EFFICIENT GAUSSIAN PROCESS MODELING USING EXPERIMENTAL DESIGN-BASED SUBAGGING

Yibo Zhao, Yasuo Amemiya and Ying Hung

Rutgers University, IBM T. J. Watson Research Center and Rutgers University

Supplementary Material

S1 Lemmas

Lemma 1. *LHD-based block bootstrap mean is unbiased, i.e.,*

$$\boldsymbol{E}^*_{N,\omega}(\bar{y}^*_N) = \bar{y}_n.$$

Proof of Lemma 1: Since the data points are equally distributed over all the blocks, we have $\boldsymbol{E}^*_{N,\omega}(\bar{y}^*_N) = m^{-d}\sum_{i_1,\ldots,i_d}\bar{y}_{i_1,\ldots,i_d} = \bar{y}_n.\square$

Lemma 2. *Let* $\bar{y}_{\boldsymbol{i}} = \frac{1}{\mathcal{B}_n(\boldsymbol{i})}\sum_{\boldsymbol{x}_s\in\mathcal{B}_n(\boldsymbol{i})} y_s$, $\forall\boldsymbol{i} = (i_1,\ldots,i_d)$. *Assuming (A.1), (A.2) and* $m = o(n^{1/d})$, *we have*

$$\frac{n}{m^{2d}}\sum_{i_1,\ldots,i_d}(\bar{y}_{i_1,\ldots,i_d} - \mu)^2 - \tau_n^2 \xrightarrow{\text{P}} 0,$$

where $\tau_n^2 = \frac{1}{n}\sum_{s,t=1}^n \boldsymbol{Cov}(Y_s(\boldsymbol{x}_s), Y_t(\boldsymbol{x}_t)).$

Proof of Lemma 2: Let $A_n = \frac{n}{m^{2d}} \sum_{i_1,\dots,i_d} (\bar{y}_{i_1,\dots,i_d} - \mu)^2$. We can show that $\boldsymbol{Cov}(A_n, A_n) = 0$ and $\boldsymbol{E}(A_n) = \tau_n^2$.

$$\begin{aligned} \boldsymbol{Cov}(A_n, A_n) &= \boldsymbol{Cov}(\frac{n}{m^{2d}} \sum_{i_1,\dots,i_d} (\bar{y}_{i_1,\dots,i_d} - \mu)^2, \frac{n}{m^{2d}} \sum_{i_1,\dots,i_d} (\bar{y}_{i_1,\dots,i_d} - \mu)^2) \\ &= \frac{1}{n^2} \sum_{\boldsymbol{i}} \sum_{\boldsymbol{x}_{s_1}, \boldsymbol{x}_{s_2}, \boldsymbol{x}_{t_1}, \boldsymbol{x}_{t_2} \in \mathcal{B}_n(\boldsymbol{i})} \boldsymbol{Cov}\{(y_{s_1} - \mu)(y_{s_2} - \mu), (y_{t_1} - \mu)(y_{t_2} - \mu)\} \\ &\quad + \frac{1}{n^2} \sum_{\boldsymbol{i} \neq \boldsymbol{j}} \sum_{\boldsymbol{x}_{s_1}, \boldsymbol{x}_{s_2} \in \mathcal{B}_n(\boldsymbol{i})} \sum_{\boldsymbol{x}_{t_1}, \boldsymbol{x}_{t_2} \in \mathcal{B}_n(\boldsymbol{j})} \boldsymbol{Cov}\{(y_{s_1} - \mu)(y_{s_2} - \mu), (y_{t_1} - \mu)(y_{t_2} - \mu)\} \end{aligned}$$

By expanding two terms above separately, we have $\boldsymbol{Cov}(A_n, A_n) = O(\frac{1}{n} + \frac{m^d}{n}) \to 0$ as $m = o(n^{1/d})$. In addition, we have

$$\boldsymbol{E}(A_n) - \tau_n^2 = \frac{1}{n} \sum_{\boldsymbol{i} \neq \boldsymbol{j}} \sum_{\boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}), \boldsymbol{x}_t \in \mathcal{B}_n(\boldsymbol{j})} \sigma^2 \psi(y(\boldsymbol{x}_s), y(\boldsymbol{x}_t)) = o(1)$$

Thus, $A_n - \tau_n^2 \xrightarrow{\text{P}} 0$. □

Lemma 3. *Assume (A.1)- (A.2), then*

$$n{\tau_N^*}^2/m^{d-1} - \tau_n^2 \xrightarrow{\text{P}} 0,$$

where ${\tau_N^*}^2 = \boldsymbol{Cov}^*_{N,\omega}(\bar{y}^*_N, \bar{y}^*_N)$.

Proof of Lemma 3: Based on the definition of $n{\tau_N^*}^2/m^{d-1}$, we have

$$n{\tau_N^*}^2/m^{d-1} = \frac{n}{m^d} \boldsymbol{Cov}^*_{N,\omega}(\bar{y}_{\boldsymbol{i}_1^*}, \bar{y}_{\boldsymbol{i}_1^*}) + 2\frac{n(m-1)}{m^d} \boldsymbol{Cov}^*_{N,\omega}(\bar{y}_{\boldsymbol{i}_1^*}, \bar{y}_{\boldsymbol{i}_2^*}).$$

For the first term on the right, we have

$$\frac{n}{m^d} \boldsymbol{Cov}^*_{N,\omega}(\bar{y}_{\boldsymbol{i}_1^*}, \bar{y}_{\boldsymbol{i}_1^*}) = \frac{n}{m^{2d}} \sum_{i_1,\dots,i_d} (\bar{y}_{i_1,\dots,i_d} - \mu)^2 - \frac{n}{m^d} (\bar{y}_n - \mu)^2 = A_n - B_n.$$

By Lemma 2, we have $A_n - \tau_n^2 \xrightarrow{\text{P}} 0$. For $B_n = \frac{n}{m^d}(\bar{y}_n - \mu)^2$, by the central limit theorem for $\bar{y}_n$, we have $B_n \xrightarrow{\text{P}} 0$. Next, it suffices to show that $\frac{n(m-1)}{m^d}\boldsymbol{Cov}^*_{N,\omega}(\bar{y}_{\boldsymbol{i}_1^*}, \bar{y}_{\boldsymbol{i}_2^*})$ converges to 0 in probability under P. The following double summation $\sum_{i_1,\ldots,j_d,j_1,\ldots,j_d}$ are taken over $\boldsymbol{i} = (i_1,\ldots,i_d)$ and $\boldsymbol{j} = (j_1,\ldots,j_d)$ such that $\mathcal{B}_n(\boldsymbol{i})$ and $\mathcal{B}_n(\boldsymbol{j})$ are not equal and are selected together.

$$\begin{aligned}
\frac{n(m-1)}{m^d}\boldsymbol{Cov}^*_{N,\omega}(\bar{y}_{\boldsymbol{i}_1^*}, \bar{y}_{\boldsymbol{i}_2^*}) &= \frac{n(m-1)}{m^{2d}}\frac{1}{m^d-1-d(m-1)}\sum_{\boldsymbol{i}\neq\boldsymbol{j}}(\bar{y}_{\boldsymbol{i}}-\mu)(\bar{y}_{\boldsymbol{j}}-\mu) \\
&\quad+\frac{n(m-1)}{m^d}[1-\frac{2m^d}{m\{m^d-1-d(m-1)\}}](\bar{y}_n-\mu)^2 \\
&= C_n + D_n.
\end{aligned}$$

Similar to A_n and B_n, we can show that $C_n \stackrel{\mathrm{P}}{\longrightarrow} 0$ and $D_n \stackrel{\mathrm{P}}{\longrightarrow} 0$. The result follows immediately. $\square$

Lemma 4. *Under (A.1)-(A.3), for each* $\boldsymbol{\phi} \in \Theta$,

$$\begin{aligned}
&\lim_{n\to\infty} P\Big[P^*_{N,\omega}\big(|N^{-1}\sum_{s=1}^{N} q_s^*(\cdot,\omega,\boldsymbol{\phi}) + N^{-1}r_N^*(\cdot,\omega,\boldsymbol{\phi}) \\
&-n^{-1}\sum_{s=1}^{n} q_s(\omega,\boldsymbol{\phi}) - n^{-1}r_n(\omega,\boldsymbol{\phi})| > \delta\big) > \xi\Big] = 0.
\end{aligned}$$

Proof of Lemma 4: Rewrite the bootstrapped likelihood function as $I_1 + I_2 + I_3$, where $I_1 = N^{-1}\sum_{s=1}^{N}\{q_s^*(\cdot,\omega,\boldsymbol{\phi}) - \boldsymbol{E}^*q_s^*(\cdot,\omega,\boldsymbol{\phi})\}$, $I_2 = \{N^{-1}\sum_{s=1}^{N}\boldsymbol{E}^*q_s^*(\cdot,\omega,\boldsymbol{\phi}) - n^{-1}\sum_{s=1}^{n}q_s(\omega,\boldsymbol{\phi})\}$, $I_3 = N^{-1}r_N^*(\cdot,\omega,\boldsymbol{\phi}) - n^{-1}r_n(\omega,\boldsymbol{\phi})$. By Lemma 3, $I_2 \equiv 0$. For I_3, it can be shown that $n^{-1}r_n(\omega,\boldsymbol{\phi}) \to 0$ in P and $N^{-1}r_N^*(\cdot,\omega,\boldsymbol{\phi}) \to 0$, prob-$P^*_{N,\omega}$ prob-P. For notation simplicity, we omit $\boldsymbol{\theta}$ in the following discussion. The expectation and variance of

$n^{-1}r_n(\omega,\boldsymbol{\phi})$ are:

$$
\begin{aligned}
&|\boldsymbol{E}\{n^{-1}r_n(\omega,\boldsymbol{\phi})\}| \\
\leq\ & \frac{1}{2n\sigma^2(1+g)}\lambda_{\max}(E_n)\lambda_{\max}(D_n^{-1}) + |\log\{1+\lambda_{\max}^n(E_n)|D_n^{-1}\}| \\
=\ & o(1)
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{Var}(n^{-1}r_n(\omega,\boldsymbol{\phi})) &\leq \frac{1}{4(1+g)^2\sigma^4n^2}\boldsymbol{Var}\{\sum_{i=1}^n(\sum_{j=1}^n u_{ij}\varepsilon_j)^2\} \\
&\leq \frac{c_n}{4(1+g)^2\sigma^4n^2}\sum_{i=1}^n\sum_{j=1}^n \boldsymbol{Var}(\varepsilon_j^2) = o(1)
\end{aligned}
$$

where ε_j is the i^{th} entry of $D_n^{-1}(\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta})$ and $\boldsymbol{u}_i = (u_{ij})$ is the i^{th} row of U_n; $c_n = \max_i\{\sum_{j=1}^n u_{ij}^2\}$.

In addition, as $\lambda_{\max}(E_N^*) \leq \lambda_{\max}(E_n)$ and $\lambda_{\max}({D_N^*}^{-1}) \leq \lambda_{\max}(D_n^{-1})$, we have

$$
\begin{aligned}
& \frac{1}{2\sigma^2(1+g^*)}(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\boldsymbol{\beta})^T {D_N^*}^{-1}E_N^*{D_N^*}^{-1}(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\boldsymbol{\beta}) \\
\leq\ & \frac{1}{2\sigma^2}\lambda_{\max}(E_n)\lambda_{\max}(D_n^{-1})\|\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\boldsymbol{\beta}\|_2^2.
\end{aligned}
$$

According to Lemma 6 below, we have $N^{-1}\|\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\boldsymbol{\beta}\|_2^2 - n^{-1}\|\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}\|_2^2 \to 0$ prob-$P_{N,\omega}^*$ prob-P. Similarly, we can bound $\log|I_N + {U_N^*}^T{D_N^*}^{-1}U_N^*|$. As $\lambda_{\max}(E_n) \to 0$, we have $\frac{1}{N}r_N^*(\cdot,\omega,\boldsymbol{\phi}) \to 0$, prob-$P_{N,\omega}^*$ prob-P.

So when n is sufficiently large, we only need to show that $\lim_{n\to\infty} P\big[P_{N,\omega}^*(|I_1| >$

$\delta) > \xi\big] = 0$. By Chebyshev's inequality,

$$P^*_{N,\omega}(|I_1| > \delta) \le \frac{1}{\delta^2}\boldsymbol{Var}^*_{N,\omega}(\bar{q}^*_N(\cdot,\omega,\boldsymbol{\phi})).$$

By Lemma 1, $r^{-1}\boldsymbol{Var}^*_{N,\omega}(\bar{q}^*_N(\cdot,\omega,\boldsymbol{\phi})) = O_p(1)$, together with the fact that $N = n/m^{d-1}$,

$$\begin{aligned} P\big[P^*_{N,\omega}(|I_1| > \delta) > \xi\big] &\le P\big[\frac{n}{m^{d-1}}\frac{1}{\delta^2}\boldsymbol{Var}^*_{N,\omega}(\bar{q}^*_N(\cdot,\omega,\boldsymbol{\phi})) > \xi\frac{n}{m^{d-1}}\big] \\ &= O(m^{2d-2}/n^2) \to 0. \end{aligned}$$

□

The next lemma further extends Lemma 4 to the uniform weak law of large numbers for the LHD-based block bootstrap likelihood functions.

Lemma 5. *(Uniform Weak Law of Large Numbers) Under (A.1)-(A.5),* $\forall$ $\delta, \xi > 0$,

$$\lim_{n\to\infty} P\Big[P^*_{N,\omega}(\sup_{\boldsymbol{\phi}\in\Theta}|N^{-1}\sum_{s=1}^{N} q^*_s(\cdot,\omega,\boldsymbol{\phi}) + N^{-1}r^*_N(\cdot,\omega,\boldsymbol{\phi}) \\ -n^{-1}\sum_{s=1}^{n} q_s(\omega,\boldsymbol{\phi}) - n^{-1}r_n(\omega,\boldsymbol{\phi})| > \delta) > \xi\Big] = 0.$$

Proof of Lemma 5: By Lemma 4, $|n^{-1}r_n(\omega,\boldsymbol{\phi}) - N^{-1}r^*_N(\cdot,\omega,\boldsymbol{\phi})|$ can be arbitrarily small as n is large enough uniformly over Θ. We only need to show that

$$\lim_{n\to\infty} P\big[P^*_{N,\omega}(\sup_{\boldsymbol{\phi}\in\Theta}|N^{-1}\sum_{s=1}^{N} q^*_s(\cdot,\omega,\boldsymbol{\phi}) - n^{-1}\sum_{s=1}^{n} q_s(\omega,\boldsymbol{\phi})| > \delta) > \xi\big] = 0.$$

Given $\epsilon > 0$ that will be selected later, let $\{\eta(\boldsymbol{\phi}_j,\epsilon), j = 1,\ldots,K\}$ be a finite cover of Θ, where $\eta(\boldsymbol{\phi}_i, \epsilon) = \{\boldsymbol{\phi} \in \Theta : |\boldsymbol{\phi} - \boldsymbol{\phi}_j| < \epsilon\}$. Then

$$\begin{aligned} & \sup_{\boldsymbol{\phi}} |N^{-1} \sum_{s=1}^{N} q_s^*(\cdot, \omega, \boldsymbol{\phi}) - n^{-1} \sum_{s=1}^{n} q_s(\omega, \boldsymbol{\phi})| \\ = \quad & \max_{j=1}^{K} \sup_{\boldsymbol{\phi} \in \eta(\boldsymbol{\phi}_j, \epsilon)} |\bar{q}_N^*(\cdot, \omega, \boldsymbol{\phi}) - \bar{q}_n(\omega, \boldsymbol{\phi})|. \end{aligned}$$

It follows that $\forall\ \delta > 0$ with fixed ω,

$$\begin{aligned} & P_{N,\omega}\big(\sup_{\boldsymbol{\phi} \in \Theta} |\bar{q}_N^*(\cdot, \omega, \boldsymbol{\phi}) - \bar{q}_n(\omega, \boldsymbol{\phi})| > \delta \big) \\ \leq \quad & \sum_{j=1}^{K} P_{N,\omega}\big(\sup_{\boldsymbol{\phi} \in \eta(\boldsymbol{\phi}_j, \epsilon)} |\bar{q}_N^*(\cdot, \omega, \boldsymbol{\phi}) - \bar{q}_n(\omega, \boldsymbol{\phi})| > \delta \big). \end{aligned}$$

For $\forall\ \boldsymbol{\phi} \in \eta(\boldsymbol{\phi}_j, \epsilon)$, by Global Lipschitz condition,

$$|\bar{q}_N^*(\cdot, \omega, \boldsymbol{\phi}) - \bar{q}_n(\omega, \boldsymbol{\phi})| \leq |\bar{q}_N^*(\cdot, \omega, \boldsymbol{\phi}_j) - \bar{q}_n(\omega, \boldsymbol{\phi}_i)| + N^{-1} \sum_{s=1}^{N} L_s^* \epsilon + n^{-1} \sum_{s=1}^{n} L_s \epsilon,$$

where L_s^* is the bootstrapped Lispchitz coefficient.

By Markov inequality and the fact that $\sup_n \{n^{-1} \sum_{s=1}^n \boldsymbol{E} L_s\} = O(1)$, we have $P(n^{-1} \sum_{s=1}^n L_s > \delta/3) \leq 3\epsilon\Delta/\delta \leq \xi/3$, where Δ is a large constant. If we choose $\epsilon < \xi\delta/(9\Delta)$, we have

$$\begin{aligned} & P\big[P_{N,\omega}^*(\sup_{\boldsymbol{\phi} \in \eta(\boldsymbol{\phi}_j, \epsilon)} |\bar{q}_N^*(\cdot, \omega, \boldsymbol{\phi}) - \bar{q}_n(\omega, \boldsymbol{\phi})| > \delta) > \xi\big] \\ \leq \quad & P\big[P_{N,\omega}^*(|\bar{q}_N^*(\cdot, \omega, \boldsymbol{\phi}_j) - \bar{q}_n(\omega, \boldsymbol{\phi}_j)| > \delta) > \xi/3\big] \\ & + P\big[P_{N,\omega}^*(N^{-1} \sum_{s=1}^{N} L_s^* \epsilon > \delta/3) > \xi/3\big] + P[n^{-1} \sum_{s=1}^{n} L_s \epsilon > \delta/3] \\ = \quad & I_1 + I_2 + I_3. \end{aligned}$$

According to Lemma 4, $I_1 \leq \xi/3$ when n is large enough. By Markov's inequality,

$$P^*_{N,\omega}(N^{-1}\sum_{s=1}^{N} L^*_s\epsilon > \delta/3) \leq N^{-1}\sum_{s=1}^{N} \boldsymbol{E}^* L^*_s/(\delta/3\epsilon) = n^{-1}\sum_{s=1}^{n} L_s/(\delta/3\epsilon).$$

The last equality is because of Lemma 1. Thus, $I_2 < \xi/3$ as well as I_3. □

S2 Consistency of the LHD-based block bootstrap mean

Before studying the asymptotic performance of MLEs, we first focus on understanding properties of the LHD-based block bootstrap mean, which is an important foundation to the theoretical development of $\hat{\boldsymbol{\phi}}^*_N$ later.

The LHD-based block bootstrap can be formulated mathematically as follows. Given the underlying probability space $(\Omega, \mathcal{F}, P)$ of a Gaussian process, a sample of size n with settings $\boldsymbol{x}_1(\omega), ..., \boldsymbol{x}_n(\omega)$ and responses $y(\boldsymbol{x})$'s are observed from a given realization $\omega \in \Omega$. Let $(\Lambda, \mathcal{G})$ be a measurable space on the realization. For each $\omega \in \Omega$, denote $P^*_{N,\omega}$ as the probability measure induced by the m-run LHD-based block bootstrap on $(\Lambda, \mathcal{G})$. The proposed bootstrap is a method to generate new dataset on $(\Lambda, \mathcal{G}, P^*_{N,\omega})$ conditional on the n original observations. Let $\tau_t : \Lambda \to \{1, ..., n\}$ denote a random index generated by the LHD-based block bootstrap. So, τ_t is the tth index in the intersect index of observations and $\{\mathcal{B}_n(\boldsymbol{i}_1^*), ..., \mathcal{B}_n(\boldsymbol{i}_m^*)\}$, where $(\boldsymbol{i}_1^*, ..., \boldsymbol{i}_m^*)$ is a randomly generated m-run LHD. Therefore, for $(\lambda, \omega) \in \Lambda \times \Omega$, we have the tth bootstrap sample: $\boldsymbol{x}_t^*(\lambda, \omega) \equiv \boldsymbol{x}_{\tau_t(\lambda)}(\omega)$.

Suppose $\{Y(\boldsymbol{x}_t), t \in R\}$ follows a GP with mean μ. Given n observations, the sample estimation of mean μ is

$$\bar{y}_n = \frac{1}{n}\sum_{s=1}^{n} y_s,$$

and the LHD-based block bootstrap mean with N samples is given by

$$\bar{y}_N^* = \frac{1}{N}\sum_{s=1}^{N} y_s^*.$$

With a slight abuse of notation, we replace the notation of random variable Y by its realization y unless otherwise specified. The following lemma shows the asymptotic consistency of the LHD-based block bootstrap mean.

Lemma 6. *Under (A.1)-(A.2), if $m \to \infty$ and $m = o(n^{1/d})$, then*

$$\sup_x |P_{N,\omega}^*(\sqrt{n/m^{d-1}}(\bar{y}_N^* - \bar{y}_n)/\tau_n \leq x) - P(\sqrt{n}(\bar{y}_n - \mu)/\tau_n \leq x)| \xrightarrow{\text{P}} 0,$$

when $n \longrightarrow \infty$.

Note that $\boldsymbol{E}(\cdot)$ and $\boldsymbol{Cov}(\cdot,\cdot)$ denote the expectation and variance under P while $\boldsymbol{E}_{N,\omega}^*(\cdot)$ and $\boldsymbol{Cov}_{N,\omega}^*(\cdot,\cdot)$ denote the expectation and variance under $P_{N,\omega}^*$.

Proof of Lemma 6: It suffices to show that (1) $\boldsymbol{E}_{N,\omega}^*(\bar{y}_N^*) = \bar{y}_n$; (2) $n{\tau_N^*}^2/m^{d-1} - \tau_n^2 \xrightarrow{\text{P}} 0$; and (3) $\sup_x |P_{N,\omega}^*((\bar{y}_N^* - \boldsymbol{E}_{N,\omega}^*(\bar{y}_N^*))/\tau_N^* \leq x) -$

$\Phi(x)| \xrightarrow{\text{P}} 0$, where $\Phi(\cdot)$ denotes standard normal distribution function and $\tau_N^{*\,2} = \boldsymbol{Cov}^*_{N,\omega}(\bar{y}^*_N, \bar{y}^*_N)$.

Lemmas 1 and 3 imply the results in (1) and (2). Note that $\bar{y}^*_N = \frac{1}{m}\sum_{j=1}^m \bar{y}_{\boldsymbol{i}^*_j}$ and $(\bar{y}_{\boldsymbol{i}^*_1}, \ldots, \bar{y}_{\boldsymbol{i}^*_m})$ follows Latin Hypercube sampling distribution. According to Loh (1996), we have the Berry-Essen type of bound for Latin Hypercube sampling

$$\sup_x |P^*_{N,\omega}((\bar{y}^*_N - \bar{y}_n)/\tau^*_N \le x) - \Phi(x)| \le c^* m^{-1/2},$$

where c^* is a constant that depends only on d, given $\boldsymbol{E}^*_{N,\omega}\|\bar{y}_{\boldsymbol{i}^*_1}\|^3 < \infty$. So we only need to show that $\boldsymbol{E}^*_{N,\omega}\|\bar{y}_{\boldsymbol{i}^*_1}\|^3$ is bounded uniformly in probability under P. Since $\boldsymbol{E}^*_{N,\omega}\|\bar{y}_{\boldsymbol{i}_1}\|^3 = \frac{1}{m^d}\sum_{\boldsymbol{i}} \bar{y}^3_{\boldsymbol{i}}$ and according to Minkowski's inequality, it follows that

$$\frac{1}{m^d}\sum_{\boldsymbol{i}} \boldsymbol{E}\{\bar{y}^3_{\boldsymbol{i}}\} \le \frac{1}{m^d}\sum_{\boldsymbol{i}} \frac{1}{|\mathcal{B}_n(\boldsymbol{i})|^3}\{\sum_{\boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i})} \boldsymbol{E}(y_s)\}^3 < \infty.$$

□

S3 Proof of Theorem 1

To investigate the asymptotic properties of the estimators from LHD-based block bootstrap, we decompose the likelihood function into blocks. For each block, denote $\boldsymbol{y}_{\boldsymbol{i}} = (y_s(\boldsymbol{x}_s), \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))$, $\boldsymbol{X}_{\boldsymbol{i}} = (\boldsymbol{x}_s, \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))^T$, $R_{\boldsymbol{i},\boldsymbol{j}}(\boldsymbol{\theta}) = \big[\psi(y(\boldsymbol{x}_s), y(\boldsymbol{x}_t); \boldsymbol{\theta}), \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}), \boldsymbol{x}_t \in \mathcal{B}_n(\boldsymbol{j})\big]$ and $\boldsymbol{z}_{\boldsymbol{i}} = R^{-1/2}_{\boldsymbol{i},\boldsymbol{i}}(\boldsymbol{\theta})(\boldsymbol{y}_{\boldsymbol{i}} -$

$\boldsymbol{X}_{\boldsymbol{i}}\boldsymbol{\beta})$. Then, we can rewrite the penalized log-likelihood function $n^{-1}\ell(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi})$ as

$$\begin{aligned} Q_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) = & -(2n\sigma^2)^{-1} \textstyle\sum_{s=1}^n z_s^2 - (2n)^{-1} \sum_{s=1}^n \log(\lambda_s) \\ & -(2n)^{-1} \textstyle\sum_{s=1}^n \log(\sigma^2) + n^{-1} r_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) \\ & - \textstyle\sum_{s=1}^p p_\lambda(|\beta_s|) \\ = & \; n^{-1} \textstyle\sum_{s=1}^n q_s(\omega, \boldsymbol{\phi}) + n^{-1} r_n(\omega, \boldsymbol{\phi}) - \sum_{s=1}^p p_\lambda(|\beta_s|) \end{aligned} \tag{S3.1}$$

where $\{\lambda_s, s = 1, \ldots, n\} = \{\text{eigenvalues of } |R_{\boldsymbol{i},\boldsymbol{i}}(\boldsymbol{\theta})|, \boldsymbol{i} = (i_1, \ldots, i_d)\}$ with $(i_1, \ldots, i_d)$ in lexicographical order and eigenvalues from the largest to the smallest. Note that $r_n(\omega, \boldsymbol{\phi}) = \ell(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) - \sum_{s=1}^n q_s(z_s, \boldsymbol{\phi})$ contains all terms involving the off block-diagonal terms. Define $D_n(\boldsymbol{\theta}) = \text{diag}(R_{\boldsymbol{i},\boldsymbol{i}}(\boldsymbol{\theta}))$ and $E_n(\boldsymbol{\theta}) = R_n(\boldsymbol{\theta}) - D_n(\boldsymbol{\theta})$. Assuming that $E_n(\boldsymbol{\theta}) = U_n(\boldsymbol{\theta}) U_n^T(\boldsymbol{\theta})$, we have

$$\begin{aligned} r_n(\omega, \boldsymbol{\phi}) \;=\; & \frac{1}{2\sigma^2(1+g)} (\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta})^T D_n^{-1}(\boldsymbol{\theta}) E_n(\boldsymbol{\theta}) D_n^{-1}(\boldsymbol{\theta}) (\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}) \\ & + \frac{1}{2} \log |I_n + U_n^T(\boldsymbol{\theta}) D_n^{-1}(\boldsymbol{\theta}) U_n(\boldsymbol{\theta})|, \end{aligned}$$

where $g = \text{trace}(E_n(\boldsymbol{\theta}) D_n^{-1}(\boldsymbol{\theta}))$.

The MLE is obtained by $\hat{\boldsymbol{\phi}}_n = \arg\max_{\boldsymbol{\phi}} Q_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi})$. Analogue to the decomposition for $Q_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi})$, the log-likelihood function for LHD-based block bootstrap samples can be written as

$$
\begin{aligned}
Q^*_N(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \boldsymbol{\phi}) &= N^{-1}\sum_{s=1}^{N} q^*_s(\cdot, \omega, \boldsymbol{\phi}) + N^{-1} r^*_N(\cdot, \omega, \boldsymbol{\phi}) \\
&\quad - \sum_{s=1}^{p} p_\lambda(|\beta_s|) \qquad \text{(S3.2)}
\end{aligned}
$$

where $r^*_N(\cdot, \omega, \boldsymbol{\phi})$ contains all terms involving the off block-diagonal terms with bootstrapped samples. Specifically,

$$
\begin{aligned}
& r^*_N(\cdot, \omega, \boldsymbol{\phi}) \\
= \; & \frac{1}{2\sigma^2(1+g^*)}(\boldsymbol{y}^*_N - \boldsymbol{X}^*_N\boldsymbol{\beta})^T {D^*_N}^{-1}(\boldsymbol{\theta}) E^*_N(\boldsymbol{\theta}) {D^*_N}^{-1}(\boldsymbol{\theta})(\boldsymbol{y}^*_N - \boldsymbol{X}^*_N\boldsymbol{\beta}) \\
& + \frac{1}{2}\log|I_N + {U^*_N}^T(\boldsymbol{\theta}){D^*_N}^{-1}(\boldsymbol{\theta})U^*_N(\boldsymbol{\theta})|,
\end{aligned}
$$

where $D^*_N(\boldsymbol{\theta}) = \mathrm{diag}(R_{\boldsymbol{i}^*_j, \boldsymbol{i}^*_j}(\boldsymbol{\theta}), j = 1, \ldots, m)$ and $E^*_N(\boldsymbol{\theta}) = R^*_N(\boldsymbol{\theta}) - D^*_N(\boldsymbol{\theta})$ with $E^*_N(\boldsymbol{\theta}) = U^*_N(\boldsymbol{\theta}){U^*_N}^T(\boldsymbol{\theta})$; $g^* = \mathrm{trace}(E^*_N(\boldsymbol{\theta}){D^*_N}^{-1}(\boldsymbol{\theta}))$. The bootstrapped version of $\hat{\boldsymbol{\phi}}_n$ is $\hat{\boldsymbol{\phi}}^*_N = \arg\max_{\boldsymbol{\phi}} Q^*_N(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \boldsymbol{\phi})$. Theoretical properties of the LHD-based block bootstrap likelihood function (S3.2) are established in lemmas 4 and 5, which leads to a proof of convergence properties of the bootstrap estimator $\hat{\boldsymbol{\phi}}^*_N$. Lemma 4 first established the pointwise weak law of large numbers for the LHD-based block bootstrap likelihood functions. Lemma 5 further extends Lemma 4 to the uniform weak law of large numbers for the LHD-based block bootstrap likelihood functions.

Proof of Theorem 1: Based on Lemma 5, we have

$$\lim_{n\to\infty} P[P^*_{N,w}(\sup_{\boldsymbol{\phi}\in\Theta} |Q_n - Q^*_N| > \delta) > \xi] = 0,$$

where Q_n and Q^*_N are given in (S3.1) and (S3.2). With the full preparation of the likelihood convergence developed in Lemmas 4 and 5, the convergence of bootstrap parameter estimation follows immediately given the existence of $\hat{\boldsymbol{\phi}}_n$ and $\hat{\boldsymbol{\phi}}^*_N$.

Denote $\bar{q}^*_N(\cdot,\omega,\boldsymbol{\phi}) = N^{-1}\sum_{i=1}^N q^*_i(\cdot,\omega,\boldsymbol{\phi})$ and $\bar{q}_n(\omega,\boldsymbol{\phi}) = n^{-1}\sum_{i=1}^n q_i(\omega,\boldsymbol{\phi})$. By (A.6), $q^*_s(\cdot,\omega,\cdot) : \Lambda\times\Theta \to R$ and $r^*_N(\cdot,\omega,\cdot) : \Lambda\times\Theta \to R$ are measurable-$\mathcal{G}$ for each $\boldsymbol{\phi}\in\Theta$. In addition, $q^*_s(\lambda,\omega,\cdot)$ and $r^*_N(\lambda,\omega,\cdot)$ are continuous on Θ for all λ. Thus, we have $\hat{\boldsymbol{\phi}}^*_N(\cdot,\omega)$ exists as a measurable-$\mathcal{G}$ function by Jennrich (1969).

Following the procedure in Goncalves and White (2004), for any subsequence $\{n'\}$, given that $\hat{\boldsymbol{\phi}}_{n'}$ is identifiable and unique, there exists a further subsequence $\{n''\}$ such that $\hat{\boldsymbol{\phi}}_{n''}$ is identifiably unique with respect to $\{Q_{n''}\}$ for all $\omega\in F$ in some $F\in\mathcal{F}$ with $P(F) = 1$. By condition (A.6), there exists $G\in\mathcal{F}$ with $P(G) = 1$ such that for all $\omega\in G$, $\{Q^*_{N''}(\cdot,\omega,\boldsymbol{\phi})\}$ (N'' is corresponding bootstrapped sample size of n'') is a sequence of random function on $(\Lambda,\mathcal{G},P^*_{N,\omega})$ continuous on Θ for all $\lambda\in\Lambda$. Hence, by White (1996), for fixed $\omega\in G$, there exists $\hat{\boldsymbol{\phi}}^*_{N''}(\cdot,\omega) : \Lambda\to\Theta$ measurable-$\mathcal{G}$ and $\hat{\boldsymbol{\phi}}^*_{N''}(\cdot,\omega) = \arg\max_{\boldsymbol{\phi}} Q^*_{N''}(\cdot,\omega,\boldsymbol{\phi})$. By the uniform weak law of large numbers for $Q_N^*(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \boldsymbol{\phi})$ obtained from Lemma 5, we have $Q_{N''}^*(\cdot, \omega, \boldsymbol{\phi}) - Q_{n''}(\omega, \boldsymbol{\phi}) \to 0$ as $n'' \to \infty$ $prob - P_{N,\omega}^*$ $prob - P$ uniformly on Θ, where we write $\hat{Q}_N^* \to 0$ $prob - P_{N,\omega}^*, prob - P$ if, for any $\epsilon > 0$ and $\delta > 0$, $\lim_{n\to\infty} P\{P_{N,\omega}^*(|\hat{Q}_N^* > \epsilon| > \delta)\} = 0$ and omit $prob - P_{N,\omega}^*, prob - P$ in the text for notation simplicity. Hence, there exists a further subsequence $\{n'''\}$ such that $Q_{N'''}^*(\cdot, \omega, \boldsymbol{\phi}) - Q_{n'''}(\omega, \boldsymbol{\phi}) \to 0$ as $n'' \to \infty$ $prob - P_{N,\omega}^*$ $prob - P$ for all ω in some $H \in \mathcal{F}$ with $P(H) = 1$. Choose $\omega \in F \cap G \cap H$, by White (1996), we have $\hat{\boldsymbol{\phi}}_{N'''}^* - \hat{\boldsymbol{\phi}}_{n'''} \to 0$ as $n''' \to \infty$ $prob - P_{N,\omega}^*$ $prob - P$. Since this is true for any subsequence $\{n'\}$, we have $P(F \cap G \cap H) = 1$. Thus, $\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n \to 0$ $prob - P_{N,\omega}^*, prob - P$. Then $\hat{\boldsymbol{\phi}}_N = \frac{1}{K}\sum_{i=1}^K \hat{\boldsymbol{\phi}}_N^*(i) - \hat{\boldsymbol{\phi}}_n \to 0$ $prob - P_{N,\omega}^*, prob - P$. □

S4 Proof of Theorem 2

Proof. Define $B = Var\{n^{-1/2}\sum_{s=1}^n \nabla q_s(\cdot, \omega, \boldsymbol{\phi}_0)\}$. We first show that $\sqrt{n/m^{d-1}} B^{-1/2} \nabla Q_N^*(\cdot, \omega, \hat{\boldsymbol{\phi}}_n) \to N(0, I)$. Denote $\bar{h}_N^*(\boldsymbol{\phi}) = N^{-1}\sum_{s=1}^N \nabla q_s^*(z_s^*, \boldsymbol{\phi})$

and $\bar{h}_n(\boldsymbol{\phi}) = n^{-1}\sum_{s=1}^{n} \nabla q_s(z_s, \boldsymbol{\phi})$. We have

$$\begin{aligned}
\sqrt{n/m^{d-1}}[\bar{h}_N^*(\hat{\boldsymbol{\phi}}_n) - \bar{h}_n(\hat{\boldsymbol{\phi}}_n)] &= \sqrt{n/m^{d-1}}[\bar{h}_N^*(\hat{\boldsymbol{\phi}}_n) - \bar{h}_N^*(\boldsymbol{\phi}^0)] \\
&\quad +\sqrt{n/m^{d-1}}[\bar{h}_N^*(\boldsymbol{\phi}^0) - \bar{h}_n(\boldsymbol{\phi}^0)] \\
&\quad +\sqrt{n/m^{d-1}}[\bar{h}_n(\boldsymbol{\phi}^0) - \bar{h}_n(\hat{\boldsymbol{\phi}}_n)] \\
&= J_1 + J_2 + J_3.
\end{aligned}$$

Since $\bar{h}_n$ and $\bar{h}_N^*$ are functions whose secondary derivative are continuous, $J_1 + J_3 \to 0$ as $\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0 \to 0$ by Theorem 3.1 in Chu (2011). Moreover, the two terms in J_2 are both evaluated at $\boldsymbol{\phi}_0$ which is a fixed value, then by Lemma 6, we have $B^{-1/2}J_2 \to N(0, I)$.

By condition (A.10) and follow a similar proof as Lemma 5, we have

$$\nabla^2 Q_N^*(\cdot, \omega, \boldsymbol{\phi}) - \nabla^2 Q_n(\omega, \boldsymbol{\phi}) \to 0 \ \ prob - P_{N,\omega}^*, prob - P.$$

Let $\hat{H}_n(\omega) = \nabla^2 Q_n(\omega, \hat{\boldsymbol{\phi}}_n)$. According to White (1996), given the result $\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n \to 0 \ prob - P_{N,\omega}^*, prob - P$ and assumption (A.8), we have

$$\begin{aligned}
\sqrt{N}(\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n) &= -\hat{H}_n^{-1}(\omega)\sqrt{N}\nabla Q_N^*(\cdot, \omega, \hat{\boldsymbol{\phi}}_n) + o_{P_{N,\omega}^*}(1) \\
&= -H_n(\boldsymbol{\phi}_0)^{-1}(\omega)\sqrt{N}\nabla Q_N^*(\cdot\omega, \hat{\boldsymbol{\phi}}_n) + o_{P_{N,\omega}^*}(1).
\end{aligned}$$

Given the fact that

$$\sqrt{n/m^{d-1}}B^{-1/2}\nabla Q_N^*(\cdot, \omega, \hat{\boldsymbol{\phi}}_n) \to N(0, I) \ \ prob - P_{N,\omega}^*, prob - P.$$

we have

$$B^{-1/2}H_n(\boldsymbol{\phi}_0)\sqrt{N}(\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n) \to N(0, I).$$

For $\boldsymbol{\beta}_{10}$, B and H can be written as $\mathbf{J}(\boldsymbol{\beta}_{10})$ and $\mathbf{J}(\boldsymbol{\beta}_{10}) + \mathbf{G}(\boldsymbol{\beta}_{10})$. For $\hat{\boldsymbol{\beta}}_{N,1}^*$, we have

$$\sqrt{N}[\mathbf{J}(\boldsymbol{\beta}_{10}) + \mathbf{G}(\boldsymbol{\beta}_{10})]\{\hat{\boldsymbol{\beta}}_{N,1}^* - \hat{\boldsymbol{\beta}}_{n,1}\} \to N(0, J(\boldsymbol{\beta}_{10})).$$

For sub-bagging estimator $\hat{\boldsymbol{\beta}}_{N,1} = \sum_{i=1}^K \hat{\boldsymbol{\beta}}_{N,1}^*(i)$, we have

$$\sqrt{KN}[\mathbf{J}(\boldsymbol{\beta}_{10}) + \mathbf{G}(\boldsymbol{\beta}_{10})]\{\hat{\boldsymbol{\beta}}_{N,1} - \hat{\boldsymbol{\beta}}_{n,1}\} \to N(0, J(\boldsymbol{\beta}_{10})),$$

then the result follows.

S5 Proof of Theorem 3

Using the same technique before, we decompose the log-likelihood by blocks and rewrite the likelihood of $\boldsymbol{\beta}$ based on the OSE approach as follows:

$$\begin{aligned} Q_n(\boldsymbol{\beta}) &= n^{-1}\sum_{s=1}^{n} q_s(\omega, \boldsymbol{\beta}, \hat{\boldsymbol{\theta}}_n^{(0)}, \hat{\sigma_n^2}^{(0)}) + n^{-1} r_n(\omega, \boldsymbol{\beta}, \hat{\boldsymbol{\theta}}_n^{(0)} \hat{\sigma_n^2}^{(0)}) \\ &\quad - \sum_{j=1}^{p} p'_\lambda(|\hat{\beta}_j^{(0)}|)|\beta_j|. \end{aligned}$$

The likelihood based on subsampled data can be written as:

$$
\begin{aligned}
Q_N^*(\boldsymbol{\beta}) &= N^{-1}\sum_{s=1}^{N} q_s^*(\omega, \boldsymbol{\beta}, \hat{\boldsymbol{\theta}}_N^{*(0)}, \hat{\sigma_N^2}^{*(0)}) + N^{-1} r_N^*(\omega, \boldsymbol{\beta}, \hat{\boldsymbol{\theta}}_N^{*(0)}, \hat{\sigma_N^2}^{*(0)}) \\
&\quad - \sum_{j=1}^{p} p'_\lambda(\hat{\beta}_j^{*(0)}|)|\beta_j|.
\end{aligned}
$$

By the fact that $\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n \to 0$ and the results in Lemma 2, Lemma 3 and Lemma 6 still hold, we have $\hat{\boldsymbol{\phi}}_{N,OSE}^* - \hat{\boldsymbol{\phi}}_{n,OSE} \to 0$. Then follows the same technique in the proof of Theorem 2, the result follows. □