Wild Bootstrap for Mean Response Inference in Functional Linear Regression Models

Hyemin Yeon[1], Xiongtao Dai[2], and Daniel Nordman[3]

[1] *George Mason University,* [2] *University of California, Berkeley,* [3] *Iowa State University*

Supplementary Material

Abstract: In this supplement, we provide the technical details behind the wild bootstrap consistency in functional linear regression models as provided in Section 3 of the main paper (Yeon et al., 2025). After introducing some preliminary discussions in Section S1, the proof of Theorem 1 is comprehensively studied in Section S2. Specifically, Section S2.1 devotes to reviewing perturbation theory while Section S2.2 establishes the consistency of the wild bootstrap estimator. Sections S2.3-S2.5 study the convergences related to the wild bootstrap consistency in Theorem 1, which are combined for the main proof in Section S2.6. Additional simulation results and extra figures used in real data analysis are also given in Sections S3-S4, respectively.

Key words and phrases: functional principal components analysis, heteroscedasticity, infinite dimensionality, resampling, scalar-on-function regression, stabilized volatility method

S1 Preliminaries

We begin by providing some preliminary results. First, in Section S1.1, we introduce convergence concepts in conditional probability and conditional distribution; this helps deriving wild bootstrap consistency. Next, theoretical conditions are discussed in Section S1.2, while main differences from the paired bootstrap results by Yeon et al. (2024a) are given in Section S1.3. Simplified conditions under polynomial decay rates are provided in Section S1.4. In what follows, for brevity, we may not specifically add the citation of the main paper as Yeon et al. (2025).

S1.1 Convergences in conditional probability

When establishing bootstrap theory, we need to handle convergences in conditional probability distribution with respect to $\mathsf{P}(\cdot|\mathcal{X}_n, \mathcal{Y}_n, X_0)$ or $\mathsf{P}(\cdot|\mathcal{X}_n, \mathcal{Y}_n)$. Yeon et al. (2024a) develop a version of such conditional probability theory for the ease of the proofs in their supplement Yeon et al. (2024b); computation properties between these convergent terms in conditional probability hold in a way similar to those in original probability. We will follow the framework therein.

To set up the convergence notions in conditional probability, let $\{A_n^*\}$ denote a sequence of random variables, where A_n^* follows conditional probability $\mathsf{P}^* \equiv \mathsf{P}(\cdot|\mathcal{F}_n)$ depending on the sub σ-field $\mathcal{F}_n$ for each $n \in \mathbb{N}$, and $\{a_n\}$ be a sequence of positive real numbers. Next, let $\hat{F}_n$ be the conditional distribution function of A_n^* defined as

$\hat{F}_n(x) \equiv \mathsf{P}^*(A_n^* \leq x)$ for $x \in \mathbb{R}$. Also, A represents a (real-world) random variable with distribution function F, i.e., $F(x) = \mathsf{P}(A \leq x)$ for $x \in \mathbb{R}$, where F does not depend on any sub σ-field $\mathcal{F}_n$. Finally, C_F denotes the set of all continuity points of F.

Definition S1 (Big O and little o in conditional distribution)**.**

(a) We say that $\{A_n^*/a_n\}$ is bounded in conditional probability, denoted as $A_n^* = O_{\mathsf{P}^*}(a_n)$, if

$$\forall \kappa > 0, \quad \lim_{M\to\infty} \limsup_{n\to\infty} \mathsf{P}(\mathsf{P}^*(|A_n^*/a_n| \geq M) > \kappa) = 0$$

(b) We say that $\{A_n^*/a_n\}$ converges to zero in conditional probability, denoted as $A_n^* = o_{\mathsf{P}^*}(a_n)$, if

$$\forall \kappa, \eta > 0, \quad \limsup_{n\to\infty} \mathsf{P}(\mathsf{P}^*(|A_n^*/a_n| > \eta) > \kappa) = 0.$$

Definition S2 (Weak convergences in conditional distribution)**.**

(a) The sequence $\{A_n^*\}$ (pointwise) converges to A in conditional distribution, denoted as $A_n^* \xrightarrow{\mathsf{d}*} A$, if for each $x \in C_F$, $\hat{F}_n(x) \xrightarrow{\mathsf{P}} F(x)$, i.e.,

$$\forall \kappa > 0, x \in C_F, \quad \limsup_{n\to\infty} \mathsf{P}(|\hat{F}_n(x) - F(x)| > \kappa) = 0.$$

(b) The sequence $\{A_n^*\}$ uniformly converges to A in bootstrap distribution in weak sense, if

$$\forall \kappa > 0, \quad \limsup_{n\to\infty} \sup_{x\in C_F} \mathsf{P}(|\hat{F}_n(x) - F(x)| > \kappa) = 0.$$

(c) The sequence $\{A_n^*\}$ uniformly converges to A in bootstrap distribution in strong sense, if

$$\sup_{x \in C_F} |\hat{F}_n(x) - F(x)| \xrightarrow{\mathsf{P}} 0.$$

The computational properties between convergent terms in original probability mostly hold for convergent quantities in conditional probability (Yeon et al., 2024a). In what follows, we frequently use these results without mentioning any reference.

S1.2 Technical assumptions

We below list the technical assumptions for establishing the consistency of the WB distributional approximation, in addition to the following fundamental identifiability condition:

(A0) $\ker \Gamma = \{0\}$, where $\ker \Gamma \equiv \{x \in \mathbb{H} : \Gamma x = 0\}$.

Similar conditions arise for paired bootstrap (cf. Yeon et al., 2024a). To briefly explain, the first four conditions (A1)-(A4) on the regressors are common in the FPCR literature for providing the consistency of the FPCR estimator $\hat{\beta}_{h_n}$ in (2.4); the next four conditions $B(u)$, (S), (C), (D) are needed to derive the central limit theorems for the statistical quantities in (2.7) (cf. Yeon et al., 2024a); and the last two conditions (W1)-(W2) are mild assumptions on the multipliers for defining WB errors.

(A1) $\sup_{j \in \mathbb{N}} \gamma_j^{-2} \mathsf{E}[\langle X - \mathsf{E}[X], \phi_j \rangle^4] < \infty$;

(A2) γ_j is a convex positive function of j if $j \geq J$ for some large integer $J \geq 1$;

(A3) $\sup_{j\in\mathbb{N}} \gamma_j j \log j < \infty$;

(A4) $h_n^{-1} + n^{-1/2} h_n^2 \log h_n + n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2} \to 0$ as $n \to \infty$, where $\{\delta_j\}_{j=1}^{\infty}$ is the sequence of eigengaps defined as $\delta_1 \equiv \gamma_1 - \gamma_2$ and $\delta_j \equiv \min\{\gamma_j - \gamma_{j+1}, \gamma_{j-1} - \gamma_j\}$ for integer $j \geq 2$.

$B(u)$ $\sup_{j\in\mathbb{N}} j^{-1} m(j,u) \langle \beta, \phi_j \rangle^2 < \infty$, where $m(j,u) \equiv \max\{j^u, \sum_{l=1}^{j} \delta_l^{-2}\}$ is a function of integer $j \geq 1$ for a generic constant $u > 0$;

(S) $h_n s_{h_n}(X_0)^{-1} = O_{\mathsf{P}}(1)$;

(C) there exists $\delta \in (0,2]$ such that $\sup_{j\in\mathbb{N}} \lambda_j^{-(2+\delta)/2} \mathsf{E}[|\langle X\varepsilon, \psi_j \rangle|^{2+\delta}] < \infty$, where (λ_j, ψ_j) is the j-th eigenvalue-eigenfunction pair of Λ;

(D) $\sup_{j\in\mathbb{N}} \gamma_j^{-1} \|\Lambda^{1/2} \phi_j\|^2 < \infty$.

(R) $\tau \equiv \lim_{n\to\infty} h_n / g_n \geq 1$.

(L) for $\delta \in (0,2]$ in Condition (C), it holds that $\mathsf{E}[\|X\|^{4+2\delta}] < \infty$,
$n^{-\delta/2} h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} = O(1)$ (which implies $n^{-\delta/2} h_n^{(2+\delta)/2} = o(1)$).

(W1) the multipliers $\mathcal{W}_n \equiv \{W_i\}_{i=1}^n$ are independent copies of W with $\mathsf{E}[W] = 0$ and $\mathsf{E}[W^2] = 1$ and independent of the data sample $\{(X_i, Y_i)\}_{i=1}^n$;

(W2) for $\delta \in (0,2]$ in Condition (C), it holds that $\mathsf{E}[|W|^{2+\delta}] < \infty$.

We discuss these technical assumptions in greater detail.

Conditions (A1)-(A4) are assumed on the regressor X, which are common in the functional principal component regression (FPCR) literature. Condition (A1) implies that the regressor X has finite fourth moment, while Conditions (A2)-(A4) are related to the decay behavior of eigenvalues $\{\gamma_j\}_{j=1}^{\infty}$ and eigengaps $\{\delta_j\}_{j=1}^{\infty}$. The latter three conditions are mild and imposed to prove the propositions involving perturbation theory for functional data, though these conditions can be relaxed. See also Remarks 1-2 of Cardot et al. (2007) for discussion on Conditions (A1)-(A2). In the functional linear regression model (FLRM) literature, it is common that to assume Conditions (A1)-(A4) along with the model identifiability condition (A0) (Cardot et al., 2007, Khademnoe and Hosseini-Nasab, 016a, Yeon et al., 2023, 2024a) or even stronger moment assumptions about X for ease of exposition, such as polynomial decay rates on $\{\gamma_j\}_{j=1}^{\infty}$ or $\{\delta_j\}_{j=1}^{\infty}$ (Cai and Hall, 2006, Crambes and Mas, 2013, Hall and Horowitz, 2007, Imaizumi and Kato, 2018, Khademnoe and Hosseini-Nasab, 016b, Lei, 2014).

We impose Condition $B(u)$ for balancing the decay rates of eigengaps $\{\delta_j\}_{j=1}^{\infty}$ and the Fourier coefficients $\{\langle \beta, \phi_j \rangle\}_{j=1}^{\infty}$ of the slope function β. Such conditions are commonly assumed to eliminate a non-random form of bias due to truncation in FPCR estimators (cf. Cardot et al., 2007, Crambes and Mas, 2013); further literature also considers a balancing condition, similar to $B(u)$, upon further supposing polynomial decay rates for $\{\delta_j\}_{j=1}^{\infty}$ and $\{\langle \beta, \phi_j \rangle\}_{j=1}^{\infty}$ (cf. Hall and Horowitz, 2007, Imaizumi and

Kato, 2018, Khademnoe and Hosseini-Nasab, 016a, Lin and Lin, 2025).

Conditions (S), (C), (D) are further required, especially for a central limit theorem (CLT) under heteroscedasticity. Condition (S) provides an asymptotic lower bound for the scaling $s_{h_n}(X_0)$, which is a mild assumption; see Yeon et al. (2023) for further motivation and discussion with its heteroscedastic version. Condition (C) ensures a finite $(2+\delta)$-th moment of the product $X\varepsilon$, which is also sufficient for the existence of the covariance operator $\Lambda \equiv \mathsf{E}[(X\varepsilon)^{\otimes 2}]$. Condition (D) is required for the heteroscedastic case due to the occurrence of two different operators Γ and Λ; this assumption balances the decay rates of $\{\gamma_j\}_{j=1}^{\infty}$ and $\{\langle \Lambda\phi_j, \phi_j\rangle\}_{j=1}^{\infty}$.

In regression problems, it is common for bootstrap procedures to involve additional tuning parameters (bandwidths or truncation levels) (Ferraty et al., 2010, Franke and Hardle, 1992, González-Manteiga and Martínez-Calvo, 2011, Yeon et al., 2023, 2024a), which holds also with FLRMs here. In addition to the truncation h_n defining the original FPCR estimator $\hat{\beta}_{h_n}$ from data in (2.7) of the main paper, akin to the residual bootstrap by Yeon et al. (2023) and paired bootstrap by Yeon et al. (2024a), the proposed wild bootstrap method involves a further bootstrap-level truncation g_n, which is used to construct a FPCR estimator $\hat{\beta}_{g_n} \equiv \hat{\Gamma}_{g_n}^{-1}\hat{\Delta}_n$ for purposes of playing the role of the true parameter in the bootstrap world. It is possible to choose the tuning parameter g_n to be h_n (so that the original FPCR estimator $\hat{\beta}_{h_n}$ is used to mimic the true parameter in bootstrap), though more flexibility in bootstrap performance is possible by considering a bootstrap-level tuning parameter g_n separately from the truncation h_n of the original-data statistic $\hat{\beta}_{h_n}$. Generally, this bootstrap-level truncation g_n should satisfy the truncation Condition (A4) in order for $\hat{\beta}_{g_n}$ to be a consistent estimator of the true slope β and, further, g_n should not be smaller than h_n. The latter aspect is also common in bootstrap (Ferraty et al., 2010, Franke and Hardle, 1992, González-Manteiga and Martínez-Calvo, 2011, Yeon et al., 2023, 2024a). Namely, for valid bootstrap approximations, we require the original-data h_n and bootstrap-level truncations g_n to satisfy the relative growth condition, as given in Condition (R). Indeed, Yeon et al. (2024a) have shown that, if Condition (R) is not satisfied (i.e., $\tau \in (0, 1)$), there can arise cases where $\hat{\beta}_{g_n}$ fails to serve as an adequate bootstrap version of β and, consequently, all bootstrap methods can become invalid (cf. Proposition 2 therein).

Lastly, the wild bootstrap method, as well as the paired bootstrap by Yeon et al. (2024a), require one further condition to deal with heteroscedastic errors: Condition (L). This condition is technical, though fairly mild, and used for verifying distributional convergence of bootstrap (cf. Yeon et al., 2024a).

For completeness, we leave the central limit theorem for mean response under heteroscedasticity, provided by Yeon et al. (2024a).

Theorem S1 (Yeon et al., 2024a, Theorem 1)**.** *Suppose that Conditions (A1)-(A4), (S), (C), (D) (including Condition (A4) for k_n) hold along with and Condition $B(u)$ for some $u > 7$. We further suppose that $n^{-1/2}h_n^{7/2}(\log h_n)^4 = o(1)$ and $n = O(m(h_n, u))$*

hold. Then, as $n \to \infty$, we have

$$\sup_{y\in\mathbb{R}} \left| \mathsf{P}\left(\sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}} \{\hat{\mu}_{h_n}(X_0) - \mu(X_0)\} \middle| X_0 \right) - \Phi(y) \right| \xrightarrow{\mathsf{P}} 0,$$

where $\mu(X_0)$ and $\hat{\mu}(X_0)$ are the true and estimated mean responses respectively defined in Equations (2.5) and (2.6) of the main paper, and Φ stands for the standard normal distribution function.

S1.3 Technical differences between our WB consistency theorem and the PB consistency result of Yeon et al. (2024a)

As suggested by the reviewers, we contrast proofs behind the WB and PB consistency results, which arise from the aforementioned differences between resampling mechanisms.

Regarding the WB bias term related to U_n^* in Equation (S2.4) of the supplement (see Proposition S1 there), we directly obtain a specific rate for $\|(zI - \Gamma)^{-1/2}U_n^*\|$ (Lemmas S5–S6 in the supplement). Due to dependence of the distribution ε_i^* on residuals, we need such a careful treatment for the term U_n^*. In contrast, for a similar PB bias term, Yeon et al. (2024b, Lemma S37) simply applies inequality $\|(zI-\Gamma)^{-1/2}U_n^*\| \le \|(zI-\Gamma)^{-1/2}\|_\infty \|U_n^*\|$ to bound it, which could arise from its fully nonparametric nature. See sufficient conditions for Proposition S1 in the supplement and Yeon et al. (2024b, Proposition S8) for comparison between them.

This bias term shows another important difference. For the WB bias term, the consistency for $\hat{\beta}_{k_n}$ is required, because the resampling scheme is based on the residuals $\hat{\varepsilon}_{i,k_n} \equiv Y_i - \bar{Y} - \langle \hat{\beta}_k, X_i - \bar{X} \rangle$, where $\bar{Y} \equiv n^{-1} \sum_{i=1}^n Y_i$ and $\bar{X} \equiv n^{-1} \sum_{i=1}^n X_i$. However, for the PB bias term, the truncation k_n is not involved at all, but rather the consistency for $\hat{\beta}_{g_n}$ plays a more crucial role.

Regarding variance term, which is also related to U_n^*, the derivation of the Lyapounov condition is more involved to separate the important term $\mathsf{E}^*[\|\Lambda_{h_n}^{-1/2} X_i \varepsilon_i^*\|^{2+\delta}]$ into other specific forms. Specifically, different types of Jensen's inequality are used multiple times. This is again because of its resampling nature depending on the residuals; see the proof of Proposition S2 in the supplement. In contrast, for the PB case, the corresponding term $\mathsf{E}^*[\|\Lambda_{h_n}^{-1/2} X_i \varepsilon_{i,g_n}^*\|^{2+\delta}]$, e.g., Equation (40) in the proof of Yeon et al. (2024a, Proposition 4), can be decomposed by using Jensen's inequality only once.

S1.4 Truncation interpretation

For easier interpretation of the bootstrap results in Theorem 1, it can be helpful to recast the technical conditions involved in terms of more specific decay rates for eigengaps $\delta_j \asymp j^{-a}$ and projections $|\langle \beta, \phi_j \rangle| \asymp j^{-b}$ with some constants $a > 2$ and $b > 1$ with $a + 2 < 2b$ (implying $b > 2$). Similar polynomial decay assumptions are common in the literature (cf. Imaizumi and Kato, 2018, Khademnoe and Hosseini-Nasab, 016a,b, Lei, 2014, Lin and Lin, 2025) and permit conditions related to eigengaps $\{\delta_j\}_{j=1}^{\infty}$, eigenvalues $\{\gamma_j\}_{j=1}^{\infty}$, and the slope function β to be translated to conditions about a and b. In particular, this approach has also been used to characterize the optimal truncation h_n in order for minimizing the mean-squared errors of a FPCR estimator $\hat{\beta}_{h_n}$ (Hall and Horowitz, 2007) or its projection $\langle \hat{\beta}_{h_n}, X_0 - \mathsf{E}[X] \rangle$ (Cai and Hall, 2006), which allows an understanding of how truncations h_n for bootstrap compare relatively in FLRMs.

In terms of polynomial decay assumptions, the conclusions of Theorem 1 remain valid if we replace "Conditions (A2)-(A4), (A4) for g_n, k_n, $h_n^{-1} + n^{-1/2} h_n^{7/2} (\log h_n)^3 = o(1)$, and $n = O(m(h_n, u))$ with Condition $B(u)$ for some $u > 7$" by the following simpler conditions (P1)-(P3) for the truncation levels k_n, h_n, and g_n:

(P1) $k_n \to \infty$ with $k_n^{v_k} = O(n)$ for some $v_k > 2a + 1$;

(P2) $g_n \to \infty$ with $g_n^{v_g} = O(n)$ for some $v_g > 2a + 1$;

(P3) $n \asymp h_n^v$ for some $v \in (\max\{7, (2a+1)\}, a + 2b - 1)$.

The previous work by Yeon et al. (2023) contains a formal statement on the residual bootstrap consistency under such polynomial decay rates, and the same argument extends to the wild bootstrap method. As an example, with $a = 2.5$ and $b = 3$, the growth rate $h_n \asymp n^{1/8}$ is one sufficient condition in order for each bootstrap approximation to be valid.

S2 Technical details for multiplier wild bootstrap

This section treats the asymptotic properties of the wild bootstrap (WB) estimator $\hat{\beta}^*_{h_n}$ given in Theorem 1 in the main paper. We follow the same notation as there. The bootstrap consistency result in Theorem 1 is based on the following decompositions:

$$\hat{\beta}_{h_n} - \beta = (\hat{\Gamma}^{-1}_{h_n} - \Gamma^{-1}_{h_n})U_n + \Gamma^{-1}_{h_n}U_n + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta + \Pi_{h_n}\beta - \beta, \tag{S2.1}$$

$$\hat{\beta}^*_{h_n} - \hat{\beta}_{g_n} = (\hat{\Gamma}^{-1}_{h_n} - \Gamma^{-1}_{h_n})U^*_n + \Gamma^{-1}_{h_n}U^*_n + \hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n}, \tag{S2.2}$$

where $\Gamma^{-1}_{h_n} \equiv \sum_{j=1}^{h_n} \gamma_j^{-1}\phi_j^{\otimes 2}, \hat{\Gamma}^{-1}_{h_n} \equiv \sum_{j=1}^{h_n} \hat{\gamma}_j^{-1}\hat{\phi}_j^{\otimes 2}$ are the pseudo-inverses of the covariance operators Γ and $\hat{\Gamma}_n$, respectively; $\Pi_{h_n} \equiv \sum_{j=1}^{h_n} \phi_j^{\otimes 2}, \hat{\Pi}_{h_n} \equiv \sum_{j=1}^{h_n} \hat{\phi}_j^{\otimes 2}$ are the projection operators onto the linear spaces spanned by the first h_n eigenfunctions $\{\phi_j\}_{j=1}^{h_n}$ and $\{\hat{\phi}_j\}_{j=1}^{h_n}$, respectively; and

$$U_n \equiv n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon}) = n^{-1}\sum_{i=1}^{n} X_i\varepsilon_i - \bar{X}\bar{\varepsilon}, \tag{S2.3}$$

$$U^*_n \equiv n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(\varepsilon^*_i - \bar{\varepsilon}^*) = n^{-1}\sum_{i=1}^{n} X_i\varepsilon^*_i - \bar{X}\bar{\varepsilon}^* \tag{S2.4}$$

are the cross-covariance functions between $\{X_i\}_{i=1}^n$ and either errors $\{\varepsilon_i\}_{i=1}^n$ or $\{\varepsilon^*_i\}_{i=1}^n$ with $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$, $\bar{\varepsilon} \equiv n^{-1}\sum_{i=1}^n \varepsilon_i$. Here, (γ_j, ϕ_j) and $(\hat{\gamma}_j, \hat{\phi}_j)$ again denote the j-th eigenvalue-eigenfunction pair of the covariance operators Γ and $\hat{\Gamma}_n$, respectively, and for $x \in \mathbb{H}$, $x^{\otimes 2} \equiv x \otimes x$ denotes the squared tensor product of x. The non-random bias $\hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n}$ in the bootstrap world is further decomposed as

$$\hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n} = (\hat{\Pi}_{h_n} - \Pi_{h_n} + \Pi_{h_n} - I)\hat{\beta}_{g_n}$$

$$= (\hat{\Pi}_{h_n} - \Pi_{h_n})(\hat{\beta}_{g_n} - \beta) + (\hat{\Pi}_{h_n} - \Pi_{h_n})\beta$$
$$+ (\Pi_{h_n} - I)(\hat{\beta}_{g_n} - \beta) + (\Pi_{h_n} - I)\beta. \tag{S2.5}$$

Under heteroscedasticity, Yeon et al. (2024a) develop and prove a CLT for projection estimator based on the decomposition (S2.1) and the convergence to zero of the last term in (S2.2). It hence suffices to derive the convergences of only the first two terms in (S2.2).

We start with reviewing the concepts of convergences in conditional probability to handle bootstrap terms and perturbation theory for calculus between operators on an infinite dimensional space in Section S2.1. The consistency of the WB estimator $\hat{\beta}^*_{h_n}$ is next provided in Section S2.2. In the remaining sections, we study the convergence results for proving Theorem 1 in the main paper, where the convergences of WB bias, variance, and scaling terms are respectively given in Sections S2.3, S2.4, and S2.5.

In what follows, we assume that both response and regressor have zero mean, i.e., $\mathsf{E}[Y] = 0$ and $\mathsf{E}[X] = 0$, implying $\alpha = 0$, unless otherwise stated. We will discuss why such zero-mean condition is reasonable later when we prove the main theorems.

S2.1 Perturbation theory

Convergence results in functional data analysis heavily depend on the so-called perturbation theory or functional calculus. In perturbation theory, we study how small perturbation causes difference in operators. For general introduction, we refer to (Dunford and Schwartz, 1988, Chapter VII), (Gohberg et al., 1990, Chapter I), or (Hsing and Eubank, 2015, Chapter 5). See also Cardot et al. (2007), Yeon et al. (2023, 2024a) for related theoretical developments.

To describe the results from perturbation theory, we define the contour integral for functions taking operators as their values. Let $\mathcal{B}_j = \{z \in \mathbb{C} : |z - \gamma_j| \leq \delta_j/2\}$ be the oriented circle in the complex plane $\mathbb{C}$ and set $\mathcal{C}_{h_n} = \bigcup_{j=1}^{h_n} \mathcal{B}_j$ their union. We learn from functional calculus for bounded linear operators that

$$\Pi_{h_n} = \frac{1}{2\pi\iota} \sum_{j=1}^{h_n} \int_{\mathcal{B}_j} (zI - \Gamma)^{-1} dz, \tag{S2.6}$$

$$\Gamma_{h_n}^{-1} = \frac{1}{2\pi\iota} \sum_{j=1}^{h_n} \int_{\mathcal{B}_j} z^{-1}(zI - \Gamma)^{-1} dz, \tag{S2.7}$$

where $\Pi_{h_n} \equiv \sum_{j=1}^{h_n} \phi_j^{\otimes 2}$ denotes the projection operator onto the space spanned by first h_n eigenfunctions $\{\phi_j\}_{j=1}^{h_n}$ of Γ and $\Gamma_{h_n}^{-1}$ is a finite approximation of $\Gamma^{-1} \equiv \sum_{j=1}^{\infty} \gamma^{-1}\phi_j^{\otimes 2}$. The sample counterparts $\hat{\Pi}_{h_n}$ and $\hat{\Gamma}_{h_n}^{-1}$ are similary defined from the sample covariance operator $\hat{\Gamma}_n$ with the corresponding random contours $\hat{\mathcal{B}}_j = \{z \in \mathbb{C} : |z - \hat{\gamma}_j| \leq \delta_j/2\}$ and $\hat{\mathcal{C}}_{h_n} = \bigcup_{j=1}^{h_n} \hat{\mathcal{B}}_j$. For further purposes, we use the following notations:

$$G_n(z) = (zI - \Gamma)^{-1/2}(\hat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1/2};$$

$$K_n(z) = (zI - \Gamma)^{1/2}(zI - \hat{\Gamma}_n)^{-1}(zI - \Gamma)^{1/2};$$

$$\mathcal{E}_j = (\|G_n(z)\|_\infty < 1/2, \forall z \in \mathcal{B}_j);$$

$$\mathcal{A}_{h_n} = \{\forall j \in \{1, \ldots, h_n\}, |\hat{\gamma}_j - \gamma_j| < \delta_j/2\}.$$

The key lemmas related to these notations are provided below; the proofs are found in Cardot et al. (2007), Yeon et al. (2023, 2024a).

Lemma S1 (Second resolvent identity). *The difference between the resolvents of* $\hat{\Gamma}_n$ *and* Γ *can be written as*

$$(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1} = (zI - \hat{\Gamma}_n)^{-1}(\hat{\Gamma}_n - \Gamma)(zI - \Gamma)^{-1}$$
$$= (zI - \Gamma)^{-1}(\hat{\Gamma}_n - \Gamma)(zI - \hat{\Gamma}_n)^{-1},$$

and hence,

$$(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1} = (zI - \Gamma)^{-1/2} K_n(z) G_n(z) (zI - \Gamma)^{-1/2}.$$

Lemma S2. *Suppose that* γ_j *is a convex function of* j *at least for sufficiently large* j*. Also, suppose that* $\sup_{j\in\mathbb{N}} \mathsf{E}[\xi_j^4] < \infty$*. Then, for sufficiently large* j*, we have the following.*

(a) $\mathsf{E}\left[\sup_{z\in\mathcal{B}_j} \|G_n(z)\|_\infty^2\right] \leq Cn^{-1}(j\log j)^2$*;*

(b) $\mathsf{E}\left[\sup_{z\in\mathcal{B}_j} \left\|(zI - \Gamma)^{-1/2}X\right\|^2\right] \leq Cj\log j$*;*

(c) $\mathsf{E}\left[\sup_{z\in\mathcal{B}_j} \|(zI - \Gamma)^{-1/2}X\|^4\right] \leq C(j\log j)^2$*; and*

(d) $\mathsf{E}\left[\sup_{z\in\mathcal{B}_j} \left\|(zI - \Gamma)^{-1/2}U_n\right\|^2\right] \leq Cn^{-1}\delta_j^{-1}$*.*

Lemma S3. *Suppose the same assumptions of Lemma S2. Then, we have that*

$$\sup_{z\in\mathcal{B}_j} \|K_n(z)\|_\infty \mathbb{I}_{\mathcal{E}_j} \leq 2 \quad \textit{almost surely}$$

and $\mathsf{P}(\mathcal{E}_j^c) \leq Cn^{-1/2} j \log j$, *where the latter implies that for any non-negative quantity* Q_j *(that can be either random or fixed quantities and can depend on* n *or not), we have*

$$\sum_{j=1}^{h_n} Q_j \mathbb{I}_{\mathcal{E}_j^c} = O_{\mathsf{P}(\cdot|X_0)}\left(n^{-1/2} \sum_{j=1}^{h_n} j \log j\right).$$

Lemma S4.

(a) We observe that

$$\hat{\Pi}_{h_n} - \Pi_{h_n} = \frac{1}{2\pi\iota} \int_{\mathcal{C}_{h_n}} \{(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1}\} dz + r_{1n} \mathbb{I}_{\mathcal{A}_{h_n}^c},$$
$$\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1} = \frac{1}{2\pi\iota} \int_{\mathcal{C}_{h_n}} z^{-1} \{(zI - \hat{\Gamma}_n)^{-1} - (zI - \Gamma)^{-1}\} dz + r_{2n} \mathbb{I}_{\mathcal{A}_{h_n}^c},$$

where

$$r_{1n} = \hat{\Pi}_{h_n} - \frac{1}{2\pi\iota} \int_{\mathcal{C}_{h_n}} (zI - \hat{\Gamma}_n)^{-1} dz,$$
$$r_{2n} = \hat{\Gamma}_{h_n}^{-1} - \frac{1}{2\pi\iota} \int_{\mathcal{C}_{h_n}} z^{-1} (zI - \hat{\Gamma}_n)^{-1} dz.$$

(b) Suppose the same assumptions of Lemma S2. We then have that

$$\mathsf{P}(\mathcal{A}_{h_n}^c) \leq C_1 n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2} + C_2 n^{-1/2} \sum_{j=1}^{h_n} j \log j.$$

Hence, for any non-negative quantity Q_n *(that can be either random or fixed quantities and can depend on* n *or not), we have*

$$Q_n \mathbb{I}_{\mathcal{A}_{h_n}^c} = O_{\mathsf{P}(\cdot|X_0)}\left(n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2} + n^{-1/2} \sum_{j=1}^{h_n} j \log j\right).$$

Since the terms related to the sets $\mathcal{E}_j$ and $\mathcal{A}_{h_n}$ will be negligible under Condition (A4), thanks to Lemmas S3-S4, we may not mention such terms in what follows.

The following two lemmas find an upper bound for the term related to U_n^* from (S2.4), which is similar to the one in the fourth part of Lemma S2 In what follows, we write $\mathsf{P}^* \equiv \mathsf{P}(\cdot|\mathcal{D}_n)$ and $\mathsf{E}^* \equiv \mathsf{E}[\cdot|\mathcal{D}_n]$ for the bootstrap probability and bootstrap expectation operator, respectively, conditional on the data $\mathcal{D}_n \equiv \{(X_i, Y_i)\}_{i=1}^n$.

Lemma S5. *Suppose the same assumptions of Lemma S2 and write $\tilde{U}_n^* \equiv n^{-1}\sum_{i=1}^n X_i\varepsilon_i^*$. Then, we have*

$$\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\tilde{U}_n^*\|^2\right] \le Q_{1jn} + Q_{2jn}\|\hat{\beta}_{k_n}-\beta\|^2,$$

where Q_{1jn} and Q_{2jn} are non-negative random variables such that

$$\mathsf{E}[Q_{1jn}] \le \begin{cases} Cn^{-1}\delta_j^{-1} & \text{in general when } \mathsf{E}[|X\varepsilon\|^2] < \infty \\ Cn^{-1}j\log j & \text{if either } \mathsf{E}[\varepsilon^2|X] = \sigma_\varepsilon^2 \in (0,\infty) \text{ or } \mathsf{E}[\varepsilon^4] < \infty, \end{cases}$$

and $\mathsf{E}[Q_{2jn}] \le Cn^{-1}j\log j$.

Proof. Note that

$$\begin{aligned} &\|(zI-\Gamma)^{-1/2}\tilde{U}_n^*\|^2 \\ =&n^{-2}\sum_{i=1}^n \|(zI-\Gamma)^{-1/2}X_i\|^2(\varepsilon_i^*)^2 \\ &+ n^{-2}\sum_{i=1}^n \langle (zI-\Gamma)^{-1/2}X_i, (zI-\Gamma)^{-1/2}X_{i'}\rangle\varepsilon_i^*\varepsilon_{i'}^*, \end{aligned}$$

which implies that

$$\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\tilde{U}_n^*\|^2$$

$$\begin{aligned}\leq & n^{-2}\sum_{i=1}^{n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2(\varepsilon_i^*)^2\\ & + n^{-2}\sum_{i=1}^{n}\sup_{z\in\mathcal{B}_j}\langle(zI-\Gamma)^{-1/2}X_i,(zI-\Gamma)^{-1/2}X_{i'}\rangle\varepsilon_i^*\varepsilon_{i'}^*.\end{aligned}$$

Since $\mathsf{E}^*[(\varepsilon_i^*)^2]=\hat{\varepsilon}_{i,k_n}^2$ for each i and $\mathsf{E}^*[\varepsilon_i^*\varepsilon_{i'}^*]=0$ if $i\neq i'$, we have

$$\begin{aligned}&\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\tilde{U}_n^*\|^2\right]\leq n^{-2}\sum_{i=1}^{n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\hat{\varepsilon}_{i,k_n}^2\\ \leq & 2n^{-2}\sum_{i=1}^{n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\varepsilon_i^2+2n^{-2}\sum_{i=1}^{n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\langle\hat{\beta}_{k_n}-\beta,X_i\rangle^2\\ = & 2n^{-2}\sum_{i=1}^{n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\varepsilon_i^2 \qquad \text{(S2.8)}\\ & + 2\left\langle\left(n^{-2}\sum_{i=1}^{n}\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2X_i^{\otimes 2}\right)(\hat{\beta}_{k_n}-\beta),\hat{\beta}_{k_n}-\beta\right\rangle \qquad \text{(S2.9)}\end{aligned}$$

The term in (S2.8) is differently bounded depending on the error assumption. We first consider general case with $\mathsf{E}[\|X\varepsilon\|^2]<\infty$. Recall by Hsing and Eubank (2015, Equation (5.3)), for $z\in\mathcal{B}_j$,

$$\|(zI-\Gamma)^{-1/2}\|_\infty=\left(\min_{l\in\mathbb{N}}|z-\gamma_l|^{1/2}\right)^{-1}=|z-\gamma_j|^{-1/2}=(\delta_j/2)^{-1/2}.$$

This implies that

$$\begin{aligned}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\varepsilon^2\right]&\leq\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\|_\infty\|X\varepsilon\|^2\right]\\&\leq 2\delta_j^{-1}\mathsf{E}[\|X\varepsilon\|^2]=2\mathrm{tr}(\mathsf{var}[X\varepsilon])\delta_j^{-1}.\end{aligned}$$

Next, under homoscedasticity with $\mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2$, we have that

$$\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\varepsilon^2\right]=\sigma_\varepsilon^2\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\right]\leq Cj\log j.$$

Last, under heteroscedasticity with $\mathsf{E}[\varepsilon^4] < \infty$, by Cauchy-Schwarz inequality, we have that

$$\begin{aligned}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\varepsilon^2\right] &\leq \sqrt{\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^4\right]}\sqrt{\mathsf{E}[\varepsilon^4]}\\ &\leq Cj\log j.\end{aligned}$$

We thus have that

$$\begin{aligned}&\mathsf{E}\left[n^{-2}\sum_{i=1}^n\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\varepsilon_i^2\right]\\ &\leq\begin{cases}Cn^{-1}\delta_j^{-1} & \text{in general when } \mathsf{E}[|X\varepsilon\|^2]<\infty\\ Cn^{-1}j\log j & \text{if either } \mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2\in(0,\infty) \text{ or } \mathsf{E}[\varepsilon^4]<\infty.\end{cases}\end{aligned}$$

The term in (S2.9) is bounded as

$$\begin{aligned}&\mathsf{E}\left[\left\|n^{-2}\sum_{i=1}^n\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2X_i^{\otimes 2}\right\|\right]\\ \leq& n^{-1}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\|X_i\|^2\right]\\ \leq& n^{-1}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^4\right]^{1/2}\mathsf{E}[\|X_i\|^4]^{1/2}\\ \leq& Cn^{-1}j\log j\end{aligned}$$

by the third part of Lemma S2. This completes the proof. □

Lemma S6. *Suppose the same assumptions of Lemma S2. For U_n^* in (S2.4), we have* *that*

$$\mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}U_n^*\|^2\right] \le Q_{1jn} + Q_{2jn}\|\hat{\beta}_{k_n}-\beta\|^2 + R_nQ_{2jn},$$

where Q_{1jn}*,* Q_{2jn}*,* Q_{3jn}*, and* R_n *are non-negative random variables such that*

$$\mathsf{E}[Q_{1jn}] \le \begin{cases} Cn^{-1}\delta_j^{-1} & \text{in general when } \mathsf{E}[|X\varepsilon\|^2]<\infty \\ Cn^{-1}j\log j & \text{if either } \mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2\in(0,\infty) \text{ or } \mathsf{E}[\varepsilon^4]<\infty, \end{cases}$$

$\mathsf{E}[Q_{2jn}] \le Cn^{-1}j\log j$*,* $\mathsf{E}[Q_{3jn}] \le Cj\log j$*, and* $R_n = O_{\mathsf{P}}(n^{-1})$*.*

Proof. Note that

$$\|(zI-\Gamma)^{-1/2}U_n^*\|^2 \le 2\|(zI-\Gamma)^{-1/2}\tilde{U}_n^*\|^2 + 2(\bar{\varepsilon}^*)^2\|(zI-\Gamma)^{-1/2}\bar{X}\|^2. \tag{S2.10}$$

The first term in (S2.10) is bounded based on Lemma S5. For the second term in (S2.10), note that

$$\mathsf{E}^*[(\bar{\varepsilon}^*)^2] = n^{-2}\sum_{i=1}^n \mathsf{E}^*[(\varepsilon_i^*)^2] + n^{-2}\sum_{i\neq i'}\mathsf{E}^*[\varepsilon_i^*\varepsilon_{i'}^*] = n^{-2}\sum_{i=1}^n\hat{\varepsilon}_{i,k_n}^2 = O_{\mathsf{P}}(n^{-1})$$

if $\|\hat{\beta}_{k_n}-\beta\| \xrightarrow{\mathsf{P}} 0$, since

$$\begin{aligned} n^{-1}\sum_{i=1}^n\hat{\varepsilon}_{i,k_n}^2 &= n^{-1}\sum_{i=1}^n(\varepsilon_i - \langle X_i, \hat{\beta}_{k_n}-\beta\rangle)^2 \\ &\le n^{-1}\sum_{i=1}^n\varepsilon_i^2 + n^{-1}\sum_{i=1}^n\|X_i\|^2\|\hat{\beta}_{k_n}-\beta\|^2 + 2n^{-1}\sum_{i=1}^n\|X_i\varepsilon_i\|\|\hat{\beta}_{k_n}-\beta\| \\ &= \{\mathsf{E}[\varepsilon^2]+o_{\mathsf{P}}(1)\} + \{\mathsf{E}[\|X\|^2]+o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) + \{\mathsf{E}[\|X\varepsilon\|]+o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) \\ &= O_{\mathsf{P}}(1). \end{aligned}$$

Meanwhile, by Jensen's inequality, we have

$$\begin{aligned}
&\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}\bar{X}\|^2\right]\\
\leq& n^{-1}\sum_{i=1}^{n}\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X_i\|^2\right]=\mathsf{E}\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}X\|^2\right]\\
\leq& Cj\log j.
\end{aligned}$$

This completes the proof. □

S2.2 Consistency of WB estimator

In this section, we study the convergence of the WB estimator $\hat{\beta}^*_{h_n}$. Throughout the section, we suppose Conditions (A0)-(A4), (D) in the main paper hold along with Condition (A4) for k_n and g_n.

Lemma S7. *If $n^{-1}\sum_{j=1}^{h_n}(j\log j)^2=O(1)$ (which is implied by $n^{-1/2}h_n^2\log h_n=o(1)$), then $\langle(\hat{\Gamma}^{-1}_{h_n}-\Gamma^{-1}_{h_n})U^*_n,X_0\rangle=o_{\mathsf{P}^*}(1)$.*

Proof. We observe from Lemma S1 that

$$\begin{aligned}
\hat{\Gamma}^{-1}_{h_n}-\Gamma^{-1}_{h_n}&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}z^{-1}\left\{(zI-\hat{\Gamma}_n)^{-1}-(zI-\Gamma)^{-1}\right\}dz+r_{2n}\mathbb{I}_{\mathcal{A}^c_{h_n}}\\
&=\frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j}z^{-1}(zI-\Gamma)^{-1/2}K_n(z)G_n(z)(zI-\Gamma)^{-1/2}dz+r_{2n}\mathbb{I}_{\mathcal{A}^c_{h_n}}.
\end{aligned}$$

This implies that $\|\langle(\hat{\Gamma}^{-1}_{h_n}-\Gamma^{-1}_{h_n})U^*_n\|\leq C\sum_{j=1}^{h_n}A_j+\|r_{2n}\|_\infty\|U^*_n\|\mathbb{I}_{\mathcal{A}^c_{h_n}}$, where

$$A_j=\int_{\mathcal{B}_j}\frac{1}{|z|}\|(zI-\Gamma)^{-1/2}\|_\infty\|K_n(z)\|_\infty\|G_n(z)\|_\infty\|(zI-\Gamma)^{-1/2}U^*_n\|dz.$$

Note that for all $z \in \mathcal{B}_j$, $|z| \geq \gamma_j - \delta_j/2 \geq \gamma_j/2$. By Equation (5.3) of Hsing and Eubank (2015), for $z \in \mathcal{B}_j$, we have

$$\|(zI - \Gamma)^{-1/2}\|_\infty = \left(\min_{l \in \mathbb{N}} |z - \gamma_l|^{1/2}\right)^{-1} = |z - \gamma_j|^{-1/2} = (\delta_j/2)^{-1/2}.$$

Note that

$$\begin{aligned}
&\mathsf{E}^*[A_j \mathbb{I}_{\mathcal{E}_j}] \\
\leq & C \operatorname{diam}(\mathcal{B}_j)\delta_j^{-1} V_{jn} \sup_{z \in \mathcal{B}_j} \|(zI - \Gamma)^{-1/2}\|_\infty \mathsf{E}^* \left[\sup_{z \in \mathcal{B}_j} \|(zI - \Gamma)^{-1/2} U_n^*\|\right] \\
\leq & C\delta_j^{-1/2} V_{jn} \mathsf{E}^* \left[\sup_{z \in \mathcal{B}_j} \|(zI - \Gamma)^{-1/2} U_n^*\|^2\right]^{1/2},
\end{aligned}$$

where $V_{jn} \equiv \sup_{z \in \mathcal{B}_j} \|K_n(z)\|_\infty \mathbb{I}_{\mathcal{E}_j} \sup_{z \in \mathcal{B}_j} \|G_n(z)\|_\infty$. Then, by Lemmas S2-S3 and S6,

$$\begin{aligned}
&\mathsf{E}^* \left[\sum_{j=1}^{h_n} A_j \mathbb{I}_{\mathcal{E}_j}\right] \\
\leq & C \sum_{j=1}^{h_n} \delta_j^{-1/2} V_{jn} \mathsf{E}^* \left[\sup_{z \in \mathcal{B}_j} \|(zI - \Gamma)^{-1/2} U_n^*\|^2\right]^{1/2} \\
= & \begin{cases} O_{\mathsf{P}}\left(n^{-1} \sum_{j=1}^{h_n} \delta_j^{-1} j \log j\right) & \text{in general when } \mathsf{E}[|X\varepsilon\|^2] < \infty \\ O_{\mathsf{P}}\left(n^{-1} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2}\right) & \text{if either } \mathsf{E}[\varepsilon^2|X] = \sigma_\varepsilon^2 \in (0,\infty) \text{ or } \mathsf{E}[\varepsilon^4] < \infty. \end{cases}
\end{aligned}$$

A Cauchy–Schwarz inequality implies

$$\begin{aligned}
\left(n^{-1} \sum_{j=1}^{h_n} \delta_j^{-1} j \log j\right)^2 &\leq \left(n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2}\right) \left\{n^{-1} \sum_{j=1}^{h_n} (j \log j)^2\right\} \\
&= o\left(n^{-1} \sum_{j=1}^{h_n} (j \log j)^2\right),
\end{aligned}$$

$$\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)^2 \le \left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-1}j\log j\right)\left\{n^{-1}\sum_{j=1}^{h_n}(j\log j)^2\right\}$$
$$= o\left(\left\{n^{-1}\sum_{j=1}^{h_n}(j\log j)^2\right\}^{3/2}\right).$$

by Condition (A4). Therefore, in both cases, we have

$$\mathsf{E}^*\left[\sum_{j=1}^{h_n}A_j\mathbb{I}_{\mathcal{E}_j}\right] = o_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}(j\log j)^2\right).$$

The proof is finally complete by Lemmas S2-S3. □

Lemma S8. *We have*

$$\mathsf{E}^*[\|\Gamma_{h_n}^{-1}U_n^*\|^2] = O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\|\Lambda^{1/2}\phi_j\|^2\right) + O_{\mathsf{P}}\left(n^{-2}\sum_{j=1}^{h_n}\gamma_j^{-2}\right).$$

Proof. We first note that

$$\Gamma_{h_n}^{-1}U_n^* = \Gamma_{h_n}^{-1}\overline{X\varepsilon^*} - \Gamma_{h_n}^{-1}\bar{X}\bar{\varepsilon}^*. \tag{S2.11}$$

The first term in (S2.11) is bounded as follows. Note that

$$\|\Gamma_{h_n}^{-1}\overline{X\varepsilon^*}\|^2 = \sum_{j=1}^{h_n}\gamma_j^{-2}\langle\overline{X\varepsilon^*},\phi_j\rangle^2$$

and

$$\mathsf{E}^*[\langle\overline{X\varepsilon^*},\phi_j\rangle^2] = n^{-2}\sum_{i=1}^{n}\mathsf{E}^*[\langle X_i\varepsilon_i^*,\phi_j\rangle^2] = n^{-2}\sum_{i=1}^{n}\langle X_i^{\otimes 2}\phi_j,\phi_j\rangle\mathsf{E}^*[(\varepsilon_i^*)^2]$$
$$= n^{-2}\sum_{i=1}^{n}\langle X_i^{\otimes 2}\phi_j,\phi_j\rangle\hat{\varepsilon}_{i,k_n}^2 = n^{-1}\langle\overline{(X\hat{\varepsilon})^{\otimes 2}}\phi_j,\phi_j\rangle$$
$$= n^{-1}\langle(\overline{(X\hat{\varepsilon})^{\otimes 2}}-\Lambda)\phi_j,\phi_j\rangle + n^{-1}\langle\Lambda\phi_j,\phi_j\rangle,$$

$\overline{(X\hat{\varepsilon})^{\otimes 2}} \equiv n^{-1}\sum_{i=1}^{n}(X_i\hat{\varepsilon}_{i,k_n})^{\otimes 2}$. Recall that $\|\overline{(X\hat{\varepsilon})^{\otimes 2}} - \Lambda\|_\infty \xrightarrow{\mathsf{P}} 0$ if $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$ as $n \to \infty$ from Yeon et al. (2024b, Lemma S23), which implies that

$$\sum_{j=1}^{h_n} \gamma_j^{-2}\left\{n^{-1}\langle(\overline{(X\hat{\varepsilon})^{\otimes 2}} - \Lambda)\phi_j, \phi_j\rangle\right\} \leq o_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\right) = o_{\mathsf{P}}(1)$$

by Condition (A4). We therefore have that

$$\mathsf{E}^*[\|\Gamma_{h_n}^{-1}\overline{X\varepsilon^*}\|^2] = O_{\mathsf{P}}\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\|\Lambda^{1/2}\phi_j\|^2\right).$$

The second term is bounded as

$$\mathsf{E}^*[\|\Gamma_{h_n}^{-1}\bar{X}\bar{\varepsilon}^*\|^2] \leq \|\Gamma_{h_n}^{-1}\|_\infty^2 O_{\mathsf{P}}(n^{-2}) = O_{\mathsf{P}}\left(n^{-2}\sum_{j=1}^{h_n}\gamma_j^{-2}\right)$$

thanks to Lemma S9. □

Theorem S2 (Consistency of the WB FPCR estimator)**.** *Suppose that Conditions (A0)-(A4), (D) the main paper hold along with along with Condition (A4) for k_n and g_n. Then, the bootstrap FPCR estimator $\hat{\beta}_{h_n}^*$ converges to the slope function β in the bootstrap probability, i.e., $\|\hat{\beta}_{h_n}^* - \beta\| = o_{\mathsf{P}^*}(1)$.*

Proof. This is proved by using the same argument as the proofs of Yeon et al. (2024b, Theorems S3-S4) based on the decompositions (S2.2) and (S2.5) and Lemmas S7-S8 along with Yeon et al. (2024b, Lemma S20). □

S2.3 WB bias term

In this section, we derive the convergence of the WB bias term related to $(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n^*$, the first term in the decomposition (S2.2). This is the only one term that should be studied because Yeon et al. (2024a) already verify the convergence of the non-random bias term in the bootstrap world, i.e., the term related to the last term $\hat{\Pi}_{h_n}\hat{\beta}_{g_n} - \hat{\beta}_{g_n}$ in (S2.2) (cf. Yeon et al., 2024b, Section S3.3.1).

Proposition S1. *Suppose Conditions (A0)-(A4), (S) in the main paper. As $n \to \infty$, we further suppose either*

(a) $n^{-1/2}h_n^{-1/2}\left(\sum_{j=1}^{h_n}\delta_j^{-1}j\log j\right)^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2} \to 0$ *in general when* $\mathsf{E}[\|X\varepsilon\|^2] < \infty$*; or*

(b) $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2 \to 0$ *if either* $\mathsf{E}[\varepsilon^2|X] \equiv \sigma_\varepsilon^2 \in (0,\infty)$ *or* $\mathsf{E}[\varepsilon^4] < \infty$.

We then have that $\sqrt{n/s_{h_n}(X_0)}\langle(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n^*, X_0\rangle = o_{\mathsf{P}^*(\cdot|X_0)}(1)$*. Both convergence rates hold if* $n^{-1/2}h_n^{7/2}(\log h_n)^4 = O(1)$*.*

Proof. We observe from Lemma S1 that

$$\begin{aligned}\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1} &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}\left\{(zI-\hat{\Gamma}_n)^{-1} - (zI-\Gamma)^{-1}\right\}dz + r_{2n}\mathbb{I}_{\mathcal{A}_{h_n}^c}\\ &= \frac{1}{2\pi\iota}\sum_{j=1}^{h_n}\int_{\mathcal{B}_j} z^{-1}(zI-\Gamma)^{-1/2}K_n(z)G_n(z)(zI-\Gamma)^{-1/2}dz + r_{2n}\mathbb{I}_{\mathcal{A}_{h_n}^c}.\end{aligned}$$

This implies that $|\langle(\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})U_n^*, X_0\rangle| \le C\sum_{j=1}^{h_n}A_j + \|r_{2n}\|_\infty\|U_n^*\|\|X_0\|\mathbb{I}_{\mathcal{A}_{h_n}^c}$, where

$$A_j = \int_{\mathcal{B}_j}\frac{1}{|z|}\|(zI-\Gamma)^{-1/2}X_0\|\|K_n(z)\|_\infty\|G_n(z)\|_\infty\|(zI-\Gamma)^{-1/2}U_n^*\|dz.$$

Note that

$$\mathsf{E}^*[A_j\mathbb{I}_{\mathcal{E}_j}|X_0]$$

$$\begin{aligned}
&\leq C \operatorname{diam}(\mathcal{B}_j)\delta_j^{-1} \sup_{z\in\mathcal{B}_j} \|K_n(z)\|_\infty \mathbb{I}_{\mathcal{E}_j} \sup_{z\in\mathcal{B}_j} \|G_n(z)\|_\infty \\
&\quad \times \sup_{z\in\mathcal{B}_j} \|(zI-\Gamma)^{-1/2}X_0\| \mathsf{E}^* \left[\sup_{z\in\mathcal{B}_j} \|(zI-\Gamma)^{-1/2}U_n^*\| \right] \\
&\leq C V_{1jn} V_{2jn} \mathsf{E}^* \left[\sup_{z\in\mathcal{B}_j} \|(zI-\Gamma)^{-1/2}U_n^*\|^2 \right]^{1/2},
\end{aligned}$$

where

$$V_{1jn} \equiv \sup_{z\in\mathcal{B}_j} \|(zI-\Gamma)^{-1/2}X_0\| \quad \text{and} \quad V_{2jn} \equiv \sup_{z\in\mathcal{B}_j} \|K_n(z)\|_\infty \mathbb{I}_{\mathcal{E}_j} \sup_{z\in\mathcal{B}_j} \|G_n(z)\|_\infty.$$

Then,

$$\begin{aligned}
\mathsf{E}^* \left[\sum_{j=1}^{h_n} A_j \mathbb{I}_{\mathcal{E}_j} \middle| X_0 \right] &\leq C \sum_{j=1}^{h_n} V_{1jn} V_{2jn} \mathsf{E}^* \left[\sup_{z\in\mathcal{B}_j} \|(zI-\Gamma)^{-1/2}U_n^*\|^2 \right]^{1/2} \\
&\leq C \left(\sum_{j=1}^{h_n} V_{1jn}^2 \right)^{1/2} \left(\sum_{j=1}^{h_n} V_{2jn}^2 \mathsf{E}^* \left[\sup_{z\in\mathcal{B}_j} \|(zI-\Gamma)^{-1/2}U_n^*\|^2 \right] \right)^{1/2}.
\end{aligned}$$

By Lemma S2 and Lemma S3,

$$\mathsf{E}\left[\sum_{j=1}^{h_n} V_{1jn}^2 \right] \leq C \sum_{j=1}^{h_n} (j\log j)^2.$$

Also, note by Lemma S6 and the independence between $\{(X_i, Y_i)\}_{i=1}^n$ and X_0 that

$$\begin{aligned}
&\mathsf{E}\left[\sum_{j=1}^{h_n} V_{2jn}^2 Q_{1jn} \right] \\
&\leq \begin{cases} Cn^{-2} \sum_{j=1}^{h_n} \delta_j^{-1} (j\log j)^2 & \text{in general when } \mathsf{E}[\|X\varepsilon\|^2] < \infty \\ Cn^{-2} \sum_{j=1}^{h_n} (j\log j)^3 & \text{if either } \mathsf{E}[\varepsilon^2|X] = \sigma_\varepsilon^2 \in (0,\infty) \text{ or } \mathsf{E}[\varepsilon^4] < \infty, \end{cases}
\end{aligned}$$

$\mathsf{E}\left[\sum_{j=1}^{h_n} V_{2jn}^2 Q_{2jn}\right] \leq Cn^{-2}\sum_{j=1}^{h_n}(j\log j)^3$, and $\mathsf{E}\left[\sum_{j=1}^{h_n} V_{2jn}^2 Q_{3jn}\right] \leq C\sum_{j=1}^{h_n}(j\log j)^3$.

This implies that

$$\begin{aligned}
&\sum_{j=1}^{h_n} V_{2jn}^2 \mathsf{E}^*\left[\sup_{z\in\mathcal{B}_j}\|(zI-\Gamma)^{-1/2}U_n^*\|^2\right] \\
&\leq \sum_{j=1}^{h_n} V_{2jn}^2 Q_{1jn} + \sum_{j=1}^{h_n} V_{2jn}^2 Q_{2jn}\|\hat{\beta}_{k_n}-\beta\|^2 + \sum_{j=1}^{h_n} V_{2jn}^2 Q_{3jn} R_n \\
&= \begin{cases} O_{\mathsf{P}}\left(n^{-2}\sum_{j=1}^{h_n}\delta_j^{-1}(j\log j)^2\right) & \text{in general when } \mathsf{E}[\|X\varepsilon\|^2]<\infty \\ O_{\mathsf{P}}\left(n^{-2}\sum_{j=1}^{h_n}(j\log j)^3\right) & \text{if either } \mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2\in(0,\infty) \text{ or } \mathsf{E}[\varepsilon^4]<\infty, \end{cases}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathsf{E}^*\left[\sqrt{\frac{n}{s_{h_n}(X_0)}}\sum_{j=1}^{h_n} A_j \mathbb{I}_{\mathcal{E}_j}\Big| X_0\right] \\
&= \begin{cases} O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\left\{\sum_{j=1}^{h_n}\delta_j^{-1}(j\log j)^2\right\}^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}\right) \\ \qquad\qquad\qquad \text{in general when } \mathsf{E}[\|X\varepsilon\|^2]<\infty \\ O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^3\right\}^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}\right) \\ \qquad\qquad\qquad \text{if either } \mathsf{E}[\varepsilon^2|X]=\sigma_\varepsilon^2\in(0,\infty) \text{ or } \mathsf{E}[\varepsilon^4]<\infty, \end{cases}
\end{aligned}$$

and by Lemmas S3-S4, we conclude the first assertion. Note that

$$n^{-1/4}h_n^{1/4}\left(\sum_{j=1}^{h_n}\delta_j^{-1}\right)^{1/2} \leq \left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\right)^{1/4} \to 0$$

by Cauchy-Schwarz inequality. It then follows that

$$n^{-1/2}h_n^{-1/2}\left\{\sum_{j=1}^{h_n}\delta_j^{-1}(j\log j)^2\right\}^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}$$

$$\begin{aligned}&\leq n^{-1/2}h_n^2(\log h_n)^2\left(\sum_{j=1}^{h_n}\delta_j^{-1}\right)^{1/2}=n^{-1/4}h_n^{7/4}(\log h_n)^2n^{-1/4}h_n^{-1/4}\left(\sum_{j=1}^{h_n}\delta_j^{-1}\right)^{1/2}\\&=o(\{n^{-1/2}h_n^{7/2}(\log h_n)^4\}^{1/2})\end{aligned}$$

and

$$\begin{aligned}n^{-1/2}h_n^{-1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^3\right\}^{1/2}\left\{\sum_{j=1}^{h_n}(j\log j)^2\right\}^{1/2}&\leq n^{-1/2}h_n^3(\log h_n)^{5/2}\\&\leq o(n^{-1/2}h_n^{7/2}(\log h_n)^4).\end{aligned}$$

This proves the last assertion. □

S2.4 Weak convergence of WB variance term

This section devote to deriving the weak convergence of the WB variance term related to $\Gamma_{h_n}^{-1}U_n^*$, the second term in the decomposition (S2.2). After a few lemmas about negligible terms are provided, the weak convergence is derived by verifying the Lyapunov condition.

Lemma S9. *If $\|\hat{\beta}_{k_n}-\beta\|=o_{\mathsf{P}}(1)$, then we have that $\mathsf{E}^*[\|\bar{X}\bar{\varepsilon}^*\|^2]=O_{\mathsf{P}}(n^{-2})$, where $\bar{X}\equiv n^{-1}\sum_{i=1}^n X_i$ and $\bar{\varepsilon}^*\equiv n^{-1}\sum_{i=1}^n\varepsilon_i^*$.*

Proof. Note that

$$(\bar{\varepsilon}^*)^2=n^{-2}\sum_{i=1}^n(\varepsilon_i^*)^2+n^{-2}\sum_{i\neq i'}\varepsilon_i^*\varepsilon_{i'}^*,$$

which implies that $\mathsf{E}^*[(\bar{\varepsilon}^*)^2] = n^{-2}\sum_{i=1}^n \hat{\varepsilon}_{i,k_n}^2$. Since

$$\begin{aligned}
n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{i,k_n}^2 &= n^{-1}\sum_{i=1}^n (\varepsilon_i - \langle X_i, \hat{\beta}_{k_n} - \beta\rangle)^2 \\
&\le n^{-1}\sum_{i=1}^n \varepsilon_i^2 + n^{-1}\sum_{i=1}^n \|X_i\|^2\|\hat{\beta}_{k_n} - \beta\|^2 + 2n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i\|\|\hat{\beta}_{k_n} - \beta\| \\
&= \{\mathsf{E}[\varepsilon^2] + o_{\mathsf{P}}(1)\} + \{\mathsf{E}[\|X\|^2] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) + \{\mathsf{E}[\|X\varepsilon\|] + o_{\mathsf{P}}(1)\}o_{\mathsf{P}}(1) \\
&= O_{\mathsf{P}}(1),
\end{aligned}$$

we have that $\mathsf{E}^*[(\bar{\varepsilon}^*)^2] = O_{\mathsf{P}}(n^{-1})$. Finally, since $\bar{X} = O_{\mathsf{P}}(n^{-1/2})$ (cf. Horváth and Kokoszka, 2012, Theorem 2.3), we conclude that $\mathsf{E}^*[\|\bar{X}\bar{\varepsilon}^*\|^2] = \|\bar{X}\|^2\mathsf{E}^*[(\bar{\varepsilon}^*)^2] = O_{\mathsf{P}}(n^{-2})$.

□

Lemma S10. *Suppose that Condition (A4) in the main paper hold. Then, we have*

$$\mathsf{E}^*\left[\frac{n}{s_{h_n}(X_0)}\langle \bar{X}\bar{\varepsilon}^*, \Gamma_{h_n}^{-1}X_0\rangle^2 \Big| X_0\right] = o_{\mathsf{P}}(1),$$

where $\bar{X} \equiv n^{-1}\sum_{i=1}^n X_i$ *and* $\bar{\varepsilon}^* \equiv n^{-1}\sum_{i=1}^n \varepsilon_i^*$.

Proof. Note from Jensen's inequality that

$$n^{-1/2}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1} \le n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\gamma_j^{-1} \le \left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\right)^{1/2} \le \left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\right)^{1/2}$$

From Lemma S9, since $\mathsf{E}[\|\Gamma_{h_n}^{-1}X_0\|^2] = \sum_{j=1}^{h_n}\gamma_j^{-1}$, we have that

$$\begin{aligned}
&\mathsf{E}^*\left[\frac{n}{s_{h_n}(X_0)}\langle \bar{X}\bar{\varepsilon}^*, \Gamma_{h_n}^{-1}X_0\rangle^2 \Big| X_0\right] \\
\le&\{h_n s_{h_n}(X_0)^{-1}\}(n\mathsf{E}^*[\|\bar{X}\bar{\varepsilon}^*\|^2])(h_n^{-1}\|\Gamma_{h_n}^{-1}X_0\|^2)
\end{aligned}$$

$$=O_{\mathsf{P}}(1)O_{\mathsf{P}}(n^{-1})O_{\mathsf{P}}\left(h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)$$
$$=O_{\mathsf{P}}\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)=o_{\mathsf{P}}(n^{-1/2})$$

by Condition (A4). □

We denote the sample cross covariance function between the regressors $\{X_i\}_{i=1}^n$ and the bootstrap errors $\{\varepsilon_i^*\}_{i=1}^n$ by

$$U_n^* \equiv n^{-1}\sum_{i=1}^n (X_i-\bar{X})(\varepsilon_i^*-\bar{\varepsilon}^*) = n^{-1}\sum_{i=1}^n X_i\varepsilon_i^* - \bar{X}\bar{\varepsilon}^*, \tag{S2.12}$$

where $\bar{X}\equiv n^{-1}\sum_{i=1}^n X_i$ and $\bar{\varepsilon}^*\equiv n^{-1}\sum_{i=1}^n \varepsilon_i^*$. Suppose that $\|\hat{\beta}_{k_n}-\beta\| \xrightarrow{\mathsf{P}} 0$ as $n\to\infty$.

Proposition S2. *Suppose that the assumptions in Theorem 1 in the main paper hold. Then, the WB variance term $\sqrt{n/s_{h_n}(X_0)}\langle \Gamma_{h_n}^{-1}U_n^*, X_0\rangle$ uniformly converges to the standard normal distribution in strong sense, i.e.,*

$$\sup_{y\in\mathbb{R}}\left|\mathsf{P}^*\left(\sqrt{\frac{n}{s_{h_n}(X_0)}}\langle \Gamma_{h_n}^{-1}U_n^*, X_0\rangle \le y \middle| X_0\right) - \Phi(y)\right| \xrightarrow{\mathsf{P}} 0,$$

where Φ denotes the cumulative distribution function of the standard normal distribution.

Proof. First, we know that the latter term in (S2.12) related to $\bar{X}\bar{\varepsilon}^*$ is negligible due to Lemma S10. Also, the condition $n^{-\delta/2}h_n^{\delta/2}\sum_{j=1}^{h_n}\lambda_j^{-(2+\delta)/2}=O(1)$ implies the convergence $n^{-\delta}h_n^{(2+\delta)/2}\to 0$ as follows. Since $\sum_{j=1}^\infty \lambda_j<\infty$, we have $j\lambda_j\to 0$ due to Olivier's lemma (cf. Hardy, 1921, Section 173). This then implies $\lambda_j\le cj^{-1}$ for some $c\in(0,\infty)$

and

$$\sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} \geq c \sum_{j=1}^{h_n} j^{(2+\delta)/2} \geq c h_n^{(4+\delta)/2},$$

further indicating that

$$c n^{-\delta/2} h_n^{(4+2\delta)/2} \leq n^{-\delta/2} h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} = O(1).$$

To prove the main part, we write $Z_{i,n}^* \equiv \langle X_i \varepsilon_i^*, \Gamma_{h_n}^{-1} X_0 \rangle$ so that $\mathsf{E}^*[Z_{i,n}^* | X_0] = 0$ and $\mathsf{E}^*[(Z_{i,n}^*)^2 | X_0] = \langle X_i \hat{\varepsilon}_{i,k_n}, \Gamma_{h_n}^{-1} X_0 \rangle^2$.

We first derive that

$$\frac{n^{-1} \hat{v}_n^2}{s_{h_n}(X_0)} \xrightarrow{\mathsf{P}} 1. \tag{S2.13}$$

since

$$\begin{aligned} n^{-1} \hat{v}_n^2 &\equiv n^{-1} \sum_{i=1}^{n} \mathsf{E}^*[(Z_{i,n}^*)^2 | X_0] = n^{-1} \sum_{i=1}^{n} \langle X_i \hat{\varepsilon}_{i,k_n}, \Gamma_{h_n}^{-1} X_0 \rangle^2 \\ &= n^{-1} \sum_{i=1}^{n} \langle (X_i \hat{\varepsilon}_{i,k_n})^{\otimes 2} \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \rangle \\ &= \left\langle \left(n^{-1} \sum_{i=1}^{n} (X_i \hat{\varepsilon}_{i,k_n})^{\otimes 2} - \Lambda \right) \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle + s_{h_n}(X_0) \end{aligned}$$

by Yeon et al. (2024b, Lemma S30).

We will now verify the Lyapunov condition

$$\hat{\mathcal{L}}_n \xrightarrow{\mathsf{P}} 0 \tag{S2.14}$$

with Lyapunov term defined as $\hat{\mathcal{L}}_n = \hat{\mathcal{L}}_{n,\delta} \equiv \hat{v}_n^{-(2+\delta)} \sum_{i=1}^{n} \mathsf{E}^*[|Z_{i,n}^*|^{2+\delta} | X_0]$. The Lyapunov term $\mathcal{L}_n$ is expanded as

$$\hat{\mathcal{L}}_n = \hat{v}_n^{-(2+\delta)} \sum_{i=1}^n \mathsf{E}^*[|Z_{i,n}^*|^{2+\delta}|X_0] \leq \hat{v}_n^{-(2+\delta)} \sum_{i=1}^n \mathsf{E}^*[\|\Lambda_{h_n}^{-1/2} X_i \varepsilon_i^*\|^{2+\delta}|X_0] \|\Lambda_{h_n}^{1/2} \Gamma_{h_n}^{-1} X_0\|^{2+\delta}$$
$$= \left(\frac{\|\Lambda_{h_n}^{1/2}\Gamma_{h_n}^{-1} X_0\|^2}{n^{-1}\hat{v}_n^2} \right)^{(2+\delta)/2} n^{-\delta/2} \mathsf{E}^* \left[n^{-1} \sum_{i=1}^n \|\Lambda_{h_n}^{-1/2} X_i \varepsilon_i^*\|^{2+\delta} \Big| X_0 \right], \qquad \text{(S2.15)}$$

where

$$\frac{\|\Lambda_{h_n}^{-1/2}\Gamma_{h_n}^{-1} X_0\|^2}{n^{-1}\hat{v}_n^2} \leq \frac{s_{h_n}(X_0)}{n^{-1}\hat{v}_n^2} = 1 + o_{\mathsf{P}}(1) = O_{\mathsf{P}}(1)$$

as $n \to \infty$, since $\|\Lambda_{h_n}^{-1/2}\Gamma_{h_n}^{-1} X_0\|^2 = \langle \Lambda_{h_n} \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \rangle \leq \langle \Lambda \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \rangle = s_{h_n}(X_0)$.

To handle the latter term in (S2.15), note by Lyapunov inequality that

$$\left(h_n^{-1} \sum_{j=1}^{h_n} a_j^2 \right)^{1/2} \leq \left(h_n^{-1} \sum_{j=1}^{h_n} a_j^{2+\delta} \right)^{1/(2+\delta)},$$

meaning that $\left(\sum_{j=1}^{h_n} a_j^2\right)^{(2+\delta)/2} \leq h_n^{\delta/2} \sum_{j=1}^{h_n} a_j^{2+\delta}$. We then have

$$\begin{aligned}
&\mathsf{E}^*[\|\Lambda_{h_n}^{-1/2} X_i \varepsilon_i^*\|^{2+\delta}] = \|\Lambda_{h_n}^{-1/2} X_i \hat{\varepsilon}_{i,k_n}\|^{2+\delta} \mathsf{E}[|W|^{2+\delta}] \\
=&\mathsf{E}[|W|^{2+\delta}] \left(\sum_{j=1}^{h_n} \lambda_j^{-1} \langle X \hat{\varepsilon}_{i,k_n}, \psi_j \rangle^2 \right)^{(2+\delta)/2} \\
\leq&\mathsf{E}[|W|^{2+\delta}] h_n^{\delta/2} \sum_{j=1} \lambda_j^{-(2+\delta)/2} |\langle X_i \hat{\varepsilon}_{i,k_n}, \psi_j \rangle|^{2+\delta} \\
\leq&\mathsf{E}[|W|^{2+\delta}] 2^{1+\delta} h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} |\langle X_i (\hat{\varepsilon}_{i,k_n} - \varepsilon_i), \psi_j \rangle|^{2+\delta} \\
&+ \mathsf{E}[|W|^{2+\delta}] 2^{1+\delta} h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} |\langle X_i \varepsilon_i, \psi_j \rangle|^{2+\delta}
\end{aligned} \qquad \text{(S2.16)}$$

from Jensen's inequality with form

$$\left(\frac{x+y}{2} \right)^a \leq \frac{x^a + y^a}{2}, \quad \forall a > 0, x, y \geq 0.$$

Since $\hat{\varepsilon}_{i,k_n} - \varepsilon_i = -\langle \hat{\beta}_{k_n} - \beta, X_i \rangle$, the first term in (S2.16) is

$$
\begin{aligned}
& h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} |\langle X_i(\hat{\varepsilon}_{i,k_n} - \varepsilon_i), \psi_j \rangle|^{2+\delta} \\
= & h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} |\langle X_i^{\otimes 2}(\hat{\beta}_{k_n} - \beta), \psi_j \rangle|^{2+\delta} = h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} |\langle \hat{\beta}_{k_n} - \beta, X_i^{\otimes 2} \psi_j \rangle|^{2+\delta} \\
\leq & h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} \|\hat{\beta}_{k_n} - \beta\|^{2+\delta} \|X_i\|^{4+2\delta}.
\end{aligned}
$$

Again, for $\delta \in (0, 2]$, by Lyapunov inequality we have that

$$
\left(h_n^{-1} \sum_{j=1}^{h_n} a_j^{2+\delta} \right)^{1/(2+\delta)} \leq \left(h_n^{-1} \sum_{j=1}^{h_n} a_j^4 \right)^{1/4},
$$

meaning that $h_n^{\delta/2} \sum_{j=1}^{h_n} a_j^{2+\delta} \leq h_n^{(2+\delta)/4} \left(\sum_{j=1}^{h_n} a_j^4 \right)^{(2+\delta)/4}$. The second term in (S2.16) is then bounded as

$$
h_n^{\delta/2} \sum_{j=1} \lambda_j^{-(2+\delta)/2} |\langle X_i \varepsilon_i, \psi_j \rangle|^{2+\delta} \leq h_n^{(2+\delta)/4} \left(\sum_{j=1}^{h_n} \lambda_j^{-2} \langle X_i \varepsilon_i, \psi_j \rangle^4 \right)^{(2+\delta)/4}.
$$

It follows that the latter part in (S2.15) is bounded as

$$
\begin{aligned}
& n^{-\delta/2} \mathsf{E}^* \left[n^{-1} \sum_{i=1}^{n} \|\Lambda_{h_n}^{-1/2} X_i \varepsilon_i^*\|^{2+\delta} \Big| X_0 \right] \\
\leq & \mathsf{E}[|W|^{2+\delta}] 2^{1+\delta} \left(n^{-\delta/2} h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} \right) \left(n^{-1} \sum_{i=1}^{n} \|X_i\|^{4+2\delta} \right) \|\hat{\beta}_{k_n} - \beta\|^{2+\delta} \qquad \text{(S2.17)} \\
& + \mathsf{E}[|W|^{2+\delta}] 2^{1+\delta} n^{-\delta/2} h_n^{(2+\delta)/4} n^{-1} \sum_{i=1}^{n} \left(\sum_{j=1}^{h_n} \lambda_j^{-2} \langle X_i \varepsilon_i, \psi_j \rangle^4 \right)^{(2+\delta)/4}.
\end{aligned}
$$

The first term in (S2.17) converges to zero in probability since $n^{-\delta/2} h_n^{\delta/2} \sum_{j=1}^{h_n} \lambda_j^{-(2+\delta)/2} = O(1)$, $\mathsf{E}[\|X\|^{4+2\delta}] < \infty$, and $\|\hat{\beta}_{k_n} - \beta\| \xrightarrow{\mathsf{P}} 0$. In addition, since $\sup_{j \in \mathbb{N}} \lambda_j^{-2} \mathsf{E}[\langle X\varepsilon, \psi_j \rangle^4] <$

∞, by Lyapunov inequality,

$$\mathsf{E}\left[\left(\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i,\psi_j\rangle^4\right)^{(2+\delta)/4}\right]\le\left(\mathsf{E}\left[\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i,\psi_j\rangle^4\right]\right)^{(2+\delta)/4}\le Ch_n^{(2+\delta)/4},$$

and hence, the second term in (S2.17) is bounded as

$$\mathsf{E}\left[n^{-\delta/2}h_n^{(2+\delta)/4}n^{-1}\sum_{i=1}^{n}\left(\sum_{j=1}^{h_n}\lambda_j^{-2}\langle X_i\varepsilon_i,\psi_j\rangle^4\right)^{(2+\delta)/4}\right]\le Cn^{-\delta/2}h_n^{(2+\delta)/2}.$$

Therefore, as $n\to\infty$, since $n^{-\delta/2}h_n^{(2+\delta)/2}\to 0$, then the second term converges to zero in probability, which verifies (S2.14).

Finally, by combining Slutsky's theorem, Polya's theorem (Athreya and Lahiri, 2006, Theorem 9.1.4), a subsequence argument (cf. Billingsley, 1999, Theorem 20.5), Lemma S10, and (S2.13), (S2.14), we conclude the desired result. □

S2.5 WB scaling term

We will derive the consistency of the WB scaling $\hat{s}^*_{h_n}(X_0)=\hat{s}^{WB*}_{h_n}(X_0)$ for the sample scaling $\hat{s}_{h_n}(X_0)$ (or the population scaling $s_{h_n}(X_0)$). We follow the same argument as what was used by Yeon et al. (2024b, Section S3.4).

We first explain that it suffices to show the consistency when dropping the means $\bar{X}$ from $\hat{s}_{h_n}(X_0)$ and $\bar{X}$ from $\hat{s}^*_{h_n}(X_0)$. For this, we consider another scaling $\hat{r}^*_{h_n}(X_0)$, akin to $\hat{s}^*_{h_n}(X_0)$, by replacing the sample mean $\bar{X}$ in $\hat{s}_{h_n}(X_0)$ with the population mean $\mathsf{E}[X]$ as

$$\hat{r}^*_{h_n}(X_0)\equiv\langle\hat{\Lambda}^*_n\hat{\Gamma}^{-1}_{h_n}(X_0-\mathsf{E}[X]),\hat{\Gamma}^{-1}_{h_n}(X_0-\mathsf{E}X)\rangle$$

$$= \|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(X_0 - \mathsf{E}X)\|^2.$$

Then, we show the asymptotic equivalence between two sample scaling terms $\hat{s}^*_{h_n}(X_0)$ and $\hat{r}^*_{h_n}(X_0)$. We note the following inequality

$$\left|\frac{\hat{s}^*_{h_n}(X_0)}{\hat{r}^*_{h_n}(X_0)} - 1\right| \leq \frac{\|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(\bar{X} - \mathsf{E}[X])\|^2}{\|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(X_0 - \mathsf{E}[X])\|^2} + 2\frac{\|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(\bar{X} - \mathsf{E}[X])\|}{\|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(X_0 - \mathsf{E}[X])\|},$$

where

$$\begin{aligned}
&\frac{\|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(\bar{X} - \mathsf{E}[X])\|^2}{\|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(X_0 - \mathsf{E}[X])\|^2} \\
=&h_n^{-1}\|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(\bar{X} - \mathsf{E}[X])\|^2\{h_n s_{h_n}(X_0)^{-1}\}\frac{s_{h_n}(X_0)}{\hat{r}_{h_n}(X_0)}.
\end{aligned}$$

We will derive

$$h_n^{-1}\|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(\bar{X} - \mathsf{E}[X])\|^2 = o_{\mathsf{P}^*}(1). \tag{S2.18}$$

Then, by Condition (S), if the ratio $s_{h_n}(X_0)/\hat{r}^*_{h_n}(X_0)$ is bounded above, we can deduce that two bootstrap scaling terms$\hat{s}^*_{h_n}(X_0)$ and $\hat{r}^*_{h_n}(X_0)$ are equivalent. The convergence in (S2.18) indeed holds, as

$$\begin{aligned}
&h_n^{-1}\|(\hat{\Lambda}_n^*)^{1/2}\hat{\Gamma}_{h_n}^{-1}(\bar{X} - \mathsf{E}[X])\|^2 \\
\leq&CO_{\mathsf{P}^*}(1)h_n^{-1}\{(\|\hat{\Gamma}^{-1} - \Gamma_{h_n}^{-1}\|_\infty\|\bar{X} - \mathsf{E}[X]\|)^2 + \|\Gamma_{h_n}^{-1}(\bar{X} - \mathsf{E}[X])\|^2\} \\
=&O_{\mathsf{P}^*}\left(\left(n^{-1}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1}j\log j\right)^2 + n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right) \\
=&o_{\mathsf{P}^*}(n^{-1/2}h_n^{-1/2}),
\end{aligned}$$

where the first inequality can be obtained by using the triangular inequality; the second inequality is due to Yeon et al. (2024b, Lemma S20) (which requires Conditions (A0)-(A4)), $\bar{X} - \mathsf{E}[X] = O_{\mathsf{P}}(n^{-1/2})$ (Hsing and Eubank, 2015, Chapter 7), and $\mathsf{E}[\|\Gamma_{h_n}^{-1}(\bar{X} - \mathsf{E}[X])\|^2] = n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}$; and the last inequality comes from Cauchy-Schwarz inequality along with Condition (A4) because

$$\begin{aligned}
&\left\{n^{-1}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1}(j\log j)^2\right\}^2\\
\leq& n^{-1/2}h_n^{-1/2}\left(n^{-1}\sum_{j=1}^{h_n}\delta_j^{-2}\right)\left\{n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}(j\log j)^2\right\}\\
=&o_{\mathsf{P}}(n^{-1/2}h_n^{-1/2})
\end{aligned}$$

and

$$n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1} \leq n^{-1/2}h_n^{-1/2}\left(n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-2}\right)^{1/2} = o_{\mathsf{P}}(n^{-1/2}h_n^{-1/2}).$$

Under the assumptions for Theorem 1 in the main paper, since the ratio $\hat{r}_{h_n}^*(X_0)/s_{h_n}(X_0)$ converges to 1 (conditionally on X_0) by Proposition S3, we have

$$\frac{\hat{s}_{h_n}^*(X_0)}{\hat{r}_{h_n}^*(X_0)} = 1 + o_{\mathsf{P}^*(\cdot|X_0)}(1).$$

Now, without loss of generality, it suffices to show the consistency of the ratio of $\hat{s}_{h_n}^*(X_0)$ over $\hat{s}_{h_n}(X_0)$ to 1 conditionally on X_0, upon dropping the the sample mean $\bar{X}$ from $\hat{s}_{h_n}(X_0)$ and $\hat{s}_{h_n}^*(X_0)$ under mean-zero condition $\mathsf{E}[Y] = 0$ and $\mathsf{E}[X] = 0$. Then, the WB scaling $\hat{s}_{h_n}^*(X_0) = \hat{s}_{h_n}^{WB*}(X_0)$ can be decomposed as follows:

$$\hat{s}_{h_n}^*(X_0) = \langle\hat{\Lambda}_n^*\hat{\Gamma}_{h_n}^{-1}X_0, \hat{\Gamma}_{h_n}^{-1}X_0\rangle$$

$$
\begin{aligned}
=&\langle\hat{\Lambda}_n\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0,\hat{\Gamma}_{h_n}^{-1}X_0\rangle+\langle\hat{\Lambda}_n^*\Gamma_{h_n}^{-1}X_0,\hat{\Gamma}_{h_n}^{-1}X_0\rangle\\
=&\langle\hat{\Lambda}_n^*\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle+\langle\hat{\Lambda}_n^*\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0,\Gamma_{h_n}^{-1}X_0\rangle\\
&+\langle\hat{\Lambda}_n^*\Gamma_{h_n}^{-1}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle+\langle\hat{\Lambda}_n^*\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle\\
=&\langle\hat{\Lambda}_n^*\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle\\
&+2\langle\hat{\Lambda}_n^*\Gamma_{h_n}^{-1}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle+\langle\hat{\Lambda}_n^*\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle\\
=&\langle(\hat{\Lambda}_n^*-\Lambda)\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle+\langle\Lambda\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle\\
&+2\langle(\hat{\Lambda}_n^*-\Lambda)\Gamma_{h_n}^{-1}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle+2\langle\Lambda\Gamma_{h_n}^{-1}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle\\
&+\langle(\hat{\Lambda}_n^*-\Lambda)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle+\langle\Lambda\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&\left|\frac{\hat{s}_{h_n}^*(X_0)}{s_{h_n}(X_0)}-1\right|\\
\le& s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_n^*-\Lambda)\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle| &\text{(S2.19)}\\
&+s_{h_n}(X_0)^{-1}|\langle\Lambda\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle| &\text{(S2.20)}\\
&+2s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_n^*-\Lambda)\Gamma_{h_n}^{-1}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle| &\text{(S2.21)}\\
&+2s_{h_n}(X_0)^{-1}|\langle\Lambda\Gamma_{h_n}^{-1}X_0,\{\hat{\Gamma}_{h_n}^{-1}-\Gamma_{h_n}^{-1}\}X_0\rangle| &\text{(S2.22)}\\
&+s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_n^*-\Lambda)\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\rangle| &\text{(S2.23)}
\end{aligned}
$$

Also, with $\varepsilon_i^*=Y_i^*-\langle\hat{\beta}_{g_n},X_i\rangle$, we see by putting $\hat{\varepsilon}_{i,k_n,g_n}^*=\varepsilon_i^*-\langle\hat{\beta}_{k_n}^*-\hat{\beta}_{g_n},X_i\rangle$ that

$$
\hat{\Lambda}_n^*-\Lambda=n^{-1}\sum_{i=1}^n(X_i\hat{\varepsilon}_{i,k_n,g_n}^*)^{\otimes 2}-(\overline{X\hat{\varepsilon}^*})^{\otimes 2}-\Lambda
$$

$$= \tilde{\Lambda}_n^* - \Lambda + n^{-1}\sum_{i=1}^n X_i^{\otimes 2}\langle \hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i\rangle^2$$
$$- 2n^{-1}\sum_{i=1}^n (X_i\varepsilon_i^* \otimes X_i)\langle \hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i\rangle - \overline{X\hat{\varepsilon}^*}^{\otimes 2} \tag{S2.24}$$

where $\tilde{\Lambda}_n^* \equiv n^{-1}\sum_{i=1}^n (X_i\varepsilon_i^*)^{\otimes 2}$ and

$$\overline{X\hat{\varepsilon}^*} \equiv n^{-1}\sum_{i=1}^n X_i\hat{\varepsilon}_{i,k_n,g_n}^* = n^{-1}\sum_{i=1}^n X_i\varepsilon_i^* - n^{-1}\sum_{i=1}^n \langle \hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i\rangle X_i$$
$$= \overline{X\varepsilon^*} - (\tilde{\Gamma}_n - \Gamma)(\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}) - \Gamma(\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}) \tag{S2.25}$$

with $\overline{X\varepsilon^*} \equiv n^{-1}\sum_{i=1}^n X_i\varepsilon_i^*$. Throughout the section, we suppose that the assumptions in Theorem 1 in the main paper hold.

Lemma S11. *For $\tilde{\Lambda}_n^* \equiv n^{-1}\sum_{i=1}^n (X_i\varepsilon_i^*)^{\otimes 2}$, we have $\mathsf{E}^*[\|\tilde{\Lambda}_n^* - \Lambda\|_{HS}] = o_{\mathsf{P}}(1)$.*

Proof. On the other hand, we observe

$$\tilde{\Lambda}_n^* - \Lambda = n^{-1}\sum_{i=1}^n X_i^{\otimes 2}\hat{\varepsilon}_{i,k_n}^2 W_i^2 - \Lambda$$
$$= n^{-1}\sum_{i=1}^n \{(X_i\varepsilon_i)^{\otimes 2}W_i^2 - \Lambda\} + n^{-1}\sum_{i=1}^n X_i^{\otimes 2}\langle \hat{\beta}_{k_n} - \beta, X_i\rangle^2 W_i^2 \tag{S2.26}$$
$$- 2n\sum_{i=1}^n [(X_i\varepsilon_i) \otimes \{X_i^{\otimes 2}(\hat{\beta}_{k_n} - \beta)\}]W_i^2.$$

The terms $\{(X_i\varepsilon_i)^{\otimes 2}W_i^2 - \Lambda\}_{i=1}^n$ in the first sum of (S2.26) are iid with zero mean and finite variance as

$$\mathsf{E}[(X_i\varepsilon_i)^{\otimes 2}W_i^2 - \Lambda] = \Lambda\mathsf{E}[W_i^2] - \Lambda = 0,$$

$$\mathsf{E}[\|(X_i\varepsilon_i)^{\otimes 2}W_i^2 - \Lambda)\|_{HS}^2] \leq C\mathsf{E}[\|X\varepsilon\|^4]\mathsf{E}[W^4] < \infty.$$

It then follows that

$$\mathsf{E}\left[\left\|n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)^{\otimes 2}W_i^2-\Lambda\}\right\|_{HS}^2\right]=O(n^{-1}),$$

which implies that

$$\mathsf{E}^*\left[\left\|n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)^{\otimes 2}W_i^2-\Lambda\}\right\|_{HS}\right]=O_{\mathsf{P}}(n^{-1/2}).$$

The last two terms in (S2.26) also converge to zero as

$$\begin{aligned}&\mathsf{E}^*\left[\left\|n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}\langle\hat{\beta}_{k_n}-\beta,X_i\rangle^2W_i^2\right\|_{HS}\right]\\&\le\left(n^{-1}\sum_{i=1}^{n}\|X_i\|^4\right)\mathsf{E}[W^2]\|\hat{\beta}_{k_n}-\beta\|^2\xrightarrow{\mathsf{P}}0\end{aligned}$$

and

$$\begin{aligned}&\mathsf{E}^*\left[\left\|n^{-1}\sum_{i=1}^{n}[(X_i\varepsilon_i)\otimes\{X_i^{\otimes 2}(\hat{\beta}_{k_n}-\beta)\}]W_i^2\right\|_{HS}\right]\\&\le\left(n^{-1}\sum_{i=1}^{n}\|X_i\varepsilon_i\|^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^{n}\|X_i\|^4\right)^{1/2}\mathsf{E}[W^2]\|\hat{\beta}_{k_n}-\beta\|\xrightarrow{\mathsf{P}}0.\end{aligned}$$

Thus, the proof is complete. □

Lemma S12. *For* $\hat{\Lambda}_n^*\equiv n^{-1}\sum_{i=1}^{n}(X_i\hat{\varepsilon}^*_{i,k_n,g_n})^{\otimes 2}$, *we have* $\|\hat{\Lambda}_n^*-\Lambda\|_\infty=o_{\mathsf{P}^*}(1)$

Proof. We use the decomposition of $\hat{\Lambda}_n^*-\Lambda$ in (S2.24) for the proof. The first term in (S2.24) converges to zero by Lemma S11. The second term in (S2.24) also converges to zero as

$$\left\|n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}\langle\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n},X_i\rangle^2\right\|\le n^{-1}\sum_{i=1}^{n}\|X_i\|^4\|\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n}\|^2=O_{\mathsf{P}}(1)o_{\mathsf{P}^*}(1)=o_{\mathsf{P}^*}(1)$$

due to Theorem S2 along with the fact that $n^{-1}\sum_{i=1}^n \|X_i\|^4 = \mathsf{E}[\|X\|^4]+o_{\mathsf{P}}(1) = O_{\mathsf{P}}(1)$.

The third term in (S2.24) is bounded as

$$\begin{aligned}
&\left\| n^{-1}\sum_{i=1}^n \{(X_i\varepsilon_i^*)\otimes X_i\}\langle \hat\beta_{k_n}^* - \hat\beta_{g_n}, X_i\rangle \right\|_\infty \\
\le & n^{-1}\sum_{i=1}^n \|X_i\varepsilon^*\|\|X_i\|^2\|\hat\beta_{k_n}^* - \hat\beta_{g_n}\| \\
\le & \left(n^{-1}\sum_{i=1}^n \|X_i\varepsilon^*\|^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^n \|X_i\|^4\right)\|\hat\beta_{k_n}^* - \hat\beta_{g_n}\|.
\end{aligned}$$

It then follows that

$$\left\| n^{-1}\sum_{i=1}^n \{(X_i\varepsilon_i^*)\otimes X_i\}\langle \hat\beta_{k_n}^* - \hat\beta_{g_n}, X_i\rangle \right\|_\infty = O_{\mathsf{P}^*}(1)O_{\mathsf{P}}(1)o_{\mathsf{P}^*}(1) = o_{\mathsf{P}^*}(1)$$

since

$$\begin{aligned}
\mathsf{E}^*\left[n^{-1}\sum_{i=1}^n \|X_i\varepsilon^*\|^2\right] &= n^{-1}\sum_{i=1}^n \|X_i\hat\varepsilon_{i,k_n}\|^2 = n^{-1}\sum_{i=1}^n \|X_i(\varepsilon_i - \langle \hat\beta_{k_n} - \beta, X_i\rangle)\|^2 \\
&\le 2n^{-1}\sum_{i=1}^n \|X_i\varepsilon_i\|^2 + 2n^{-1}\sum_{i=1}^n \|X_i\|^4\|\hat\beta_{k_n} - \beta\|^2 = O_{\mathsf{P}}(1).
\end{aligned}$$

Finally, the last term in (S2.24) converges to zero as

$$\begin{aligned}
\mathsf{E}^*[\|\overline{X\varepsilon^*}\|^2] = \mathsf{E}^*\left[\left\|n^{-1}\sum_{i=1}^n X_i\varepsilon_i^*\right\|^2\right] &= \left\|n^{-1}\sum_{i=1}^n X_i\hat\varepsilon_{i,k_n}\right\|^2 \\
&\le 2\left\|n^{-1}\sum_{i=1}^n X_i\varepsilon_i\right\|^2 + 2\left\|n^{-1}\sum_{i=1}^n X_i\langle X_i, \hat\beta_{k_n} - \beta\rangle\right\|^2 \\
&\le 2\left\|n^{-1}\sum_{i=1}^n X_i\varepsilon_i\right\|^2 + 4\left\|(\tilde\Gamma_n - \Gamma)(\hat\beta_{k_n} - \beta)\right\|^2 + 4\|\Gamma(\hat\beta_{k_n} - \beta)\|^2 \\
&= O_{\mathsf{P}}(n^{-1}) + O_{\mathsf{P}}(n^{-1}\|\hat\beta_{k_n} - \beta\|^2) + O_{\mathsf{P}}(\|\hat\beta_{k_n} - \beta\|^2)
\end{aligned}$$

$$= o_{\mathsf{P}}(1),$$

which implies that $\overline{X\hat{\varepsilon}^*} = o_{\mathsf{P}^*}(1)$ due to the decomposition (S2.25). □

Lemma S13. *For the term in (S2.19), if* $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2} = O(1)$, *then we have*

$$s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_n^* - \Lambda)\{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\}X_0, \{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\}X_0\rangle| = o_{\mathsf{P}^*}(1).$$

Proof. This follows since

$$\begin{aligned}
&s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_n^* - \Lambda)\{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\}X_0, \{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\}X_0\rangle| \\
=&O_{\mathsf{P}}(h_n^{-1})o_{\mathsf{P}^*}(1)O_{\mathsf{P}}\left(\left\{n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right\}^2\right) \\
=&o_{\mathsf{P}^*}\left(\left\{n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right\}^2\right)
\end{aligned}$$

by Lemma S12 and Yeon et al. (2024b, Lemma S20). □

Lemma S14. *For the term in (S2.20), if* $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2} = o(1)$, *then we have*

$$s_{h_n}(X_0)^{-1}|\langle\Lambda\{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\}X_0, \{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\}X_0\rangle| = o_{\mathsf{P}^*}(1).$$

Proof. This follows since

$$s_{h_n}(X_0)^{-1}|\langle\Lambda\{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\}X_0, \{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\}X_0\rangle|$$

$$
\begin{aligned}
&= O_{\mathsf{P}}(h_n^{-1})O(1)O_{\mathsf{P}}\left(\left\{n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right\}^2\right)\\
&= O_{\mathsf{P}}\left(\left\{n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right\}^2\right)
\end{aligned}
$$

by Yeon et al. (2024b, Lemma S20). □

Lemma S15. *For the term in (S2.21), if* $n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2} = O(1)$, *then we have*

$$
s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_n^* - \Lambda)\Gamma_{h_n}^{-1}X_0, \{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\}X_0\rangle| = o_{\mathsf{P}^*}(1).
$$

Proof. We first recall that $Q_n \equiv (\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1})X_0 = O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)$ by Yeon et al. (2024b, Lemma S20) By the decompositions in (S2.24) and (S2.26), we observe that

$$
\begin{aligned}
& s_{h_n}(X_0)^{-1}|\langle(\hat{\Lambda}_n^* - \Lambda)\Gamma_{h_n}^{-1}X_0, Q_n\rangle| \\
\leq & s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)^{\otimes 2}W_i^2 - \Lambda\}\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right| \qquad \text{(S2.27)}\\
& + s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}\langle\hat{\beta}_{k_n} - \beta, X_i\rangle^2 W_i^2\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right| \qquad \text{(S2.28)}\\
& + 2s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}[\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}_{k_n} - \beta, X_i\rangle]W_i^2\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right| \qquad \text{(S2.29)}\\
& + s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}\langle\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i\rangle^2\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right| \qquad \text{(S2.30)}\\
& + 2s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i^*)\otimes X_i\}\langle\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i\rangle\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right| \qquad \text{(S2.31)}
\end{aligned}
$$

$$+ s_{h_n}(X_0)^{-1}\left|\left\langle \overline{X\hat{\varepsilon}^*}^{\otimes 2}\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right| \tag{S2.32}$$

We check the convergences of each term in the above display.

For the term in (S2.27), recall from the proof of Lemma S11 that

$$\mathsf{E}^*\left[\left\|n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)^{\otimes 2}W_i^2-\Lambda\}\right\|_{HS}\right]=O_{\mathsf{P}}(n^{-1/2}).$$

This implies that

$$\begin{aligned}
&\mathsf{E}^*\left[s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}\{(X_i\varepsilon_i)^{\otimes 2}W_i^2-\Lambda\}\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right|\,\middle|X_0\right]\\
=&O_{\mathsf{P}}(h_n^{-1})O_{\mathsf{P}}(n^{-1/2})O_{\mathsf{P}}\left(\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right)O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\\
=&O_{\mathsf{P}}\left(\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right)O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\\
=&o_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)
\end{aligned}$$

under Condition (A4) (which implies $n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}=o(1)$).

The term in (S2.28) is bounded as

$$\begin{aligned}
&s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}\langle\hat{\beta}_{k_n}-\beta, X_i\rangle^2 W_i^2\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right|\\
=&s_{h_n}(X_0)^{-1}\left|n^{-1}\sum_{i=1}^{n}W_i^2\langle X_i^{\otimes 2}\hat{\beta}_{k_n}-\beta,\hat{\beta}_{k_n}-\beta\rangle\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i, Q_n\rangle\right|\\
=&s_{h_n}(X_0)^{-1}\left|\left\langle\left(n^{-1}\sum_{i=1}^{n}W_i^2X_i^{\otimes 2}\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i,Q_n\rangle\right)\hat{\beta}_{k_n}-\beta,\hat{\beta}_{k_n}-\beta\right\rangle\right|\\
\le&s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}W_i^2\|X_i\|^3|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle|\right)\|Q_n\|\|\hat{\beta}_{k_n}-\beta\|^2.
\end{aligned}$$

We note from Cauchy-Schwarz inequality that

$$\begin{aligned}
&\mathsf{E}\left[n^{-1}\sum_{i=1}^{n} W_i^2\|X_i\|^3|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle|\right] \\
=&n^{-1}\sum_{i=1}^{n}\mathsf{E}[\|X_i\|^3|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle|] \\
\leq&n^{-1}\sum_{i=1}^{n}\mathsf{E}[\|X_i\|^4]^{1/2}\mathsf{E}[\|X_i\|^2\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2]^{1/2} \\
\leq&n^{-1}\mathsf{E}[\|X\|^4]^{1/2}\sum_{i=1}^{n}\mathsf{E}[\|X_i\|^2\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2]^{1/2}.
\end{aligned}$$

Since the FPC scores $\{\xi_j\}$ are uncorrelated random variables with mean zero and variance one, we have from the independence between $\mathcal{X}_n \equiv \{X_i\}_{i=1}^n$ and X_0 that

$$\begin{aligned}
\mathsf{E}[\|X_i\|^2\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle^2] &= \sum_{j=1}^{h_n}\gamma_j^{-1}\mathsf{E}[\|X_i\|^2\langle X_i, \phi_j\rangle^2] \\
&\leq \sum_{j=1}^{h_n}\gamma_j^{-1}\mathsf{E}[\|X_i\|^4]^{1/2}\mathsf{E}[\langle X_i, \phi_j\rangle^4]^{1/2} \leq Ch_n. \qquad \text{(S2.33)}
\end{aligned}$$

It then holds that $\mathsf{E}\left[n^{-1}\sum_{i=1}^{n} W_i^2\|X_i\|^3|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle|\right] = O(h_n^{1/2})$, which implies that

$$\mathsf{E}^*\left[n^{-1}\sum_{i=1}^{n} W_i^2\|X_i\|^3|\langle X_i, \Gamma_{h_n}^{-1}X_0\rangle|\Big|X_0\right] = O_{\mathsf{P}}(h_n^{1/2}).$$

It follows that

$$\begin{aligned}
&\mathsf{E}^*\left[s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n} X_i^{\otimes 2}\langle\hat{\beta}_{k_n}-\beta, X_i\rangle^2 W_i^2\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right|\Big|X_0\right] \\
=&O_{\mathsf{P}}(h_n^{-1})O_{\mathsf{P}}(h_n^{1/2})O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\|\hat{\beta}_{k_n}-\beta\|^2 \\
=&O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\|\hat{\beta}_{k_n}-\beta\|^2.
\end{aligned}$$

The term in (S2.29) is similarly bounded as

$$
\begin{aligned}
& s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}[\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}_{k_n}-\beta, X_i\rangle]W_i^2\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right| \\
=& s_{h_n}(X_0)^{-1}\left|n^{-1}\sum_{i=1}^{n}W_i^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i\varepsilon_i, Q_n\rangle\langle\hat{\beta}_{k_n}-\beta, X_i\rangle\right| \\
=& s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}W_i^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i\varepsilon_i, Q_n\rangle X_i, \hat{\beta}_{k_n}-\beta\right\rangle\right| \\
\leq& s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}W_i^2|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle|\|X_i\varepsilon_i\|\|X_i\|\right)\|Q_n\|\|\hat{\beta}_{k_n}-\beta\|.
\end{aligned}
$$

We note from Cauchy-Schwarz inequality and (S2.33) that

$$
\begin{aligned}
& \mathsf{E}\left[n^{-1}\sum_{i=1}^{n}W_i^2|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle|\|X_i\varepsilon_i\|\|X_i\|\right] \\
=& n^{-1}\sum_{i=1}^{n}\mathsf{E}[|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle|\|X_i\varepsilon_i\|\|X_i\|] \\
\leq& n^{-1}\sum_{i=1}^{n}\mathsf{E}[\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2]^{1/2}\mathsf{E}[\|X_i\varepsilon_i\|^2]^{1/2} \\
=& O(h_n^{1/2}),
\end{aligned}
$$

which implies that $\mathsf{E}^*\left[n^{-1}\sum_{i=1}^{n}W_i^2|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle|\|X_i\varepsilon_i\|\|X_i\|\middle|X_0\right] = O_{\mathsf{P}}(h_n^{1/2})$. It then holds that

$$
\begin{aligned}
& \mathsf{E}^*\left[s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}[\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}_{k_n}-\beta, X_i\rangle]W_i^2\Gamma_{h_n}^{-1}X_0, Q_n\right\rangle\right|\middle|X_0\right] \\
=& O_{\mathsf{P}}(h_n^{-1})O_{\mathsf{P}}(h_n^{1/2})O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\|\hat{\beta}_{k_n}-\beta\| \\
=& o_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\|\hat{\beta}_{k_n}-\beta\|.
\end{aligned}
$$

The terms in (S2.30) and (S2.31) are respectively bounded in a similar way to the bounds for (S2.28) and (S2.29) as

$$
\begin{aligned}
& s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n} X_i^{\otimes 2}\langle\hat\beta^*_{k_n}-\hat\beta_{g_n},X_i\rangle^2\Gamma_{h_n}^{-1}X_0,Q_n\right\rangle\right| \\
=& s_{h_n}(X_0)^{-1}\left|n^{-1}\sum_{i=1}^{n}\langle X_i^{\otimes 2}\hat\beta^*_{k_n}-\hat\beta_{g_n},\hat\beta^*_{k_n}-\hat\beta_{g_n}\rangle\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i,Q_n\rangle\right| \\
=& s_{h_n}(X_0)^{-1}\left|\left\langle\left(n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i,Q_n\rangle\right)\hat\beta^*_{k_n}-\hat\beta_{g_n},\hat\beta^*_{k_n}-\hat\beta_{g_n}\right\rangle\right| \\
\leq& s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}\|X_i\|^3|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle|\right)\|Q_n\|\|\hat\beta^*_{k_n}-\hat\beta_{g_n}\|^2 \\
=& O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\|\hat\beta^*_{k_n}-\hat\beta_{g_n}\|^2
\end{aligned}
$$

and

$$
\begin{aligned}
& s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}[\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat\beta^*_{k_n}-\hat\beta_{g_n},X_i\rangle]\Gamma_{h_n}^{-1}X_0,Q_n\right\rangle\right| \\
=& s_{h_n}(X_0)^{-1}\left|n^{-1}\sum_{i=1}^{n}\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i\varepsilon_i,Q_n\rangle\langle\hat\beta^*_{k_n}-\hat\beta_{g_n},X_i\rangle\right| \\
=& s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i\varepsilon_i,Q_n\rangle X_i,\hat\beta^*_{k_n}-\hat\beta_{g_n}\right\rangle\right| \\
\leq& s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle|\|X_i\varepsilon_i\|\|X_i\|\right)\|Q_n\|\|\hat\beta^*_{k_n}-\hat\beta_{g_n}\| \\
=& O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\|\hat\beta^*_{k_n}-\hat\beta_{g_n}\|.
\end{aligned}
$$

We will prove the term in (S2.32) converges to zero as

$$
s_{h_n}(X_0)^{-1}\left|\left\langle\overline{X\hat\varepsilon^*}^{\otimes 2}\Gamma_{h_n}^{-1}X_0,Q_n\right\rangle\right| = s_{h_n}(X_0)^{-1}|\langle\overline{X\hat\varepsilon^*},\Gamma_{h_n}^{-1}X_0\rangle||\langle\overline{X\hat\varepsilon^*},Q_n\rangle|
$$

$$= o_{\mathsf{P}^*}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)$$

based on the following decomposition of $\overline{X\hat{\varepsilon}^*}$:

$$\begin{aligned}
\overline{X\hat{\varepsilon}^*} &\equiv n^{-1}\sum_{i=1}^{n}X_i\hat{\varepsilon}^*_{i,k_n,g_n} = n^{-1}\sum_{i=1}^{n}X_i\varepsilon_i^* - n^{-1}\sum_{i=1}^{n}\langle\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n},X_i\rangle X_i\\
&= n^{-1}\sum_{i=1}^{n}W_iX_i\hat{\varepsilon}_{i,k_n} - \tilde{\Gamma}_n(\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n})\\
&= n^{-1}\sum_{i=1}^{n}W_iX_i\varepsilon_i - n^{-1}\sum_{i=1}^{n}W_iX_i^{\otimes 2}(\hat{\beta}_{k_n}-\beta) - n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}(\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n}). \quad \text{(S2.34)}
\end{aligned}$$

Since $\{W_iX_i\varepsilon_i\}_{i=1}^n$ are iid with zero mean and finite variance, we have

$$\mathsf{E}\left[\left\|n^{-1}\sum_{i=1}^{n}W_iX_i\varepsilon_i\right\|^2\right] = O(n^{-1}),$$

which implies that $\mathsf{E}^*\left[\|n^{-1}\sum_{i=1}^{n}W_iX_i\varepsilon_i\|\right] = O_{\mathsf{P}}(n^{-1/2})$. This implies that

$$\begin{aligned}
&\mathsf{E}^*\left[s_{h_n}(X_0)^{-1/2}\left|\left\langle n^{-1}\sum_{i=1}^{n}W_iX_i\varepsilon_i, \Gamma_{h_n}^{-1}X_0\right\rangle\right|\Big|X_0\right]\\
=&O_{\mathsf{P}}(h_n^{-1/2})O_{\mathsf{P}}(n^{-1/2})O_{\mathsf{P}}\left(\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right)\\
=&O_{\mathsf{P}}\left(\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right) = o_{\mathsf{P}}(1)
\end{aligned}$$

under Condition (A4) and

$$\begin{aligned}
&\mathsf{E}^*\left[s_{h_n}(X_0)^{-1/2}\left|\left\langle n^{-1}\sum_{i=1}^{n}W_iX_i\varepsilon_i, Q_n\right\rangle\right|\Big|X_0\right]\\
=&O_{\mathsf{P}}(h_n^{-1/2})O_{\mathsf{P}}(n^{-1/2})O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)
\end{aligned}$$

$$=O_{\mathsf{P}}(n^{-1/2})O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right).$$

Similarly, since $\{W_iX_i^{\otimes 2}\}_{i=1}^n$ are iid with zero mean and finite variance, we have

$$\mathsf{E}\left[\left\|n^{-1}\sum_{i=1}^n W_iX_i^{\otimes 2}\right\|_{HS}^2\right]=O(n^{-1}),$$

which implies that $\mathsf{E}^*\left[\left\|n^{-1}\sum_{i=1}^n W_iX_i^{\otimes 2}\right\|_{HS}\right]=O_{\mathsf{P}}(n^{-1/2})$. This implies that

$$\begin{aligned}&\mathsf{E}^*\left[s_{h_n}(X_0)^{-1/2}\left|\left\langle n^{-1}\sum_{i=1}^n W_iX_i^{\otimes 2}(\hat\beta_{k_n}-\beta),\Gamma_{h_n}^{-1}X_0\right\rangle\right|\Big|X_0\right]\\ =&O_{\mathsf{P}}(h_n^{-1/2})O_{\mathsf{P}}(n^{-1/2})O_{\mathsf{P}}\left(\left(\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right)\|\hat\beta_{k_n}-\beta\|\\ =&O_{\mathsf{P}}\left(\left(n^{-1}h_n^{-1}\sum_{j=1}^{h_n}\gamma_j^{-1}\right)^{1/2}\right)\|\hat\beta_{k_n}-\beta\|=o_{\mathsf{P}}(1)\end{aligned}$$

under Condition (A4) and

$$\begin{aligned}&\mathsf{E}^*\left[s_{h_n}(X_0)^{-1/2}\left|\left\langle n^{-1}\sum_{i=1}^n W_iX_i^{\otimes 2}(\hat\beta_{k_n}-\beta),Q_n\right\rangle\right|\Big|X_0\right]\\ =&O_{\mathsf{P}}(h_n^{-1/2})O_{\mathsf{P}}(n^{-1/2})o_{\mathsf{P}}(1)O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right)\\ =&O_{\mathsf{P}}(n^{-1/2})o_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n}\delta_j^{-1/2}(j\log j)^{3/2}\right).\end{aligned}$$

Finally, the term related to the last term in (S2.34) is divided into two parts and its convergence is proved in a different way. We first have

$$\begin{aligned}&s_{h_n}(X_0)^{-1/2}\left|\left\langle n^{-1}\sum_{i=1}^n X_i^{\otimes 2}(\hat\beta_{k_n}^*-\hat\beta_{g_n}),\Gamma_{h_n}^{-1}X_0\right\rangle\right|\\ =&s_{h_n}(X_0)^{-1/2}\left|\left\langle\hat\beta_{k_n}^*-\hat\beta_{g_n},n^{-1}\sum_{i=1}^n X_i^{\otimes 2}\Gamma_{h_n}^{-1}X_0\right\rangle\right|\end{aligned}$$

$$
\begin{aligned}
=& s_{h_n}(X_0)^{-1/2}\left|\left\langle \hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}, n^{-1}\sum_{i=1}^n \langle X_i, \Gamma_{h_n}^{-1} X_0\rangle X_i \right\rangle\right| \\
=& O_{\mathsf{P}}(h_n^{-1/2}) O_{\mathsf{P}}(h_n^{1/2}) \|\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}\| = o_{\mathsf{P}^*}(1)
\end{aligned}
$$

since

$$
\begin{aligned}
\mathsf{E}\left[\left\| n^{-1}\sum_{i=1}^n \langle X_i, \Gamma_{h_n}^{-1} X_0\rangle X_i \right\|\right]^2 \le& \mathsf{E}\left[\left\| n^{-1}\sum_{i=1}^n \langle X_i, \Gamma_{h_n}^{-1} X_0\rangle X_i \right\|^2\right] \\
\le& n^{-1}\sum_{i=1}^n \mathsf{E}[\langle X_i, \Gamma_{h_n}^{-1} X_0\rangle^2 \|X_i\|^2] = O(h_n)
\end{aligned}
$$

thanks to (S2.33) and Jensen's inequality. Since $n^{-1}\sum_{i=1}^n X_i^{\otimes 2} = O_{\mathsf{P}}(1)$, we also have

$$
\begin{aligned}
& s_{h_n}(X_0)^{-1/2}\left|\left\langle n^{-1}\sum_{i=1}^n X_i^{\otimes 2}(\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}), Q_n \right\rangle\right| \\
=& O_{\mathsf{P}}(h_n^{-1/2}) O_{\mathsf{P}}(1) O_{\mathsf{P}}\left(n^{-1/2}\sum_{j=1}^{h_n} \delta_j^{-1/2}(j\log j)^{3/2}\right)\|\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}\| \\
=& o_{\mathsf{P}^*}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n} \delta_j^{-1/2}(j\log j)^{3/2}\right)\|\hat{\beta}^*_{k_n} - \hat{\beta}_{g_n}\|.
\end{aligned}
$$

This completes the proof. $\square$

Lemma S16. *For the term in (S2.22), if* $\sup_{j\in\mathbb{N}} \gamma_j^{-1}\|\Lambda\phi_j\|^2 < \infty$ *(implied by Condition (D) in the main paper), we have*

$$
s_{h_n}(X_0)^{-1}|\langle \Lambda\Gamma_{h_n}^{-1} X_0, \{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\} X_0\rangle| = O_{\mathsf{P}}\left(n^{-1/2}h_n^{-1/2}\sum_{j=1}^{h_n} \delta_j^{-1/2}(j\log j)^{3/2}\right).
$$

Proof. The desired result is obtained as

$$
s_{h_n}(X_0)^{-1}|\langle \Lambda\Gamma_{h_n}^{-1} X_0, \{\hat{\Gamma}_{h_n}^{-1} - \Gamma_{h_n}^{-1}\} X_0\rangle|
$$

$$
\begin{aligned}
&= O_{\mathsf{P}}(h_n^{-1}) O_{\mathsf{P}}(h_n^{1/2}) \left(n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2} \right) \\
&= O_{\mathsf{P}} \left(n^{-1/2} h_n^{-1/2} \sum_{j=1}^{h_n} \delta_j^{-1/2} (j \log j)^{3/2} \right)
\end{aligned}
$$

by Yeon et al. (2024a, Lemma S20) and the fact that

$$
\mathsf{E}[\|\Lambda \Gamma_{h_n}^{-1} X_0\|^2] = \sum_{j=1}^{h_n} \gamma_j^{-1} \|\Lambda \phi_j\|^2 \leq C h_n
$$

implied by Condition (D). □

Lemma S17. *For the term in (S2.23), we have*

$$
s_{h_n}(X_0)^{-1} |\langle (\hat{\Lambda}_n^* - \Lambda) \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \rangle| = o_{\mathsf{P}^*}(1).
$$

Proof. Similarly to the proof of Lemma S15, by the decompositions in (S2.24) and (S2.26), we observe that

$$
\begin{aligned}
& s_{h_n}(X_0)^{-1} |\langle (\hat{\Lambda}_n^* - \Lambda) \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \rangle| \\
\leq & s_{h_n}(X_0)^{-1} \left| \left\langle n^{-1} \sum_{i=1}^{n} \{ (X_i \varepsilon_i)^{\otimes 2} W_i^2 - \Lambda \} \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| && \text{(S2.35)} \\
& + s_{h_n}(X_0)^{-1} \left| \left\langle n^{-1} \sum_{i=1}^{n} X_i^{\otimes 2} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 W_i^2 \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| && \text{(S2.36)} \\
& + 2 s_{h_n}(X_0)^{-1} \left| \left\langle n^{-1} \sum_{i=1}^{n} [\{ (X_i \varepsilon_i) \otimes X_i \} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle] W_i^2 \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| && \text{(S2.37)} \\
& + s_{h_n}(X_0)^{-1} \left| \left\langle n^{-1} \sum_{i=1}^{n} X_i^{\otimes 2} \langle \hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i \rangle^2 \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| && \text{(S2.38)} \\
& + 2 s_{h_n}(X_0)^{-1} \left| \left\langle n^{-1} \sum_{i=1}^{n} \{ (X_i \varepsilon_i^*) \otimes X_i \} \langle \hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i \rangle \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| && \text{(S2.39)}
\end{aligned}
$$

$$+ s_{h_n}(X_0)^{-1} \left| \left\langle \overline{X\hat{\varepsilon}^*}^{\otimes 2} \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| \tag{S2.40}$$

We check the convergences of each term in the above display.

For the term in (S2.35), recall from the proof of Lemma S11 that

$$\mathsf{E}^* \left[\left\| n^{-1} \sum_{i=1}^n \{(X_i \varepsilon_i)^{\otimes 2} W_i^2 - \Lambda\} \right\|_{HS} \right] = O_{\mathsf{P}}(n^{-1/2}).$$

This implies that

$$\begin{aligned} &\mathsf{E}^* \left[s_{h_n}(X_0)^{-1} \left| \left\langle n^{-1} \sum_{i=1}^n \{(X_i \varepsilon_i)^{\otimes 2} W_i^2 - \Lambda\} \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| \Big| X_0 \right] \\ =& O_{\mathsf{P}}(h_n^{-1}) O_{\mathsf{P}}(n^{-1/2}) O_{\mathsf{P}} \left(\sum_{j=1}^{h_n} \gamma_j^{-1} \right) \\ =& O_{\mathsf{P}} \left(n^{-1/2} h_n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-1} \right) = o_{\mathsf{P}}(1) \end{aligned}$$

because under Condition (A4), we have

$$\begin{aligned} n^{-1/2} h_n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-1} \leq n^{-1/2} h_n^{-1/2} \sum_{j=1}^{h_n} \gamma_j^{-1} &\leq \left(n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-2} \right)^{1/2} \\ &\leq \left(n^{-1} \sum_{j=1}^{h_n} \delta_j^{-2} \right)^{1/2} = o(1). \end{aligned}$$

Second, the term in (S2.36) is bounded as

$$\begin{aligned} &s_{h_n}(X_0)^{-1} \left| \left\langle n^{-1} \sum_{i=1}^n X_i^{\otimes 2} \langle \hat{\beta}_{k_n} - \beta, X_i \rangle^2 W_i^2 \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| \\ =& s_{h_n}(X_0)^{-1} \left| n^{-1} \sum_{i=1}^n W_i^2 \langle X_i^{\otimes 2} \hat{\beta}_{k_n} - \beta, \hat{\beta}_{k_n} - \beta \rangle \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2 \right| \\ =& s_{h_n}(X_0)^{-1} \left| \left\langle \left(n^{-1} \sum_{i=1}^n W_i^2 X_i^{\otimes 2} \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2 \right) \hat{\beta}_{k_n} - \beta, \hat{\beta}_{k_n} - \beta \right\rangle \right| \end{aligned}$$

$$\leq s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}W_i^2\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\right)\|\hat{\beta}_{k_n}-\beta\|^2.$$

We note from (S2.33) that

$$\mathsf{E}\left[n^{-1}\sum_{i=1}^{n}W_i^2\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\right]=n^{-1}\sum_{i=1}^{n}\mathsf{E}[\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2]=O(h_n),$$

which implies that $\mathsf{E}^*\left[n^{-1}\sum_{i=1}^{n}W_i^2\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\middle|X_0\right]=O_{\mathsf{P}}(h_n)$. It follows that

$$\mathsf{E}^*\left[s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}\langle\hat{\beta}_{k_n}-\beta,X_i\rangle^2W_i^2\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\right\rangle\right|\middle|X_0\right]=O_{\mathsf{P}}(\|\hat{\beta}_{k_n}-\beta\|^2).$$

The term in (S2.37) is similarly bounded as

$$\begin{aligned}
&s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}[\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}_{k_n}-\beta,X_i\rangle]W_i^2\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\right\rangle\right|\\
=&s_{h_n}(X_0)^{-1}\left|n^{-1}\sum_{i=1}^{n}W_i^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i\varepsilon_i,\Gamma_{h_n}^{-1}X_0\rangle\langle\hat{\beta}_{k_n}-\beta,X_i\rangle\right|\\
=&s_{h_n}(X_0)^{-1}\left|n^{-1}\sum_{i=1}^{n}W_i^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\langle\hat{\beta}_{k_n}-\beta,X_i\varepsilon_i\rangle\right|\\
\leq&s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}W_i^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\|X_i\varepsilon_i\|\right)\|\hat{\beta}_{k_n}-\beta\|.
\end{aligned}$$

Similarly to (S2.33), since the FPC scores $\{\xi_j\}$ are uncorrelated random variables with mean zero and variance one, we have from the independence between $\mathcal{X}_n\equiv\{(X_i,\varepsilon_i)\}_{i=1}^n$ and X_0 that

$$\begin{aligned}
\mathsf{E}[\|X_i\varepsilon_i\|\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2]&=\sum_{j=1}^{h_n}\gamma_j^{-1}\mathsf{E}[\|X_i\varepsilon_i\|\langle X_i,\phi_j\rangle^2]\\
&\leq\sum_{j=1}^{h_n}\gamma_j^{-1}\mathsf{E}[\|X_i\varepsilon_i\|^2]^{1/2}\mathsf{E}[\langle X_i,\phi_j\rangle^4]^{1/2}\leq Ch_n. \qquad \text{(S2.41)}
\end{aligned}$$

It then holds that

$$\mathsf{E}\left[n^{-1}\sum_{i=1}^{n} W_i^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\|X_i\varepsilon_i\|\right] = n^{-1}\sum_{i=1}^{n}\mathsf{E}[\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\|X_i\varepsilon_i\|] = O(h_n),$$

which implies that $\mathsf{E}^*\left[n^{-1}\sum_{i=1}^{n} W_i^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\|X_i\varepsilon_i \,\middle|X_0\right] = O_{\mathsf{P}}(h_n)$. It then holds that

$$\begin{aligned}
&\mathsf{E}^*\left[s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}[\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}_{k_n}-\beta,X_i\rangle]W_i^2\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\right\rangle\right|\,\middle|X_0\right]\\
=&O_{\mathsf{P}}(\|\hat{\beta}_{k_n}-\beta\|)
\end{aligned}$$

The terms in (S2.38) and (S2.39) are respectively bounded in a similar way to the bounds for (S2.36) and (S2.37) as

$$\begin{aligned}
&s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}\langle\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n},X_i\rangle^2\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\right\rangle\right|\\
=&s_{h_n}(X_0)^{-1}\left|n^{-1}\sum_{i=1}^{n}\langle X_i^{\otimes 2}\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n},\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n}\rangle\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\right|\\
=&s_{h_n}(X_0)^{-1}\left|\left\langle\left(n^{-1}\sum_{i=1}^{n}X_i^{\otimes 2}\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\right)\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n},\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n}\right\rangle\right|\\
\leq&s_{h_n}(X_0)^{-1}\left(n^{-1}\sum_{i=1}^{n}\|X_i\|^2\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle^2\right)\|\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n}\|^2\\
=&O_{\mathsf{P}}(1)\|\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n}\|^2
\end{aligned}$$

and

$$\begin{aligned}
&s_{h_n}(X_0)^{-1}\left|\left\langle n^{-1}\sum_{i=1}^{n}[\{(X_i\varepsilon_i)\otimes X_i\}\langle\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n},X_i\rangle]\Gamma_{h_n}^{-1}X_0,\Gamma_{h_n}^{-1}X_0\right\rangle\right|\\
=&s_{h_n}(X_0)^{-1}\left|n^{-1}\sum_{i=1}^{n}\langle X_i,\Gamma_{h_n}^{-1}X_0\rangle\langle X_i\varepsilon_i,\Gamma_{h_n}^{-1}X_0\rangle\langle\hat{\beta}^*_{k_n}-\hat{\beta}_{g_n},X_i\rangle\right|
\end{aligned}$$

$$
\begin{aligned}
=& s_{h_n}(X_0)^{-1} \left| n^{-1} \sum_{i=1}^{n} \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2 \langle \hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}, X_i \varepsilon_i \rangle \right| \\
\leq & s_{h_n}(X_0)^{-1} \left(n^{-1} \sum_{i=1}^{n} \langle X_i, \Gamma_{h_n}^{-1} X_0 \rangle^2 \|X_i \varepsilon_i\| \right) \|\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}\| \\
=& O_{\mathsf{P}}(1) \|\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}\|.
\end{aligned}
$$

With the same argument as the one in the proof of Lemma S15, the last term in (S2.40) is bounded as

$$
s_{h_n}(X_0)^{-1} \left| \left\langle \overline{X\hat{\varepsilon}^*}^{\otimes 2} \Gamma_{h_n}^{-1} X_0, \Gamma_{h_n}^{-1} X_0 \right\rangle \right| = s_{h_n}(X_0)^{-1} \langle \overline{X\hat{\varepsilon}^*}, \Gamma_{h_n}^{-1} X_0 \rangle^2 \leq A_n^* + B_n^* + C_n^*
$$

where

$$
\begin{aligned}
E^*[A_n^* | X_0] &= O_{\mathsf{P}} \left(n^{-1} h_n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-1} \right), \\
E^*[B_n^* | X_0] &= O_{\mathsf{P}} \left(n^{-1} h_n^{-1} \sum_{j=1}^{h_n} \gamma_j^{-1} \right) \|\hat{\beta}_{k_n} - \beta\|^2, \\
C_n^* &= O_{\mathsf{P}}(1) \|\hat{\beta}_{k_n}^* - \hat{\beta}_{g_n}\|^2.
\end{aligned}
$$

This completes the proof. □

Proposition S3. *Suppose that the assumptions in Theorem 1 in the main paper hold. Then, the bootstrap scaling $s_{h_n}^*(X_0)$ and the sample scaling $\hat{s}_{h_n}(X_0)$ are asymptotic equivalent in the sense that*

$$
\frac{\hat{s}_{h_n}^*(X_0)}{\hat{s}_{h_n}(X_0)} = 1 + o_{\mathsf{P}^*}(1).
$$

Proof. This is proved by Lemmas S13-S17 along with decomposition described in Equations (S2.19)-(S2.23). □

S2.6 WB consistency

In this section, we provide the proof of Theorem 1, the main results for the proposed WB in FLRMs.

Proof of Theorem 1. We now consider a general case when the means of response and regressor may not be zero. The conditional mean difference in the bootstrap statistic is decomposed as follows:

$$\hat{\mu}^*_{h_n}(X_0) - \hat{\mu}_{g_n}(X_0) = \bar{Y}^* - \bar{Y} - \langle \hat{\beta}^*_{h_n} - \hat{\beta}_{g_n}, \bar{X} - \mathsf{E}[X]\rangle + \langle \hat{\beta}^*_{h_n} - \hat{\beta}_{g_n}, X_0 - \mathsf{E}[X]\rangle. \tag{S2.42}$$

To find an upper bound for the first term, recall that the bootstrap mean and variance of the bootstrap response are

$$\mathsf{E}^*[Y_i^*] = \hat{\alpha}_{g_n} + \langle \hat{\beta}_{g_n}, X_i\rangle = \bar{Y} + \langle \hat{\beta}_{g_n}, X_i - \bar{X}\rangle,$$

$$\mathsf{var}[Y_i^*] = \hat{\varepsilon}^2_{i,k_n}.$$

Then, the average $\bar{Y}^* \equiv n^{-1}\sum_{i=1}^n Y_i^*$ of the bootstrap responses has mean and variance as $\mathsf{E}^*[\bar{Y}^*] = \hat{\alpha}_{g_n} + \langle \hat{\beta}_{g_n}, \bar{X}\rangle = \bar{Y}$ and (by independence between bootstrap responses)

$$\mathsf{var}^*[\bar{Y}^*] = n^{-2}\sum_{i=1}^n \mathsf{var}^*[Y_i^*] = n^{-1}\left(n^{-1}\sum_{i=1}^n \hat{\varepsilon}^2_{i,k_n}\right).$$

If $\|\hat{\beta}_{k_n} - \beta\| = o_{\mathsf{P}}(1)$, the average of the residuals bounded as

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_{i,k_n}^2 &= n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y} - \langle \hat{\beta}_{k_n} - \beta, X_i - \bar{X}\rangle - \langle \beta, X_i - \bar{X}\rangle)^2 \\ &\leq 3n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 + 3n^{-1}\sum_{i=1}^{n}\langle \hat{\beta}_{k_n} - \beta, X_i - \bar{X}\rangle^2 + 3n^{-1}\sum_{i=1}^{n}\langle \beta, X_i - \bar{X}\rangle^2 \\ &= O_{\mathsf{P}}(1), \end{aligned}$$

because $n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \xrightarrow{\mathsf{P}} \mathsf{var}[Y]$, $n^{-1}\sum_{i=1}^{n}\langle \beta, X_i - \bar{X}\rangle^2 \xrightarrow{\mathsf{P}} \mathsf{var}[\langle \beta, X\rangle] = \langle \Gamma\beta, \beta\rangle$, and

$$\begin{aligned} \sum_{i=1}^{n}\langle \hat{\beta}_{k_n} - \beta, X_i - \bar{X}\rangle^2 = \sum_{i=1}^{n}\langle (X_i - \bar{X})^{\otimes 2}\hat{\beta}_{k_n} - \beta, \hat{\beta}_{k_n} - \beta\rangle &= \langle \hat{\Gamma}_n(\hat{\beta}_{k_n} - \beta), \hat{\beta}_{k_n} - \beta\rangle \\ &\leq (\|\hat{\Gamma}_n - \Gamma\|_\infty + \|\Gamma\|_\infty)\|\hat{\beta}_{k_n} - \beta\|^2 = o_{\mathsf{P}}(1). \end{aligned}$$

This implies that $\mathsf{var}^*[\bar{Y}^*] = O_{\mathsf{P}}(n^{-1})$. Chebyshev inequality can then conclude that $\bar{Y}^* - \bar{Y} = O_{\mathsf{P}^*}(n^{-1/2})$. The second term in (S2.42) is bounded as

$$|\langle \hat{\beta}^*_{h_n} - \hat{\beta}_{g_n}, \bar{X} - \mathsf{E}[X]\rangle| \leq \|\hat{\beta}^*_{h_n} - \hat{\beta}_{g_n}\|\|\bar{X} - \mathsf{E}[X]\| = o_{\mathsf{P}^*}(n^{-1/2}),$$

since $\bar{X} - \mathsf{E}[X] = O_{\mathsf{P}}(n^{-1/2})$ holds by a CLT for general Hilbert spaces (e.g., Hsing and Eubank, 2015, Theorem 7.7.6) and $\|\hat{\beta}^*_{h_n} - \hat{\beta}_{g_n}\| = o_{\mathsf{P}^*}(1)$ can happen by the consistency result in Section S2.2.

To sum up, upon scaling, the first two terms (S2.42) are negligible, indicating the weak convergences of the conditional mean statistic and the projection statistic are equivalent. In other words, it suffices to show the bootstrap consistency for the projection with centered data $\{(X_i - \mathsf{E}[X], Y_i - \mathsf{E}[Y])\}_{i=1}^n$. Equivalently, we will derive the bootstrap consistency for projection $\langle \hat{\beta}^*_{h_n} - \hat{\beta}_{g_n}, X_0 \rangle$ when both means are zero, i.e., $\mathsf{E}[X] = 0$ and $\mathsf{E}[Y] = 0$ (implying $\alpha = 0$).

Finally, under the zero-mean condition, using Slutsky's theorem, Polya's theorem (Athreya and Lahiri, 2006, Theorem 9.1.4), and a subsequence argument (cf. Billingsley, 1999, Theorem 20.5) for bootstrap distributions, Theorem 1 in the main paper follows from Propositions S1-S3 and the results in Yeon et al. (2024b, Section S3.3.1) along with the decomposition (S2.2). $\square$

S3 Additional simulation results

Section S3.1 contains the detailed algorithm used for coverage simulation studies given in Section 4 of the main paper, while Section S3.2 provides extra simulation results regarding interval widths and studentization. In Section S3.3, we discuss the effects of less consequential truncation levels k_n and g_n.

S3.1 Simulation algorithm

This section contains the details of the algorithm behind the simulation studies in Section 4 of the main paper. We simulate the new regressor X_0 as well as the data samples $\{(X_i, Y_i)\}_{i=1}^n$ at each Monte Carlo iteration, where the Monte Carlo simulation size M and the bootstrap resample size Q are given as $M = 1000 = Q$. The (symmetrized) confidence intervals are constructed by the methods as described in the main paper: CLT, residual bootstrap (RB), paired bootstrap (PB), WB with the three multipliers described in the main paper. See Yeon et al. (2023, 2024a) for the construction of RB and PB intervals. Here, the intervals from PB or WB are computed either with or without studentization.

The simulation procedure is described as follows. As mentioned in Section 4 of the main paper, we focus on the inference of projection $\mu(X_0) = \langle \beta, X_0 \rangle$ under the zero-mean condition that $\mathsf{E}[X] = 0$ and $\mathsf{E}[Y] = 0$ in our simulation studies. In such cases, we conduct the inference of the projection $\mu(X_0) = \langle \beta, X_0 \rangle$ by using the original $\langle \hat{\beta}_{h_n}, X_0 \rangle$ and bootstrap $\langle \hat{\beta}^*_{h_n}, X_0 \rangle$ projection estimators by forcing intercepts $\alpha, \hat{\alpha}_{h_n}, \hat{\alpha}^*_{h_n}$ or sample means $\bar{X}, \bar{X}^*$ to be zero. For each $m = 1, \ldots, M$, perform the following.

1. (Data generation) Simulate X_i and ε_i with $X_i \overset{\mathsf{d}}{=} X$ and $\varepsilon_i \overset{\mathsf{d}}{=} \varepsilon$ respectively, where the pairs $\{(X_i, \varepsilon_i)\}_{i=1}^n$ are independent but ε may be dependent of X, and compute the response $Y_i = \langle \beta, X_i \rangle + \varepsilon_i$ for $i = 1, \ldots, n$. Also, simulate the new regressor $X_0 \overset{\mathsf{d}}{=} X$ such that X_0 is independent of the data $\{(X_i, Y_i)\}_{i=1}^n$.

2. (Residual bootstrap)

   (a) (Residuals) Compute the residuals $\hat{\varepsilon}_{i,k_n} = Y_i - \langle \hat{\beta}_{k_n}, X_i \rangle$ for $i = 1, \ldots, n$.

   (b) (Residual bootstrap) To approximate the residual bootstrap distribution, do the following for $q = 1, \ldots, Q$.

i. Draw independent bootstrap errors $\{\varepsilon_{q,i}^{RB*}\}_{i=1}^{n}$ from the uniform distribution on the centered residuals $\{\hat{\varepsilon}_{i,k_n} - \bar{\hat{\varepsilon}}_{k_n}\}_{i=1}^{n}$.

ii. Compute the bootstrap responses $Y_{q,i}^{RB*} = \langle \hat{\beta}_{g_n}, X_i \rangle + \varepsilon_{q,i}^{RB*}$, and construct the bootstrap estimate $\hat{\beta}_{q,h_n}^{RB*}$ based on the bootstrap samples $\{(X_i, Y_{q,i}^{RB*})\}_{i=1}^{n}$.

iii. Compute the bootstrap statistic defined in Yeon et al. (2023) (cf. Theorem 3 therein):

$$\hat{T}_{q,n}^{RB*}(X_0) \equiv \sqrt{\frac{n}{\hat{t}_{h_n}(X_0)}}[\langle \hat{\beta}_{q,h_n}^{RB*}, X_0 \rangle - \langle \hat{\beta}_{g_n}, X_0 \rangle].$$

3. (De-biased paired bootstrap) Compute the bias correction term $\hat{U}_{n,g_n} = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})(\hat{\varepsilon}_{i,g_n} - \bar{\hat{\varepsilon}}_{g_n})$, where $\hat{\varepsilon}_{i,g_n} = Y_i - \langle \hat{\beta}_{g_n}, X_i \rangle$ are the residuals for $i = 1, \ldots, n$ with its average $\bar{\hat{\varepsilon}}_{g_n} \equiv n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_{i,g_n}$. To approximate the paired bootstrap distribution, do the following for $q = 1, \ldots, Q$.

(a) Draw independent bootstrap pairs $\{(X_i^{PB*}, Y_i^{PB*})\}_{i=1}^{n}$ from the uniform distribution on the samples $\{(X_i, Y_i)\}_{i=1}^{n}$.

(b) Compute the bias corrected bootstrap estimate $\hat{\beta}_{q,h_n}^{PB*}$ and the bootstrap scaling $\hat{s}_{q,h_n}^{PB*}(X_0)$ based on the bootstrap samples $\{(X_i^{PB*}, Y_i^{PB*})\}_{i=1}^{n}$ and $\hat{\beta}_{q,h_n}^{PB*}$.

(c) Compute the bootstrap statistics either with or without studentization as described in Yeon et al. (2024a) (cf. Equations (14) and (16) therein):

$$\hat{T}_{q,n}^{PB*}(X_0) \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}}[\langle \hat{\beta}_{q,h_n}^{PB*}, X_0 \rangle - \langle \hat{\beta}_{g_n}, X_0 \rangle],$$

$$\hat{T}^{PB*}_{q,n,std}(X_0) \equiv \sqrt{\frac{n}{\hat{s}^{PB*}_{q,h_n}(X_0)}}[\langle \hat{\beta}^{PB*}_{q,h_n}, X_0\rangle - \langle \hat{\beta}_{g_n}, X_0\rangle].$$

4. (Multiplier wild bootstrap) To approximate the wild bootstrap distribution, do the following for $q = 1, \dots, Q$.

   (a) Draw independent multipliers $\mathcal{W}_n \equiv \{W_i\}_{i=1}^n$ with moments $\mathsf{E}^*[W_i] = 0$ and $\mathsf{E}^*[W_i^2] = 1$ and construct independent bootstrap errors $\{\varepsilon^{WB*}_{q,i}\}_{i=1}^n$ as $\varepsilon^{WB*}_i \equiv W_i \hat{\varepsilon}_{i,k_n}$.

   (b) Compute the bootstrap responses $Y^{WB*}_{q,i} = \langle \hat{\beta}_{g_n}, X_i\rangle + \varepsilon^{WB*}_{q,i}$, and construct the bootstrap estimate $\hat{\beta}^{WB*}_{q,h_n}$ and the bootstrap scaling $\hat{s}^{WB*}_{q,h_n}(X_0)$ based on the bootstrap samples $\{(X_i, Y^{WB*}_i)\}_{i=1}^n$ and $\hat{\beta}^{WB*}_{q,h_n}$.

   (c) Compute the bootstrap statistics either with or without studentization defined in (3.12) of the main paper (Yeon et al., 2025, Section 2):

$$\hat{T}^{WB*}_{q,n}(X_0) \equiv \sqrt{\frac{n}{\hat{s}_{h_n}(X_0)}}[\langle \hat{\beta}^{WB*}_{q,h_n}, X_0\rangle - \langle \hat{\beta}_{g_n}, X_0\rangle],$$
$$\hat{T}^{WB*}_{q,n,std}(X_0) \equiv \sqrt{\frac{n}{\hat{s}^{WB*}_{q,h_n}(X_0)}}[\langle \hat{\beta}^{WB*}_{q,h_n}, X_0\rangle - \langle \hat{\beta}_{g_n}, X_0\rangle].$$

5. For all cases, construct the following confidence intervals for $\langle \beta, X_0\rangle$

   (a) (CLT) The (symmetrized) confidence interval for $\langle \beta, X_0\rangle$ based on CLT is

$$CI_{CLT} = \left[\langle \hat{\beta}_{h_n}, X_0\rangle - \sqrt{\frac{\hat{s}_{h_n}(X_0)}{n}} z_{1-\alpha/2}, \langle \hat{\beta}_{h_n}, X_0\rangle + \sqrt{\frac{\hat{s}_{h_n}(X_0)}{n}} z_{1-\alpha/2}\right],$$

   where $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution.

(b) (Residual bootstrap) Compute the $1-\alpha/2$ quantile of $\{|\hat{T}^{RB*}_{q,n}(X_0)|\}_{q=1}^{Q}$, say u. Then, the symmetrized confidence interval for $\langle \beta, X_0 \rangle$ is

$$CI_{RB} = \left[\langle \hat{\beta}_{h_n}, X_0 \rangle - \sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}} u, \langle \hat{\beta}_{h_n}, X_0 \rangle + \sqrt{\frac{\hat{t}_{h_n}(X_0)}{n}} u \right]$$

(c) (Paired bootstrap) Replace $\hat{T}^{RB*}_{q,n}(X_0)$ and $\hat{t}_{h_n}(X_0)$ by either $\hat{T}^{PB*}_{q,n}(X_0)$ or $\hat{T}^{PB*}_{q,n,std}(X_0)$ and $\hat{s}_{h_n}(X_0)$, respectively, in the procedure (b) to obtain the symmetrized confidence intervals CI_{PB} and $CI_{PB,std}$ either with or without studentization.

(d) (Wild bootstrap) Replace $\hat{T}^{PB*}_{q,n}(X_0)$ by either $\hat{T}^{WB*}_{q,n}(X_0)$ or $\hat{T}^{WB*}_{q,n,std}(X_0)$ in the procedure (c) to obtain the symmetrized confidence intervals CI_{WB} and $CI_{WB,std}$ either with or without studentization.

6. Let $\widehat{CI}(X_0)$ denote one of the intervals constructed above. Compute $I_m = \mathbb{I}(\langle \beta, X_0 \rangle \in \widehat{CI}(X_0))$.

The (unconditional) empirical coverage probability $1-\alpha$ is then approximated by $M^{-1}\sum_{m=1}^{M} I_m$. We compute the average widths as the averaged difference between the end points of the constructed intervals as above.

S3.2 Additional simulation results

Figures 1-2 in the main paper exhibit the empirical coverage probabilities of the constructed confidence intervals over the truncation level $h_n \in \{1, \ldots, 15\}$, when $\tau^2 = 0.1$

under either homoscedasticity or heteroscedasticity and when $\tau^2 = 2$ only under heteroscedasticity, respectively; in the meantime, Figures S1-S2 shows the average widths of the intervals under the same scenarios, which might be helpful to understand the behaviors of the intervals.

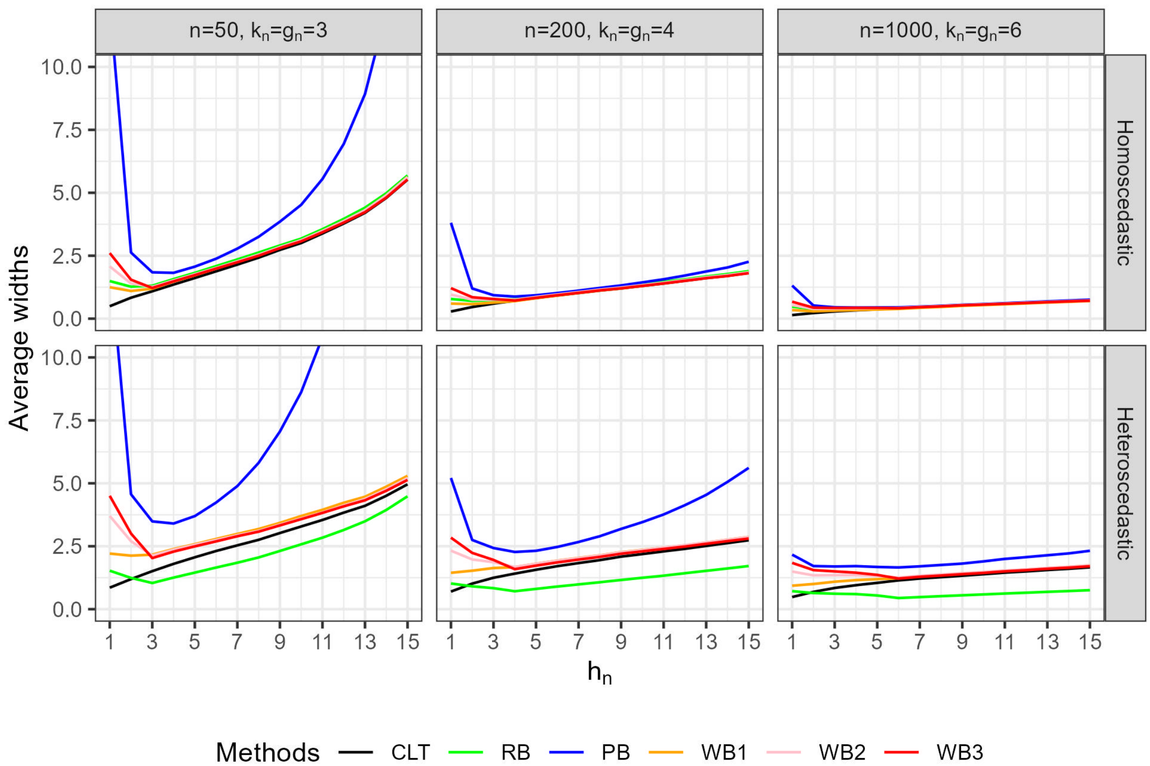

Figure S1: Average widths of (individual) confidence intervals for the projection $\langle \beta, X_0 \rangle$ constructed by different methods over the truncation level $h_n \in \{1, \ldots, 15\}$, where $\tau^2 = 0.1$ and the other truncation levels $k_n = [1.5 \cdot n^{1/v_k}]$ $(v_k = 2a + 1.1)$, $g_n = k_n$ are differently selected depending on the sample sizes.

To see the effect of studentization, in Figures S3-S4, we provide the empirical cov-

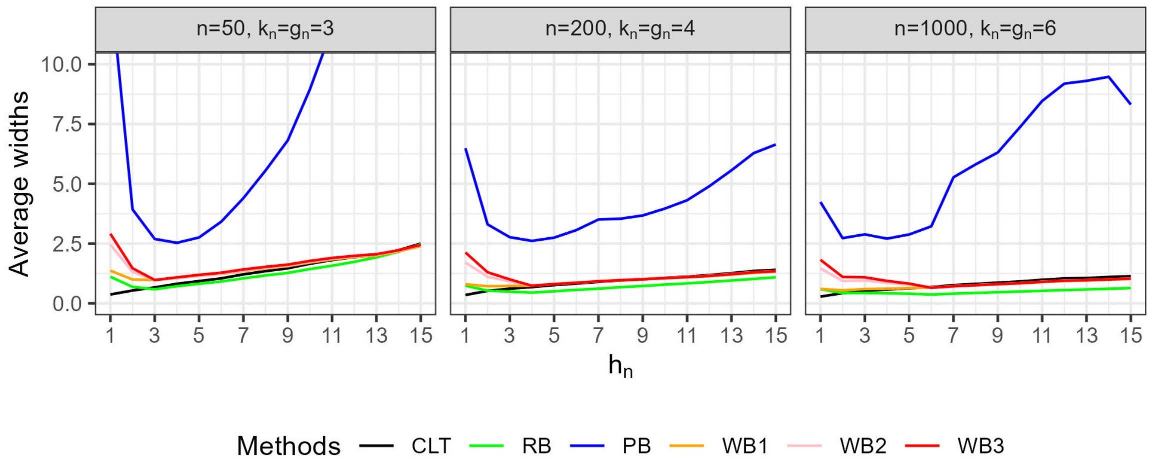

Figure S2: Average widths of (individual) confidence intervals for the projection $\langle \beta, X_0 \rangle$ constructed by different methods over the truncation level $h_n \in \{1, \dots, 15\}$ under heteroscedastic error models, where $\tau^2 = 2$ and the other truncation levels $k_n = [1.5 \cdot n^{1/v_k}]$ ($v_k = 2a + 1.1$), $g_n = k_n$ are differently selected depending on the sample sizes.

erage rates and average widths of the intervals either with (lower panels) or without (upper panels) studentization. The first impressive observation is that WB might perform poorly depending on the choice of multipliers, even when the sample size is large. For example, Figure S3 shows that WB1 (WB with two-point multipliers) and WB3 (WB with Mammen's multipliers) tend to overcover and undercover, respectively, without studentization. However, studentization seems to be able to minimize the influence of multiplier choices in WB; the performances of studentized WB methods in Figure S3 are close to each other. It it worthwhile mentioning that the WB2 (WB with normal multipliers) nevertheless works quite well regardless of studentization.

On the other hand, studentization in PB could affect differently depending on the error distribution; it might hurt the coverage of PB intervals for some error distribution cases. For less skewed errors (first and third rows in Figure S3), studentization for PB intervals seems generally a good choice. In contrast, if the error skewness is large (second and fourth rows in Figure S3), non-studentized PB intervals might work better for coverage accuracy. In this case, studentization step seems to become more unstable for PB.

S3.3 Effects of k_n and g_n

The recommendation of using $g_n = k_n$ and h_n larger than g_n stems from a finding of our previous study of RB, Yeon et al. (2023). There (e.g., Yeon et al., 2023, Figure 2), we found that excessively large g_n can harm the performance of the RB interval, more so than large k_n or h_n. In addition, theoretically, an important consideration in choosing k_n and g_n is that the resulting estimators $\hat{\beta}_{k_n}$ and $\hat{\beta}_{g_n}$ should be consistent for β. This is implied by Conditions (A1)–(A4), where the last condition (A4) specifically determines the growth rates for k_n and g_n. While we did not find a theoretical sweet spot for the choices of these truncation levels, we point out that such theoretical analyses regarding effects of tuning parameters are largely under-investigated even in central limit theorem for functional regression. We leave this to future research, but nevertheless, we have added additional numerical studies that are currently represented in Figure S5 of the supplement.

We have now more deeply illustrated the effects of varying k_n and g_n. To do so, we conducted new simulation studies by fixing either k_n or g_n and varying the other truncation level over a large range. When we fix either truncation levels, we set these as 3 when $n = 50$, 4 when $n = 200$ and 6 when $n = 1000$, mimicking the setup for Figure 2.

To explain the findings, we first consider the cases with fixed k_n but varying g_n (upper panels of Figure S5). When $n = 50$, we found some coverage improvements using larger g_n than k_n. In this case, the WB intervals with $g_n = 15$ tend to obtain the nominal level quite accurately. However, we do not interpret that larger g_n is good for the coverage of the WB intervals generally, as this yields extreme overcoverage as the sample size gets large. When $n = 1000$, the WB intervals with large g_n tend to be far too conservative (always cover), as an aspect not illustrated in our original simulation studies (Figures 1–2). In contrast, the WB intervals with g_n close to k_n behave more reasonably across sample sizes, including large cases. For all sample sizes, smaller g_n compared to k_n leads to undercoverage when h_n is also small.

When k_n is varied but g_n is fixed (lower panels of Figure S5), our observation is quite simpler: as the given value of g_n increases, the WB intervals tend to lose coverage. Incorporating both small and large sample cases, there may be some optimal choice of k_n, which looks close to the fixed g_n.

Overall, g_n far from k_n could yield either over- or under-coverage issues, while g_n close to k_n tends to lead to reasonable behaviors of the WB intervals, which can provide a better guide for the subsequent selection of h_n.

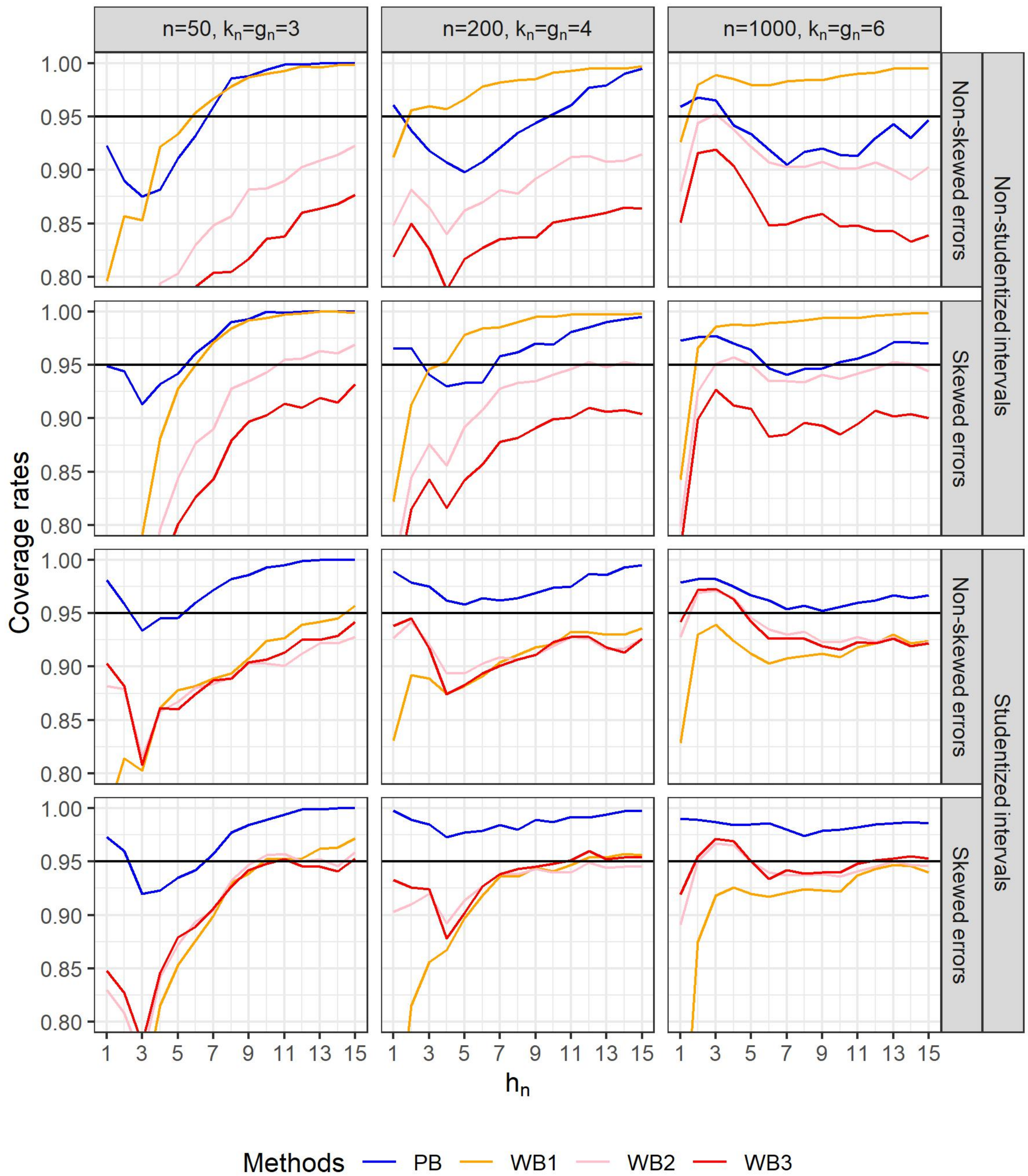

Figure S3: Empirical coverage probabilities of non-studentized versus studentized confidence intervals for the projection $\langle \beta, X_0 \rangle$ constructed by different methods over the truncation level $h_n \in \{1, \ldots, 15\}$, where $\tau^2 \in \{0.1, 2\}$ and the other truncation levels $k_n = [1.5 \cdot n^{1/v_k}]$ ($v_k = 2a + 1.1$), $g_n = k_n$ are differently selected depending on the sample sizes.

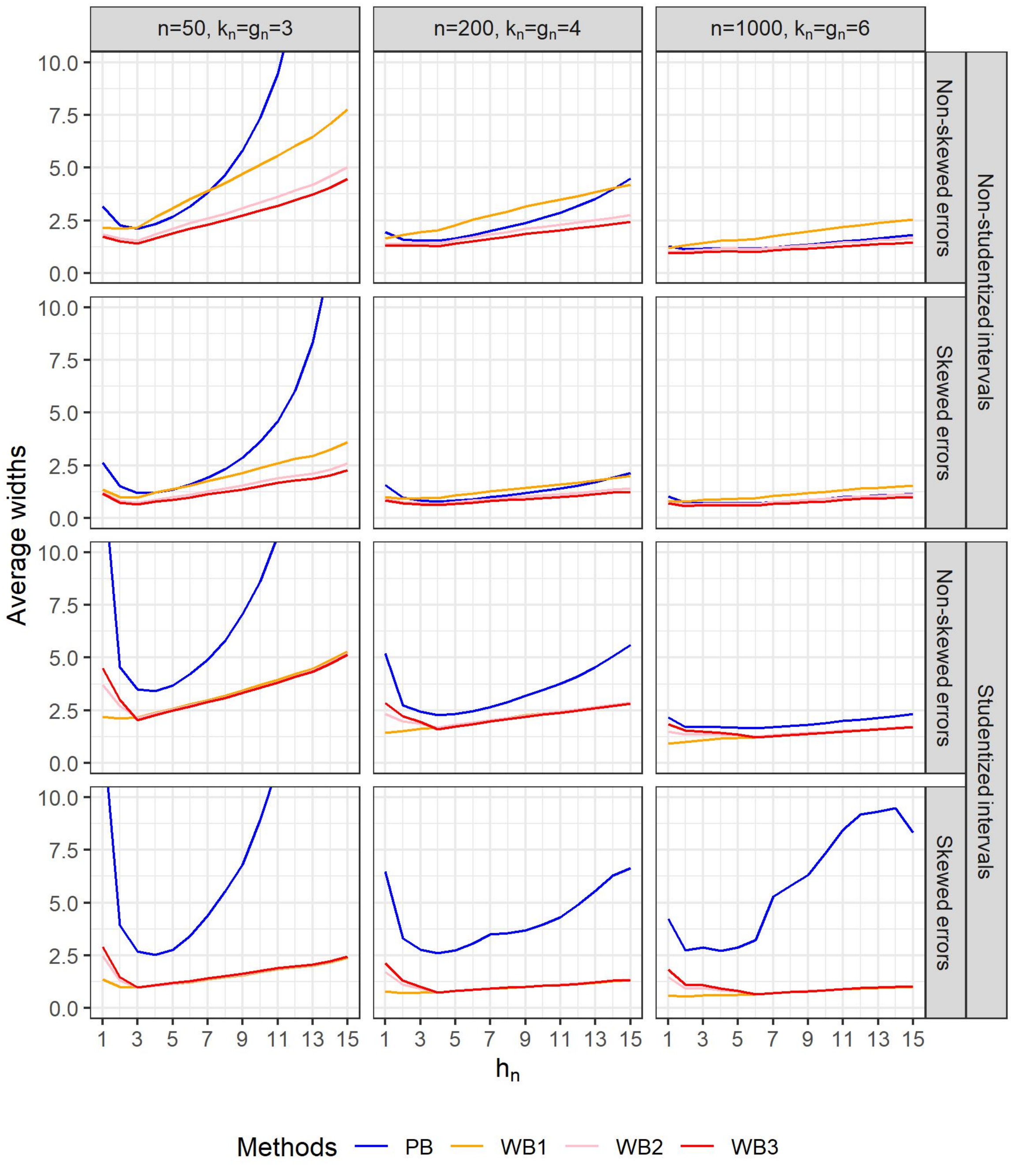

Figure S4: Average widths of non-studentized versus studentized confidence intervals for the projection $\langle \beta, X_0 \rangle$ constructed by different methods over the truncation level $h_n \in \{1, \ldots, 15\}$, where $\tau^2 \in \{0.1, 2\}$ and the other truncation levels $k_n = [1.5 \cdot n^{1/v_k}]$ $(v_k = 2a + 1.1)$, $g_n = k_n$ are differently selected depending on the sample sizes.

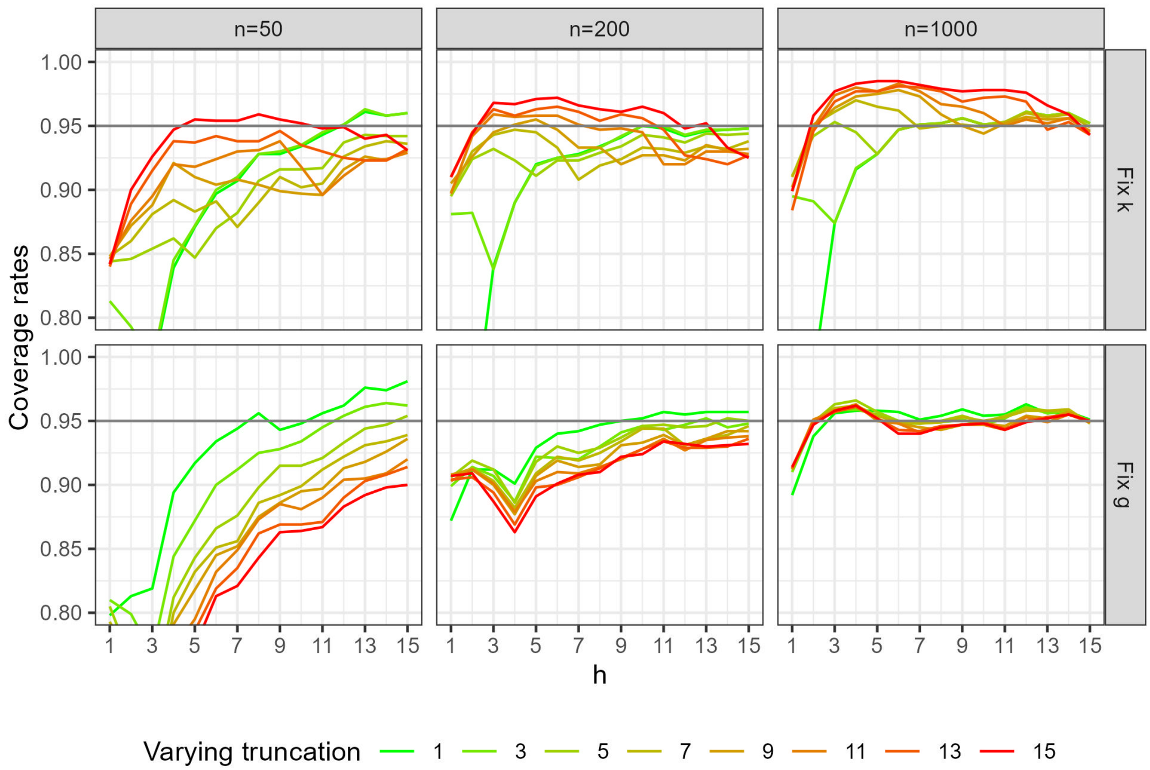

Figure S5: Empirical coverage probabilities of WB confidence intervals with standard normal multipliers, when either k_n or g_n is fixed while the other is varied. The data generation process is the same as that for Figure 2.

S4 Extra figures for real data analysis

We provide extra figures used in real data analysis in Section 5 of the main paper. Figure S6 shows the observed temperature curves for each region in the United States (US) with their respective averages being colored in blue. In Figure S7, estimated standard deviations $\{\hat{\sigma}_{l,h_n}\}_{l=1}^4$ over $h_n \in \{1, \ldots, 20\}$ are given, where $\hat{\sigma}_{l,h_n}^2 \equiv n_l^{-1} \sum_{i \in \mathcal{I}_l} \hat{\varepsilon}_{i,h_n}^2$ with $\mathcal{I}_l, n_l = |\mathcal{I}_l|$ denoting the index set and the sample size of l-th region, respectively. Also, the kernel-estimated densities of the residuals $\{\varepsilon_{i,k_n}\}_{i \in \mathcal{I}_l}$ in each region $l \in \{1, \ldots, 4\}$ with $k_n = 6$ are represented in Figure S8, where West particularly shows severe heteroscedasticity and skewness. Finally, Figure S9 collect the confidence intervals for the centered projections $\{\langle \beta, X_0 - \mathsf{E}[X] \rangle\}_{l=1}^4$ from all three bootstrap methods with $k_n = 6 = g_n$ and over $h_n \in \{3, \ldots, 20\}$.

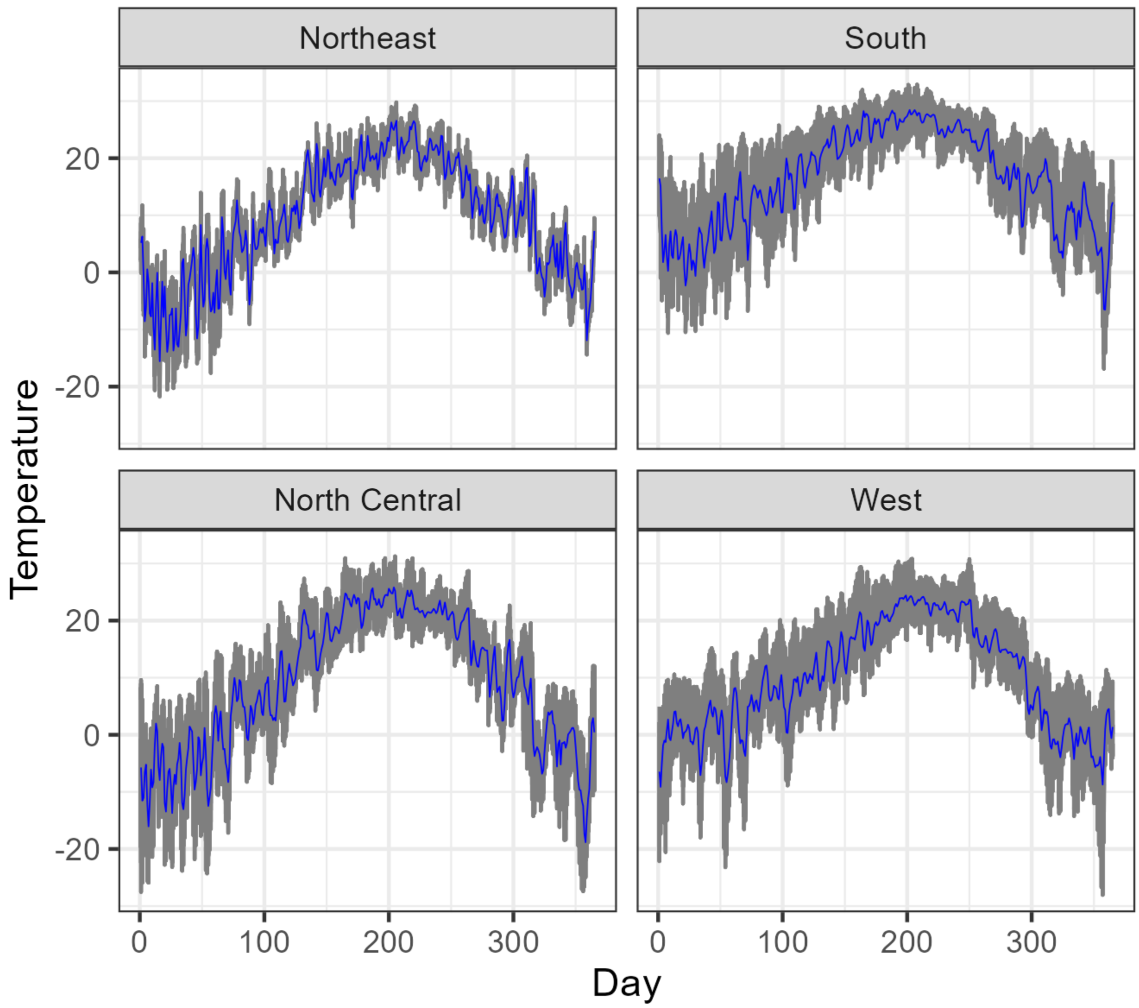

Figure S6: The temperature curves of 47 states in the US in 2022. The blue curves represent the average temperatures of each region.

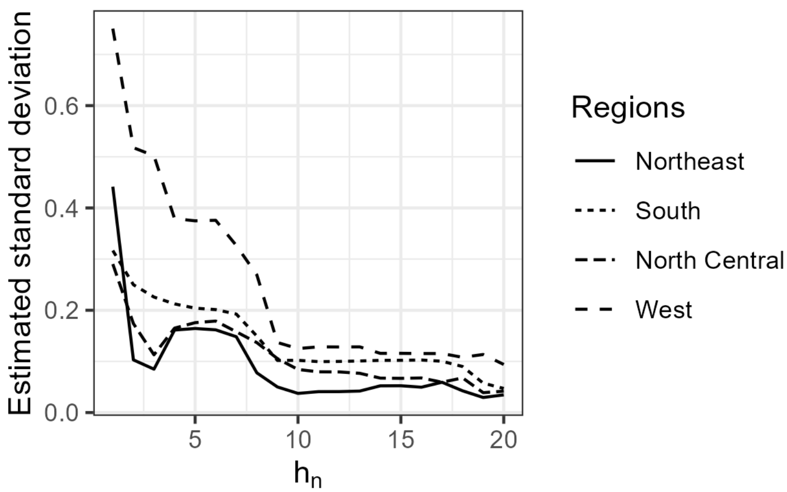

Figure S7: Estimated standard deviations for each region over different truncations $h_n \in \{1, \ldots, 20\}$.

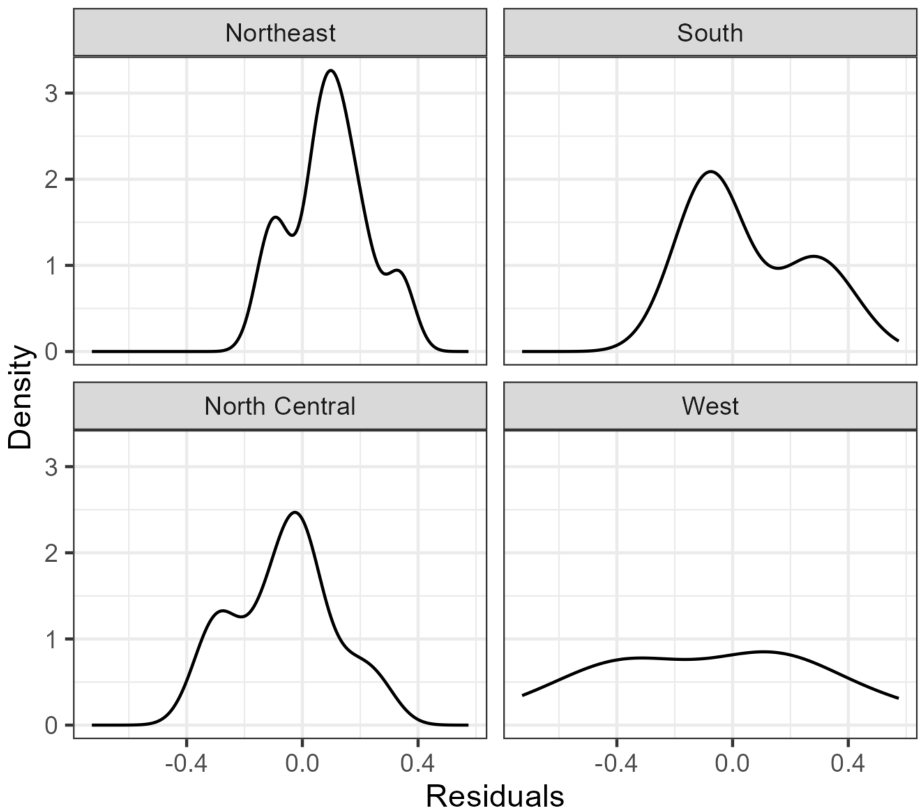

Figure S8: Density estimates of residuals with $k_n = 6$ for each region from the US weather dataset.

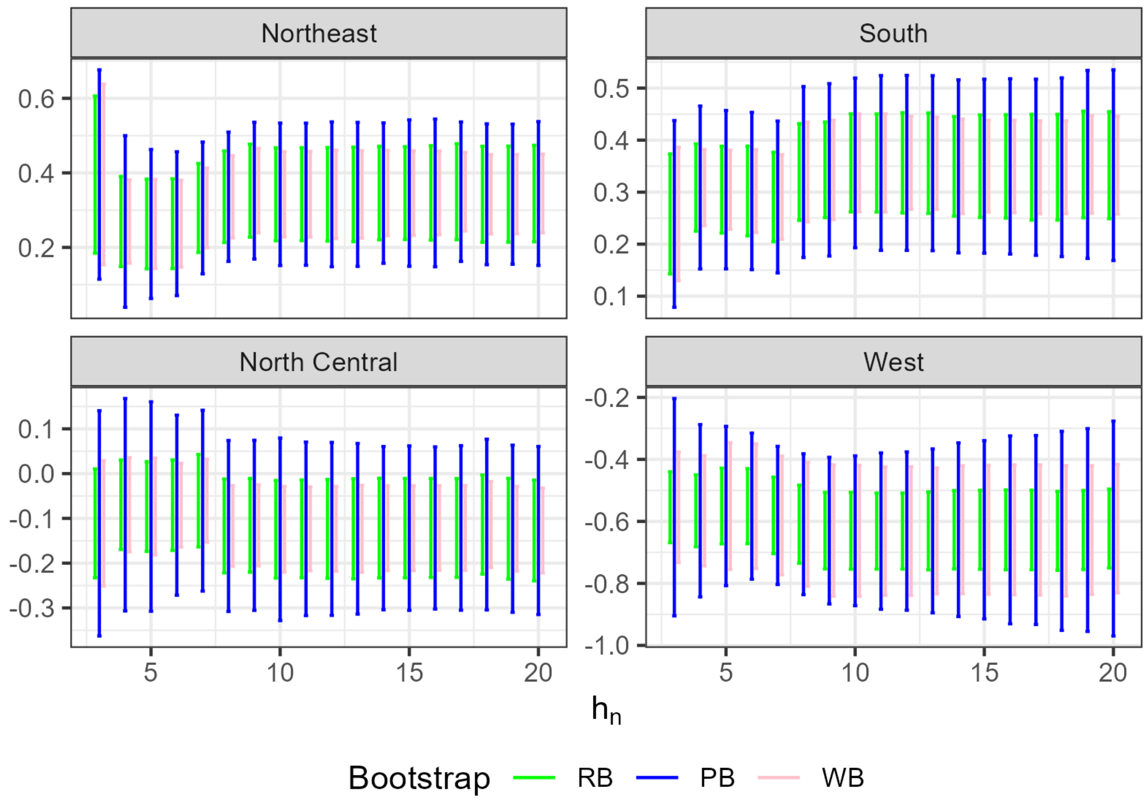

Figure S9: Bootstrap intervals for centered projections $\{\langle \beta, X_{0,l} - \mathsf{E}[X] \rangle\}_{l=1}^{4}$ for each region from the US weather dataset with $k_n = 6 = g_n$ and different $h_n \in H \equiv \{3, \ldots, 20\}$. The intervals from $h_n \in \{1, 2\}$ are dropped due to too large widths of the corresponding PB intervals.

References

Athreya, K. B. and S. N. Lahiri (2006). *Measure Theory and Probability Theory*. Springer Science & Business Media.

Billingsley, P. (1999). *Probability and Measure*. John Wiley & Sons.

Cai, T. T. and P. Hall (2006). Prediction in functional linear regression. *Annals of Statistics 34*(5), 2159–2179.

Cardot, H., A. Mas, and P. Sarda (2007). CLT in functional linear regression models. *Probability Theory and Related Fields 138*(3-4), 325–361.

Crambes, C. and A. Mas (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli 19*(5B), 2627–2651.

Dunford, N. and J. T. Schwartz (1988). *Linear Operators, Part 1: General Theory*, Volume 10. John Wiley & Sons.

Ferraty, F., I. Van Keilegom, and P. Vieu (2010). On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics 37*(2), 286–306.

Franke, J. and W. Hardle (1992). On bootstrapping kernel spectral estimates. *Annals of Statistics*, 121–145.

Gohberg, I., S. Goldberg, and M. Kaashoek (1990). *Classes of Linear Operators Vol. I*, Volume 49. Birkhäuser.

González-Manteiga, W. and A. Martínez-Calvo (2011). Bootstrap in functional linear regression. *Journal of Statistical Planning and Inference 141*(1), 453–461.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics 35*(1), 70–91.

Hardy, G. H. (1921). *A Course of Pure Mathematics* (3rd ed.). Cambridge University Press.

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications*, Volume 200. Springer Science & Business Media.

Hsing, T. and R. Eubank (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Volume 997. John Wiley & Sons.

Imaizumi, M. and K. Kato (2018). PCA-based estimation for functional linear regression with functional responses. *Journal of Multivariate Analysis 163*, 15–36.

Khademnoe, O. and S. M. E. Hosseini-Nasab (2016a). On asymptotic distribution of prediction in functional linear regression. *Statistics 50*(5), 974–990.

Khademnoe, O. and S. M. E. Hosseini-Nasab (2016b). On properties of percentile bootstrap confidence intervals for prediction in functional linear regression. *Journal of Statistical Planning and Inference 170*, 129–143.

REFERENCES

Lei, J. (2014). Adaptive global testing for functional linear models. *Journal of the American Statistical Association 109*(506), 624–634.

Lin, Y. and Z. Lin (2025). Hypothesis testing for functional linear models via bootstrapping. *Bernoulli 31*(4), 3309–3330.

Yeon, H., X. Dai, and D. J. Nordman (2023). Bootstrap inference in functional linear regression models with scalar response. *Bernoulli 29*(4), 2599 – 2626.

Yeon, H., X. Dai, and D. J. Nordman (2024a). Bootstrap inference in functional linear regression models with scalar response under heteroscedasticity. *Electronic Journal of Statistics 18*(2), 3590 – 3627.

Yeon, H., X. Dai, and D. J. Nordman (2024b). Supplement to "bootstrap inference in functional linear regression models with scalar response under heteroscedasticity". *Iowa State University Digital Repository*, <https://dr.lib.iastate.edu/handle/20.500.12876/jrl8QQqr>.

Yeon, H., X. Dai, and D. J. Nordman (2025). Wild bootstrap for mean response inference in functional linear regression models. *In preparation*.