.** First, we verify that a locally log-Lipschitz prior meets Conditions 2.1 and 2.3, providing an upper bound of ϕ_{Π_β}.

Definition 2.1. A *locally log-Lipschitz prior* is defined as a prior distribution on β that there exists $L = L_n > 0$ for which the inequality $|\log\pi(\beta) - \log\pi(\beta_0)| \leqslant L\|\beta - \beta_0\|$ holds for β such that $\|\beta - \beta_0\| \leqslant \sigma_0\overline{\lambda}^{1/2}\sqrt{p\log n}$.

The following proposition shows that a locally log-Lipschitz prior meets Condition 2.3. It also provides an upper bound of ϕ_{Π_β} with a locally log-Lipschitz prior.

Proposition 2.1. *For a locally log-Lipschitz prior Π_β with a log-Lipschitz constant L, the inequality $\phi_{\Pi_\beta}(c\sqrt{p\log n}) \leqslant cL\sigma_0\overline{\lambda}^{1/2}\sqrt{p\log n}$ holds for any $c > 0$. Further, the prior Π_β satisfies Condition 2.3 provided that $\sigma_0 L\overline{\lambda}^{1/2}/\sqrt{n} \leqslant 1/2$.*

For the verification of Condition 2.1, we focus on the following two subclasses of locally log-Lipschitz priors. Let $B := \|\beta_0\|$.

(Isotropic prior) An *isotropic prior* is a prior of the form $\pi(\beta) = \rho(\|\beta\|)/\int \rho(\|\beta\|)d\beta$ with a probability density ρ on $\mathbb{R}_+$ such that ρ is strictly positive on $[0, B+\sigma_0\overline{\lambda}^{-1/2}\sqrt{p\log n}]$ and continuously differentiable on $[0, B+\sigma_0\overline{\lambda}^{-1/2}\sqrt{p\log n}]$, and such that there exists a positive constant m for which $\int_0^\infty x^k\rho(x)dx$ is bounded above by $\mathrm{e}^{mk\log k}$ for any $k \in \mathbb{N}$;

(Product prior) A *product prior* of log-Lipschitz priors is of the form $\pi(\beta) = \prod_{i=1}^p \pi_i(\beta_i)$ of which $\log \pi_i$ for each i is strictly positive on $[0, B+\sigma_0\overline{\lambda}^{-1/2}\sqrt{p\log n}]$ and $\widetilde{L}$-Lipschitz for some $\widetilde{L} > 0$.

For expositional simplicity, we use the following assumption to verify that isotropic or product priors meet Condition 2.1.

Assumption 2.3. *There exists a universal positive constant c for which we have, for every $n \geqslant 2$, $\log\{\sqrt{\det(X^\top X)}/\sigma_0^p\} \leqslant c\log n$.*

This assumption ensures that $X^\top X/\sigma_0^2$ is not much expanded in n. The assumption is satisfied in case that the column vectors of X/σ_0 are nearly orthonormal in $\mathbb{R}^p$. The assumption is also satisfied in all applications we present.

The following proposition ensures that isotropic or product priors are locally log-Lipschitz priors satisfying Condition 2.1.

Proposition 2.2. *Under Assumption 2.3, an isotropic prior and a product prior of log-Lipschitz priors satisfy Condition 2.1. An isotropic prior is a locally log-Lipschitz prior with locally log-Lipschitz constant L such that*

$$L \leqslant c_1 B \max_{x:0\leqslant x\leqslant B+\sigma_0\overline{\lambda}^{-1/2}\sqrt{p\log n}} |d\log\rho/dx(x)|$$

for some positive constant c_1 depending only on m and c appearing in the definition of ρ and Assumption 2.3. In particular, for a standard Gaussian distribution, $L \leqslant c_1B^2$. A product prior of log-Lipschitz priors with a log-Lipschitz constant $\widetilde{L}$ is a locally log-Lipschitz with $L = \widetilde{L}p^{1/2}$.

Next, we provide discussions on Condition 2.2. We consider following two cases:

(Plug-in) $\Pi_{\sigma^2} = \Pi_{\widehat{\sigma}^2_{\mathrm{u}}}$ with $\widehat{\sigma}^2_{\mathrm{u}}(Y) := \|Y - X(X^\top X)^{-1}X^\top Y\|^2/(n-p)$;

(Full Bayes) Π_β is a standard Gaussian distribution and Π_{σ^2} is the inverse Gamma distribution $\mathrm{IG}(\mu_1,\mu_2)$ with shape parameter $\mu_1 > 1/2$ and scale parameter $\mu_2 > 1/2$.

The following two propositions provide possible choices of δ_1, δ_2, and δ_3. For simplicity, we assume that σ_0 is a constant independent of n. Let $\widetilde{\delta}_1 := \delta_1 - 2\|r\|^2/\{\sigma_0^2(n-p)\} - 1/(n-p) > 0$.

Proposition 2.3. *Suppose that $n \geqslant cp$ for some $c > 1$. Then, there exist positive constants c_1 and c_2 depending only on c, C_3 and q appearing in Assumption 2.2 such that*

$$\mathbb{P}\left(|\widehat{\sigma}^2_{\mathrm{u}}/\sigma_0^2 - 1| \geqslant \delta_1\right) \leqslant \begin{cases} c_1\max\{n^{-4/q}\delta_1^{-q/2}, n^{1-q/2}\widetilde{\delta}_1^{-q}\} & \text{under Assumption 2.2 (a)},\\ c_1\exp(-c_2 n\max\{\delta_1^2,\widetilde{\delta}_1^2\}) & \text{under Assumption 2.2 (b)}.\end{cases}$$

Proposition 2.4. *Suppose that $n \geqslant cp$ for some $c > 1$. Then, there exist positive constants c_1 and c_2 depending only on c, μ_1, μ_2, C_3 and q appearing in Assumption 2.2 such that the inequality*

$$\Pi_{\sigma^2}(\sigma^2 : |\sigma^2/\sigma_0^2 - 1| > \delta_1 \mid Y) \leqslant c_1(n\widetilde{\delta}_1)^{-1}$$

holds with probability at least

$$\begin{cases} 1 - c_1 \max\{n^{-4/q}\delta_1^{-q/2}, n^{1-q/2}\widetilde{\delta}_1^{-q}\} & \text{under Assumption 2.2 (a)}, \\ 1 - c_1 \exp\{-c_2 n \max\{\delta_1^2, \widetilde{\delta}_1^2\}\} & \text{under Assumption 2.2 (b)}. \end{cases}$$

For a better understanding, Table 1 summarizes the results using the asymptotics in n in both cases when $n \geqslant cp$ for some $c > 0$ and $\|r\|^2/n = o(n^{-1/2})$.

TABLE 1. Possible orders of $\delta_1, \delta_2, \delta_3$ with respect to n: κ is arbitrary

Assumption 2.2 and prior	δ_1	δ_2	δ_3
(a) and plug-in	$n^{-1/2+\kappa/q}$	0	$\max\{n^{-\kappa/2}, n^{1-\kappa}\}$
(a) and full Bayes	$n^{-1/2+\kappa/q}$	$n^{-1/2-\kappa/q}$	$\max\{n^{-\kappa/2}, n^{1-\kappa}\}$
(b) and plug-in	$n^{-1/2}\sqrt{\log n}$	0	n^{-1}
(b) and full Bayes	$n^{-1/2}\sqrt{\log n}$	$n^{-1/2}(\log n)^{-1/2}$	n^{-1}

Remark 2.1 (Compatibility with the result in [49])**.** In the case with the Gaussian prior for β and a sub-Gaussian distributon for ε, the possible covergence rate of δ_1 was investigated by Yoo and Ghosal [49]; See Proposition 4.1 in [49]. For the choice of δ_1, our results in Propositions 2.3 and 2.4 are compatible with their result up to a logarithmic factor.

2.2. **Berry–Esseen type bounds on posterior distributions.** Before providing several applications of the main theorem, we present an important ingredient for the proof of Theorem 2.1, namely, the Berry–Esseen type bound on posterior distributions. This bound substantially improves on the previous work related to the BvM theorem, as discussed below. For $R > 0$, let

$$H(R) := \left\{ Y \in \mathbb{R}^n : \|X(\widehat{\beta}(Y) - \beta_0)\| \leqslant R\frac{\sqrt{p\log n}\sigma_0}{4} \right\} \cap \{Y \in \mathbb{R}^n : \Pi_{\sigma^2}(|\sigma^2/\sigma_0^2 - 1| \geqslant \delta_1 \mid Y) \leqslant \delta_2\}.$$

For two probability measures P and Q, $\|P - Q\|_{\mathrm{TV}}$ denotes the total variation between P and Q.

Proposition 2.5 (Berry–Esseen type bounds on posterior distributions)**.** *Under Conditions 2.1–2.3, there exist positive constants c_1 and c_2 depending only on C_1, C_2, C_3 such that for every $n \geqslant 2$ and for $Y \in H(c_1)$, the inequality*

$$\left\| \Pi_\beta(\cdot \mid Y) - \mathcal{N}(\widehat{\beta}, \sigma_0^2(X^\top X)^{-1}) \right\|_{\mathrm{TV}} \leqslant \phi_{\Pi_\beta}\left(c_1\sqrt{p\log n}\right) + c_1(\delta_1 p \log n + \delta_2 + \mathrm{e}^{-c_2 p \log n})$$

holds.

Proposition 2.6. *Assume that Assumption 2.1 holds. Then, there exist positive constants c_1 and c_2 depending only on C_2, C_3, and q appearing in Assumptions 2.1-2.2 such that*

$$\mathbb{P}(Y \notin H(c_1)) \leqslant \begin{cases} c_1 p^{1-q/2}(\log n)^{-q/2} + \delta_3 & \text{under Assumption 2.2 (a)}, \\ c_1 \exp(-c_2 p \log n) + \delta_3 & \text{under Assumption 2.2 (b)}. \end{cases}$$

Remark 2.2 (Critical dimension of the Bernstein–von Mises theorem)**.** The previous propositions clarify the critical dimension under which the BvM theorem holds. We compare our result with the results on the critical dimension by [7, 25, 43]. In the case that $\|\beta_0\|$ is independent of n and $\sigma_0\overline{\lambda}^{1/2} \sim n^{-1/2}$, the comparison is demonstrated using a locally log-Lipschitz prior with locally log-Lipschitz constant L independent of n. The followings are the summary of the previous works:

- Ghosal [25] showed that when the distribution of i.i.d. error terms ε_is has a smooth density function with known variance, the BvM theorem holds if $p^4 \log p = o(n)$ and some additional assumptions hold;
- Bontemps [7] showed that when the distribution of i.i.d. error terms is Gaussian with known variance, the BvM theorem holds if $p \log n = o(n)$ [2];
- Spokoiny [43] showed that when the high-dimensional local asymptotic normality holds, the BvM theorem holds if $p^3 = o(n)$. See also [37].

Our result (Propositions 2.1, 2.3, 2.5, and 2.6) improves the previous work from the following viewpoints:

- In the case with known variance, our result shows that when the distribution of i.i.d. error terms has the third order moment, the BvM theorem (for quasi-posterior distributions) holds if $p \log n = o(n)$. Comparing our result with [25] indicates that borrowing the Gaussian likelihood conducts a substantial improvement on the critical dimension. When the distribution of error terms is Gaussian, our result is consistent to [7];
- Our result covers the case with unknown variance. This makes the substantial difference between our result and [7]. Taking the care of the case with unknown variance is important in application including nonparametric regression models. In the case with unknown variance, our result shows that when the distribution of i.i.d. error terms is sub–Gaussian, the BvM theorem holds for β if $p^2(\log n)^3 = o(n)$. Comparing our result with [43], our result provides a substantial improvement on the critical dimension.

3. Applications

In this section, we consider applications of the general results developed in the previous sections to quantifying coverage errors of Bayesian credible sets in Gaussian white noise models, linear inverse problems, and (possibly non-Gaussian) nonparametric regression models.

3.1. **Gaussian white noise model.** We first consider a Gaussian white noise model and analyze coverage errors of Castillo-Nickl and L^∞-credible bands. Consider a Gaussian white noise model

$$dY(t) = f_0(t)dt + \frac{1}{\sqrt{n}} dW(t), \ t \in [0, 1],$$

where dW is a canonical white noise and f_0 is an unknown function. We assume that f_0 is in the Hölder–Zygmund space $B^s_{\infty,\infty}$ with smoothness level s for some $s > 0$. It will be convenient to

[2] Bontemps [7] uses the condition that there exists M_n such that $M_n = O(n)$ and $p \log p = o(M_n)$ and it seems to exclude the setting that $p \log n = o(n)$ if p is of polynomial order with respect to n because we cannot set $M_n = p \log n$. However, replacing the asymptotic evaluation at the final step in the proof of Proposition 9 in [7] by the finite sample evaluation and multiplying M_n by a sufficiently large positive constant, the setting that $p \log n = o(n)$ is allowed.

define the Hölder–Zygmund space $B^s_{\infty,\infty}$ using a wavelet basis. Let $S > s$ be an integer and fix sufficiently large $J_0 = J_0(S)$. Let $\{\phi_{J_0,k} : 0 \leqslant k \leqslant 2^{J_0} - 1\} \cup \{\psi_{l,k} : J_0 \leqslant l, 0 \leqslant k \leqslant 2^l - 1\}$ be an S-regular Cohen–Daubechies–Vial (CDV) wavelet basis of $L^2[0,1]$. Then the Hölder–Zygmund space $B^s_{\infty,\infty}$ is defined by

$$B^s_{\infty,\infty} = \left\{ f : \|f\|_{B^s_{\infty,\infty}} := \max_{0\leqslant k\leqslant 2^{J_0}-1} |\langle \phi_{J_0,k}, f\rangle| + \sup_{J_0\leqslant l<\infty, 0\leqslant k\leqslant 2^l-1} 2^{l(s+1/2)} |\langle \psi_{l,k}, f\rangle| < \infty \right\},$$

where $\langle \cdot, \cdot \rangle$ denotes the $L^2[0,1]$ inner product, i.e., $\langle f, g\rangle := \int_{[0,1]} f(t)g(t)dt$. For the notational convention, let $\psi_{J_0-1,k} := \phi_{J_0,k}$ for $0 \leqslant k \leqslant 2^{J_0} - 1$ and let $\mathcal{I}(J) := \{(l,k) : J_0 \leqslant l \leqslant J-1, 0 \leqslant k \leqslant 2^l - 1\} \cup \{(l,k) : l = J_0 - 1, 0 \leqslant k \leqslant 2^{J_0} - 1\}$ for $J > J_0$.

3.1.1. *Castillo–Nickl credible bands.* The Castillo–Nickl credible band is defined as

$$\mathcal{C}(\widehat{f}, R) := \left\{ f : \max_{(l,k)\in\mathcal{I}(J)} \frac{|\langle f - \widehat{f}, \psi_{l,k}\rangle|}{\sqrt{l}} \leqslant R \right\} \quad \text{for } R > 0,$$

where $J = J_n > J_0$ is taken in such a way that $2^{J_n} = (n/\log n)^{1/(2s+1)} u_n$ for a divergent sequence u_n and $\widehat{f} := \sum_{(l,k)\in\mathcal{I}(J)} \psi_{l,k} \int \psi_{l,k} dY$. For a given prior Π_f on f, we call $\mathcal{C}(\widehat{f}, \widehat{R}_\alpha)$ the $(1-\alpha)$-Castillo–Nickl credible band, where the radius $\widehat{R}_\alpha$ is chosen in such a way that $\Pi_f\{\mathcal{C}(\widehat{f}, \widehat{R}_\alpha) \mid Y\} = 1 - \alpha$. We consider a sieve prior Π_f on $L^\infty[0,1]$ induced from a prior Π_β on $\mathbb{R}^{2^J}$ via the map

$$(\beta_{J_0,0}, \beta_{J_0,1}, \ldots, \beta_{J-1,2^{J-1}-1}) \mapsto \sum_{(l,k)\in\mathcal{I}(J)} \psi_{l,k}(\cdot)\beta_{l,k}.$$

The following theorem establishes bounds on coverage errors of Castillo–Nickl credible bands. Let $\tau_\infty := \|f_0 - \sum_{(l,k)\in\mathcal{I}(J)} \psi_{l,k}\beta_{0,lk}\|_\infty$ with $\beta_{0,lk} = \langle f_0, \psi_{l,k}\rangle$ for $(l,k) \in \mathcal{I}(J)$.

Proposition 3.1. *Under Conditions 2.1 and 2.3 for Π_β that corresponds to Π_f and under the assumption that $\tau_\infty \leqslant C_2' \sqrt{2^J (\log n)/n}$ for some $C_2' > 0$, there exist positive constants c_1, c_2, c_3 depending only on C_1 appearing in Condition 2.1 and C_2' such that the following hold: For $n \geqslant 2$ satisfying $\|f_0\|_{B^s_{\infty,\infty}} \leqslant u_n$, we have*

$$|\mathbb{P}\{f_0 \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\} - (1-\alpha)| \leqslant \phi_{\Pi_\beta}(c_1 \sqrt{2^J \log n}) + c_1 \left(\frac{\sqrt{n}\tau_\infty}{2^{J/2}} J^{1/2} + \mathrm{e}^{-c_2 2^J \log n} \right).$$

In addition, provided that the right hand side above is smaller than $\alpha/2$, for sufficiently large n depending only on α and u_n, the L^∞-diameter of $\mathcal{C}(\widehat{f}, \widehat{R}_\alpha)$ is bounded above as

$$\|\mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\|_\infty := \sup_{f,g\in\mathcal{C}(\widehat{f},\widehat{R}_\alpha)} \|f - g\|_\infty \leqslant c_3 \sqrt{2^J (\log n)/n}$$

with probability at least $1 - c_1 \exp(-c_2 2^J \log n)$.

Remark 3.1 (Rate of convergence)**.** We discuss asymptotic forms of the result using a locally log-Lipschitz prior with locally log-Lipschitz constant $L = L_n$. Since $\sum_{(l,k)\in\mathcal{I}(J)} |\beta_{0,lk}|^2 \lesssim u_n^2$ and $\tau_\infty \lesssim \sum_{l\geqslant J} 2^{-ls} \sup_k 2^{l(s+1/2)} |\beta_{0,lk}| \lesssim 2^{-Js} u_n$, we have

$$|\mathbb{P}\{f_0 \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\} - (1-\alpha)| \lesssim L_n \left(\frac{n}{\log n}\right)^{-s/(2s+1)} u_n^{1/2} + \frac{\log n}{u_n^{s+1/2}}. \tag{5}$$

In particular, for the standard Gaussian prior, we have

$$|\mathbb{P}\{f_0 \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\} - (1-\alpha)| \lesssim \left(\frac{n}{\log n}\right)^{-s/(2s+1)} u_n^{3/2} + \frac{\log n}{u_n^{s+1/2}},$$

since $L_n \lesssim u_n$ from Proposition 2.2.

Remark 3.2 (Coverage errors for the surrogate function)**.** Consider coverage errors for the surrogate function $\mathbb{E}[\widehat{f}]$. In this case, we can set $\tau_\infty = 0$ and thus we have

$$|\mathbb{P}\{\mathbb{E}[\widehat{f}] \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\} - (1-\alpha)| \lesssim L_n \left(\frac{n}{\log n}\right)^{-s/(2s+1)} u_n^{1/2}.$$

From this, we see that the Bayesian credible band has the polynomial decay of the coverage error with respect to $\mathbb{E}[\widehat{f}]$. Hall [28] showed that the bootstrap confidence band has the polynomial decay of the coverage error with respect to $\mathbb{E}[\widehat{f}]$. Our result shows that the Bayesian credible band is comparable to the bootstrap confidence band.

In [10], Castillo and Nickl also consider multi-scale sets using an admissible sequence $w = (w_1, w_2, \ldots)$:

$$\left\{ f : \sup_{(l,k)\in\mathcal{I}_\infty} \frac{|\langle f - \widehat{f}_\infty, \psi_{l,k}\rangle|}{w_l} \leqslant R \right\} \quad \text{for } R > 0,$$

where $\mathcal{I}_\infty := \{(l,k) : J_0 \leqslant l < \infty, 0 \leqslant k \leqslant 2^l - 1\} \cup \{(l,k) : l = J_0 - 1, 0 \leqslant k \leqslant 2^{J_0} - 1\}$ and $\widehat{f}_\infty := \sum_{(l,k)\in\mathcal{I}_\infty} \psi_{l,k} \int \psi_{l,k} dY$. Here we call a sequence such that $w_l/\sqrt{l} \nearrow \infty$ an admissible sequence. In what follows, we will bound the coverage error and the L^∞-diameter of Bayesian credible sets of the form

$$\mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha) := \left\{ f : \sup_{(l,k)\in\mathcal{I}_\infty} \frac{|\langle f - \widehat{f}_\infty, \psi_{l,k}\rangle|}{w_l} \leqslant \widehat{R}_\alpha \right\},$$

where the radius $\widehat{R}_\alpha$ is taken in such a way that $\Pi_f\{\mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha) \mid Y\} = 1 - \alpha$.

The following proposition provides the coverage error of multi-scale credible bands using a sieve prior on $\mathbb{R}^{2^{J'}}$, where J' is taken in a way that $2^{J'} = (n/\log n)^{1/(2s+1)}$. Let $u_n' := w_{J'}/\sqrt{J'}$, $\overline{w} := \inf_{l \geqslant J'} w_l$, and $\underline{w} := \max_{l<J'} w_l$. For simplicity, we assume that $\max_{l<J'}\{\sqrt{l}/w_l\} \leqslant 1$.

Proposition 3.2. *Under Conditions 2.1 and 2.3 for Π_β that corresponds to Π_f, there exist positive constants c_1, c_2, c_3 depending only on C_1 such that the followings hold: For $n \geqslant 2$ satisfying $\|f_0\|_{B^s_{\infty,\infty}} \leqslant u_n'$ and for any $\delta > 0$, we have*

$$|\mathbb{P}(f_0 \in \mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha)) - (1-\alpha)| \leqslant \phi_{\Pi_\beta}(c_1\sqrt{2^{J'}\log n}) + c_1(\mathrm{e}^{-c_2 2^{J'}\log n} + \sqrt{n}\underline{w}\delta\sqrt{\log n} + \mathrm{e}^{-c_2 n \overline{w}^2\delta^2}).$$

Further, provided that the right hand side above is smaller than $\alpha/2$, for sufficiently large n depending only on α, L^∞-diameter of $\mathcal{C}_w(\widehat{f}_\infty, \widehat{R}_\alpha)$ is bounded as

$$\|\mathcal{C}_w(\widehat{f}, \widehat{R}_\alpha)\|_\infty \leqslant c_3\sqrt{2^{J'}(\log n)/n}\max\{\underline{w}, u_n'\}$$

with probability at least $1 - c_1\exp(-c_2 2^{J'}\log n) - c_1\exp -c_2 n\overline{w}^2\delta^2$.

Remark 3.3 (Choices of δ and w)**.** From proposition 3.2, if $n\overline{w}^2\delta^2\log n\to\infty$ and $n\underline{w}^2\delta^2\log n\to 0$, then the coverage error vanishes, which suggests that an admissible sequence should depend on n when using a sieve prior. Given a divergent sequence $\{u_l : J_0-1\leqslant l\}$ in priori, typical choices of δ and w are

$$\delta = 1/(\sqrt{n}u_{J'}^{s+1/2}) \text{ and } w_l = \begin{cases} \sqrt{l} & \text{for } l < J'; \\ u_l\sqrt{l} & \text{for } l \geqslant J'. \end{cases}$$

Using these choices, the same asymptotic result as that of Proposition 3.1 is obtained.

3.1.2. *L^∞-credible bands.* Focusing on the simple case in which the smoothness s is in $(0,1]$, L^∞-credible bands are constructed using Haar scaling functions: for $l\in\mathbb{N}$, let $\{\phi_{l,0},\dots,\phi_{l,2^l-1}\}$ be $\phi_{l,k}(\cdot) := 2^{l/2}1_{(k/2^l,(k+1)/2^l]}(\cdot)$.

Let $\mathcal{C}(\widehat{f},R)$ be given as

$$\mathcal{C}(\widehat{f},R) := \left\{ f : \max_{0\leqslant k\leqslant 2^J-1} \frac{|\langle f-\widehat{f},\phi_{J,k}\rangle|}{\sqrt{J}} \leqslant R \right\},\ R>0,$$

where $J=(n/\log n)^{1/(2s+1)}u_n$ with a divergent sequence u_n and $\widehat{f} := \sum_{k=0}^{2^J-1}\phi_{J,k}\int\phi_{J,k}dY$. For a prior Π_f of f and for $\alpha\in(0,1)$, $\widehat{R}_\alpha$ is chosen in the way that $\Pi_f(\mathcal{C}(\widehat{f},\widehat{R}_\alpha)\mid Y)=1-\alpha$. We consider a sieve prior Π_f on $L^\infty[0,1]$ induced from a prior Π_β on $\mathbb{R}^{2^J}$ via the same map as in the previous subsection.

The following theorem provides the coverage error of L^∞-credible bands. In the following, $\tau_\infty := \|f_0-\sum_{k=0}^{2^J-1}\beta_{0,Jk}\phi_{J,k}\|_\infty$, where $\beta_0=(\beta_{0,J0},\dots,\beta_{0,J(2^J-1)})$ with $\beta_{0,Jk} := \langle\phi_{J,k},f_0\rangle$ for $k=0,\dots,2^J-1$.

Proposition 3.3. *Under Conditions 2.1 and 2.3 for Π_β that corresponds to Π_f and under the assumption that $\tau_\infty\leqslant C_2'\sqrt{2^J(\log n)/n}$ for some $C_2'>0$, there exist positive constants $c_1,\dots,c_4$ depending only on C_1 and C_2' such that the followings hold: For $n\geqslant 2$ satisfying $\|f_0\|_{B^s_{\infty,\infty}}\leqslant u_n$, we have*

$$|\mathbb{P}(f_0\in\mathcal{C}(\widehat{f},\widehat{R}_\alpha))-(1-\alpha)| \leqslant \phi_{\Pi_\beta}(c_1\sqrt{2^J\log n}) + c_1\left(\frac{\sqrt{n}\tau_\infty}{2^{J/2}}J^{1/2} + \mathrm{e}^{-c_2 2^J\log n}\right).$$

Further, provided that the right hand side above is smaller than $\alpha/2$, for sufficiently large n depending only on α, the L^∞-diameter of $\mathcal{C}(\widehat{f},\widehat{R}_\alpha)$ is bounded as

$$c_3\sqrt{2^J(\log n)/n} \leqslant \|\mathcal{C}(\widehat{f},\widehat{R}_\alpha)\|_\infty \leqslant c_4\sqrt{2^J(\log n)/n}$$

with probability at least $1-c_1\exp(-c_2 2^J\log n)$.

Remark 3.4 (Comparison to Proposition 3.1)**.** Compared to Proposition 3.1, Proposition 3.3 provides the lower bound of the L^∞-diameter.

3.2. **Linear inverse problem.** The second application is the frequentist evaluation of the coverage error of the credible bands based on an indirect observation in Gaussian white noise model:

$$dY(t) = K(f_0)(t)dt + \frac{1}{\sqrt{n}}dW(t),$$

where K is a known linear operator and f_0 is included in the s-Hölder–Zygmund space as described in the previous section. To this end, we introduce the wavelet-vaguelette decomposition $\{\psi_{l,k}, v^{(1)}_{l,k}, v^{(2)}_{l,k}, \kappa_l : (l,k) \in \mathcal{I}_\infty\}$ of K, where recall that $\mathcal{I}_\infty := \{(l,k) : J_0 \leqslant l < \infty, 0 \leqslant k \leqslant 2^l - 1\} \cup \{(l,k) : l = J_0 - 1, 0 \leqslant k \leqslant 2^{J_0} - 1\}$: $\{\psi_{l,k}\}$ is the wavelet basis (with the same notational convention used in the previous subsection), $\{v^{(1)}_{l,k}\}$ and $\{v^{(2)}_{l,k}\}$ are near-orthogonal functions, and $\{\kappa_{lk}\}$ is the quisi-singular values such that $K(\psi_{l,k}) = \kappa_{l,k} v^{(2)}_{l,k}$, $(l,k) \in \mathcal{I}_\infty$. For details, see [1, 20, 32, 30] and references therein. Our results cover both the mildly ill-posed and the severely ill-posed cases for $\{\kappa_{l,k}\}$: $\kappa_{l,k} \sim 2^{-rl}$ (mildly ill-posed); $\kappa_{l,k} \sim 2^{-r2^l}$ (severely ill-posed).

We use the Castillo–Nickl credible band for f. Let $\mathcal{C}(\widehat{f}, R)$ be

$$\mathcal{C}(\widehat{f}, R) := \left\{ f : \max_{(l,k)\in\mathcal{I}(J)} \frac{\kappa_{l,k}|\langle f - \widehat{f}, \psi_{l,k}\rangle|}{\sqrt{l}} \leqslant R \right\}, \ R > 0,$$

where $\widehat{f} := \sum_{(l,k)\in\mathcal{I}(J)} \psi_{l,k}\kappa^{-1}_{l,k} \int v^{(1)}_{l,k} dY$. The choice of J is as follows:

$2^J = (n/\log n)^{1/(2s+2r+1)} u_n$ for an positive sequence u_n in the mildly-ill posed case;

$2^J = c \log n$ for some $1/(2r) < c < 1/r$ in the severely-ill posed case.

We use a prior Π_f induced from Π_β on $\mathbb{R}^{2^J}$ via $\{v^{(1)}_{l,k}\}$, and $\widehat{R}_\alpha$ is chosen in the way that $\Pi_f(\mathcal{C}(\widehat{f}, \widehat{R}_\alpha) \mid Y) = 1 - \alpha$.

The following theorem provides the coverage error of Castillo–Nickl credible band in linear inverse problems. In the following, $\tau'_\infty := \|K(f_0) - K(\sum_{(l,k)\in\mathcal{I}} \beta_{0,lk}\psi_{l,k})\|_\infty$, where $\beta_{0,lk} := \langle f, \psi_{l,k}\rangle$.

Proposition 3.4. *Under Conditions 2.1 and 2.3 for Π_β that corresponds to Π_f and under the assumption that $\tau'_\infty \leqslant C'_2\sqrt{2^J(\log n)/n}$ for some $C'_2 > 0$, there exist positive constants c_1, c_2, c_3 depending only on C_1 appearing in Condition 2.1 and C'_2 such that the followings hold: For $n \geqslant 2$ satisfying $\|f_0\|_{B^s_{\infty,\infty}} \leqslant u_n$, we have*

$$\left|\mathbb{P}\left\{f_0 \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\right\} - (1-\alpha)\right| \leqslant \phi_{\Pi_\beta}\left(c_1\sqrt{2^J \log n}\right) + c_1\left(\frac{\sqrt{n}\tau'_\infty}{2^{J/2}} J^{1/2} + \mathrm{e}^{-c_2 2^J \log n}\right).$$

Further, provided that the right hand side above is smaller than $\alpha/2$, for sufficiently large n depending only on α, L^∞-diameter of $\mathcal{C}(\widehat{f}, \widehat{R}_\alpha)$ is bounded as

$$\|\mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\|_\infty \leqslant c_3 \kappa^{-1}_{J-1, 2^{J-1}-1}\sqrt{2^J(\log n)/n}$$

with probability at least $1 - c_1\exp(-c_2 2^J \log n)$.

Remark 3.5 (Rate of convergence)**.** The asymptotic form is demonstrated using a locally log-Lipschitz prior with locally log-Lipschitz constant $L = L_n$. We have

$$\begin{aligned} &\left|\mathbb{P}\left\{f_0 \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)\right\} - (1-\alpha)\right| \\ &\lesssim \begin{cases} L_n \left(\frac{n}{\log n}\right)^{-s/(2s+2r+1)} u_n^{1/2} + \frac{\log n}{u_n^{s+r+1/2}}, & \text{in the mildly ill-posed case;} \\ L_n \left(\frac{n}{\log n}\right)^{-s/(2s+2r+1)} + (\log n)^{-s}, & \text{in the severely ill-posed case,} \end{cases} \end{aligned}$$

and

$$\|\mathcal{C}(\widehat{f},\widehat{R}_\alpha)\|_\infty \lesssim \begin{cases} \left(\frac{n}{\log n}\right)^{-s/(2s+2r+1)} u_n^{r+1/2}, & \text{in the mildly ill-posed case;} \\ (\log n)^{-s}, & \text{in the severely ill-posed case,} \end{cases}$$

with probability at least $1 - c_1 \exp(-c_2 2^J \log n)$.

3.3. **Nonparametric regression model.** The third application of the main theorem is the frequentist evaluation of coverage errors of credible bands in nonparametric regression models:

$$Y_i = f_0(T_i) + \varepsilon_i,\ i = 1, \ldots, n,$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ is a vector of i.i.d. error terms with mean zero and variance σ_0^2 and $\{T_i : i = 1, \ldots, n\}$ are i.i.d. samples on $[0,1]$. For simplicity, ε and $\{T_i : i = 1, \ldots, n\}$ are independent and σ_0 does not depend on n.

Using p basis functions $\{\psi_j^p(\cdot) : 1 \leqslant j \leqslant p\}$, we consider the credible bands for f of the form

$$\mathcal{C}(\widehat{f},\widehat{R}_\alpha) = \left\{ f : \left\| \frac{f(\cdot) - \widehat{f}(\cdot)}{\|\psi^p(\cdot)\|} \right\|_\infty \leqslant \widehat{R}_\alpha \right\},$$

where $\widehat{f}(\cdot) := \sum_{j=1}^p \psi_j^p(\cdot)\widehat{\beta}_j$ with $\widehat{\beta} := \mathrm{argmin}_\beta \sum_{i=1}^n |Y_i - \sum_{j=1}^p \psi_j^p(T_i)\beta_j|^2$, $\widehat{R}_\alpha$ is taken in a way that $\Pi_f\{\mathcal{C}(\widehat{f},\widehat{R}_\alpha) \mid Y\} = 1-\alpha$, and $\psi^p(\cdot) := (\psi_1^p(\cdot), \ldots, \psi_p^p(\cdot))^\top$. We consider a prior Π_f of f induced from a sieve prior Π_β on $\mathbb{R}^p$ by the map

$$(\beta_1, \ldots, \beta_p) \mapsto \sum_{j=1}^p \beta_j \psi_j^p(\cdot).$$

Due to the randomness of $\{T_i\}$, it is necessary to develop bounds of the coverage error and L^∞-diameter taking the randomness of $\{T_i\}$ into consideration. To this end, we modify Conditions 2.1 and 2.3 and add conditions on the basis function. Let $\widetilde{\psi}^p(\cdot) := \psi^p(\cdot)/\|\psi^p(\cdot)\|$, and $\xi_p := \sup_{t\in[0,1]} \|\psi^p(t)\|$. Let $\beta_0 := \mathrm{argmin}_\beta \mathbb{E}|f_0(T_1) - \psi^p(T_1)^\top\beta|^2$. For $R > 0$, let

$$\widetilde{B}(R) := \{\beta : \|\beta - \beta_0\| \leqslant n^{-1/2}R\}, \quad \text{and} \quad \widetilde{\phi}_{\Pi_\beta}(R) := 1 - \inf_{\beta,\widetilde{\beta}\in\widetilde{B}(R)} \{\pi(\beta)/\pi(\widetilde{\beta})\}.$$

Condition 3.1. *There exists a positive constant C_1 such that $\pi(\beta_0) \geqslant \mathrm{e}^{-C_1 p \log n}$.*

Condition 3.2. *The inequality $\widetilde{\phi}_{\Pi_\beta}(1/\sqrt{n}) \leqslant 1/2$ holds.*

Condition 3.3. *There exist strictly positive constants $\underline{b}$ and $\overline{b}$ such that the eigenvalues of the $p \times p$ matrix $(\mathbb{E}\psi_i^p(T_1)\psi_j^p(T_1))$ are included in $[\underline{b}^2, \overline{b}^2]$.*

Condition 3.4. *There exist positive constants C_4 and C_5 such that the inequalities*

$$\log \xi_p \leqslant C_4 \log p \quad \text{and} \quad \log \sup_{t\neq t'\in[0,1]} \{\|\widetilde{\psi}^p(t) - \widetilde{\psi}^p(t')\|/|t-t'|\} \leqslant C_5 \log p$$

hold.

Remark 3.6 (Comments on the assumptions)**.** Conditions 3.1 and 3.2 are the versions of Conditions 2.1 and 2.3 in the case that $\{T_i\}s$ are random. Condition 3.3 is a standard assumption. Condition 3.4 is substantially weak; for example, the assumption holds for Fourier series, Spline series, CDV wavelets, and local polynomial partition series; see [5].

The following proposition provides both the coverage error and the L^∞-diameter of $\mathcal{C}(\widehat{f}, \widehat{R}_\alpha)$. Let $\tau_2 := \sqrt{\mathbb{E}|f_0(T_1) - \psi^p(T_1)^\top \beta_0|^2}$ and $\tau_\infty := \|f_0(\cdot) - \psi^p(\cdot)^\top \beta_0\|_\infty$. Further, let

$$\tau := \left\| \frac{f_0(\cdot) - \psi^p(\cdot)^\top \beta_0}{\|\psi^p(\cdot)\|} \right\|_\infty .$$

Proposition 3.5. *Under Conditions 3.1-3.4 and 2.2, there exists positive constants c_1, c_2, c_3 depending only on $C_1, \dots, C_5$, $\underline{b}, \overline{b}$, and q appearing in Conditions3.1-3.4 and 2.2 and Assumption 2.2 such that the followings hold: For $n \geqslant 2$ and any sufficiently small $\delta > 0$, we have*

$$|\mathbb{P}(f_0 \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)) - (1-\alpha)| \leqslant \widetilde{\phi}_{\Pi_\beta}(c_1\sqrt{p \log n}) + \delta_2 + \delta_3 + c_1(n^{-2\delta} + \delta_1 p \log n + \omega_n + \gamma_n), \quad (6)$$

where

$$\gamma_n := \frac{n}{\log n} \frac{\tau_2^2}{p} + \max\left\{1, \left(p\xi_p^2/n\right)^{1/2}\right\} \tau_\infty n^\delta \log p + \sqrt{n}\tau\sqrt{\log p}$$

and

$$\omega_n := \begin{cases} n^\delta (\log n)^{1/2} \max\left\{ \left(\xi_p^2/n\right)^{1/2} n^{1/q} \log n, \left(\xi_p^2/n\right)^{1/6} (\log n)^{2/3} \right\} & \text{under Assumption 2.2 (a)} \\ n^\delta (\log n)^{7/6} \left(\xi_p^2/n\right)^{1/6} & \text{under Assumption 2.2 (b).} \end{cases}$$

Further, provided that the right hand side in (6) is smaller than $\alpha/2$, for sufficiently large p depending only on α, with probability at least $1 - \delta_3 - c_1\{\sqrt{n}\tau\sqrt{\log p} + \exp(-c_2 p \log n)\}$, we have

$$\sup_{f,g \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)} \|f - g\|_\infty \leqslant c_3 \sqrt{\xi_p^2 (\log p)/n}.$$

Remark 3.7 (Choices of ξ_p, τ_2, τ_∞, and τ)**.** For typical basis functions including Fourier series, spline series, and CDV wavelets, $\xi_p \lesssim \sqrt{p}$; see Section 3 in [5]. For $S(> s)$-regular CDV wavelets, in the case that f_0 is in the Hölder–Zygmund space with smoothness level s, $\tau_2 \sim \tau_\infty \sim p^{-s}$. For the other series and the other function classes, bounds on τ_2 and τ_∞ are available from the approximation theorem; see [19] and Section 3 in [5]. Typical choice of τ is $\tau_\infty/\sqrt{p}$; For the Haar wavelet, we have $\tau \sim \tau_\infty/\sqrt{p}$, since $\tau \leqslant \tau_\infty/\inf_{t\in[0,1]} \|\psi^p(t)\|$. For periodic S-regular wavelets, we also have $\tau \sim \tau_\infty/\sqrt{p}$ as shown in Appendix E.3.

Remark 3.8 (Rate of convergence)**.** Consider the case with an unknown variance. Assume that there exists a constant $s > 1/2$ such that $\tau_2 \sim \tau_\infty \sim p^{-s}$, $\tau \sim p^{-s-1/2}$, and $\xi_p \lesssim \sqrt{p}$. Assume also that the error distribution is sub-Gaussian. Note that the assumption that $s > 1/2$ is usual in nonparametric regression with an unknown variance; see Assumption A.1 in [49]. Consider that we put a locally log-Lipschitz prior with locally log-Lipschitz constant $L = L_n$ on β and use an estimate $\widehat{\sigma}^2 = \widehat{\sigma}^2_{\mathrm{u}}$. Then, taking $p \sim (n/\log n)^{1/(2s+1)} u_n$ with a divergent sequence u_n, we have

$$|P(f_0 \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)) - (1-\alpha)| \tag{7}$$

$$\lesssim L_n \left(\frac{n}{\log n}\right)^{-s/(2s+1)} u_n^{1/2} + \left(\frac{n}{\log n}\right)^{-(s-1/2)/(2s+1)} u_n \log n + \frac{\log n}{u_n^{s+1/2}} \tag{8}$$

and

$$\sup_{f,g \in C(\widehat{f}, \widehat{R}_\alpha)} \|f-g\|_\infty \lesssim \left(\frac{n}{\log n}\right)^{-s/(2s+1)} u_n^{1/2}$$

with probability $1 - c_1/n$.

One of the important aspects of this result is that it admits a general sieve prior for β. From (8), the diminishing rate of the coverage error with respect to the prior distribution, that is, the first and second terms on the right hand side in (8) is unchanged whenever $L_n \lesssim \sqrt{p}$ up to a logarithmic factor and u_n.

4. Proof of Theorem 2.1

In this section, we provide the proof of the main theorem.

4.1. Technical lemmas.

4.1. Technical lemmas. Before the proof, we state pivotal ingredients of the proof except the Berry–Esseen type bound on posterior distributions: the high-dimensional CLT on hyper-rectangles, the anti-concentration inequality on hyper-rectangles, Anderson's lemma, and concentration inequality for Gaussian maxima.

The high-dimensional CLT on hyper-rectangles is stated as follows: let $Z_1, \dots, Z_n$ be independent p-dimensional random vectors with mean zero. We denote the j-th coordinate of Z_i by Z_{ij}. Let $\widetilde{Z_1}, \dots, \widetilde{Z_n}$ be independent centered p-dimensional Gaussian vectors such that each $\widetilde{Z}_i$ has the same covariance matrix as Z_i. Let $\mathcal{A}^{\mathrm{re}}$ be the class of all hyper-rectangles in $\mathbb{R}^p$: for any $A \in \mathcal{A}^{\mathrm{re}}$, A is of the form $A = \{\beta \in \mathbb{R}^p : \underline{a}_i \leqslant \beta_i \leqslant \overline{a}_i,\ 1 \leqslant \forall i \leqslant p\}$ for $(\underline{a}_1, \dots, \underline{a}_p)^\top \in \mathbb{R}^p$ and for $(\overline{a}_1, \dots, \overline{a}_p)^\top \in \mathbb{R}^p$. Assume that the following three conditions hold:

- H1. There exists $b > 0$ such that $n^{-1}\sum_{i=1}^n \mathbb{E}|Z_{ij}|^2 \geqslant b$ for $1 \leqslant \forall j \leqslant p$;
- H2. There exists a sequence $B_n \geqslant 1$ such that $n^{-1}\sum_{i=1}^n \mathbb{E}|Z_{ij}|^{2+k} \leqslant B_n^4$ for $1 \leqslant \forall j \leqslant p$ and for $k = 1, 2$;
- H3. We assume either one of the following two conditions:
  - (a) There exists $q > 0$ such that $\mathbb{E}[\{\max_{j=1,\dots,p} |Z_{ij}|/B_n\}^q] \leqslant 2$ for $1 \leqslant \forall i \leqslant n$;
  - (b) $\mathbb{E}[\exp\{|Z_{ij}|/B_n\}] \leqslant 2$ for $1 \leqslant \forall i \leqslant n$ and for $1 \leqslant \forall j \leqslant p$.

Lemma 4.1 (High Dimensional CLT for Hyperrectangles; Proposition 2.1 in [16])**.** *Assume that Conditions H1 and H2 hold. Let*

$$\rho = \rho_n := \sup_{A \in \mathcal{A}^{\mathrm{re}}} \left| \mathbb{P}\left(\sum_{i=1}^n Z_i/\sqrt{n} \in A\right) - \mathbb{P}\left(\sum_{i=1}^n \widetilde{Z}_i/\sqrt{n} \in A\right) \right|.$$

Then, there exists a positive constant $\widetilde{c}_1$ *for which we have*

$$\rho \leqslant \begin{cases} \widetilde{c}_1 \left(\frac{B_n^2 \log^7(pn)}{n}\right)^{1/6} + \widetilde{c}_1 \left(\frac{B_n^2 \log^3(pn)}{n^{1-2/q}}\right)^{1/3} & \text{under Condition H3 (a);} \\ \widetilde{c}_1 \left(\frac{B_n^2 \log^7(pn)}{n}\right)^{1/6} & \text{under Condition H3 (b).} \end{cases}$$

Here, $\widetilde{c}_1$ *depends only on* b *appearing in Condition H1 and* q *appearing in Condition H3.*

We use the anti-concentration inequality on hyper-rectangles. Let $\sigma_j^2 := \mathbb{E}Z_j^2 > 0$ for $1 \leqslant \forall j \leqslant p$ and let $\underline{\sigma} := \min\{\sigma_j\}$, $\overline{\sigma} := \max\{\sigma_j\}$.

Lemma 4.2 (Anti-concentration inequality on hyper-rectangles; Theorem in [35])**.** *There exists a universal positive constant* $\widetilde{c}_2$ *for which we have, for every* $z = (z_1, \dots, z_p)^\top \in \mathbb{R}^p$ *and* $R > 0$,

$$\gamma := \gamma(R) := \mathbb{P}(\widetilde{Z}_j \leqslant z_j + R \text{ for } 1 \leqslant \forall j \leqslant p) - \mathbb{P}(\widetilde{Z}_j \leqslant z_j \text{ for } 1 \leqslant \forall j \leqslant p) \leqslant \widetilde{c}_2 \frac{R}{\underline{\sigma}} \sqrt{\log p}.$$

Remark 4.1 (Comment on the anti-concentration inequality)**.** The point here is that the above inequality allows for general hyper-rectangles. Remark that when we focus on max-rectangles, the anti-concentration inequality above is less sharp than the anti-concentration inequality on max-rectangles obtained by [15]; see also [13] in the sense that the former uses $\sqrt{\log p}$ while the latter uses $\mathbb{E}[\max_{i=1,\dots,p} \widetilde{Z}_j]$.

The followings are Anderson's lemma and the concentration inequality for Gaussian maxima.

Lemma 4.3 (Anderson's lemma; Collorary 3 in [2])**.** *Let* Σ *and* $\widetilde{\Sigma}$ *be two nonnegative definite and symmetric* $p \times p$ *matrices. Let* Y *and* $\widetilde{Y}$ *be random vectors from* $\mathcal{N}(0, \Sigma)$ *and from* $\mathcal{N}(0, \widetilde{\Sigma})$, *respectively. Let* $\mathcal{C}$ *be a convex symmetric set in* $\mathbb{R}^p$. *If* $\Sigma - \widetilde{\Sigma}$ *is nonnegative definite, then* $\mathbb{P}(Y \in \mathcal{C}) \leqslant \mathbb{P}(\widetilde{Y} \in \mathcal{C})$.

Lemma 4.4 (Concentration inequality for Gaussian maxima; Theorem 2.5.8. in [27])**.** *Let* $\{N_i\}_{i=1}^p$ *be i.i.d. random variables from the standard Gaussian distribution. Then, we have*

$$\mathbb{P}(|\max_{i=1,\dots,p} |N_i| - \mathbb{E} \max_{i=1,\dots,p} |N_i|| \geqslant R) \leqslant \mathrm{e}^{-R^2/2}, \ R > 0.$$

4.2. **Proof of Theorem 2.1.** We only prove the theorem under Assumption 2.2 (a). The proof under Assumption 2.2 (b) is completed replacing Lemma 4.1 (a) by Lemma 4.1 (b).

The proof is divided into two parts: The former part is to present an upper bound of the coverage error $|\mathbb{P}(\beta_0 \in I(\widehat{\beta}(Y), \widehat{R}_\alpha)) - (1-\alpha)|$. The latter part is to evaluate the max-diameter of $I(\widehat{\beta}(Y), \widehat{R}_\alpha)$.

Upper bound for the coverage error. At the first step, we show that $\widehat{R}_\alpha$ concentrates on the $(1-\alpha)$-quantile of some distribution with a high probability. From Proposition 2.5, we have

$$|\Pi_\beta(I(\widehat{\beta}(Y), \widehat{R}_\alpha) \mid Y) - \mathcal{N}(I(\widehat{\beta}(Y), \widehat{R}_\alpha) \mid \widehat{\beta}(Y), \sigma_0^2(X^\top X)^{-1})| \leqslant \overline{\omega} \ \text{ for } Y \in H,$$

where $\overline{\omega}$ is the upper bound in Proposition 2.5 and recall that

$$H = \{Y : \|X(\widehat{\beta}(Y) - \beta_0)\| \leqslant c_1 \sqrt{p \log n} \sigma_0 / 4\} \cap \{Y : \Pi_{\sigma^2}(|\sigma^2/\sigma_0^2 - 1| \geqslant \delta_1 \mid Y) \leqslant \delta_2\}.$$

Let G be the cumulative distribution function of $\sigma_0 \max\{|e_{(p),i}^\top (X^\top X)^{-1} X^\top N|/w_i\}$, where $e_{(d),i}$ is the d-dimensional unit vector whose i-th component is 1, and N is the random vector from the standard p-dimensional Gaussian distribution. Since

$$\mathcal{N}(I(\widehat{\beta}(Y), \widehat{R}_\alpha) \mid \widehat{\beta}(Y), \sigma_0^2 (X^\top X)^{-1}) = \mathbb{P}(\sigma_0 \max_{i=1,\dots,p} \{|e_{(p),i}^\top (X^\top X)^{-1} X^\top N|/w_i\} \leqslant \widehat{R}_\alpha \mid Y),$$

we have $|(1-\alpha) - G(\widehat{R}_\alpha)| \leqslant \overline{\omega}$ for $Y \in H$. Letting G^{-1} be the quantile function of G yields

$$G^{-1}(1-\alpha-\overline{\omega}) \leqslant \widehat{R}_\alpha \leqslant G^{-1}(1-\alpha+\overline{\omega}) \quad \text{for } Y \in H, \tag{9}$$

which completes the first step.

At the second step, we derive an upper bound of $\mathbb{P}(\beta_0 \in I(\widehat{\beta}(Y), \widehat{R}_\alpha)) - (1-\alpha)$. The lower bound is obtained in the same way. Because the inequality $\widehat{R}_\alpha \leqslant G^{-1}(1-\alpha+\overline{\omega})$ holds for $Y \in H$, we have

$$\begin{aligned}
&\mathbb{P}(\beta_0 \in I(\widehat{\beta}(Y), \widehat{R}_\alpha)) - (1-\alpha) \\
&\leqslant \mathbb{P}(Y \in \{Y : \max_{i=1,\dots,p} \{|e_{(p),i}^\top (X^\top X)^{-1} X^\top Y|/w_i\} \leqslant \widehat{R}_\alpha\} \cap H) - (1-\alpha) + \mathbb{P}(Y \notin H) \\
&\leqslant \mathbb{P}\left(\max_{i=1,\dots,p} \{|e_{(p),i}^\top (X^\top X)^{-1} X^\top (\varepsilon + r)|/w_i\} \leqslant G^{-1}(1-\alpha+\overline{\omega}) \right) - (1-\alpha) + \rho + \mathbb{P}(Y \notin H) \\
&\leqslant \gamma(\|(X^\top X)^{-1} X^\top r\|_\infty) + \rho + \mathbb{P}(Y \notin H),
\end{aligned}$$

where both ρ and $\gamma = \gamma(\|(X^\top X)^{-1} X^\top r\|_\infty)$ are constants appearing in Lemmas 4.1 and 4.2 in the case that $Z := (X^\top X)^{-1} X^\top \varepsilon$. From Proposition 2.6, we have the upper bound of $\mathbb{P}(Y \notin H)$. Noting that ρ is independent of rescaling of Z and replacing Z by $Z/\sigma_0 \underline{\lambda}^{1/2}$, we can take $b = 1$ and $B_n = \sqrt{p} (\mathbb{E}|\varepsilon_1/\sigma_0|^q)^{1/q} (\overline{\lambda}/\underline{\lambda})^{1/2}$, since we have

$$n^{-1} \sum_{i=1}^n \mathbb{E}|Z_{ij}|^2 := n^{-1} \sum_{i=1}^n \mathbb{E}|e_{(p),j}^\top (X^\top X)^{-1} X^\top e_{(n),i} \varepsilon_i|^2 \geqslant \sigma_0^2 \underline{\lambda}$$

and since it follows from

$$|Z_{ij}/(\varepsilon_i/\sigma_0)| \leqslant \|e_{(n),i}^\top X (X^\top X)^{-1}\| / \sqrt{\underline{\lambda}} \leqslant (\overline{\lambda}/\underline{\lambda}^{1/2}) \|X^\top e_{(n),i}\| \leqslant (\overline{\lambda}/\underline{\lambda}) \sqrt{p}$$

that for $1 \leqslant \forall j \leqslant p$,

$$n^{-1} \sum_{i=1}^n \mathbb{E}|Z_{ij}|^3 \leqslant (\sqrt{p}\overline{\lambda}/\underline{\lambda})^3 \mathbb{E}(\varepsilon_1/\sigma_0)^3 \quad \text{and } n^{-1} \sum_{i=1}^n \mathbb{E}|Z_{ij}|^4 \leqslant (\sqrt{p}\overline{\lambda}/\underline{\lambda})^4 \mathbb{E}(\varepsilon_1/\sigma_0)^4.$$

Thus, from Lemmas 4.1 and 4.2, we obtain the following bounds on ρ and γ: For some $\widetilde{c}_1 > 0$ depending only on q,

$$\begin{aligned}
\rho &\leqslant \widetilde{c}_1 \left\{ \left(\frac{p \log^7(pn)}{n} \frac{\overline{\lambda}}{\underline{\lambda}} \right)^{1/6} + \left(\frac{p \log^3(pn)}{n^{1-2/q}} \frac{\overline{\lambda}}{\underline{\lambda}} \right)^{1/3} \right\}, \\
\gamma &\leqslant \widetilde{c}_1 \frac{\|(X^\top X)^{-1} X^\top r\|_\infty}{\sigma_0 \underline{\lambda}} \sqrt{\log p},
\end{aligned}$$

which completes the second step and thus completes the evaluation of the coverage error.

Estimate of the max-diameter. At the first step, we bound the max diameter using the quantile function F^{-1} of $\max_{i=1,\dots,p}|N_i|$. From the triangle inequality, we have, for $Y \in H$,

$$G^{-1}(1-\alpha-\varpi) \leqslant \sup_{\beta_1,\beta_2 \in I(\widehat{\beta},\widehat{R}_\alpha)} \max_{i=1,\dots,p} |\beta_{1,i}-\beta_{2,i}| \leqslant 2G^{-1}(1-\alpha+\varpi).$$

Lemma 4.3 yields, for $R>0$,

$$\mathbb{P}\left(\max_{i=1,\dots,p}|N_i| \leqslant \frac{\overline{w}R}{\sigma_0 \underline{\lambda}^{1/2}}\right) \leqslant \mathbb{P}\left(\max_{i=1,\dots,p}|\widetilde{N}_i| \leqslant R\right) \leqslant \mathbb{P}\left(\max_{i=1,\dots,p}|N_i| \leqslant \frac{\underline{w}R}{\sigma_0 \overline{\lambda}^{1/2}}\right),$$

where $\widetilde{N} := \sigma_0 W(X^\top X)^{-1}X^\top N$ and $W=\mathrm{diag}(w_1,\dots,w_p)$. Therefore, for any $\beta \in (0,1)$ we have

$$\left(\frac{\underline{w}}{\sigma_0 \overline{\lambda}^{1/2}}\right) G^{-1}(1-\beta) \leqslant F^{-1}(1-\beta) \leqslant \left(\frac{\overline{w}}{\sigma_0 \underline{\lambda}^{1/2}}\right) G^{-1}(1-\beta), \tag{10}$$

which completes the first step.

At the second step, we will show that for sufficiently large p depending only on α, $F^{-1}(1-\alpha\pm\varpi) \sim \sqrt{\log p}$ with probability at least $\mathbb{P}(Y \in H)$. First, we will show that $F^{-1}(1-\alpha+\varpi) \lesssim \sqrt{\log p}$. From Lemma 4.4, taking sufficiently large p depending only on α yields

$$\mathbb{P}(\max_{i=1,\dots,p}|N_i| - \mathbb{E}\max_{i=1,\dots,p}|N_i| \geqslant \widetilde{c}_2\sqrt{\log p}) \leqslant \exp(-\widetilde{c}_2^2 \log p/2) < \alpha - \alpha/2 < \alpha - \varpi,$$

for some positive constant $\widetilde{c}_2$. Therefore, noting that

$$\begin{aligned} F^{-1}(1-\alpha+\varpi) &:= \inf\{R : \mathbb{P}(\max_{i=1,\dots,p}|N_i| \geqslant R) \leqslant \alpha-\varpi\} \\ &= \inf\{R : \mathbb{P}(\max_{i=1,\dots,p}|N_i| - \mathbb{E}\max_{i=1,\dots,p}|N_i| \geqslant R - \mathbb{E}\max_{i=1,\dots,p}|N_i|) \leqslant \alpha-\varpi\}, \end{aligned}$$

we have $F^{-1}(1-\alpha+\varpi) \lesssim \sqrt{\log p}$.

Second, we show that $F^{-1}(1-\alpha-\varpi) \gtrsim \sqrt{\log p}$. From the Paley–Zygmund inequality, we have, for $\theta \in [0,1]$,

$$\mathbb{P}(\max_{i=1,\dots,p}|N_i| \geqslant \theta\mathbb{E}\max_{i=1,\dots,p}|N_i|) \geqslant (1-\theta)^2(\mathbb{E}\max_{i=1,\dots,p}|N_i|)^2/\mathbb{E}(\max_{i=1,\dots,p}|N_i|)^2.$$

Here, it follows that

$$\mathbb{E}\{\max_{i=1,\dots,p}|N_i|\}^2 \leqslant \{\mathbb{E}\max_{i=1,\dots,p}|N_i|\}^2 + \sqrt{2\pi}\mathbb{E}\max_{i=1,\dots,p}|N_i| + 2 \tag{11}$$

because we have, for any δ in $(0,1)$,

$$\begin{aligned} &\mathbb{E}[\{\max_{i=1,\dots,p}|N_i|\}^2] \\ &= \int_{[0,\mathbb{E}\max_{i=1,\dots,p}|N_i|+\delta]} \mathbb{P}(\max_{i=1,\dots,p}|N_i|^2 \geqslant t)dt + \int_{[\mathbb{E}\max_{i=1,\dots,p}|N_i|+\delta,\infty)} \mathbb{P}(\max_{i=1,\dots,p}|N_i|^2 \geqslant t)dt \\ &\leqslant (\mathbb{E}\max_{i=1,\dots,p}|N_i|+\delta)^2 + \int_{[\mathbb{E}\max_{i=1,\dots,p}|N_i|+\delta,\infty)} 2te^{-(t-\mathbb{E}\max_{i=1,\dots,p}|N_i|^2/2}dt \\ &\leqslant (\mathbb{E}\max_{i=1,\dots,p}|N_i|+\delta)^2 + \sqrt{2\pi}\mathbb{E}\max_{i=1,\dots,p}|N_i| + 2, \end{aligned}$$

where the second inequality follows from the concentration inequality of the maxima of a Gaussian process. Since it follows that $(\mathbb{E}\max_{i=1,\dots,p}|N_i|)^2/\mathbb{E}(\max_{i=1,\dots,p}|N_i|)^2 \to 1$ as $p\to\infty$ from (11), we have, for sufficiently large p and for some θ depending only on α,

$$\mathbb{P}(\max_{i=1,\dots,p}|N_i| \geqslant \theta\mathbb{E}\max_{i=1,\dots,p}|N_i|) \geqslant \alpha+\overline{\omega}$$

and thus we obtain $F^{-1}(1-\alpha-\overline{\omega}) \gtrsim \sqrt{\log p}$, which completes the second step and hence completes the proof. □

Appendix A. Proof of Proposition 2.5

A.1. **Technical Lemmas.** We present here some technical lemmas that will be used to prove Proposition 2.5.

Lemma A.1 (Scheffé's lemma). *Let Q_1 and Q_2 be probability measures on a measurable space with a common dominating measure μ. Let $q_1 = dQ_1/d\mu$ and $q_2 = dQ_2/d\mu$. Then*

$$\|Q_1 - Q_2\|_{\mathrm{TV}} = \frac{1}{2}\int |q_1(x)-q_2(x)|d\mu(x) = \int (q_1(x)-q_2(x))_+ d\mu(x),$$

Proof. See, e.g., p.84 in [45]. □

Lemma A.2 (Posterior contraction of a marginal prior distribution). *Recall that $B(R) = \{\beta\in\mathbb{R}^p : \|X(\beta-\beta_0)\| \leqslant \sigma_0 R\}$ for $R>0$. Under Conditions 2.1 and 2.3, there exist positive constants $\widetilde{c}_1$ and $\widetilde{c}_2$ depending only on C_1 in Condition 2.1 such that for a sufficiently large $R>0$, the inequality*

$$\Pi_\beta(\beta\notin B(R)\mid Y,\sigma^2) \leqslant 4\exp\{\widetilde{c}_1 p\log n - \widetilde{c}_2(\sigma_0^2/\sigma^2)R^2\} \tag{12}$$

holds for $Y\in H$, where recall that

$$H := \{Y : \|X(\widehat{\beta}(Y)-\beta_0)\| \leqslant R\sigma_0/4\} \cap \{Y : \Pi_{\sigma^2}(|\sigma^2/\sigma_0^2 - 1| \geqslant \delta_1 \mid Y) \leqslant \delta_2\}.$$

Proof. We use the following lower bounds on the small ball probability of a prior distribution:

Lemma A.3 (Lower bounds on the small ball probability of a prior distribution). *Let Π_β be a probability measure with a density π with respect to the p-dimensional Lebesgue measure. Recall that $\phi_{\Pi_\beta}(R) = 1-\inf_{\beta,\widetilde{\beta}\in B(R)}\{\pi(\beta)/\pi(\widetilde{\beta})\}$ for $R>0$. Then, we have, for every $R>0$,*

$$\Pi_\beta(\beta\in B(R)) \geqslant \frac{\{1-\phi_{\Pi_\beta}(R)\}(\pi\mathrm{e}R)^{p/2}}{2(p/2+1)^{p/2+1/2}}\frac{\pi(\beta_0)\sigma_0^p}{\sqrt{\det(X^\top X)}}.$$

Proof of Lemma A.3. Observe that

$$\Pi_\beta(\beta\in B(R)) = \int_{B(R)}\pi(\beta)d\beta \geqslant \inf_{\beta\in B(R)}\left\{\frac{\pi(\beta)}{\pi(\beta_0)}\right\}\pi(\beta_0)\int_{B(R)}\mathrm{d}\beta.$$

Changing variables, we have that

$$\int_{B(R)}d\beta = \frac{(\sigma_0^2R^2)^{p/2}}{\sqrt{\det(X^\top X)}}\int_{\|\beta\|\leqslant 1}d\beta = \frac{(\sigma_0^2R^2)^{p/2}\pi^{p/2}}{\sqrt{\det(X^\top X)}\Gamma(p/2+1)},$$

where $\Gamma(\cdot)$ is the Gamma function. Using the bound

$$\Gamma(p/2+1) \leqslant \frac{\sqrt{2\pi}}{\mathrm{e}}(p/2+1)^{p/2+1/2}\mathrm{e}^{-p/2}\mathrm{e}^{1/18}$$

(see section 5.6.1. in [36]), we have that

$$\int_{B(R)} \mathrm{d}\beta \geqslant \frac{(\sigma_0^2 \pi \mathrm{e} R^2)^{p/2}\mathrm{e}^{17/18}}{\sqrt{2\pi}\sqrt{\det(X^\top X)}(p/2+1)^{p/2+1/2}}.$$

Since $\mathrm{e}^{17/18}/\sqrt{2\pi} \geqslant 1/2$, we obtain the desired inequality. □

Return to the proof of Lemma A.2. Letting $P := X(X^\top X)^{-1}X^\top$, we have

$$\Pi_\beta(\beta \in B \mid Y, \sigma^2) = \frac{\int_{B^c} \mathrm{e}^{-\langle P(\varepsilon+r), X(\beta-\beta_0)\rangle/\sigma^2 - \|X(\beta-\beta_0)\|^2/(2\sigma^2)}\pi(\beta)d\beta}{\int \mathrm{e}^{-\langle P(\varepsilon+r), X(\beta-\beta_0)\rangle/\sigma^2 - \|X(\beta-\beta_0)\|^2/(2\sigma^2)}\pi(\beta)d\beta}. \tag{13}$$

Since $cx^2 + c^{-1}y^2 \geqslant 2xy$ for $x, y, c > 0$, we have, for any $c > 1$,

$$\begin{aligned}
&\int_{B^c} \mathrm{e}^{-\langle P(\varepsilon+r), X(\beta-\beta_0)\rangle/\sigma^2 - \|X(\beta-\beta_0)\|^2/(2\sigma^2)}\pi(\beta)d\beta \\
&\quad\leqslant \int_{B^c} \mathrm{e}^{\|P(\varepsilon+r)\|\|X(\beta-\beta_0)\|/\sigma^2 - \|X(\beta-\beta_0)\|^2/(2\sigma^2)}\pi(\beta)d\beta \\
&\quad\leqslant \int_{B^c} \mathrm{e}^{\{c\|P(\varepsilon+r)\|^2 + c^{-1}\|X(\beta-\beta_0)\|^2\}/(2\sigma^2) - \|X(\beta-\beta_0)\|^2/(2\sigma^2)}\pi(\beta)d\beta \\
&\quad\leqslant \exp\{c\|P(\varepsilon+r)\|^2/(2\sigma^2) - (1-c^{-1})(\sigma_0^2/\sigma^2)R^2/2\}.
\end{aligned} \tag{14}$$

Letting $\widetilde{R} = 1/\sqrt{\pi e n}$, we have

$$\begin{aligned}
&\int \mathrm{e}^{-\langle P(\varepsilon+r), X(\beta-\beta_0)\rangle/\sigma^2 - \|X(\beta-\beta_0)\|^2/(2\sigma^2)}\pi(\beta)d\beta \\
&\quad\geqslant \int_{B(\widetilde{R})} \mathrm{e}^{-\langle P(\varepsilon+r), X(\beta-\beta_0)\rangle/\sigma^2 - \|X(\beta-\beta_0)\|^2/(2\sigma^2)}\pi(\beta)d\beta \\
&\quad\geqslant \int_{B(\widetilde{R})} \mathrm{e}^{-\{c\|P(\varepsilon+r)\|^2 + c^{-1}\|X(\beta-\beta_0)\|^2\}/(2\sigma^2) - \|X(\beta-\beta_0)\|^2/(2\sigma^2)}\pi(\beta)d\beta \\
&\quad\geqslant \exp\{-c\|P(\varepsilon+r)\|^2/(2\sigma^2) - (1+c^{-1})(\sigma_0^2/\sigma^2)\widetilde{R}^2/2\}\Pi_\beta(B(\widetilde{R})).
\end{aligned} \tag{15}$$

We have

$$\begin{aligned}
&\int \mathrm{e}^{-\langle P(\varepsilon+r), X(\beta-\beta_0)\rangle/\sigma^2 - \|X(\beta-\beta_0)\|^2/(2\sigma^2)}\pi(\beta)d\beta \\
&\geqslant \frac{1-\phi_{\Pi_\beta}(\widetilde{R})}{2}\mathrm{e}^{p\log n/2 - p\log p - C_1 p\log n}\mathrm{e}^{-c\|P(\varepsilon+r)\|^2/(2\sigma^2) - (1+c^{-1})(\widetilde{R}^2/2)(\sigma_0^2/\sigma^2)} \\
&\geqslant 4^{-1}\exp\{p\log n/2 - p\log p - C_1 p\log n - c\|P(\varepsilon+r)\|^2/(2\sigma^2) - (1+c^{-1})(\sigma_0^2/\sigma^2)\widetilde{R}^2/2\},
\end{aligned} \tag{16}$$

where the first inequality follows from (15) and from Lemma A.3 and the second inequality follows from Condition 2.3.

Combining (14) and (16) with (13), we have, for $Y \in H$,

$$\begin{aligned}
&\int_{B^c} \mathrm{e}^{-\|Y-X\beta\|^2/(2\sigma^2)}\pi(\beta)d\beta \Big/ \int \mathrm{e}^{-\|Y-X\beta\|^2/(2\sigma^2)}\pi(\beta)d\beta \\
&\leqslant 4\exp[(C_1+1/2)p\log n + \{(1+c^{-1})/(2n)\}(\sigma_0^2/\sigma^2) - \{(1-c^{-1})/2 - c/16\}(\sigma_0^2/\sigma^2)R^2].
\end{aligned} \tag{17}$$

Taking $c = 3$ completes the proof. $\square$

Lemma A.4. *Let A be an $n \times n$ symmetric positive semidefinite matrix such that $\|A\|_{\mathrm{op}} \leqslant 1$ and $\mathrm{rank}(A) < n$. Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ be a vector of i.i.d. random variables with mean zero and unit variance.*

(a) *If in addition Assumption 2.2 (a) holds for an integer $q \geqslant 2$ and $C_3 > 0$, then there exists a positive constant $\widetilde{c}_1$ depending only on q and C_3 such that, for every $R > \sqrt{\mathrm{rank}(A)}$,*

$$\mathbb{P}\left(\varepsilon^\top A\varepsilon \geqslant R^2\right) \leqslant \widetilde{c}_1\, \mathrm{rank}(A)/(R - \sqrt{\mathrm{rank}(A)})^q.$$

(b) *If instead Assumption 2.2 (b) holds for $C_3 > 0$, then there exists a positive constant $\widetilde{c}_1$ depending only on C_3 such that, for every $R > 0$,*

$$\mathbb{P}\left(|\varepsilon^\top A\varepsilon - \mathbb{E}[\varepsilon^\top A\varepsilon]| > R^2\right) \leqslant 2\exp\{-\widetilde{c}_1 \min\left(R^4/\|A\|_{\mathrm{HS}}^2, R^2\right)\},$$

where $\|\cdot\|_{\mathrm{HS}}$ denotes Hilbert–Schmidt norm.

Proof. For Case (a), see Corollary 5.1 in [4]. The inequality in Case (b) is called the Hanson-Wright inequality; for a proof, we refer to [29] and [41]. $\square$

A.2. **Proof of Proposition 2.5.** Before the proof, we prepare additional notations for the sake of notational simplicity. Let $\widetilde{\mathcal{N}} := \mathcal{N}(\widehat{\beta}(Y), \sigma_0^2(X^\top X)^{-1})$. Let $B := B(c_1\sqrt{p\log n})$ and $H := H(c_1)$ for a sufficiently large $c_1 > 0$ depending on C_1 and C_2. Let $\Pi_\beta^B(d\beta \mid Y)$ be the probability measure defined by

$$\Pi_\beta^B(d\beta \mid Y) := 1_{\beta\in B}\Pi_\beta(d\beta \mid Y) \Big/ \int_B \Pi_\beta(d\widetilde{\beta} \mid Y)$$

and let $\widetilde{\mathcal{N}}^B$ be the probability measure defined by

$$\widetilde{\mathcal{N}}^B(d\beta) := 1_{\beta\in B}\widetilde{\mathcal{N}}(d\beta) \Big/ \int_B \widetilde{\mathcal{N}}(d\beta).$$

Let $\Pi_\beta(\cdot \mid Y, \sigma^2)$ be the distribution defined by

$$\Pi_\beta(d\beta \mid Y, \sigma^2) := \mathrm{e}^{-\|Y-X\beta\|^2/(2\sigma^2)}\pi(\beta)d\beta \Big/ \int \mathrm{e}^{-\|Y-X\widetilde{\beta}\|^2/(2\sigma^2)}\pi(\widetilde{\beta})d\widetilde{\beta}$$

and let $\Pi_\beta^B(\cdot \mid Y, \sigma^2)$ be the distribution defined by

$$\Pi_\beta^B(d\beta \mid Y, \sigma^2) := 1_{\beta\in B}\mathrm{e}^{-\|Y-X\beta\|^2/(2\sigma^2)}\pi(\beta)d\beta \Big/ \int_B \mathrm{e}^{-\|Y-X\widetilde{\beta}\|^2/(2\sigma^2)}\pi(\widetilde{\beta})d\widetilde{\beta}.$$

In the proof, $\widetilde{c}_1, \widetilde{c}_2, \ldots$ are positive constants depending only on C_1, C_2, and c_1.

Proof outline :

First of all, we present a brief outline of the proof. From the triangle inequality, we have

$$\|\Pi_\beta(d\beta \mid Y) - \widetilde{\mathcal{N}}\|_{\mathrm{TV}} \leqslant \|\Pi_\beta(d\beta \mid Y) - \Pi_\beta(d\beta \mid Y, \sigma_0^2)\|_{\mathrm{TV}} + \|\Pi_\beta(d\beta \mid Y, \sigma_0^2) - \widetilde{\mathcal{N}}\|_{\mathrm{TV}}. \tag{18}$$

Consider the first term on the right hand side of (18). Letting $S = S(\delta_1) := \{\sigma^2 : |\sigma^2/\sigma_0^2 - 1| \leqslant \delta_1\}$, it follows that

$$\|\Pi_\beta(d\beta \mid Y) - \Pi_\beta(d\beta \mid Y, \sigma_0^2)\|_{\mathrm{TV}} \leqslant \int_S \|\Pi_\beta(d\beta \mid Y, \sigma^2) - \Pi_\beta(d\beta \mid Y, \sigma_0^2)\|_{\mathrm{TV}}\Pi_{\sigma^2}(d\sigma^2 \mid Y) + \delta_1$$

with probability at least $1-\delta_2$, from the application of Jensen's inequality to the function $x \to |x|$ and from Condition 2.2. For the bound of the first term on the rightmost hand in the above inequality, the triangle inequality yields

$$\int_S \|\Pi_\beta(d\beta \mid Y, \sigma^2) - \Pi_\beta(d\beta \mid Y, \sigma_0^2)\|_{\mathrm{TV}} \Pi_{\sigma^2}(d\sigma^2 \mid Y) \leqslant A_1 + A_2 + A_3, \tag{19}$$

where

$$\begin{aligned}
A_1 &:= \int_S \|\Pi_\beta(d\beta \mid Y, \sigma^2) - \Pi_\beta^B(d\beta \mid Y, \sigma^2)\|_{\mathrm{TV}} \Pi_{\sigma^2}(d\sigma^2 \mid Y), \\
A_2 &:= \int_S \|\Pi_\beta^B(d\beta \mid Y, \sigma^2) - \Pi_\beta^B(d\beta \mid Y, \sigma_0^2)\|_{\mathrm{TV}} \Pi_{\sigma^2}(d\sigma^2 \mid Y), \\
A_3 &:= \int_S \|\Pi_\beta^B(d\beta \mid Y, \sigma_0^2) - \Pi_\beta(d\beta \mid Y, \sigma_0^2)\|_{\mathrm{TV}} \Pi_{\sigma^2}(d\sigma^2 \mid Y).
\end{aligned}$$

Upper bounds of A_1, A_2, A_3 will be presented in (21),(22), and (23). Consider the second term on the right hand side of (18). From the triangle inequality, we have

$$\|\Pi_\beta(d\beta \mid Y, \sigma_0^2) - \widetilde{\mathcal{N}}\|_{\mathrm{TV}} \leqslant A_4 + A_5 + A_6, \tag{20}$$

where $A_4 := \|\widetilde{\mathcal{N}} - \widetilde{\mathcal{N}}^B\|_{\mathrm{TV}}$, $A_5 := \|\widetilde{\mathcal{N}}^B - \Pi_\beta^B(d\beta \mid Y, \sigma_0^2)\|_{\mathrm{TV}}$, and $A_6 := \|\Pi_\beta^B(d\beta \mid Y, \sigma_0^2) - \Pi_\beta(d\beta \mid Y)\|_{\mathrm{TV}}$. Upper bounds of A_4, A_5, A_6 will be presented in (25), (26), and (27).

Upper bound of (19):

Consider A_1 in (19). From Lemmas A.1 and A.2, taking a sufficiently large c_1 depending only on C_1 yields

$$A_1 = \int_S \Pi_\beta(\beta \notin B \mid Y, \sigma^2) \Pi_{\sigma^2}(d\sigma^2 \mid Y) \leqslant 4\mathrm{e}^{-\widetilde{c}_1 p \log n}. \tag{21}$$

Consider A_2 in (19). From Lemma A.1, we have

$$A_2 = \int_S \int \left(1 - \phi_{\Pi_\beta,2}(\beta, \sigma^2)\right)_+ \Pi_\beta^B(d\beta \mid Y, \sigma_0^2) \Pi_{\sigma^2}(d\sigma^2 \mid Y),$$

where

$$\phi_{\Pi_\beta,2}(\beta, \sigma^2) := \frac{\pi(\beta)\mathrm{e}^{-\|Y - X\beta\|^2/(2\sigma^2)}}{\int_B \mathrm{e}^{-\|Y - X\widetilde{\beta}\|^2/(2\sigma^2)} \pi(\widetilde{\beta}) d\widetilde{\beta}} \frac{\int_B \mathrm{e}^{-\|Y - X\widetilde{\beta}\|^2/(2\sigma_0^2)} \pi(\widetilde{\beta}) d\widetilde{\beta}}{\pi(\beta)\mathrm{e}^{-\|Y - X\beta\|^2/(2\sigma_0^2)}}.$$

From Cauchy–Schwarz's inequality and Assumption 2.1, we have

$$\begin{aligned}
&\mathrm{e}^{-\langle P(\varepsilon + r), X\beta_0 - X\beta\rangle/\sigma^2 - \|X\beta_0 - X\beta\|^2/(2\sigma^2)} \\
&\geqslant \mathrm{e}^{-\langle P(\varepsilon + r), X\beta_0 - X\beta\rangle/\sigma_0^2 - \|X\beta_0 - X\beta\|^2/(2\sigma_0^2)} \mathrm{e}^{-C_2 c_1 \delta_1 p \log n/\{4(1-\delta_1)\} - c_1^2 \delta_1 p \log n/(1-\delta_1)}.
\end{aligned}$$

Likewise, we have

$$\begin{aligned}
&\int_B \mathrm{e}^{-\langle P(\varepsilon + r), X\beta_0 - X\widetilde{\beta}\rangle/\sigma^2 - \|X\beta_0 - X\widetilde{\beta}\|^2/(2\sigma^2)} \pi(\widetilde{\beta}) d\widetilde{\beta} \\
&\leqslant \mathrm{e}^{C_2 c_1 \delta_1 p \log n/\{4(1-\delta_1)\} + c_1^2 \delta_1 p \log n/(1-\delta_1)} \int_B \mathrm{e}^{-\langle P(\varepsilon + r), X\beta_0 - X\beta\rangle/\sigma_0^2 - \|X\beta_0 - X\beta\|^2/(2\sigma_0^2)} \pi(\widetilde{\beta}) d\widetilde{\beta}.
\end{aligned}$$

Therefore, we have, for $\beta \in B$, $Y \in H$, and $\sigma^2 \in S$,

$$\phi_{\Pi_\beta,2}(\beta,\sigma^2) \geqslant \exp(-\widetilde{c}_2\delta_1 p\log n)$$

and thus it follows that

$$A_2 \leqslant \widetilde{c}_2\delta_1 p\log n \tag{22}$$

since $(1-\mathrm{e}^{-x})_+ \leqslant x$ for $x>0$.

Consider A_3 in (19). From Lemmas A.1 and A.2, taking a sufficiently large c_1 depending only on C_1 yields, for $Y \in H$,

$$A_3 \leqslant \Pi_\beta(\beta \notin B \mid Y, \sigma_0^2) \leqslant 4\exp(-\widetilde{c}_3 p\log n). \tag{23}$$

Therefore, inequalities (21), (22), and (23) yield

$$\int_S \|\Pi_\beta(d\beta \mid Y,\sigma^2) - \Pi_\beta(d\beta \mid Y,\sigma_0^2)\|_{\mathrm{TV}}\Pi_{\sigma^2}(d\sigma^2 \mid Y) \leqslant \widetilde{c}_4\mathrm{e}^{-\widetilde{c}_5 p\log n} + \widetilde{c}_4\delta_1 p\log n. \tag{24}$$

Upper bound of (20):

Consider A_4 in (20). From Lemmas A.1 and A.4, we have

$$A_4 = \widetilde{\mathcal{N}}(B^{\mathrm{c}}) \leqslant \exp\{-(3c_1\sqrt{p\log n}/4 - \sqrt{p})^2/2\}. \tag{25}$$

Consider A_5 in (20). From Lemma A.1, we have

$$A_5 = \int\left(1 - \frac{d\widetilde{\mathcal{N}}^B}{d\Pi_\beta^B(\cdot \mid Y,\sigma_0^2)}(\beta)\right)_+ \Pi_\beta^B(d\beta \mid Y,\sigma_0^2).$$

We denote the density of $\widetilde{\mathcal{N}}$ with respect to the Lebesgue measure by $\widetilde{\phi}$. Since we have, for $\beta \in B$,

$$\frac{d\widetilde{\mathcal{N}}^B}{d\beta}(\beta) = \frac{\widetilde{\phi}(\beta)}{\int_B \widetilde{\phi}(\widetilde{\beta})d\widetilde{\beta}} \quad\text{and}\quad \frac{d\Pi_\beta^B(\cdot \mid Y,\sigma_0^2)}{d\beta}(\beta) = \frac{\pi(\beta)\widetilde{\phi}(\beta)}{\int_B \pi(\widetilde{\beta})\widetilde{\phi}(\widetilde{\beta})d\widetilde{\beta}},$$

applying Jensen's inequality to $x \to (1-x)_+$ yields

$$\begin{aligned}\int\left(1 - \frac{d\widetilde{\mathcal{N}}^B}{d\Pi_\beta^B}(\beta \mid Y,\sigma_0^2)\right)_+ \Pi_\beta^B(d\beta \mid Y) &= \int\left(1 - \int_B \frac{\pi(\widetilde{\beta})}{\pi(\beta)}\frac{\widetilde{\phi}(\widetilde{\beta})}{\int_B \widetilde{\phi}(\beta')d\beta'}d\widetilde{\beta}\right)_+ \Pi_\beta^B(d\beta \mid Y)\\ &\leqslant \int\int_B\left(1 - \frac{\pi(\widetilde{\beta})}{\pi(\beta)}\right)_+ \frac{\widetilde{\phi}(\widetilde{\beta})}{\int_B \widetilde{\phi}(\beta')d\beta'}d\widetilde{\beta}\Pi_\beta^B(d\beta \mid Y).\end{aligned}$$

Therefore, we obtain

$$A_5 \leqslant \phi_{\Pi_\beta}(c_1\sqrt{p\log n}). \tag{26}$$

Consider A_6 in (20). From Lemmas A.1 and A.2, taking a sufficiently large $c_1 > 0$ yields

$$A_6 = \Pi_\beta(\beta \notin B \mid Y,\sigma_0^2) \leqslant 4\exp(-\widetilde{c}_6 p\log n). \tag{27}$$

Therefore, inequalities (25), (26), and (27) yield

$$\|\Pi_\beta(d\beta \mid Y,\sigma_0^2) - \widetilde{\mathcal{N}}\|_{\mathrm{TV}} \leqslant \phi_{\Pi_\beta}(c_1\sqrt{p\log n}) + \widetilde{c}_7\exp(-\widetilde{c}_8 p\log n). \tag{28}$$

Combining (24) and (28) with (19) provides the upper bound of the target total variation and thus completes the proof. □

A.3. **Proof of Proposition 2.6.** Let c be any positive number. Under Assumption 2.2 (a), it follows from Lemma A.4 (a) with $R = c\sqrt{p\log n}$ that the inequality

$$\mathbb{P}(Y \notin H(c)) \leqslant \widetilde{c}_1 p^{1-q/2}(\log n)^{-q/2} + \delta_3$$

holds for some $\widetilde{c}_1 > 0$ depending only on c, C_3, and q. Under Assumption 2.2 (b), it follows from Lemma A.4 (b) with $R = (c^2+1)p\log n$, that the inequality

$$\mathbb{P}(Y \notin H(c)) \leqslant 2\mathrm{e}^{-\widetilde{c}_2 \min\{p(\log n)^2, p\log n\}} + \delta_3$$

holds for some $\widetilde{c}_2 > 0$ depending only on c, C_3, and q. Thus, we completes the proof. □

Appendix B. Proofs of Propositions 2.1–2.4

Proof of Proposition 2.1. Let $\widetilde{B}(R) := \{\beta : \|\beta - \beta_0\| \leqslant \sigma_0 \overline{\lambda}^{1/2} R\}$ for $R > 0$. The inequality

$$\phi_{\Pi_\beta}(c\sqrt{p\log n}) < cL\sigma_0\overline{\lambda}^{1/2}\sqrt{p\log n}, \tag{29}$$

holds for any $c > 0$, since

$$\phi_{\Pi_\beta}(c\sqrt{p\log n}) \leqslant \sup_{\beta,\widetilde{\beta}\in\widetilde{B}} [1 - \exp\{-\log\{\pi(\beta)/\pi(\widetilde{\beta})\}\}] \leqslant c\sigma_0 L\overline{\lambda}^{1/2}\sqrt{p\log n},$$

where the first inequality follows because $\|X(\beta-\beta_0)\| \geqslant \overline{\lambda}^{-1/2}\|\beta - \beta_0\|$ and the second inequality follows because $1 - \mathrm{e}^{-x} \leqslant x$. Substituting $c = 1/(\sqrt{pn\log n})$ into (29), we have $\phi_{\Pi_\beta}(1/\sqrt{n}) \leqslant L\overline{\lambda}^{1/2}\sigma_0/\sqrt{n}$. Therefore, we complete the proof. □

Proof of Proposition 2.2. First, consider an isotropic prior. We have

$$\begin{aligned}
\log\pi(\beta_0) &= \log\rho(\|\beta_0\|) - \log\int\rho(\|\beta\|)d\beta \\
&= \log\rho(\|\beta_0\|) - \log\left[\{p\pi^{p/2}/\Gamma(p/2+1)\}\int_0^\infty x^{p-1}\rho(x)dx\right] \\
&\geqslant \log\inf_{x\in[0,B]}\rho(x) - \widetilde{c}_1 p\log p \\
&\geqslant \log\inf_{x\in[0,B]}\rho(x) - \widetilde{c}_1 p\log p - \widetilde{c}_1\log n + \log\{\sqrt{\det(X^\top X)}/\sigma_0^p\}
\end{aligned}$$

for some positive constant $\widetilde{c}_1$ depending only on m and c appealing in the definition of an isotropic prior and Assumption 2.3. Thus, we see that an isotropic prior satisfies Condition 2.1. To see the locally log-Lipschitz continuity, Taylor's expansion yields

$$|\log\pi(\beta_0+s_1) - \log\pi(\beta_0+s_2)| \leqslant \sup_{x:0\leqslant x\leqslant B+\sqrt{\sigma_0^2\overline{\lambda}p\log n}} |d\log\rho/dx(x)|(\|\beta_0+s_1\| - \|\beta_0+s_2\|).$$

This completes the proof for the case of an isotropic prior.

Second, consider a product prior $\pi(\beta) = \prod_{i=1}^p \pi_i(\beta_i)$. We have

$$
\begin{aligned}
\log \pi(\beta_0) &\geqslant p \log \min_i \pi_i(0) - \widetilde{L} p^{1/2} \|\beta_0\| \\
&\geqslant p \log \min_i \pi_i(0) - \widetilde{L} B p \log n \\
&\geqslant -\widetilde{L} B p (1 + o(1)) \log n - \widetilde{c}_2 \log n + \log\{\sqrt{\det(X^\top X)}/\sigma_0^p\}
\end{aligned}
$$

for some positive constant $\widetilde{c}_2$ depending only on c appearing in Assumption 2.3. Thus, we see that a product prior satisfies Condition 2.1. To see the locally log-Lipschitz continuity, the Lipschitz continuity of $\log \pi(\beta)$ yields

$$
|\log \pi(\beta) - \log \pi(\beta_0)| \leqslant \sum_{i=1}^p |\log \pi_i(\beta_i) - \log \pi_i(\beta_{0,i})| \leqslant \widetilde{L} p^{1/2} \|\beta - \beta_0\|,
$$

which completes the proof. □

Proof of Proposition 2.3. We present the proof only for the case under Assumption 2.2 (a). The proof for the case under Assumption 2.2 (b) is completed replacing Lemma A.4 (a) by Lemma A.4 (b).

First, from Lemma A.4 (a), we have $\mathbb{P}(\widehat{\sigma}_{\mathrm{u}}^2/\sigma_0^2 - 1 \geqslant \delta_1) \leqslant \widetilde{c}_1/(n-p)^{q/2-1}\widetilde{\delta}_1^q$ for some positive constant $\widetilde{c}_1$ depending only on q, because it follows that

$$
\begin{aligned}
\widehat{\sigma}_{\mathrm{u}}^2 &= \frac{\|Y - X(X^\top X)^{-1} X^\top Y\|^2}{\sigma_0^2 (n-p) \sigma_0^2} \\
&\leqslant \frac{\|\varepsilon - X(X^\top X)^{-1} X^\top \varepsilon\|^2 + 2\|r - X(X^\top X)^{-1} X^\top r\|^2 + |\varepsilon^\top u|^2}{\sigma_0^2 (n-p)} \\
&= \frac{\varepsilon^\top \widetilde{A} \varepsilon + 2\|r - X(X^\top X)^{-1} X^\top r\|^2}{\sigma_0^2 (n-p)}
\end{aligned}
$$

where

$$
u := \begin{cases} \{I - X(X^\top X)^{-1} X^\top\} r / \|\{I - X(X^\top X)^{-1} X^\top\} r\| & \text{if } \{I - X(X^\top X)^{-1} X^\top\} r \neq 0, \\ \text{arbitrary} & \text{if otherwise,} \end{cases}
$$

and $\widetilde{A} := I - X(X^\top X)^{-1} X^\top + u u^\top$.

Next, we will show that

$$
\mathbb{P}\left(\widehat{\sigma}_{\mathrm{u}}^2(Y)/\sigma_0^2 - 1 \leqslant -\delta_1\right) \leqslant \widetilde{c}_2 \frac{\max\{n^{q/4}, n\}}{\delta_1^{q/2} (n-p)^{q/2}} + \widetilde{c}_2 \frac{p^{q/2+1}}{(n-p)^q \delta_1^q} \tag{30}
$$

for some positive constant $\widetilde{c}_2$ depending only on q. Letting $\widetilde{P}$ be the projection onto the linear space spanned by $\{X, (I - X(X^\top X)^{-1} X^\top) r\}$, we have

$$
\begin{aligned}
&\mathbb{P}\left(\widehat{\sigma}_{\mathrm{u}}^2(Y)/\sigma_0^2 - 1 \leqslant -\delta_1\right) \\
&\quad \leqslant \mathbb{P}\left(\{\|\varepsilon\|^2 - \|\widetilde{P}\varepsilon\|^2\}/\{\sigma_0^2(n-p)\} \leqslant 1 - \delta_1\right) \\
&\quad \leqslant \mathbb{P}\left(\|\varepsilon\|^2/\sigma_0^2(n-p) - n/(n-p) \leqslant -\delta_1/2\right) + \mathbb{P}\left(\|\widetilde{P}\varepsilon\|^2/\sigma_0^2(n-p) \geqslant p/(n-p) + \delta_1/2\right). \qquad (31)
\end{aligned}
$$

For the upper bound of the first term on the rightmost hand side in (31), we use Rosenthal's inequality:

Lemma B.1 (Rosenthal's inequality; see [38] and [47].)**.** *For some positive constant $\widetilde{c}_3$ depending only on q, we have* $\mathbb{E}\left|\|\varepsilon/\sigma_0\|^2 - n\right|^{q/2} \leqslant \widetilde{c}_3 \max\{n^{q/4}, n\}$.

We have, from Markov's inequality and Rosenthal's inequality,

$$\mathbb{P}\left(\|\varepsilon\|^2/\{\sigma_0^2(n-p)\} - n/(n-p) \leqslant -\delta_1/2\right) \leqslant \widetilde{c}_4 \max\{n^{q/4}, n\}/\{\delta_1^{q/2}(n-p)^{q/2}\} \tag{32}$$

for some $\widetilde{c}_4 > 0$ depending only on q. For the upper bound of the second term on the rightmost hand side in (31), we use Lemma A.4 (a) with $R = \sqrt{p + (n-p)\delta_1/2}$. We thus obtain

$$\mathbb{P}\left(\|\widetilde{P}\varepsilon\|^2/\{\sigma_0^2(n-p)\} \geqslant p/(n-p) + \delta_1/2\right) \leqslant \widetilde{c}_4 n^{1-q/2}/\delta_1^{q/2}. \tag{33}$$

Combining (32) and (33) with (31) yields (30), which completes the proof for the case under Assumption A.4 (a). □

Proof of Proposition 2.4. The marginal posterior distribution of σ^2 is given by the inverse Gamma distribution $\mathrm{IG}(a^*, b^*)$, where $a^* = \mu_1 + n/2 - p/2$ and $b^* = \mu_2 + \|Y - PY\|^2/2$. The mean of this marginal posterior is $\{2\mu_2 + \|(I - X(X^\top X)^{-1}X^\top)Y\|^2\}/\{2\mu_1 + n - p - 2\}$; while the variance is $2\{2\mu_2 + \|(I - X(X^\top X)^{-1}X^\top)Y\|^2\}^2/\{2\mu_1 + n - p - 2\}^2\{2\mu_1 + n - p - 4\}$. From Chebyshev's inequality,

$$\Pi_{\sigma^2}(\sigma^2 : |\sigma^2/\sigma_0^2 - 1| \geqslant \delta_1 \mid Y) \leqslant \widetilde{c}_1 \|(I - X(X^\top X)^{-1}X^\top)Y\|^2/\{n^2(\delta_1 - |\mathbb{E}[\sigma^2/\sigma_0^2 \mid Y] - 1|)^2\} \tag{34}$$

for some positive constant $\widetilde{c}_1$ depending only on μ_1 and μ_2. We have, from the proof of Proposition 2.3, the upper bound of $\mathbb{P}(\|(I - X(X^\top X)^{-1}X^\top)Y\|^2/(n-p) - 1 \geqslant \delta_1/2)$ and thus complete the proof. □

Appendix C. Proofs for Section 3.1

C.1. **Proof of Proposition 3.1.** First, we transform a white noise model

$$dY(t) = f_0(t)dt + \frac{dW(t)}{\sqrt{n}}$$

into a Gaussian sequence model

$$Y_{l,k} = \beta_{0,lk} + r_{l,k} + \varepsilon_{l,k},\ (l,k) \in \mathcal{I},$$

where the distribution of $\varepsilon_{l,k}$ is $\mathcal{N}(0, 1/n)$ and $r_{l,k} = \tau_\infty/\sqrt{2^J}$. This transformation is done via a mapping

$$f \to \left(\int \psi_{(J_0-1)0}(t)f(t)dt, \int \psi_{(J_0-1)1}(t)f(t)dt, \ldots, \int \psi_{(J-1)(2^{J-1}-1)}(t)f(t)dt\right).$$

Therefore, if the estimate of L^∞-diameter is provided, Theorem 2.1 will complete the proof.

Next, we derive the upper bound of the L^∞-diameter of $\mathcal{C}(\widehat{f}, \widehat{R}_\alpha)$. For $f, g \in \mathcal{C}(\widehat{f}, \widehat{R}_\alpha)$, let $h := f - g$. It follows from the triangle inequality and from the approximation ability of the S-regular wavelet that $||h||_\infty \leqslant \widetilde{c}_1(A_1 + A_2 + A_3)$ for some $\widetilde{c}_1 > 0$, where

$$
\begin{aligned}
A_1 &:= 2^{J_0/2} \max_{0 \leqslant k \leqslant 2^{J_0}-1} |\langle h, \psi_{(J_0-1)k} \rangle|, \\
A_2 &:= \sum_{J_0 \leqslant l \leqslant J-1} 2^{l/2} \max_{0 \leqslant k \leqslant 2^l - 1} |\langle h, \psi_{l,k} \rangle|, \\
A_3 &:= \sum_{J \leqslant l} 2^{l/2} \max_{0 \leqslant k \leqslant 2^l - 1} |\langle h, \psi_{l,k} \rangle|.
\end{aligned}
$$

Using the radius $\widehat{R}_\alpha$, the quantities A_1 and A_2 on the right hand side are bounded as $\max\{A_1, A_2\} \leqslant \widetilde{c}_2 2^{J/2} \sqrt{J} \widehat{R}_\alpha$ for some $\widetilde{c}_2 > 0$. There exist positive constants $\widetilde{c}_3, \widetilde{c}_4, \widetilde{c}_5$ such that, for sufficiently large n depending only on α,

$$
\max\{A_1, A_2\} \leqslant \widetilde{c}_3 \sqrt{2^J/n} \sqrt{\log n}
$$

with probability at least $1 - \widetilde{c}_4 \exp(-\widetilde{c}_5 2^J \log n)$, since it follows from Theorem 2.1 that $\widehat{R}_\alpha \leqslant \widetilde{c}_6 \sigma_0$ on H for some $\widetilde{c}_6 > 0$. The quantity A_3 on the right hand side is bounded as

$$
A_3 \leqslant \widetilde{c}_7 2^{J/2} 2^{-J(s+1/2)} u_n^{1/2} = \widetilde{c}_7 \sqrt{2^J/n} \sqrt{\log n} / u_n^s
$$

for some $\widetilde{c}_7 > 0$. This complete the proof. □

C.2. **Proof of Proposition 3.2.** The proof consists of the upper bound of the coverage error and the estimate of the L^∞-diameter.

Upper bound of the coverage error. We apply Theorem 2.1 for the Gaussian sequence model

$$
Y_{l,k} = \beta_{0,lk} + r_{l,k} + \varepsilon_{l,k},\ (l,k) \in \mathcal{I}(J')
$$

and the corresponding credible band by setting $X = I_{2^{J'}}$, $w_{l,k} = w_l$ for $(l,k) \in \mathcal{I}(J')$, and $\sigma_0 = 1/\sqrt{n}$.

However, an additional treatment for the term $\sup_{l \geqslant J', 0 \leqslant k \leqslant 2^l - 1} |\beta_{0,lk} - Y_{l,k}|/w_l$ is required because $\mathcal{C}(\widehat{f}_\infty, \widehat{R}_\alpha)$ depends on $(Y_{J'+1,0}, Y_{J'+1,1}, \ldots)$, where $Y_{l,k} := \int \psi_{l,k} dY$ for $J' \leqslant l < \infty, 0 \leqslant k \leqslant 2^l - 1$. Let $Y := (Y_{J_0-1,0}, \ldots)$.

The treatment that we conduct is as follows. Let $\widetilde{H} := \{Y : \sup_{J' \leqslant l, 0 \leqslant k \leqslant 2^l - 1} |Y_{l,k} - \beta_{0,lk}|/w_l \leqslant \delta\}$. Note that we have

$$
\mathbb{P}(f_0 \in \mathcal{C}(\widehat{f}_\infty, \widehat{R}_\alpha)) \leqslant \mathbb{P}(\sup_{(l,k) \in \mathcal{I}(J')} |\beta_{0,lk} - Y_{l,k}|/w_l \leqslant \widehat{R}_\alpha)
$$

and

$$
\begin{aligned}
&\mathbb{P}(f_0 \in \mathcal{C}(\widehat{f}_\infty, \widehat{R}_\alpha)) \\
&\quad \geqslant \mathbb{P}(Y \in \{Y : \sup_{(l,k)\in\mathcal{I}(J')} |\beta_{0,lk} - Y_{l,k}|/w_l + \sup_{J'\leqslant l, 0\leqslant k\leqslant 2^l-1} |\beta_{0,lk} - Y_{l,k}|/w_l \leqslant \widehat{R}_\alpha\} \cup \widetilde{H}) \\
&\quad \geqslant \mathbb{P}(Y \in \{Y : \sup_{(l,k)\in\mathcal{I}(J')} |\beta_{0,lk} - Y_{l,k}|/w_l + \delta \leqslant \widehat{R}_\alpha\}) - \mathbb{P}(Y \notin \widetilde{H}) \\
&\quad = \mathbb{P}(Y \in \{Y : \sup_{(l,k)\in\mathcal{I}(J')} |\beta_{0,lk} - Y_{l,k} - \mathrm{sgn}(Y_{l,k} - \beta_{0,lk})\delta w_l|/w_l \leqslant \widehat{R}_\alpha\}) + \mathbb{P}(Y \notin \widetilde{H}).
\end{aligned}
$$

Therefore, setting $r := (\pm w_l \delta)$ enables us to apply Theorem 2.1 for this case. Two remarks are in order. Remark that the sign of r does not affect on the result. Remark also that Assumption 2.1 is always satisfied because Assumption 2.1 is on a bias term in the model and because there is no bias term in the Gaussian sequence model. Thus, adding $\mathbb{P}(Y \notin \widetilde{H})$ to the upper bound will complete the evaluation of the coverage error.

Consider the upper bound of $\mathbb{P}(Y \notin \widetilde{H})$. Let $\{N_{l,k} : J' < l, 0 \leqslant k \leqslant 2^l - 1\}$ be i.i.d. random variables from the standard Gaussian distribution. It follows from Lemma 4.4 and from the inequality $\mathbb{E}\max_{k\leqslant 2^l-1} |N_{l,k}| \leqslant \sqrt{2l}$ that for some $\widetilde{c}_1, \widetilde{c}_2 > 0$,

$$
\begin{aligned}
\mathbb{P}(Y \notin \widetilde{H}) &\leqslant \sum_{l\geqslant J'} \mathbb{P}(\max_k |N_{l,k}| - \mathbb{E}\max |N_{l,k}| \geqslant w_l(\delta/\sigma_0 - \sqrt{2}/u_n)) \\
&\leqslant \widetilde{c}_1 \mathrm{e}^{-\widetilde{c}_2 n \overline{w}^2 \delta^2}.
\end{aligned}
$$

This completes the evaluation of the coverage error.

Estimate of the L^∞-diameter. We derive the upper bound of the L^∞-diameter of $\mathcal{C}(\widehat{f}_\infty, \widehat{R}_\alpha)$. For $f, g \in \mathcal{C}(\widehat{f}_\infty, \widehat{R}_\alpha)$, we have

$$
\|f - g\|_\infty \leqslant 2\sum_{l<J'} 2^{l/2} \max_{k\leqslant 2^l-1} |\langle h, \psi_{l,k}\rangle| + 2\sum_{l\geqslant J'} 2^{l/2} \max_{k\leqslant 2^l-1} |\langle h, \psi_{l,k}\rangle|,
$$

where $h := f - g$. The former term on the right hand side in the above inequality is bounded as

$$
\sum_{l<J'} 2^{l/2} \max_{k\leqslant 2^l-1} |\langle h, \psi_{l,k}\rangle| \leqslant \sum_{l<J'} 2^{l/2}\sqrt{l}(w_l/\sqrt{l})\{\max_{k\leqslant 2^l-1} |\langle h, \psi_{l,k}\rangle/w_l|\} \leqslant \widetilde{c}_3 2^{J'/2}\sqrt{J'} \max_{l<J'}\{w_l/\sqrt{l}\}\widehat{R}_\alpha
$$

for some $\widetilde{c}_3 > 0$. There exist $\widetilde{c}_4, \widetilde{c}_5, \widetilde{c}_6 > 0$ such that for sufficiently large n depending only on α,

$$
\sum_{l\leqslant J'} 2^{l/2} \max_{k\leqslant 2^l-1} |\langle h, \psi_{l,k}\rangle| \leqslant \widetilde{c}_4 \sqrt{2^{J'}(\log n)/n} \max_{l<J'} \frac{w_l}{\sqrt{l}}
$$

with probability at least $1 - \widetilde{c}_5 \exp(-\widetilde{c}_6 2^{J'} \log n)$, since it follows from (9) and (10) that $\widehat{R}_\alpha \leqslant \sqrt{(\log n)/n}$ on H for some $\widetilde{c}_7 > 0$,

The latter term on the right hand side is bounded as

$$
2\sum_{l\geqslant J'} 2^{l/2} \max_{k\leqslant 2^l-1} |\langle h, \psi_{l,k}\rangle| \leqslant \widetilde{c}_8 2^{-J'(s+1/2)} u_n
$$

for some $\widetilde{c}_8 > 0$ because of the approximation property of the S-regular wavelet. This completes the proof. □

C.3. **Proof of Proposition 3.3.** The proof follows essentially the same line as that of Proposition 3.1. The only difference is to provide the lower bound of the L^∞-diameter. It follows that there exists $\widetilde{c}_1 > 0$ for which we have, for sufficiently large n such that the coverage error is bounded by $\alpha/2$, $\sup_{f,g\in\mathcal{C}(\widehat{f},\widehat{R}_\alpha)} \|f-g\|_\infty \geqslant \widetilde{c}_1\sqrt{2^J/n}$ with a high probability, since $\sup_{f,g\in\mathcal{C}(\widehat{f},\widehat{R}_\alpha)} \|f-g\|_\infty \geqslant \sqrt{2^J}\widehat{R}_\alpha$. This completes the proof. $\square$

Appendix D. Proofs for Section 3.2

Proof of Prosposition 3.4. The proof follows the line of that of Proposition 3.3. The only difference is to deal with the non-orthogonality of $\{v_{l,k}^{(1)}\}$.

When denoting by $\widetilde{Y} = (\widetilde{Y}_{l,k})_{(l,k)\in\mathcal{I}(J)}$ the coefficients of dY with respect to $\{v_{l,k}^{(1)} : (l,k) \in \mathcal{I}(J)\}$, $\widetilde{Y}$ follows

$$\widetilde{Y}_{l,k} = \kappa_{l,k}\beta_{0,lk} + \widetilde{r} + \varepsilon_{l,k},\ \ (l,k) \in \mathcal{I}(J),$$

where $\beta_0 = (\beta_{0,lk})_{(l,k)\in\mathcal{I}(J)}$ is the coefficient vector of f_0 with respect to $\{\psi_{l,k} : (l,k) \in \mathcal{I}(J)\}$, $\widetilde{r} := \tau'_\infty/\sqrt{2^J}$, and $\varepsilon = (\varepsilon_{l,k})_{(l,k)\in\mathcal{I}(J)}$ follows the Gaussian distribution with mean zero and covariance matrix Σ.

From the near-orthogonality of $\{v_{l,k}^{(1)}\}$, there exist $\underline{b}$ and $\bar{b}$ such that $\underline{b}I_{2^J}/n \preceq \Sigma \preceq \bar{b}I_{2^J}/n$, where constants $\underline{b}, \bar{b}$ depend only on the frame constants of $\{v_{l,k}^{(1)}\}$ and $\{v_{l,k}^{(2)}\}$. Let $\kappa := \mathrm{diag}(\kappa_{l,k})_{(l,k)\in\mathcal{I}(J)}$. Thus, setting $Y = \Sigma^{-1/2}\widetilde{Y}$, $X = \Sigma^{-1/2}\kappa$, $r = \Sigma^{-1/2}\widetilde{r}$, $\sigma_0 = 1$, and $w_{l,k} = \sqrt{l}\kappa_{l,k}^{-1}$, Theorem 2.1 completes the proof. $\square$

Appendix E. Proof for Section 3.3

First of all, we transform the model into the following regression model based on the approximation by p basis functions $\{\psi_j^p : 1 \leqslant j \leqslant p\}$:

$$Y = X\beta_0 + r + \varepsilon,$$

where $Y = (Y_1, \ldots, Y_n)^\top$, $X = (X_1, \ldots, X_n)^\top$ with each component X_i of which the $j(\in \{1, \ldots, p\})$-th component is $\psi_j^p(T_i)$, and $r = (r_1, \ldots, r_n)^\top$ with each component $r_i = f_0(T_i) - \psi^p(T_i)^\top\beta_0$. Here, recall that $\beta_0 \in \mathrm{argmin}\mathbb{E}|f_0(T_1) - \sum_{j=1}^p \psi_j^p(T_1)\beta_j|^2$.

E.1. **Technical lemmas.** Before the proof, we state four technical tools used in the proof. Let $N_{(n)}$ be a random n-vector from $\mathcal{N}(0, \sigma_0^2 I_n)$, and $N_{(p)}$ be a random p-vector from $\mathcal{N}(0, \sigma_0^2 I_p)$. Let $B = (B_{ij}) := (\mathbb{E}\psi_i^p(T_1)\psi_j^p(T_1))$ and recall $\widetilde{\psi}^p(\cdot) := \psi^p(\cdot)/\|\psi^p(\cdot)\|$ and $\xi_p := \|\|\psi^p(\cdot)\|\|_\infty$.

Lemma E.1 (Matrix Chernoff inequality; [44])**.** *Let $\{A_i : i = 1, \ldots, n\}$ be an i.i.d. sequence of positive semi-definite and self-adjoint $p \times p$ matrices of which the maximum eigenvalues are almost surely bounded by R. Then,*

$$\mathbb{P}(\lambda_{\min}(\textstyle\sum A_i/n) \leqslant (1-\delta)\lambda_{\min}(\mathbb{E}[A_1])) \leqslant p\{\mathrm{e}^{-\delta}/(1-\delta)^{1-\delta}\}^{n\lambda_{\min}(\mathbb{E}[A_1])/R} \ \textit{for}\ \delta \in [0,1] \ \ \textit{and}$$
$$\mathbb{P}(\lambda_{\max}(\textstyle\sum A_i/n) \leqslant (1-\delta)\lambda_{\max}(\mathbb{E}[A_1])) \leqslant p\{\mathrm{e}^{\delta}/(1+\delta)^{1-\delta}\}^{n\lambda_{\min}(\mathbb{E}[A_1])/R} \ \textit{for}\ \delta > 0,$$

where $\lambda_{\min}(\cdot)$ *and* $\lambda_{\max}(\cdot)$ *are the maximum and the minimum eigenvalues.*

Lemma E.2 (Lemma 4.2 in [5]). *Under Conditions 3.3-3.4, the equality*

$$\left\|\widetilde{\psi}^p(\cdot)^\top\sqrt{n}(\widehat{\beta}-\beta_0)-\widetilde{\psi}^p(\cdot)^\top\frac{B^{-1}X^\top\varepsilon}{\sqrt{n}}\right\|_\infty\leqslant R_1+R_2$$

holds, where R_1 *and* R_2 *are random variables such that there exist positive constants* $\widetilde{c}_1$ *and* $\widetilde{c}_2$ *depending only on* q *appearing in Assumption 2.2 (a) for which we have*

$$R_1\leqslant\begin{cases}\widetilde{c}_1\eta^2\sqrt{\frac{\xi_p^2\log p}{n}}(n^{1/q}\sqrt{\log p}+\sqrt{p}\tau_\infty) & \text{under Assumption 2.2 (a)},\\ \widetilde{c}_1\eta^2\sqrt{\frac{\xi_p^2\log p}{n}}(\sqrt{\log n}\sqrt{\log p}+\sqrt{p}\tau_\infty) & \text{under Assumption 2.2 (b)},\end{cases}$$
$$R_2\leqslant\widetilde{c}_2\eta\sqrt{\log p}\tau_\infty$$

with probability at least $1-\widetilde{c}_2/\eta$, *for any* $\eta>1$.

Remark E.1. Belloni et al. [5] provides the proof under Assumption 2.2 (a). The proof under Assumption 2.2 (b) is almost the same noting that $n^{1/q}$ comes from $\mathbb{E}[\max_{i=1,\dots,n}|\varepsilon_i|]$.

Lemma E.3 (Corollary 2.2 and Proposition 3.3 in [14]). *Under Conditions 3.3-3.4, for any* $\eta>0$, *there exists a random variable* $\widetilde{Z}\overset{d}{=}\|\widetilde{\psi}^p(\cdot)^\top B^{-1}N_{(p)}\|_\infty$ *such that the inequality*

$$\left|\left\|\widetilde{\psi}^p(\cdot)^\top\sqrt{n}\frac{B^{-1}X^\top\varepsilon}{n}\right\|_\infty-\widetilde{Z}\right|$$
$$\leqslant\begin{cases}\widetilde{c}_1\frac{n^{1/q}\log n}{\eta^{1/2}}\frac{\xi_p}{n^{1/2}}+\frac{(\log n)^{3/4}}{\eta^{1/2}}\frac{\xi_p^{1/2}}{n^{1/4}}+\frac{(\log n)^{2/3}}{\eta^{1/3}}\frac{\xi_p^{1/3}}{n^{1/6}} & \text{under Assumption 2.2 (a)},\\ \widetilde{c}_1\frac{\log n}{\eta^{1/2}}\frac{\xi_p}{n^{1/2}}+\frac{(\log n)^{3/4}}{\eta^{1/2}}\frac{\xi_p^{1/2}}{n^{1/4}}+\frac{(\log n)^{2/3}}{\eta^{1/3}}\frac{\xi_p^{1/3}}{n^{1/6}} & \text{under Assumption 2.2 (b)}\end{cases}$$

holds with probability at least $1-\widetilde{c}_2\{\eta+(\log n)/n\}$ *for some* $\widetilde{c}_1,\widetilde{c}_2>0$ *not depending on* n *and* p.

Lemma E.4. *Under Condition 3.4, the inequality* $\mathbb{E}\|\widetilde{\psi}^p(\cdot)^\top B^{-1}N_{(p)}\|_\infty\leqslant\widetilde{c}_1\sqrt{\log p}$ *holds for some positive constant* $\widetilde{c}_1$ *depending only on* C_5 *appearing in Condition 3.4.*

Proof. From Dudley's entropy integral (e.g.,see Corollary 2.2.8 in [46]),

$$\begin{aligned}&\mathbb{E}[\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2}N_{(p)}\|_\infty]\\ &\leqslant\mathbb{E}[|\widetilde{\psi}^p(0)^\top B^{-1/2}N_{(p)}|]+\mathbb{E}[\sup_{t\neq t'\in[0,1]}|\widetilde{\psi}^p(t)^\top B^{-1/2}N_{(p)}-\widetilde{\psi}^p(t')^\top B^{-1/2}N_{(p)}|]\\ &\leqslant\underline{b}+\int_0^\theta\sqrt{\log N([0,1],d_X,\delta)}d\delta,\end{aligned}$$

where $N([0,1],d_X,\delta)$ is a δ-covering number of $[0,1]$ with respect to

$$d_X(t,t'):=\{\mathbb{E}[\widetilde{\psi}^p(t)^\top B^{-1/2}N_{(p)}-\widetilde{\psi}^p(t')^\top B^{-1/2}N_{(p)}]^2\}^{1/2}$$

and $\theta:=\sup_{t\in[0,1]}d_X(t,0)$. Since θ is bounded by $2\underline{b}$,

$$\int_0^\theta\sqrt{\log N([0,1],d_X,\delta)}d\delta\leqslant\int_0^{2\underline{b}}\sqrt{\log N([0,1],d_X,\delta)}d\delta.$$

Since it follows from the bound on covering numbers of functions Lipschitz in one parameter (Theorem 2.7.11 in [46]) that we have, for some $\widetilde{c}_2 > 0$, $N([0,1], d_X, \delta) \leqslant \big(\widetilde{c}_2 p^{C_5}/\delta\big)$, the inequality

$$\int_0^{2\underline{b}} \sqrt{\log N([0,1], d_X, \delta)} d\delta \leqslant \sqrt{C_5 \log p} + \int_0^{2\underline{b}} \sqrt{\log(\widetilde{c}_2 \underline{b}/\delta)} d\delta$$

holds. Thus, we obtain the desired inequality. ☐

Lemma E.5. *Under Conditions 3.3-3.4, there exists a positive constant $\widetilde{c}_1$ not depending on n and p for which we have*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(\big|\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2} N_{(p)}\|_\infty - x\big| \leqslant R\big) \leqslant \widetilde{c}_1 R \sqrt{\log p},\ \ R > 0.$$

Proof. From Theorem 2.1 in [14],

$$\sup_{x\in\mathbb{R}} \mathbb{P}\big(\big|\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2} N_{(p)}\|_\infty - x\big| \leqslant R\big) \leqslant \widetilde{c}_1 R \mathbb{E}[\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2} N_{(p)}\|_\infty]$$

and thus Lemma E.4 completes the proof. ☐

E.2. **Proof of Proposition 3.5.** We only provide the proof under Assumption 2.2 (a). We mention that although the proof is not a direct consequence of Theorem 2.1, we can follow the same line as the proof of Theorem 2.1.

Modification of the test set H. Take $c_1 > 0$ sufficiently large. Before providing the coverage error and the L^∞-diameter, we modify the test set

$$H = \{Y : \|X(\widehat{\beta}(Y) - \beta_0)\| \leqslant c_1 \sqrt{p \log n}\} \cap \{Y : \Pi_{\sigma^2}(|\sigma^2/\sigma_0^2 - 1| \geqslant \delta_1 \mid Y) \leqslant \delta_2\}$$

in Proposition 2.5 as

$$\begin{aligned} H := \{(X,Y) :& \|X(\widehat{\beta}(Y) - \beta_0)\| \leqslant c_1 \sqrt{p \log n}, (\underline{b}/2)^2 I_p \preceq X^\top X/n \preceq (2\overline{b})^2 I_p\} \\ &\cap \{(X,Y) : \Pi_{\sigma^2}(|\sigma^2/\sigma_0^2 - 1| \geqslant \delta_1 \mid Y) \leqslant \delta_2\}. \end{aligned}$$

The probability $\mathbb{P}((X,Y) \notin H)$ is evaluated as follows:

$$\mathbb{P}((X,Y) \notin H) \leqslant A_1 + A_2 + A_3 + \delta_3,$$

where

$$\begin{aligned} A_1 &:= \mathbb{P}(\|X(X^\top X)^{-1} X^\top \varepsilon\| \geqslant c_1 \sqrt{p \log n}/2,\ (\underline{b}/2)^2 I_p \preceq X^\top X/n \preceq (2\overline{b})^2 I_p), \\ A_2 &:= \mathbb{P}(\|X(X^\top X)^{-1} X^\top r\| \geqslant c_1 \sqrt{p \log n}/2), \\ A_3 &:= \mathbb{P}(X \notin \{X : (\underline{b}/2)^2 I_p \preceq (X^\top X)/n \preceq (2\overline{b})^2 I_p\}). \end{aligned}$$

It follows from Lemma A.4 that $A_1 \leqslant \widetilde{c}_1 \mathrm{e}^{-\widetilde{c}_2 p \log n}$ for some $\widetilde{c}_1, \widetilde{c}_2 > 0$. It follows from the Markov inequality that

$$A_2 \leqslant \frac{\mathbb{E}[r^\top X(X^\top X)^{-1} X^\top r]}{p \log n} \leqslant \frac{n}{\log n} \frac{\tau_2^2}{p}.$$

It follows from Lemma E.1 that $A_3 \leqslant \widetilde{c}_1 \mathrm{e}^{-\widetilde{c}_2 p \log n}$.

Upper bound for the coverage error. At the first step, we derive a high-probability bound of $\widehat{R}_\alpha$. From Proposition 2.5, for $(X,Y) \in H$,

$$\left|\Pi_\beta\{\|\widetilde{\psi}^p(\cdot)^\top(\widehat{\beta}-\beta_0)\|_\infty \leqslant \widehat{R}_\alpha \mid Y,X\} - \mathbb{P}(\|\widetilde{\psi}^p(\cdot)^\top(X^\top X)^{-1}X^\top N_{(n)}\|_\infty \leqslant \widehat{R}_\alpha \mid Y,X)\right| \leqslant \overline{\omega},$$

where $\overline{\omega} := \widetilde{\phi}_{\Pi_\beta}(c_1\sqrt{p\log n}) + c_1\delta_1 p\log n + \delta_2 + \delta_3 + c_1 \mathrm{e}^{-c_2 p\log n}$. Here the constant c_2 is determined in Proposition 2.5. Letting G be the distribution function of $\|\widetilde{\psi}^p(\cdot)^\top(X^\top X)^{-1}X^\top N_{(n)}\|_\infty$, G^{-1} its quantile function, we have

$$G^{-1}(1-\alpha-\overline{\omega}) \leqslant \widehat{R}_\alpha \leqslant G^{-1}(1-\alpha+\overline{\omega}) \text{ for } (X,Y)\in H. \tag{35}$$

Thus, we complete the first step.

At the second step, we derive approximation bounds of $\|\widetilde{\psi}^p(\cdot)^\top(X^\top X)^{-1}X^\top N_{(n)}\|_\infty$ and $\|\widetilde{\psi}^p(\cdot)^\top(\widehat{\beta}-\beta_0)\|_\infty$ by $\sqrt{n}\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2}N_{(p)}\|_\infty$ in Kolmogorov distance. Let $\eta=\eta_n$ be arbitrary divergent sequence. Let

$$\begin{aligned}
\rho_1 &:= \sup_{R>0}\left|\mathbb{P}(\|\widetilde{\psi}^p(\cdot)^\top\sqrt{n}(\widehat{\beta}-\beta_0)\|_\infty \leqslant R) - \mathbb{P}(\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2}N_{(p)}\|_\infty \leqslant R)\right|,\\
\rho_2 &:= \sup_{R>0}\left|\mathbb{P}(\|\widetilde{\psi}^p(\cdot)^\top\sqrt{n}(X^\top X)^{-1}X^\top N\|_\infty \leqslant R) - \mathbb{P}(\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2}N_{(p)}\|_\infty \leqslant R)\right|, \text{ and}\\
\gamma(R) &:= \sup_{x>0}\mathbb{P}(|\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2}N_{(p)}\|_\infty - x| \leqslant R).
\end{aligned}$$

and consider upper bounds of ρ_1, ρ_2, and $\gamma(R)$.

There exist $\widetilde{c}_3, \widetilde{c}_4 > 0$ for which we have four inequalities

$$\begin{aligned}
&\mathbb{P}\Bigg(\sqrt{n}\left|\left\|\widetilde{\psi}^p(\cdot)^\top(\widehat{\beta}-\beta_0)\right\|_\infty - \left\|\widetilde{\psi}^p(\cdot)^\top B^{-1}X^\top\varepsilon/n\right\|_\infty\right|\\
&\qquad \geqslant \widetilde{c}_3\eta\left\{\left(\xi_p^2/n\right)^{1/2}\sqrt{\log p}(n^{1/q}\sqrt{\log p}+\sqrt{p}\tau_\infty)+\sqrt{\log p}\tau_\infty\right\}\Bigg)\\
&\leqslant \widetilde{c}_4/\eta^2,
\end{aligned} \tag{36}$$

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{n}\left|\|\widetilde{\psi}^p(\cdot)^\top(X^\top X)^{-1}X^\top N_{(n)}\|_\infty - \|\widetilde{\psi}^p(\cdot)^\top B^{-1}X^\top N_{(n)}/n\|_\infty\right| \geqslant \widetilde{c}_3\eta\left(\xi_p^2/n\right)^{1/2} n^{1/q}\log p\right)\\
&\leqslant \widetilde{c}_4/\eta^2,
\end{aligned} \tag{37}$$

$$\begin{aligned}
&\mathbb{P}\Bigg(\sqrt{n}\left|\left\|\widetilde{\psi}^p(\cdot)^\top B^{-1}X^\top\varepsilon/n\right\|_\infty - \widetilde{Z}\right|\\
&\quad \geqslant \widetilde{c}_3\eta\left\{\left(\xi_p^2/n\right)^{1/2}(n^{1/q}\log n)+\left(\xi_p^2/n\right)^{1/4}(\log n)^{3/4}\right\}+\widetilde{c}_3\eta^{2/3}\left\{\left(\xi_p^2/n\right)^{1/6}(\log n)^{2/3}\right\}\Bigg)\\
&\leqslant \widetilde{c}_4(1/\eta^2+\log n/n),
\end{aligned} \tag{38}$$

and

$$\mathbb{P}\Bigg(\sqrt{n}\bigg|\Big\|\widetilde{\psi}^p(\cdot)^\top B^{-1}X^\top N_{(n)}/n\Big\|_\infty - \widetilde{Z}\bigg|$$
$$\geqslant \widetilde{c}_3\eta\left\{\left(\xi_p^2/n\right)^{1/2}\left(n^{1/q}\log n\right)+\left(\xi_p^2/n\right)^{1/4}(\log n)^{3/4}\right\}+\widetilde{c}_3\eta^{2/3}\left\{\left(\xi_p^2/n\right)^{1/6}(\log n)^{2/3}\right\}\Bigg)$$
$$\leqslant \widetilde{c}_4(1/\eta^2+\log n/n). \tag{39}$$

The first two inequalities follows from Lemma E.2 and the last two inequalities follows from Lemma E.3.

It follows from inequalities (36) and (38) and from Lemma E.5 that for some $\widetilde{c}_5>0$,

$$\rho_1 \leqslant \widetilde{c}_5(A_4+A_5), \tag{40}$$

where

$$A_4 := \frac{1}{\eta^2}+\frac{\log n}{n}+\eta(\log p)^{1/2}\max\left\{\left(\xi_p^2/n\right)^{1/2}n^{1/q}\log n, \left(\xi_p^2/n\right)^{1/6}(\log n)^{2/3}\right\} \text{ and}$$
$$A_5 := \eta(\log p)\tau_\infty \max\left\{1, \left(p\xi_p^2/n\right)^{1/2}\right\}. \tag{41}$$

Likewise, it follows from inequalities (37) and (39) and from Lemma E.5 that for some $\widetilde{c}_5>0$,

$$\rho_2 \leqslant \widetilde{c}_5 A_4. \tag{42}$$

It follows from Lemma E.5 that for some $\widetilde{c}_5>0$,

$$\gamma(R) \leqslant \widetilde{c}_5 R\sqrt{\log p}. \tag{43}$$

Thus, we complete the second step.

Finally, we have

$$\begin{aligned}
\mathbb{P}(f_0\in\mathcal{C}(\widehat{f},\widehat{R}_\alpha)) &\leqslant \mathbb{P}\{\|\widetilde{\psi}^p(\cdot)^\top(\widehat{\beta}-\beta_0)\|_\infty \leqslant G^{-1}(1-\alpha+\varpi)+\tau\}+\mathbb{P}\{(X,Y)\notin H\}\\
&\leqslant \mathbb{P}\{\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2}N_{(p)}/\sqrt{n}\|_\infty \leqslant G^{-1}(1-\alpha+\varpi)+\tau\}+\rho_1+\mathbb{P}\{(X,Y)\notin H\}\\
&\leqslant \mathbb{P}\{\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2}N_{(p)}/\sqrt{n}\|_\infty \leqslant G^{-1}(1-\alpha+\varpi)\}\\
&\qquad +\gamma(\sqrt{n}\tau)+\rho_1+\mathbb{P}\{(X,Y)\notin H\}\\
&\leqslant \varpi+\rho_1+\rho_2+\gamma(\sqrt{n}\tau)+\mathbb{P}((X,Y)\notin H),
\end{aligned}$$

and thus from (40)-(43), taking $\eta=n^\delta$, we obtain the upper bound of $\mathbb{P}(f_0\in\mathcal{C}(\widehat{f},\widehat{R}_\alpha))-(1-\alpha)$. Likewise, we obtain the lower bound of $\mathbb{P}(f_0\in\mathcal{C}(\widehat{f},\widehat{R}_\alpha))-(1-\alpha)$, which provides the desired bound of the coverage error.

Estimate of the L^∞-diameter. We will show that $G^{-1}(1-\alpha+\varpi)\leqslant \widetilde{c}_6\sqrt{(\log p)/n}$ for some $\widetilde{c}_6>0$. It follows from the concentration inequality for the suprema of the Gaussian process, from Lemma E.4, by taking sufficiently large p depending only on α, that for sufficiently large $\widetilde{c}_7>0$,

$$\mathbb{P}(\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2}N_{(p)}\|_\infty - \mathbb{E}[\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2}N_{(p)}\|_\infty] \geqslant \widetilde{c}_7\sqrt{\log p}) \leqslant \mathrm{e}^{-\widetilde{c}_7^2\log p} < \alpha-\varpi-\rho_2.$$

Thus, observing

$$
\begin{aligned}
G^{-1}(1-\alpha+\varpi) &:= \inf\{R : \mathbb{P}(\|\widetilde{\psi}^p(\cdot)^\top (X^\top X)^{-1} X^\top N_{(n)}\|_\infty \geqslant R) \leqslant \alpha - \varpi\} \\
&\leqslant \inf\{R : \mathbb{P}(\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2} N_{(p)}/\sqrt{n}\|_\infty \geqslant R) \leqslant \alpha - \varpi - \rho_2\} \\
&= \inf\Bigg\{R : \mathbb{P}\Bigg(\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2} N_{(p)}\|_\infty/\sqrt{n} - \mathbb{E}[\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2} N_{(p)}/\sqrt{n}\|_\infty] \\
&\qquad\qquad \geqslant R - \mathbb{E}[\|\widetilde{\psi}^p(\cdot)^\top B^{-1/2} N_{(p)}/\sqrt{n}\|_\infty]\Bigg) \leqslant \alpha - \varpi - \rho_2\Bigg\}
\end{aligned}
$$

and taking sufficiently large p depending only on α, we have $G^{-1}(1-\alpha+\varpi) \lesssim \sqrt{(\log p)/n}$.

□

E.3. **Proof of the bound on τ.** In this subsection, we show that $\tau \lesssim \tau_\infty/\sqrt{p}$ for periodic $S \geqslant 2$-regular wavelets. We consider a wavelet pair (ϕ, ψ) satisfying the following three assumptions:

- There exists an integer N for which the support of ϕ is included in $[0, N]$ and the support of ψ is included in $[-N+1, N]$;
- ϕ and ψ are C^S;
- The inequality $\inf_{x\in\mathbb{R}} \sum_{k\in\mathbb{Z}} \{\psi(x-k)\}^2 > 0$ holds.

We periodize the pair (ϕ, ψ) as follows:

$$
\phi^{(\mathrm{per})}_{l,k}(t) := \sum_{m\in\mathbb{Z}} 2^{l/2}\phi(2^l t + 2^l m - k) \quad\text{and}\quad \psi^{(\mathrm{per})}_{l,k}(t) := \sum_{m\in\mathbb{Z}} 2^{l/2}\psi(2^l t + 2^l m - k)
$$

for $k = 0, \dots, 2^l - 1$ and $l = 1, \dots, J$. Taking J_0 as $2^{J_0} \geqslant 2N$, $\{\phi^{(\mathrm{per})}_{J_0,k} : k = 0, \dots, 2^{J_0} - 1\} \cup \{\psi^{(\mathrm{per})}_{l,k} : k = 0, \dots, 2^l - 1, l = J_0, \dots, J\}$ forms $p = 2^J$ basis functions based on periodic S-regular wavelets.

It suffices to show that $\inf_{t\in[0,1]} \|\psi^p(t)\| \gtrsim \sqrt{p}$. Since $2^J > 2N$ and since the support of ψ is included in $[-N+1, N]$, we have

$$
\|\psi^p(t)\|^2 \geqslant 2^J \sum_{k=0}^{2^J-1} \{\sum_{m\in\mathbb{Z}} \psi(2^J t + 2^J m - k)\}^2 = 2^J \sum_{k=0}^{2^J-1} \sum_{m\in\mathbb{Z}} \{\psi(2^J t + 2^J m - k)\}^2
$$

and the rightmost quantity in the above inequality is bounded below by $2^J \inf_{x\in\mathbb{R}} \sum_{k\in\mathbb{Z}} \{\psi(x-k)\}^2$. Thus we complete the proof.

References

[1] F. Abramovich and B. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85:115–129, 1998.

[2] T. Anderson. The integral of a symmetric unimodal function. *Proc. Amer. Math. Soc.*, 6:170–176, 1955.

[3] Y. Atchadé. On the contraction properties of some high-dimensional quisi-posterior distributions. *Ann. Statist.*, 45:2248–2273, 2017.

[4] Y. Baraud. Model selection for regression on a fixed design. *Probab. Theory Relat. Fields*, 117:467–493, 2000.

[5] A. Belloni, V. Chernozhukov, D. Chetverikov, and K. Kato. Some new asymptotic theory for least squares series: Pointwise and uniform results. *J. Econometrics*, 186:345–366, 2015.

[6] P. Bickel and M. Rosenblatt. On some global measures of the derivations of density function estimates. *Ann. Statist.*, 1:1071–1095, 1973.

[7] D. Bontemps. Bernstein–Von Mises theorems for Gaussian regression with increasing number of reguressors. *Ann. Statist.*, 39:2557–2584, 2011.

[8] S. Boucheron and E. Gassiat. A Bernstein–von mises theorem for discrete probability distributions. *Electron. J. Statist.*, 3:114–148, 2009.

[9] I. Castillo and R. Nickl. Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.*, 41:1999–2028, 2013.

[10] I. Castillo and R. Nickl. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.*, 42:1941–1969, 2014.

[11] I. Castillo and J. Rousseau. A Bernstein–von mises theorem for smooth functionals in semiparametric models. *Ann. Statist.*, 43:2353–2383, 2015.

[12] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.*, 41:2786–2819, 2013.

[13] V. Chernozhukov, D. Chetverikov, and K. Kato. Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.*, 42:1787–1818, 2014.

[14] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximation of suprema of empirical processes. *Ann. Statist.*, 42:1564–1597, 2014.

[15] V. Chernozhukov, D. Chetverikov, and K. Kato. Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields*, 162:47–70, 2015.

[16] V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. *Ann. Probab.*, 45:2309–2352, 2017.

[17] V. Chernozhukov and H. Hong. An MCMC approach to classical estimation. *J. Econometrics*, 115:293–346, 2003.

[18] G. Claeskens and I. Van Keilegom. Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.*, 31:1852–1884, 2003.

[19] R. DeVore and G. Lorentz. *Constructive Approximation.* Springer, 1993.

[20] D. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied Comput. Harmonic Analysis*, 2:101–126, 1995.

[21] J.-P. Florens and A. Simoni. Nonparametric estimation of an instrumental regression: A quasi-Baysian approach based on regularized posterior. *J. Econometrics*, 170:458–475, 2012.

[22] D. Freedman. On the Bernstein–von Mises theorem with infinite dimensional parameters. *Ann. Statist.*, 27:1119–1140, 1999.

[23] C. Gao and H. Zhou. Bernstein–von Mises theorems for functionals of the covariance matrix. *Electron. J. Statist.*, 10:1751–1806, 2016.

[24] S. Ghosal. Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli*, 5:315–331, 1999.

[25] S. Ghosal. Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *J. Multivariate Anal.*, 74:49–68, 2000.

[26] E. Giné and R. Nickl. Confidence bands in density estimation. *Ann. Statist.*, 38:1122–1170, 2010.

[27] E. Gine and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models.* Cambridge Univ. Press, 2016.

[28] P. Hall. On convergence rates of suprema. *Prob. Theory Related Fields*, 89:447–455, 1991.

[29] D. Hanson and E. Watson. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.*, 42:1079–1083, 1971.

[30] I. Johnstone. *Gaussian estimation: Sequence and wavelet models.* Unpublished manuscript.

[31] I. Johnstone. High dimensional Bernstein–von Mises: simple examples. *IMS Collections*, 6:87–98, 2010.

[32] K. Kato. Quasi-Bayesian analysis of nonparametric instrumental variables models. *Ann. Statist.*, 41:2359–2390, 2013.

[33] H. Leahu. On the Bernstein–von Mises phenomenon in the Gaussian white noise model. *Elec. J. Statist.*, 5:373–404, 2011.

[34] Y. Liao and W. Jiang. Posterior consistency of nonparametric conditional moment restricted models. *Ann. Statist.*, 39:3003–3031, 2011.

[35] F. Nazarov. On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a gaussian measure. In *Geometric Aspects of Functional Analysis (2001-2002)*, pages 169–187. Springer, 2003.

[36] F. Olver, D. Lozier, R. Boisvert, and C. Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

[37] M. Panov and V. Spokoiny. Finite sample Bernstein–von Mises theorem for semiparametric problems. *Bayesian Analysis*, 10:665–710, 2015.

[38] V. Petrov. *Limit theorems of probability theory. Sequences of independent random variables*. Oxford University Press, 1995.

[39] K. Ray. Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.*, 45:2511–2536, 2017.

[40] V. Rivoirard and J. Rousseau. Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.*, 40:1489–1523, 2012.

[41] M. Rudelson and Roman Vershynin. Hanson–Wright inequality and sub-gaussian concentration. *Electron. Commun. Probab.*, 18:1–9, 2013.

[42] N. Smirnov. On the construction of confidence regions for the density of distribution of random variables. *Doklady Akad. Nauk SSSR*, 74:189–191, 1950.

[43] V. Spokoiny. Bernstein–von Mises theorem for growing parameter dimension. arXiv:1302.3430.

[44] J. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12:389–434, 2012.

[45] A. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

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

[47] B. von Bahr and C. Esseen. Inequalities for the rth absolute moment of a sum of random variables $1 \leqslant r \leqslant 2$. *Ann. Math. Statist.*, 36:299–303, 1965.

[48] Y. Yang, A. Bhattacharya, and D. Pati. Frequentist coverage and sup-norm convergence rate in Gaussian process regression. arXiv:1708.04753.

[49] W. Yoo and S. Ghosal. Supremum norm posterior contraction and credible set for nonparametric multivariate regression. *Ann. Statist.*, 44:1069–1102, 2016.

(K. Yano) Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

E-mail address: `yano@mist.i.u-tokyo.ac.jp`

(K. Kato) Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

E-mail address: `kkato@e.u-tokyo.ac.jp`