Online Appendix for "Realized Regularized Regressions"

Aleksey Kolokolov[*] Shifan Yu[†]

This Version: May 4, 2026

This Online Appendix comprises two separate parts. Appendix A collects the proofs for all the theoretical results presented in the main text. Appendix B contains additional results for both the Monte Carlo simulations (Section 4) and empirical applications (Section 5).

[*]New Economic School, Russia. Email: akolokolov@nes.ru.

[†]Corresponding author, Oxford-Man Institute of Quantitative Finance, University of Oxford, United Kingdom. Email: shifan.yu@omi.ox.ac.uk.

Appendix A Proofs

Notation. For a vector x, $\|x\| = \sqrt{\sum_i |x_i|^2}$ is the Euclidean norm (ℓ_2-norm), $\|x\|_1 = \sum_i |x_i|$ is the ℓ_1-norm, and $\|x\|_\infty = \max_i |x_i|$ is the maximum norm (ℓ_∞-norm). For a matrix $\mathbf{A}$, $\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j |\mathbf{A}_{ij}|^2}$ is the Frobenius norm, and $\|\mathbf{A}\| = \sup_{\|x\|=1} \|\mathbf{A}x\|$ is the operator norm, which is the square root of the largest eigenvalue of the symmetric matrix $\mathbf{A}^\top \mathbf{A}$. For a symmetric positive definite matrix $\mathbf{A}$, $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote its largest and smallest eigenvalues, and it holds that $(\lambda_{\min}(\mathbf{A}))^{-1} = \|\mathbf{A}^{-1}\|$. For a square integrable function $f\colon [0,T] \to \mathbb{R}^p$, $\|f\|_{L^2} = \sqrt{\int_0^T \|f_s\|^2 ds}$ is the L^2-norm over $[0,T]$, where $\|\cdot\|$ inside the integral is the Euclidean norm on $\mathbb{R}^p$, and $\|f\|_{L^2(c)} = \langle f, f\rangle_c^{1/2}$, where $\langle f, g\rangle_c = \int_0^T f_s^\top c_s g_s ds$.

In all the sequel, the positive constants $\{K, K', ...\}$ may change from line to line (and even within a line) but never depend on n or other indices.

Localization. With a standard localization procedure (see, e.g., Lemma 4.4.9, Jacod and Protter, 2012), we impose the following stronger assumption than Assumptions 1 and 2 without loss of generality:

Assumption A.1. We have Assumptions 1 and 2 with $\tau_1 \to \infty$. Moreover, the processes β, β^J, b, $\widetilde{b}$, b', b^β, σ, $\widetilde{\sigma}$, σ', σ^β, c, and the functions δ, $\widetilde{\delta}$, δ', δ^β are bounded.

Throughout the remainder of Appendix A, references to any part of Assumptions 1 and 2 are understood as its strengthened version under Assumption A.1.

A.1 Properties of B-Splines

Fix an integer degree $d \geq 1$ and the partition $0 = v_0 < v_1 < \ldots, v_{N_n+1} = T$. Let $K_n = N_n + d + 1$ and define an extended knot sequence

$$\bar{v}_0 = \cdots = \bar{v}_d = 0, \qquad \bar{v}_{d+j} = v_j \text{ (for } j = 1, \ldots, N_n), \qquad \bar{v}_{N_n+d+1} = \cdots = \bar{v}_{K_n+d} = T. \tag{A.1}$$

We define B-spline basis function $B_t^{(k)} \equiv B_{d,t}^{(k)}$ by the Cox-de Boor recursion:

$$B_{0,t}^{(k)} = \begin{cases} 1 & \text{if } t \in [\bar{v}_{k-1}, \bar{v}_k) \\ 0 & \text{otherwise} \end{cases}, \qquad \text{for } k = 1, \ldots, K_n + d, \tag{A.2}$$

and $B_{r,t}^{(k)}$ for all $r = 1, \ldots, d$ and $k = 1, \ldots, K_n + d - r$ are calculated recursively from degree 0:

$$B_{r,t}^{(k)} = \frac{t - \bar{v}_{k-1}}{\bar{v}_{k+r-1} - \bar{v}_{k-1}} B_{r-1,t}^{(k)} + \frac{\bar{v}_{k+r} - t}{\bar{v}_{k+r} - \bar{v}_k} B_{r-1,t}^{(k+1)}, \tag{A.3}$$

where

$$
\begin{aligned}
\frac{t-\bar{v}_{k-1}}{\bar{v}_{k+r-1}-\bar{v}_{k-1}} B_{r-1,t}^{(k)} &= 0, \qquad \text{if } \bar{v}_{k+r-1}=\bar{v}_{k-1}; \\
\frac{\bar{v}_{k+r}-t}{\bar{v}_{k+r}-\bar{v}_{k}} B_{r-1,t}^{(k+1)} &= 0, \qquad \text{if } \bar{v}_{k+r}=\bar{v}_{k}.
\end{aligned} \tag{A.4}
$$

The B-spline basis functions $B_t^{(k)}$ have the following properties (de Boor, 1978):

(i) $B_t^{(k)} \geq 0$;

(ii) Partition of unity: $\sum_{k=1}^{K_n} B_t^{(k)} = 1$;

(iii) Local support: For each $k = 1, \ldots, K_n$, $B_t^{(k)}$ is nonzero only on an interval covering at most $d+1$ consecutive knot intervals. For any function f, define its support as $\operatorname{supp}(f) = \{t : f(t) \neq 0\}$, and denote by $|[a,b]| = b-a$ the interval length. Therefore, it holds that

$$
|\operatorname{supp}(B^{(k)})| \leq (d+1)H_n \asymp K_n^{-1}, \qquad \text{where } H_n = \max_{1\leq j\leq N_n+1}(v_j - v_{j-1}), \tag{A.5}
$$

by the quasi-uniformity of partitions. Equivalently, for any $t \in [0,T]$, at most $d+1$ basis functions in $\{B_t^{(k)}\}_{k=1}^{K_n}$ are nonzero, i.e., $|\{k : B_t^{(k)} > 0\}| \leq d+1$.

(iv) (Lemma A.1, Huang et al., 2004). Let $g_t = \sum_{k=1}^{K_n} B_t^{(k)}\gamma^{(k)}$ with $\gamma^{(k)} = (\gamma_1^{(k)}, \ldots, \gamma_p^{(k)})^\top \in \mathbb{R}^p$ and $\boldsymbol{\gamma} = ((\gamma^{(1)})^\top, \ldots, (\gamma^{(K_n)})^\top)^\top \in \mathbb{R}^{pK_n}$. Then it holds that for any $\boldsymbol{\gamma} \in \mathbb{R}^{pK_n}$,

$$
\|g\|_{L^2}^2 \asymp \sum_{j=1}^{p} \|g_j\|_{L^2}^2 \asymp \frac{\|\boldsymbol{\gamma}\|^2}{K_n}. \tag{A.6}
$$

A.2 Proof of Lemma 1

Under Assumption 1 (iii), c is bounded and nonsingular, i.e., there exist constants $0 < \underline{\kappa} \leq \overline{\kappa} < \infty$ such that for all $x \in \mathbb{R}^p$ and any $t \in [0,T]$, it holds that

$$
\underline{\kappa}\|x\|^2 \leq x^\top c_t x \leq \overline{\kappa}\|x\|^2, \tag{A.7}
$$

and thus implies the norm equivalence, i.e., for all measurable $f : [0,T] \to \mathbb{R}^p$,

$$
\sqrt{\underline{\kappa}}\|f\|_{L^2} \leq \|f\|_{L^2(c)} \leq \sqrt{\overline{\kappa}}\|f\|_{L^2}. \tag{A.8}
$$

Therefore, by definition of the $L^2(c)$-orthogonal projection, for any $g \in \mathbb{G}_n$,

$$
\|\beta - \Pi_n\beta\|_{L^2} \leq \frac{1}{\sqrt{\underline{\kappa}}}\|\beta - \Pi_n\beta\|_{L^2(c)} \leq \frac{1}{\sqrt{\underline{\kappa}}}\|\beta - g\|_{L^2(c)} \leq \sqrt{\frac{\overline{\kappa}}{\underline{\kappa}}}\|\beta - g\|_{L^2}. \tag{A.9}
$$

Hence it suffices to construct a single $\overline{\beta} \in \mathbb{G}_n$ that satisfies

$$\|\beta - \overline{\beta}\|_{L^2} = O_p\left(\sqrt{\frac{\log K_n}{K_n}}\right). \tag{A.10}$$

We consider a specific $\overline{\beta}$ as a local spline approximation, i.e. the quasi-interpolant (Chapter XII, de Boor, 1978): For each $k = 1, \ldots, K_n$, choose an arbitrary $\tau^{(k)} \in \mathrm{supp}(B^{(k)})$, and define

$$\overline{\beta}_t = \sum_{k=1}^{K_n} \beta_{\tau^{(k)}} B_t^{(k)}, \tag{A.11}$$

which satisfies $\overline{\beta} \in \mathbb{G}_n$ by construction, and it remains to bound $\|\beta - \overline{\beta}\|_{L^2}$. The local support property of B-splines (Appendix A.1 (iii)) implies that, whenever $B_t^{(k)} > 0$, $|t - \tau^{(k)}| \leq |\mathrm{supp}(B^{(k)})| \leq (d+1)H_n \asymp K_n^{-1}$.

We denote by β^c the continuous component of β, and $\beta^j = \beta - \beta^c$ collects the jumps:

$$\beta_t^c = \beta_0 + \int_0^t b_s^\beta ds + \int_0^t \sigma_s^\beta dW_s^\beta, \qquad \beta_t^j = \int_0^t \int_{\mathbb{R}^p} \delta^\beta(s,x)\mu^\beta(ds,dx). \tag{A.12}$$

Let $\overline{\beta}^c$ and $\overline{\beta}^j$ denote the corresponding quasi-interpolants (on the same knots) of β^c and β^j, and thus $\overline{\beta} = \overline{\beta}^c + \overline{\beta}^j$ by linearity of Eq. (A.11). By the triangle inequality,

$$\|\beta - \overline{\beta}\|_{L^2} \leq \|\beta^c - \overline{\beta}^c\|_{L^2} + \|\beta^j - \overline{\beta}^j\|_{L^2}, \tag{A.13}$$

and we bound the two terms separately.

We start with $\|\beta^c - \overline{\beta}^c\|_{L^2}$. Define the modulus of continuity of β^c:

$$\omega_{\beta^c}(h) = \sup_{\substack{s,t \in [0,T] \\ |t-s| \leq h}} \|\beta_t^c - \beta_s^c\|. \tag{A.14}$$

By the convexity of the norm and $\sum_{k=1}^{K_n} B_t^{(k)} = 1$, for any $t \in [0,T]$,

$$\|\beta_t^c - \overline{\beta}_t^c\| \leq \left\| \sum_{k=1}^{K_n} (\beta_t^c - \beta_{\tau^{(k)}}^c) B_t^{(k)} \right\| \leq \sum_{k=1}^{K_n} \|\beta_t^c - \beta_{\tau^{(k)}}^c\| B_t^{(k)} \leq \omega_{\beta^c}((d+1)H_n), \tag{A.15}$$

and thus

$$\|\beta^c - \overline{\beta}^c\|_{L^2} \leq \sqrt{T} \sup_{t \in [0,T]} \|\beta_t^c - \overline{\beta}_t^c\| \leq \sqrt{T}\omega_{\beta^c}((d+1)H_n). \tag{A.16}$$

To calculate the asymptotic order of $\omega_{\beta^c}(h)$, we decompose $\beta^c = A + M$ with

$$A_t = \beta_0 + \int_0^t b_s^\beta ds, \qquad M_t = \int_0^t \sigma_s^\beta dW_s^\beta. \tag{A.17}$$

Under Assumption A.1, for $|t-s| \le h$,

$$\omega_A(h) = \sup_{\substack{s,t\in[0,T] \\ |t-s|\le h}} \|A_t - A_s\| \le \left| \int_s^t \|b_u^\beta\| du \right| \le Kh. \tag{A.18}$$

Moreover, M is a continuous local martingale, which is a time-changed Brownian motion by the Dambis-Dubins-Schwarz theorem. By Lévy's modulus of continuity for Brownian motion, it holds that when $h \to 0$,

$$\omega_M(h) = \sup_{\substack{s,t\in[0,T] \\ |t-s|\le h}} \|M_t - M_s\| = O_p\left(\sqrt{h \log \frac{1}{h}}\right). \tag{A.19}$$

Therefore, we obtain

$$\omega_{\beta^c}(h) \le \omega_A(h) + \omega_M(h) = O_p\left(\sqrt{h \log \frac{1}{h}}\right). \tag{A.20}$$

Taking $h = (d+1)H_n \asymp K_n^{-1}$, by Eq. (A.16), it holds that

$$\|\beta^c - \overline{\beta}^c\|_{L^2} = O_p\left(\sqrt{\frac{\log K_n}{K_n}}\right). \tag{A.21}$$

Next, we verify that $\|\beta^j - \overline{\beta}^j\|_{L^2} = O_p(K_n^{-1/2})$. Define $V_{\beta^j}([a,b])$ the total variation of β^j on the interval $[a,b]$, i.e.,

$$V_{\beta^j}([a,b]) = \int_a^b \int_{\mathbb{R}^p} \|\delta^\beta(s,x)\| \mu^\beta(ds,dx). \tag{A.22}$$

Then for any $s,t \in [a,b]$, $\|\beta_t^j - \beta_s^j\| \le V_{\beta^j}([a,b])$. Therefore, whenever $B_t^{(k)} > 0$, $\|\beta_t^j - \beta_{\tau^{(k)}}^j\| \le V_{\beta^j}(\mathrm{supp}(B^{(k)}))$. By the convexity of the norm, $\sum_{k=1}^{K_n} B_t^{(k)} = 1$, for any $t \in [0,T]$,

$$\|\beta_t^j - \overline{\beta}_t^j\| \le \sum_{k\in\{k:B_t^{(k)}>0\}} \|\beta_t^j - \beta_{\tau^{(k)}}^j\| B_t^{(k)} \le \sum_{k\in\{k:B_t^{(k)}>0\}} B_t^{(k)} V_{\beta^j}(\mathrm{supp}(B^{(k)})). \tag{A.23}$$

By the Cauchy-Schwarz inequality and $|\{k : B_t^{(k)} > 0\}| \le d+1$,

$$\|\beta_t^j - \overline{\beta}_t^j\|^2 \le (d+1) \sum_{k\in\{k:B_t^{(k)}>0\}} V_{\beta^j}^2(\mathrm{supp}(B^{(k)})). \tag{A.24}$$

Therefore, we have

$$\begin{aligned} \|\beta^j - \overline{\beta}^j\|_{L^2}^2 &= \int_0^T \|\beta_t^j - \overline{\beta}_t^j\|^2 dt \le (d+1) \sum_{k=1}^{K_n} V_{\beta^j}^2(\mathrm{supp}(B^{(k)})) \int_0^T \mathbb{1}_{\{B_t^{(k)}>0\}} dt \\ &\le (d+1)|\mathrm{supp}(B^{(k)})| \sum_{k=1}^{K_n} V_{\beta^j}^2(\mathrm{supp}(B^{(k)})) \le (d+1)^2 H_n \sum_{k=1}^{K_n} V_{\beta^j}^2(\mathrm{supp}(B^{(k)})). \end{aligned} \tag{A.25}$$

Since $\mathrm{supp}(B^{(k)})$ covers at most $d+1$ knot intervals and each knot interval belongs to the supports of at most $d+1$ B-splines,

$$\begin{aligned}\sum_{k=1}^{K_n} V_{\beta^j}^2(\mathrm{supp}(B^{(k)})) &\le (d+1)^2 \sum_{j=1}^{N_n+1} V_{\beta^j}^2([v_{j-1}, v_j]) \\ &\le (d+1)^2 \left(\sum_{j=1}^{N_n+1} V_{\beta^j}([v_{j-1}, v_j])\right)^2 \le (d+1)^2 V_{\beta^j}^2([0,T]).\end{aligned} \tag{A.26}$$

Therefore, under Assumption 2, $V_{\beta^j}([0,T]) = O_p(1)$, and

$$\|\beta^j - \overline{\beta}^j\|_{L^2} \le K\sqrt{H_n} V_{\beta^j}([0,T]) = O_p(K_n^{-1/2}). \tag{A.27}$$

Combining Eqs. (A.13), (A.21) and (A.27) leads to Eq. (A.10). This completes the proof.

A.3 Proof of Theorem 1

Identifiability. We start with the identifiability of $\widehat{\boldsymbol{\gamma}}$ by verifying that $\mathbf{R}^\top\mathbf{R}$ is nonsingular. We first prove that the theoretical Gram matrix is positive definite, and then show that the theoretical and empirical Gram matrices are uniformly close over the estimation space, such that the empirical identifiability holds with $\widehat{\boldsymbol{\gamma}}$ uniquely defined; see Section 2 in Huang (2003) for further details.

Define the matrix $\widetilde{\mathbf{R}}$ as the theoretical counterpart of $\mathbf{R}$:

$$\begin{aligned}\widetilde{\mathbf{R}} &= \begin{pmatrix} \int_0^{\Delta_n} B_s^{(1)}\, dX_{1,s}^c & \cdots & \int_0^{\Delta_n} B_s^{(K_n)}\, dX_{1,s}^c & \cdots & \int_0^{\Delta_n} B_s^{(1)}\, dX_{p,s}^c & \cdots & \int_0^{\Delta_n} B_s^{(K_n)}\, dX_{p,s}^c \\ \int_{\Delta_n}^{2\Delta_n} B_s^{(1)}\, dX_{1,s}^c & \cdots & \int_{\Delta_n}^{2\Delta_n} B_s^{(K_n)}\, dX_{1,s}^c & \cdots & \int_{\Delta_n}^{2\Delta_n} B_s^{(1)}\, dX_{p,s}^c & \cdots & \int_{\Delta_n}^{2\Delta_n} B_s^{(K_n)}\, dX_{p,s}^c \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \int_{(n-1)\Delta_n}^{n\Delta_n} B_s^{(1)}\, dX_{1,s}^c & \cdots & \int_{(n-1)\Delta_n}^{n\Delta_n} B_s^{(K_n)}\, dX_{1,s}^c & \cdots & \int_{(n-1)\Delta_n}^{n\Delta_n} B_s^{(1)}\, dX_{p,s}^c & \cdots & \int_{(n-1)\Delta_n}^{n\Delta_n} B_s^{(K_n)}\, dX_{p,s}^c \end{pmatrix} \\ &= \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \mathbf{B}_s^\top dX_s^c\right)_{1\le i\le n}^\top \in \mathbb{R}^{n\times pK_n},\end{aligned} \tag{A.28}$$

and the corresponding Gram matrix is $\widetilde{\mathbf{R}}^\top\widetilde{\mathbf{R}}$. Let $g_t = (g_{1,t}, \ldots, g_{p,t})^\top$ with $g_{j,t} = \sum_{k=1}^{K_n} B_t^{(k)} v_j^{(k)}$ for any $v_j^{(k)} \in \mathbb{R}$ and $\boldsymbol{v} = (v_1^{(1)}, \ldots, v_1^{(K_n)}, \ldots, v_p^{(1)}, \ldots, v_p^{(K_n)})^\top \in \mathbb{R}^{pK_n}$, and $g \in \mathbb{G}_n$ by construction. By the properties of B-splines (Appendix A.1 (iv)) as well as bounded and nonsingular c, as $\Delta_n \to 0$,

$$\boldsymbol{v}^\top\widetilde{\mathbf{R}}^\top\widetilde{\mathbf{R}}\boldsymbol{v} \xrightarrow{\mathbb{P}} \int_0^T g_s^\top c_s g_s ds \asymp \sum_{j=1}^p \|g_j\|_{L^2}^2 \asymp \frac{\|\boldsymbol{v}\|^2}{K_n}, \qquad \text{where } \|\boldsymbol{v}\|^2 = \sum_{j=1}^p \sum_{k=1}^{K_n} (v_j^{(k)})^2, \tag{A.29}$$

which implies that $\widetilde{\mathbf{R}}^\top\widetilde{\mathbf{R}}$ is positive definite. The standard law of large numbers (LLN) result for truncated functionals (Theorem 9.2.1, Jacod and Protter, 2012) implies that, with $u_n \asymp \Delta_n^{\varpi}$ for some $0 < \varpi < 1/2$,

$$\boldsymbol{v}^\top \mathbf{R}^\top \mathbf{R} \boldsymbol{v} \xrightarrow{\mathbb{P}} \int_0^T g_s^\top c_s g_s ds, \tag{A.30}$$

as $\Delta_n \to 0$. By Lemma A.2 in Huang et al. (2004), it holds that

$$\sup_{g \in \mathbb{G}_n} \left| \frac{\boldsymbol{v}^\top \mathbf{R}^\top \mathbf{R} \boldsymbol{v}}{\boldsymbol{v}^\top \widetilde{\mathbf{R}}^\top \widetilde{\mathbf{R}} \boldsymbol{v}} - 1 \right| = o_p(1), \tag{A.31}$$

if $\Delta_n K_n \log K_n \to 0$. Therefore, the empirical Gram matrix $\mathbf{R}^\top \mathbf{R}$ is nonsingular and $\widehat{\boldsymbol{\gamma}}$ is uniquely defined with probability approaching one.

By Eqs. (A.29) and (A.31), for some $\epsilon_n = o_p(1)$ and $0 \le \epsilon_n < 1$ with probability approaching one,

$$\boldsymbol{v}^\top \mathbf{R}^\top \mathbf{R} \boldsymbol{v} \ge (1 - \epsilon_n) \boldsymbol{v}^\top \widetilde{\mathbf{R}}^\top \widetilde{\mathbf{R}} \boldsymbol{v} \asymp (1 - \epsilon_n) \frac{\|\boldsymbol{v}\|^2}{K_n}. \tag{A.32}$$

Let $\|\boldsymbol{v}\| \asymp 1$, the eigenvalue bound of Rayleigh quotient implies that $\lambda_{\min}(\mathbf{R}^\top \mathbf{R}) \ge K/K_n$, and thus $\|(\mathbf{R}^\top \mathbf{R})^{-1}\| = (\lambda_{\min}(\mathbf{R}^\top \mathbf{R}))^{-1} = O_p(K_n)$. A similar but reverse inequality to Eq. (A.32) indicates that $\lambda_{\max}(\mathbf{R}^\top \mathbf{R}) = O_p(K_n^{-1})$ and $\|\mathbf{R}\| = \sqrt{\lambda_{\max}(\mathbf{R}^\top \mathbf{R})} = O_p(K_n^{-1/2})$. The probabilistic bounds for $\|(\mathbf{R}^\top \mathbf{R})^{-1}\|$ and $\|\mathbf{R}\|$ will be used in the following proofs.

Consistency of $\widehat{\beta}$. The truncated increment of Y in Eq. (1) over the i-th interval is

$$\begin{aligned}
\mathbf{Y}_i &= \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \beta_{s-}^\top dX_s^c + \sum_{(i-1)\Delta_n \le s \le i\Delta_n} (\beta_{s-}^J)^\top \Delta X_s + \Delta_i^n Z \right) \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} \\
&= \underbrace{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \widetilde{\beta}_s^\top dX_s^c \right) \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}}_{\widetilde{\mathbf{Y}}_i} + \underbrace{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right) \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}}_{\mathbf{U}_i} \\
&\quad + \underbrace{\left(\sum_{(i-1)\Delta_n \le s \le i\Delta_n} (\beta_{s-}^J)^\top \Delta X_s \right) \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}}_{\mathbf{J}_i} + \underbrace{\Delta_i^n Z \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}}_{\mathbf{Z}_i},
\end{aligned} \tag{A.33}$$

where $\widetilde{\beta} = \mathbf{B}\boldsymbol{\gamma}$ is the spline approximation of β, and $e_t = \beta_{t-} - \widetilde{\beta}_t$.

Remark A.1 (Predictability)**.** Note that, since $H = \widetilde{\beta}$ or e is not predictable, whenever we write the stochastic integral $\int H_s^\top dX_s^c$, it is understood as $\int ({}^p H_s)^\top dX_s^c$, where ${}^p H$ is the predictable projection with ${}^p H_\tau = \mathbb{E}[H_\tau | \mathcal{F}_{\tau-}]$ for all predictable times τ; see Theorem I.2.28 in Jacod and Shiryaev (2013).

Since $\beta_{\tau-}$ is predictable under Assumption 1 (i), by Proposition I.2.18 in Jacod and Shiryaev (2013) and linearity of predictable projection, it holds that ${}^p\beta_{\tau-} = {}^p\widetilde{\beta}_\tau + {}^p e_\tau = \beta_{\tau-}$. In particular, we will apply the Itô isometry and the Burkholder-Davis-Gundy (BDG) inequality to ${}^p H$. Moreover, by Jensen's inequality, $\|{}^p H_s\|^2 \le {}^p(\|H_s\|^2)$, and thus whenever we use Itô isometry,

$$\mathbb{E}\left[\left(\int_0^T H_s^\top dX_s^c\right)^2\right] = \mathbb{E}\left[\int_0^T ({}^p H_s)^\top c_s ({}^p H_s) ds\right] \le K\mathbb{E}\left[\int_0^T \|H_s\|^2 ds\right] = K\mathbb{E}[\|H\|_{L^2}^2]. \quad \text{(A.34)}$$

by the boundedness of c under Assumption A.1.

Then, by Eq. (18),

$$\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = (\mathbf{R}^\top \mathbf{R})^{-1}\mathbf{R}^\top((\widetilde{\mathbf{Y}} - \mathbf{R}\boldsymbol{\gamma}) + \mathbf{U} + \mathbf{J} + \mathbf{Z}). \quad \text{(A.35)}$$

To prove $\|\widehat{\beta} - \beta\|_{L^2} = o_p(1)$, we verify that the following four components are all $o_p(1)$:

$$\begin{aligned} \|\widehat{\beta} - \beta\|_{L^2} &\le \|\widehat{\beta} - \widetilde{\beta}\|_{L^2} + \|e\|_{L^2} \\ &\le \underbrace{\|\mathbf{B}(\mathbf{R}^\top \mathbf{R})^{-1}\mathbf{R}^\top(\widetilde{\mathbf{Y}} - \mathbf{R}\boldsymbol{\gamma})\|_{L^2}}_{A_1} + \underbrace{\|\mathbf{B}(\mathbf{R}^\top \mathbf{R})^{-1}\mathbf{R}^\top \mathbf{U}\|_{L^2}}_{A_2} \\ &\quad + \underbrace{\|\mathbf{B}(\mathbf{R}^\top \mathbf{R})^{-1}\mathbf{R}^\top \mathbf{J}\|_{L^2}}_{A_3} + \underbrace{\|\mathbf{B}(\mathbf{R}^\top \mathbf{R})^{-1}\mathbf{R}^\top \mathbf{Z}\|_{L^2}}_{A_4} + \|e\|_{L^2}, \end{aligned} \quad \text{(A.36)}$$

where $\|e\|_{L^2} = o_p(1)$ by Lemma 1.

We start with two lemmas. Lemma A.1 is a result from the properties of B-spline basis functions and will be used to derive probabilistic bounds for the components listed above. Lemma A.2 provides some results regarding the truncation technique of Mancini (2009), and will be used repeatedly in subsequent proofs.

Lemma A.1. For any vector $\boldsymbol{r} \in \mathbb{R}^n$, $\|\mathbf{B}(\mathbf{R}^\top \mathbf{R})^{-1}\mathbf{R}^\top \boldsymbol{r}\|_{L^2} \le K\sqrt{K_n}\|\mathbf{R}^\top \boldsymbol{r}\| \le K'\|\boldsymbol{r}\|$.

Proof of Lemma A.1. It holds that, by Appendix A.1 (iv) and the Cauchy-Schwarz inequality,

$$\begin{aligned} \|\mathbf{B}(\mathbf{R}^\top \mathbf{R})^{-1}\mathbf{R}^\top \boldsymbol{r}\|_{L^2}^2 &\asymp K_n^{-1}\|(\mathbf{R}^\top \mathbf{R})^{-1}\mathbf{R}^\top \boldsymbol{r}\|^2 \le K K_n^{-1}\|(\mathbf{R}^\top \mathbf{R})^{-1}\|^2\|\mathbf{R}^\top \boldsymbol{r}\|^2 \\ &\le K' K_n \|\mathbf{R}\|^2\|\boldsymbol{r}\|^2 \le K''\|\boldsymbol{r}\|^2, \end{aligned} \quad \text{(A.37)}$$

where $\|(\mathbf{R}^\top \mathbf{R})^{-1}\| = O_p(K_n)$ and $\|\mathbf{R}\| = O_p(K_n^{-1/2})$ from Eq. (A.32). □

Lemma A.2. Let $u_n \asymp \Delta_n^{\varpi}$ for some $0 < \varpi < 1/2$. Then, as $\Delta_n \to 0$,

(i) $\mathbb{P}(\|\Delta_i^n X\| > u_n) \le K\Delta_n u_n^{-r}$, $\mathbb{P}(|\Delta_i^n Y| > u_n) \le K\Delta_n u_n^{-r}$;

(ii) $\|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} = O_p(\Delta_n u_n^{2-r})$.

Proof of Lemma A.2. (i) Under Assumption 1 (v), it holds that for $0 < u < 1$,

$$\int_{\{\|\delta\| > u\}} \lambda(dx) \le u^{-r}\int_{\mathbb{R}^p}(\|\delta(s,x)\|^r \wedge 1)\lambda(dx) \le u^{-r}\int_{\mathbb{R}^p} f_m(x)\lambda(dx) \le K u^{-r}, \quad \text{(A.38)}$$

$$\int_{\{\|\delta\| \le u\}} \|\delta(s,x)\|^2 \lambda(dx) \le u^{2-r}\int_{\mathbb{R}^p}(\|\delta(s,x)\|^r \wedge 1)\lambda(dx) \le K u^{2-r}. \quad \text{(A.39)}$$

$$
\begin{aligned}
\int_{\{\|\delta\|\leq u\}} \|\delta(s,x)\|\lambda(dx) &= \int_{\mathbb{R}^p} \|\delta(s,x)\| \mathbb{1}_{\{\|\delta(s,x)\|\leq u\}}\lambda(dx) = \int_{\mathbb{R}^p} \left(\int_0^{\|\delta(s,x)\|} \mathbb{1}_{\{y\leq u\}} dy \right) \lambda(dx) \\
&= \int_0^u \left(\int_{\mathbb{R}^p} \mathbb{1}_{\{\|\delta(s,x)\|>y\}}\lambda(dx) \right) dy - u \int_{\mathbb{R}^p} \mathbb{1}_{\{\|\delta(s,x)\|>u\}}\lambda(dx) \\
&\leq \int_0^u \left(\int_{\mathbb{R}^p} \mathbb{1}_{\{\|\delta(s,x)\|>y\}}\lambda(dx) \right) dy \\
&\leq \int_0^u y^{-r}\, dy \int_{\mathbb{R}^p} (\|\delta(s,x)\|^r \wedge 1)\lambda(dx) \\
&\leq \left(\int_{\mathbb{R}^p} f_m(x)\lambda(dx) \right) \int_0^u y^{-r} dy \leq K u^{1-r}.
\end{aligned}
\tag{A.40}
$$

Define the count of large jumps and the sum of small jumps with the cutoff level u_n:

$$
N_i^n(u_n) = \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} \mathbb{1}_{\{\|\delta(s,x)\|>u_n\}}\mu(ds,dx), \tag{A.41}
$$

$$
S_i^n(u_n) = \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\{\|\delta\|\leq u_n\}} \delta(s,x)\mu(ds,dx), \tag{A.42}
$$

and consider the following event decomposition:

$$
E_n = \left\{ \|\Delta_i^n X\| > u_n,\, \|\Delta_i^n X^c\| \leq \frac{u_n}{2},\, N_i^n\left(\frac{u_n}{2}\right) = 0 \right\}, \qquad E_n^{\complement} = \{\|\Delta_i^n X\| > u_n\} \backslash E_n. \tag{A.43}
$$

By the triangle inequality, on E_n,

$$
\left\| S_i^n\left(\frac{u_n}{2}\right) \right\| = \|\Delta_i^n X - \Delta_i^n X^c\| \geq \|\Delta_i^n X\| - \|\Delta_i^n X^c\| > \frac{u_n}{2}, \tag{A.44}
$$

so that $\{\|\Delta_i^n X\| > u_n\} \subseteq \{\|S_i^n(u_n/2)\| > u_n/2\} \cup E_n^{\complement}$. Therefore,

$$
\mathbb{P}(\|\Delta_i^n X\| > u_n) \leq \mathbb{P}\left(\|\Delta_i^n X^c\| > \frac{u_n}{2} \right) + \mathbb{P}\left(N_i^n\left(\frac{u_n}{2}\right) \geq 1 \right) + \mathbb{P}\left(\left\| S_i^n\left(\frac{u_n}{2}\right) \right\| > \frac{u_n}{2} \right), \tag{A.45}
$$

so it suffices to bound all three probabilities on the right-hand side in Eq. (A.45).

Choose any $q \geq 2$ that satisfies $q > \frac{1-\varpi r}{1/2-\varpi}$, and it holds that

$$
\mathbb{P}\left(\|\Delta_i^n X^c\| > \frac{u_n}{2} \right) \leq \frac{\mathbb{E}[\|\Delta_i^n X^c\|^q]}{(u_n/2)^q} \leq K \Delta_n^{q/2} u_n^{-q} = o(\Delta_n u_n^{-r}). \tag{A.46}
$$

by Markov's inequality and the BDG inequality. Then, by Eq. (A.38),

$$
\mathbb{P}\left(N_i^n\left(\frac{u_n}{2}\right) \geq 1 \right) \leq \mathbb{E}\left[N_i^n\left(\frac{u_n}{2}\right) \right] \leq \mathbb{E}\left[\int_{(i-1)\Delta_n}^{i\Delta_n} ds \int_{\{\|\delta\|>u_n/2\}} \lambda(dx) \right] \leq K \Delta_n u_n^{-r}. \tag{A.47}
$$

By subadditivity of the Euclidean norm,

$$
\left\| S_i^n\left(\frac{u_n}{2}\right) \right\| \leq \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\{\|\delta\|\leq u_n/2\}} \|\delta(s,x)\|\mu(ds,dx), \tag{A.48}
$$

and Markov's inequality implies that

$$\begin{aligned}\mathbb{P}\left(\left\|S_i^n\left(\frac{u_n}{2}\right)\right\| > \frac{u_n}{2}\right) &\le \mathbb{P}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\{\|\delta\|\le u_n/2\}}\|\delta(s,x)\|\mu(ds,dx) > \frac{u_n}{2}\right)\\ &\le \frac{2}{u_n}\mathbb{E}\left[\int_{(i-1)\Delta_n}^{i\Delta_n} ds\int_{\{\|\delta\|\le u_n/2\}}\|\delta(s,x)\|\lambda(dx)\right] \le K\Delta_n u_n^{-r},\end{aligned} \tag{A.49}$$

by Eq. (A.40). Combining Eqs. (A.46), (A.47) and (A.49) in Eq. (A.45), we obtain

$$\mathbb{P}(\|\Delta_i^n X\| > u_n) \le K\Delta_n u_n^{-r}. \tag{A.50}$$

We write $\Delta_i^n Y$ into

$$\Delta_i^n Y = \underbrace{\int_{(i-1)\Delta_n}^{i\Delta_n}\beta_{s-}^\top b_s ds + \int_{(i-1)\Delta_n}^{i\Delta_n}\beta_{s-}^\top\sigma_s dW_s}_{\Delta_i^n Y^{(1)}} + \underbrace{\sum_{(i-1)\Delta_n\le s\le i\Delta_n}(\beta_{s-}^J)^\top\Delta X_s}_{\Delta_i^n Y^{(2)}} + \Delta_i^n Z, \tag{A.51}$$

and by the triangle inequality,

$$\begin{aligned}\mathbb{P}(|\Delta_i^n Y| > u_n) \le\ &\mathbb{P}\left(|\Delta_i^n Y^{(1)}| > \frac{u_n}{4}\right) + \mathbb{P}\left(|\Delta_i^n Y^{(2)}| > \frac{u_n}{4}\right)\\ &+ \mathbb{P}\left(|\Delta_i^n Z^c| > \frac{u_n}{4}\right) + \mathbb{P}\left(|\Delta_i^n Z - \Delta_i^n Z^c| > \frac{u_n}{4}\right).\end{aligned} \tag{A.52}$$

By bounded β under Assumption A.1 and the BDG inequality, for any $q > 2$,

$$\mathbb{E}[|\Delta_i^n Y^{(1)}|^q] \le K\mathbb{E}\left[\left(\Delta_n^2 + \int_{(i-1)\Delta_n}^{i\Delta_n}\|\beta_s^\top\sigma_s\|^2 ds\right)^{q/2}\right] \le K'\Delta_n^{q/2}. \tag{A.53}$$

Then by Markov's inequality,

$$\mathbb{P}\left(|\Delta_i^n Y^{(1)}| > \frac{u_n}{4}\right) \le \frac{\mathbb{E}[|\Delta_i^n Y^{(1)}|^q]}{(u_n/4)^q} \le K\Delta_n^{q/2}u_n^{-q} = o(\Delta_n u_n^{-r}), \qquad \text{if } q > \frac{1-\varpi r}{1/2-\varpi}, \tag{A.54}$$

which is always possible because $0 < \varpi < 1/2$ and $0 \le r < 1$.

By bounded β^J under Assumption A.1, there exists some deterministic $M > 0$ such that

$$|\Delta_i^n Y^{(2)}| \le \sum_{(i-1)\Delta_n\le s\le i\Delta_n}\|\beta_s^J\|\|\Delta X_s\| \le M\sum_{(i-1)\Delta_n\le s\le i\Delta_n}\|\Delta X_s\|, \tag{A.55}$$

and thus

$$\begin{aligned}\mathbb{P}\left(|\Delta_i^n Y^{(2)}| > \frac{u_n}{4}\right) &\le \mathbb{P}\left(N_i^n\left(\frac{u_n}{8M}\right) \ge 1\right) + \mathbb{P}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\{\|\delta\|\le u_n/8M\}}\|\delta(s,x)\|\mu(ds,dx) > \frac{u_n}{8M}\right)\\ &\le K\Delta_n u_n^{-r},\end{aligned} \tag{A.56}$$

which follows the same steps as in Eq. (A.47) and the second inequality in Eq. (A.49).

For the last two terms in Eq. (A.52), we apply the same arguments to the one-dimensional Itô semimartingale Z under Assumption 1 (vi). Combining them, Eqs. (A.54) and (A.56) in Eq. (A.52) leads to

$$\mathbb{P}(|\Delta_i^n Y| > u_n) \leq K\Delta_n u_n^{-r}. \tag{A.57}$$

This completes the proof of Lemma A.2 (i).

(ii) We consider the following event decomposition:

$$\{\|\Delta_i^n X\| \leq u_n\} = \{\|\Delta_i^n X\| \leq u_n, \|\Delta_i^n X^c\| \leq u_n\} \cup \{\|\Delta_i^n X\| \leq u_n, \|\Delta_i^n X^c\| > u_n\}. \tag{A.58}$$

For each i, on the event $\{\|\Delta_i^n X\| \leq u_n, \|\Delta_i^n X^c\| \leq u_n\}$,

$$\|\Delta_i^n X - \Delta_i^n X^c\| \leq \|\Delta_i^n X\| + \|\Delta_i^n X^c\| \leq 2u_n, \tag{A.59}$$

and thus

$$\begin{aligned} \|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \leq\ & \underbrace{\|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\|\Delta_i^n X - \Delta_i^n X^c\| \leq 2u_n}}_{V_i^{(1)}} \\ & + \underbrace{\|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\|\Delta_i^n X\| \leq u_n} \mathbb{1}_{\|\Delta_i^n X^c\| > u_n}}_{V_i^{(2)}}. \end{aligned} \tag{A.60}$$

For $V_i^{(1)}$, since the function $x \mapsto x \wedge a$ is subadditive,

$$V_i^{(1)} \leq (\|\Delta_i^n X - \Delta_i^n X^c\| \wedge 2u_n)^2 \leq \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n) \mu(ds, dx) \right)^2. \tag{A.61}$$

Hence, using $(a+b)^2 \leq 2a^2 + 2b^2$ and $\mu = \nu + (\mu - \nu)$,

$$\begin{aligned} \mathbb{E}[V_i^{(1)}] \leq\ & 2\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)(\mu - \nu)(ds, dx) \right)^2 \right] \\ & + 2\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)\nu(ds, dx) \right)^2 \right]. \end{aligned} \tag{A.62}$$

For the first term in Eq. (A.62), by the Itô isometry for compensated Poisson integrals,

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}^p}(\|\delta(s,x)\|\wedge 2u_n)(\mu-\nu)(ds,dx)\right)^2\right]\\
&\quad=\mathbb{E}\left[\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}^p}(\|\delta(s,x)\|\wedge 2u_n)^2\nu(ds,dx)\right]\\
&\quad\leq \mathbb{E}\left[\int_{(i-1)\Delta_n}^{i\Delta_n}ds\left(\int_{\{\|\delta\|\leq 2u_n\}}\|\delta(s,x)\|^2\lambda(dx)+4u_n^2\int_{\{\|\delta\|>2u_n\}}\lambda(dx)\right)\right]\\
&\quad\leq K\Delta_n u_n^{2-r},
\end{aligned}\tag{A.63}$$

by Eqs. (A.38) and (A.39). For the second term in Eq. (A.62), by Jensen's inequality,

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\mathbb{R}^p}(\|\delta(s,x)\|\wedge 2u_n)\nu(ds,dx)\right)^2\right]\\
&\quad\leq \Delta_n\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\left[\left(\int_{\mathbb{R}^p}(\|\delta(s,x)\|\wedge 2u_n)\lambda(dx)\right)^2\right]ds\\
&\quad\leq \Delta_n\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\left[\left(\int_{\{\|\delta\|\leq 2u_n\}}\|\delta(s,x)\|\lambda(dx)+2u_n\int_{\{\|\delta\|>2u_n\}}\lambda(dx)\right)^2\right]ds\\
&\quad\leq K\Delta_n\int_{(i-1)\Delta_n}^{i\Delta_n}u_n^{2-2r}ds=K\Delta_n^2u_n^{2-2r},
\end{aligned}\tag{A.64}$$

by Eqs. (A.38) and (A.40). Combining Eqs. (A.62) to (A.64), we obtain

$$\mathbb{E}[V_i^{(1)}]\leq K\Delta_n u_n^{2-r}+K'\Delta_n^2u_n^{2-2r}\leq K\Delta_n u_n^{2-r}.\tag{A.65}$$

For $V_i^{(2)}$, on the event $\{\|\Delta_i^nX\|\leq u_n, \|\Delta_i^nX^c\|>u_n\}$,

$$\|\Delta_i^nX-\Delta_i^nX^c\|\leq u_n+\|\Delta_i^nX^c\|\leq 2\|\Delta_i^nX^c\|,\tag{A.66}$$

and thus $\mathbb{E}[V_i^{(2)}]\leq 4\mathbb{E}\left[\|\Delta_i^nX^c\|^2\mathbb{1}_{\{\|\Delta_i^nX^c\|>u_n\}}\right]$. It holds that, for any $q>2$,

$$\mathbb{E}\left[\|\Delta_i^nX^c\|^2\mathbb{1}_{\{\|\Delta_i^nX^c\|>u_n\}}\right]\leq \mathbb{E}\left[\|\Delta_i^nX^c\|^2\left(\frac{\|\Delta_i^nX^c\|}{u_n}\right)^{q-2}\right]\leq\frac{\mathbb{E}[\|\Delta_i^nX^c\|^q]}{u_n^{q-2}}\leq \Delta_n^{q/2}u_n^{2-q}.\tag{A.67}$$

by the BDG inequality. Therefore,

$$\mathbb{E}[V_i^{(2)}]\leq \Delta_n^{q/2}u_n^{2-q}=o(\Delta_nu_n^{2-r}),\qquad \text{if } q>\frac{1-\varpi r}{1/2-\varpi},\tag{A.68}$$

which is always possible because $0<\varpi<1/2$ and $0\leq r<1$.

Finally, by Eqs. (A.60), (A.65) and (A.68),

$$\mathbb{E}\left[\|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}\right] = O(\Delta_n u_n^{2-r}), \tag{A.69}$$

and thus, by Markov's inequality,

$$\|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} = O_p(\Delta_n u_n^{2-r}). \tag{A.70}$$

This completes the proof of Lemma A.2 (ii).

□

Next, we verify that A_1, A_2, A_3 and A_4 in Eq. (A.36) are all $o_p(1)$:

(i) $\boldsymbol{A_1 = o_p(1)}$. We consider a further decomposition $\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Y}}^{(1)} + \widetilde{\mathbf{Y}}^{(2)}$:

$$\widetilde{\mathbf{Y}}_i^{(1)} = \widetilde{\beta}_{(i-1)\Delta_n}^\top \Delta_i^n X^c \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}, \tag{A.71}$$

$$\widetilde{\mathbf{Y}}_i^{(2)} = \left(\int_{(i-1)\Delta_n}^{i\Delta_n} (\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n})^\top dX_s^c\right) \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}. \tag{A.72}$$

The i-th element in $\widetilde{\mathbf{Y}}^{(1)} - \mathbf{R}\boldsymbol{\gamma}$ is given by

$$\begin{aligned} &\widetilde{\beta}_{(i-1)\Delta_n}^\top \Delta_i^n X^c \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} - \widetilde{\beta}_{(i-1)\Delta_n}^\top \Delta_i^n X \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \\ = &\underbrace{\widetilde{\beta}_{(i-1)\Delta_n}^\top \Delta_i^n X^c \left(\mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} - \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}\right)}_{M_i^{(1)}} - \underbrace{\widetilde{\beta}_{(i-1)\Delta_n}^\top (\Delta_i^n X - \Delta_i^n X^c) \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}}_{M_i^{(2)}}, \end{aligned} \tag{A.73}$$

and thus

$$\|\widetilde{\mathbf{Y}}^{(1)} - \mathbf{R}\boldsymbol{\gamma}\|^2 = \sum_{i=1}^n (M_i^{(1)} - M_i^{(2)})^2 \le 2\sum_{i=1}^n (M_i^{(1)})^2 + 2\sum_{i=1}^n (M_i^{(2)})^2. \tag{A.74}$$

For $M_i^{(1)}$ and $M_i^{(2)}$, since $\|\widetilde{\beta}\|_\infty$ is bounded by de Boor's conjecture (see, e.g., de Boor, 1973; Shadrin, 2001), it holds that, by the BDG inequality and Lemma A.2 (i),

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[(M_i^{(1)})^2] &\le K \sum_{i=1}^n \mathbb{E}\left[\|\Delta_i^n X^c\|^2 \left(\mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} - \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}\right)\right] \\ &\le K \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}[\|\Delta_i^n X^c\|^2 | \mathcal{F}_{(i-1)\Delta_n}] \left(\mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} - \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}\right)\right] \\ &\le K' \Delta_n \sum_{i=1}^n (\mathbb{P}(|\Delta_i^n Y| > u_n) + \mathbb{P}(\|\Delta_i^n X\| > u_n)) \\ &\le K'' \Delta_n u_n^{-r} = o(1), \end{aligned} \tag{A.75}$$

and, by Lemma A.2 (ii),

$$\sum_{i=1}^{n}\mathbb{E}[(M_i^{(2)})^2] \le K\sum_{i=1}^{n}\mathbb{E}\left[\|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X\|\le u_n\}}\right] \le K u_n^{2-r} = o(1). \tag{A.76}$$

Therefore, $\mathbb{E}[\|\widetilde{\mathbf{Y}}^{(1)} - \mathbf{R}\boldsymbol{\gamma}\|^2] = o(1)$, and thus $\|\widetilde{\mathbf{Y}}^{(1)} - \mathbf{R}\boldsymbol{\gamma}\| = o_p(1)$ by Markov's inequality.

Next, we verify that $\|\widetilde{\mathbf{Y}}^{(2)}\| = o_p(1)$. By the Itô isometry (see Remark A.1) and bounded c,

$$\mathbb{E}[\|\widetilde{\mathbf{Y}}^{(2)}\|^2] \le K\sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}[\|\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n}\|^2]ds. \tag{A.77}$$

Since $\widetilde{\beta}$ is absolutely continuous and has bounded derivative on $[0,T]$,

$$\|\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n}\| \le (s-(i-1)\Delta_n)\|\widetilde{\beta}'\|_\infty, \qquad \text{where } \|\widetilde{\beta}'\|_\infty = \sup_{s\in[0,T]}\|\widetilde{\beta}'_s\|, \tag{A.78}$$

and $\widetilde{\beta}'_s$ stands for the derivative of $\widetilde{\beta}$ on s. By Eqs. (A.77) and (A.78),

$$\begin{aligned}\mathbb{E}[\|\widetilde{\mathbf{Y}}^{(2)}\|^2] &\le K\mathbb{E}[\|\widetilde{\beta}'\|_\infty^2]\sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n}(s-(i-1)\Delta_n)^2 ds\\ &\le K\mathbb{E}[\|\widetilde{\beta}'\|_\infty^2]\sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n}\Delta_n^2 ds = K\mathbb{E}[\|\widetilde{\beta}'\|_\infty^2]\sum_{i=1}^{n}\frac{\Delta_n^3}{3}\\ &\le K'\Delta_n^2\mathbb{E}[\|\widetilde{\beta}'\|_\infty^2] = O(\Delta_n^2 K_n^2) = o(1).\end{aligned} \tag{A.79}$$

because $\|\widetilde{\beta}'\|_\infty = O_p(K_n)$ and $\Delta_n K_n \log K_n \to 0$, and thus $\|\widetilde{\mathbf{Y}}^{(2)}\| = o_p(1)$ by Markov's inequality.

By Lemma A.1, it holds that

$$\begin{aligned}A_1 &\le \|\mathbf{B}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top(\widetilde{\mathbf{Y}}^{(1)} - \mathbf{R}\boldsymbol{\gamma})\|_{L^2} + \|\mathbf{B}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\widetilde{\mathbf{Y}}^{(2)}\|_{L^2}\\ &\le K\|\widetilde{\mathbf{Y}}^{(1)} - \mathbf{R}\boldsymbol{\gamma}\| + K'\|\widetilde{\mathbf{Y}}^{(2)}\| = o_p(1).\end{aligned} \tag{A.80}$$

(ii) $\boldsymbol{A_2 = o_p(1)}$. Conditional on a σ-field $\mathcal{G} = \sigma(e_s : 0\le s\le T)$, it holds that $\mathbb{E}[\|\mathbf{U}\|^2|\mathcal{G}] \le K\|e\|_{L^2}^2$ by the Itô isometry and bounded c. Fix $\varepsilon > 0$. For any $\eta > 0$, it holds that by Chebyshev's inequality,

$$\begin{aligned}\mathbb{P}(\|\mathbf{U}\| > \varepsilon) &\le \mathbb{P}(\|e\|_{L^2} > \eta) + \mathbb{P}(\|\mathbf{U}\| > \varepsilon, \|e\|_{L^2}\le\eta)\\ &\le \mathbb{P}(\|e\|_{L^2} > \eta) + \frac{\mathbb{E}[\|\mathbf{U}\|^2\mathbb{1}_{\{\|e\|_{L^2}\le\eta\}}]}{\varepsilon^2}\\ &\le \mathbb{P}(\|e\|_{L^2} > \eta) + \frac{\mathbb{E}[\mathbb{E}[\|\mathbf{U}\|^2|\mathcal{G}]\mathbb{1}_{\{\|e\|_{L^2}\le\eta\}}]}{\varepsilon^2}\\ &\le \mathbb{P}(\|e\|_{L^2} > \eta) + K\frac{\mathbb{E}[\|e\|_{L^2}^2\mathbb{1}_{\{\|e\|_{L^2}\le\eta\}}]}{\varepsilon^2}\\ &\le \mathbb{P}(\|e\|_{L^2} > \eta) + K\frac{\eta^2}{\varepsilon^2},\end{aligned} \tag{A.81}$$

where $\mathbb{P}(\|e\|_{L^2} > \eta) \to 0$ since $\|e\|_{L^2} = o_p(1)$ by Lemma 1. Let $\eta \downarrow 0$, and then $\|\mathbf{U}\| = o_p(1)$. Therefore, $A_2 \leq K\|\mathbf{U}\| = o_p(1)$ by Lemma A.1.

(iii) $A_3 = o_p(1)$. We define a vector $\widetilde{\mathbf{J}} \in \mathbb{R}^n$ with the i-th element

$$\widetilde{\mathbf{J}}_i = \mathbf{J}_i \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} = \left(\sum_{(i-1)\Delta_n \leq s \leq i\Delta_n} (\beta_{s-}^J)^\top \Delta X_s \right) \mathbb{1}_{\{|\Delta_i^n Y| \leq u_n\}} \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}}. \tag{A.82}$$

Because the design matrix $\mathbf{R}$ already includes the same truncation $\mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}}$ for covariate increments, we have $\mathbf{R}^\top \mathbf{J} = \mathbf{R}^\top \widetilde{\mathbf{J}}$. Hence, by Lemma A.1, it holds that

$$A_3 \leq K\sqrt{K_n}\|\mathbf{R}^\top \mathbf{J}\| = K\sqrt{K_n}\|\mathbf{R}^\top \widetilde{\mathbf{J}}\| \leq K'\|\widetilde{\mathbf{J}}\|, \tag{A.83}$$

so that it suffices to show that $\|\widetilde{\mathbf{J}}\| = o_p(1)$. By bounded β^J under Assumption A.1, it holds that

$$\begin{aligned}
|\widetilde{\mathbf{J}}_i|^2 &\leq \left(\sum_{(i-1)\Delta_n \leq s \leq i\Delta_n} \|\beta_{s-}^J\| \|\Delta X_s\| \right)^2 \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \leq K \left(\sum_{(i-1)\Delta_n \leq s \leq i\Delta_n} \|\Delta X_s\| \right)^2 \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \\
&\leq K \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} \|\delta(s,x)\| \mu(ds,dx) \right)^2 \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \\
&\leq K \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n) \mu(ds,dx) + \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\{\|\delta\| > 2u_n\}} \|\delta(s,x)\| \mu(ds,dx) \right)^2 \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \\
&\leq 2K \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n) \mu(ds,dx) \right)^2 \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \\
&\quad + 2K \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\{\|\delta\| > 2u_n\}} \|\delta(s,x)\| \mu(ds,dx) \right)^2 \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}},
\end{aligned} \tag{A.84}$$

and thus

$$\begin{aligned}
\mathbb{E}[|\widetilde{\mathbf{J}}_i|^2] \leq K\mathbb{E} &\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n) \mu(ds,dx) \right)^2 \right] \\
&+ K'\mathbb{E} \left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\{\|\delta\| > 2u_n\}} \|\delta(s,x)\| \mu(ds,dx) \right)^2 \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \right].
\end{aligned} \tag{A.85}$$

Following the same steps as in Eqs. (A.62) to (A.65), we obtain

$$\mathbb{E} \left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n) \mu(ds,dx) \right)^2 \right] \leq K\Delta_n u_n^{2-r}. \tag{A.86}$$

With bounded δ under Assumption A.1, it holds that

$$\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\{\|\delta\|>2u_n\}}\|\delta(s,x)\|\mu(ds,dx)\right)^2 \mathbb{1}_{\{\|\Delta_i^n X\|\le u_n\}} \le K(N_i^n(2u_n))^2\mathbb{1}_{\{\|\Delta_i^n X\|\le u_n\}}, \tag{A.87}$$

where $N_i^n(2u_n)$ is defined in Eq. (A.41). Therefore, the left-hand side in the above inequality must be zero unless on the event $\{N_i^n(2u_n)\ge 1,\ \|\Delta_i^n X\|\le u_n\}$. We consider the following event relations:

$$\begin{aligned}\{N_i^n(2u_n)\ge 1,\ \|\Delta_i^n X\|\le u_n\} &\subseteq \left\{N_i^n(2u_n)\ge 1,\ \|\Delta_i^n X\|\le u_n,\ N_i^n\left(\frac{u_n}{2}\right)=1\right\}\\ &\quad\cup\left\{N_i^n(2u_n)\ge 1,\ \|\Delta_i^n X\|\le u_n,\ N_i^n\left(\frac{u_n}{2}\right)\ge 2\right\}\\ &\subseteq \{N_i^n(2u_n)=1,\ \|\Delta_i^n X\|\le u_n\}\cup\left\{N_i^n\left(\frac{u_n}{2}\right)\ge 2\right\}\end{aligned} \tag{A.88}$$

On the event $\{N_i^n(2u_n)=1,\ \|\Delta_i^n X\|\le u_n\}$, by the triangle inequality,

$$\left\|\Delta_i^n X^c + S_i^n\left(\frac{u_n}{2}\right)\right\| > 2u_n - \|\Delta_i^n X\| > u_n, \tag{A.89}$$

and therefore

$$\{N_i^n(2u_n)\ge 1,\ \|\Delta_i^n X\|\le u_n\}\ \subseteq\ \underbrace{\left\{\|\Delta_i^n X^c\|>\frac{u_n}{2}\right\}}_{E_n^{(1)}}\cup\underbrace{\left\{\left\|S_i^n\left(\frac{u_n}{2}\right)\right\|>\frac{u_n}{2}\right\}}_{E_n^{(2)}}\cup\underbrace{\left\{N_i^n\left(\frac{u_n}{2}\right)\ge 2\right\}}_{E_n^{(3)}}. \tag{A.90}$$

Then, by Eq. (A.87),

$$\begin{aligned}&\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\{\|\delta\|>2u_n\}}\|\delta(s,x)\|\mu(ds,dx)\right)^2\mathbb{1}_{\{\|\Delta_i^n X\|\le u_n\}}\right]\\ &\quad\le K\mathbb{E}\left[(N_i^n(2u_n))^2\mathbb{1}_{E_n^{(1)}}\right]+K'\mathbb{E}\left[(N_i^n(2u_n))^2\mathbb{1}_{E_n^{(2)}}\right]+K''\mathbb{E}\left[(N_i^n(2u_n))^2\mathbb{1}_{E_n^{(3)}}\right].\end{aligned} \tag{A.91}$$

By the factorial moment of Poisson random variable and Eq. (A.38),

$$\begin{aligned}\mathbb{E}\left[(N_i^n(2u_n))^2\mathbb{1}_{E_n^{(3)}}\right] &\le \mathbb{E}\left[\left(N_i^n\left(\frac{u_n}{2}\right)\right)^2\mathbb{1}_{E_n^{(3)}}\right]\le 2\mathbb{E}\left[\left(N_i^n\left(\frac{u_n}{2}\right)\right)\left(N_i^n\left(\frac{u_n}{2}-1\right)\right)\mathbb{1}_{E_n^{(3)}}\right]\\ &\le 2\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n}ds\int_{\{\|\delta\|>u_n/2\}}\lambda(dx)\right)^2\right]\le K'\Delta_n^2u_n^{-2r}.\end{aligned} \tag{A.92}$$

By the Cauchy-Schwarz inequality, Eqs. (A.38) and (A.47), take some $q\ge 2$ and $q>\frac{6(1-\varpi r)}{1-2\varpi}$, then

$$\begin{aligned}\mathbb{E}\left[(N_i^n(2u_n))^2\mathbb{1}_{E_n^{(1)}}\right] &\le \left(\mathbb{E}[(N_i^n(2u_n))^4]\right)^{1/2}\left(\mathbb{P}\left(\|\Delta_i^n X^c\|>\frac{u_n}{2}\right)\right)^{1/2}\\ &\le K(\Delta_n u_n^{-r})^{1/2}(\Delta_n^{q/2}u_n^{-q})^{1/2}=K\Delta_n^{1/2+q/4}u_n^{-(r+q)/2}=o(\Delta_n^2u_n^{-2r}).\end{aligned} \tag{A.93}$$

Following a similar approach to Eq. (A.49),

$$\begin{aligned}\mathbb{E}\left[(N_i^n(2u_n))^2 \mathbb{1}_{E_n^{(2)}}\right] &\le \mathbb{E}\left[(N_i^n(2u_n))^2 \left(\frac{2}{u_n}\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\{\|\delta\|\le u_n/2\}} \|\delta(s,x)\|\mu(ds,dx)\right)\right] \\ &= \frac{2}{u_n}\mathbb{E}\left[N_i^n(2u_n)(N_i^n(2u_n)-1)\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\{\|\delta\|\le u_n/2\}} \|\delta(s,x)\|\mu(ds,dx)\right)\right] \\ &\quad + \frac{2}{u_n}\mathbb{E}\left[N_i^n(2u_n)\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\{\|\delta\|\le u_n/2\}} \|\delta(s,x)\|\mu(ds,dx)\right)\right] \\ &\le \frac{K}{u_n}(\Delta_n u_n^{-r})^2(\Delta_n u_n^{1-r}) + \frac{K'}{u_n}(\Delta_n u_n^{-r})(\Delta_n u_n^{1-r}) \le K'\Delta_n^2 u_n^{-2r}. \end{aligned} \tag{A.94}$$

by Eqs. (A.38) and (A.40), and the fact that $\{\|\delta(s,x)\| > 2u_n\}$ and $\{\|\delta(s,x)\| \le u_n/2\}$ are disjoint.

Combining Eqs. (A.92) to (A.94) in Eq. (A.91) leads to

$$\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\int_{\{\|\delta\|>2u_n\}} \|\delta(s,x)\|\mu(ds,dx)\right)^2 \mathbb{1}_{\{\|\Delta_i^n X\|\le u_n\}}\right] \le K\Delta_n^2 u_n^{-2r}. \tag{A.95}$$

Then, combining Eqs. (A.86) and (A.95) in Eq. (A.85), we have

$$\mathbb{E}[|\widetilde{\mathbf{J}}_i|^2] \le K\Delta_n u_n^{2-r} + K'\Delta_n^2 u_n^{-2r}, \tag{A.96}$$

and therefore

$$\mathbb{E}[\|\widetilde{\mathbf{J}}\|^2] \le \sum_{i=1}^n \mathbb{E}[|\widetilde{\mathbf{J}}_i|^2] \le K u_n^{2-r} + K'\Delta_n u_n^{-2r} = o(1). \tag{A.97}$$

Therefore, we have $\|\widetilde{\mathbf{J}}\| = o_p(1)$ by Markov's inequality and thus $A_3 = o_p(1)$.

(iv) $A_4 = o_p(1)$. We denote the i'-th element in $\mathbf{R}^\top \mathbf{Z} \in \mathbb{R}^{pK_n}$ for $i'(j,k) = (j-1)K_n + k$ by

$$C_j^{(k)} = \sum_{i=1}^n B_{(i-1)\Delta_n}^{(k)} \Delta_{j,i}^n X \Delta_i^n Z \mathbb{1}_{\{|\Delta_i^n Y|\le u_n\}} \mathbb{1}_{\{\|\Delta_i^n X\|\le u_n\}}. \tag{A.98}$$

By Appendix A.1 (iv),

$$\Delta_n \sum_{i=1}^n (B_{(i-1)\Delta_n}^{(k)})^2 \xrightarrow{\mathbb{P}} \int_0^T (B_s^{(k)})^2 ds \asymp K_n^{-1}. \tag{A.99}$$

Under the orthogonality condition in Assumption 3, X and Z have no co-jumps, and the (fixed-weighted) realized covariance $\sum_{i=1}^n B_{(i-1)\Delta_n}^{(k)} \Delta_{j,i}^n X \Delta_i^n Z$ has zero limit. By the standard results of realized covariance functionals (see, e.g., Theorem 10.3.2, Jacod and Protter, 2012), it holds that

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} B^{(k)}_{(i-1)\Delta_n}\Delta^n_{j,i}X\Delta^n_i Z\right)^2\right] = O(K_n^{-1}\Delta_n), \tag{A.100}$$

and thus $\mathbb{E}[(C_j^{(k)})^2] = O(K_n^{-1}\Delta_n)$. Therefore,

$$\mathbb{E}[\|\mathbf{R}^\top \mathbf{Z}\|^2] = \sum_{l=1}^{p}\sum_{k=1}^{K_n}\mathbb{E}[(C_j^{(k)})^2] = O(\Delta_n), \tag{A.101}$$

and thus $\|\mathbf{R}^\top \mathbf{Z}\| = O_p(\sqrt{\Delta_n})$. By Lemma A.1,

$$A_4 \leq K\sqrt{K_n}\|\mathbf{R}^\top \mathbf{Z}\| = O_p(\sqrt{K_n\Delta_n}) = o_p(1), \tag{A.102}$$

when $\Delta_n K_n \log K_n \to 0$.

Therefore, all components in Eq. (A.36) are $o_p(1)$, and thus $\|\widehat{\beta} - \beta\|_{L^2} = o_p(1)$. This completes the proof.

A.4 Proof of Theorem 2

We start with the following decomposition:

$$\begin{aligned}
\frac{1}{\sqrt{\Delta_n}}(\widehat{I\beta}_T - I\beta_T) &= \frac{1}{\sqrt{\Delta_n}}\left(\sum_{i=1}^{n}\widehat{\beta}_{(i-1)\Delta_n}\Delta_n - \int_0^T \mathbf{B}_s\boldsymbol{\gamma}\,ds - \int_0^T e_s ds\right)\\
&= \frac{1}{\sqrt{\Delta_n}}\left(\sum_{i=1}^{n}\mathbf{B}_{(i-1)\Delta_n}\widehat{\boldsymbol{\gamma}}\Delta_n - \int_0^T \mathbf{B}_s\boldsymbol{\gamma} ds\right) - \frac{1}{\sqrt{\Delta_n}}\int_0^T e_s ds\\
&= \frac{1}{\sqrt{\Delta_n}}\left(\sum_{i=1}^{n}\mathbf{B}_{(i-1)\Delta_n}\Delta_n\right)(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) - \frac{1}{\sqrt{\Delta_n}}\left(\int_0^T \mathbf{B}_s ds - \sum_{i=1}^{n}\mathbf{B}_{(i-1)\Delta_n}\Delta_n\right)\boldsymbol{\gamma}\\
&\quad - \frac{1}{\sqrt{\Delta_n}}\int_0^T e_s ds.
\end{aligned} \tag{A.103}$$

To derive the asymptotic distribution of the p-dimensional vector in the first term by the Cramér-Wold device, we need to show that, for any fixed linear combination with $a \in \mathbb{R}^p$,

$$\begin{aligned}
H_n &= \frac{1}{\sqrt{\Delta_n}}a^\top\left(\sum_{i=1}^{n}\mathbf{B}_{(i-1)\Delta_n}\Delta_n\right)(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\\
&\xrightarrow{\mathcal{L}-s} \int_0^T \sqrt{\widetilde{\sigma}_s^2}\left(a^\top\left(\int_0^T \mathbf{B}_s ds\right)\left(\int_0^T \mathbf{B}_s^\top c_s \mathbf{B}_s ds\right)^{-1}\mathbf{B}_s^\top\right)^\top c_s^{1/2} dW'_s,
\end{aligned} \tag{A.104}$$

where W' is a standard p-dimensional Brownian motion defined on an extension of the original probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. We also show that the second term (the bias induced by the discretization error) and the third term (the bias induced by the spline approximation error) in Eq. (A.103) are both asymptotically negligible as $\Delta_n \to 0$.

For the fixed linear combination in Eq. (A.104), we consider a further decomposition by Eq. (A.35):

$$
\begin{aligned}
H_n = &\underbrace{\frac{1}{\sqrt{\Delta_n}} a^\top \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n \right) (\mathbf{R}^\top \mathbf{R})^{-1} \mathbf{R}^\top (\widetilde{\mathbf{Y}} - \mathbf{R}\boldsymbol{\gamma})}_{H_{1,n}} + \underbrace{\frac{1}{\sqrt{\Delta_n}} a^\top \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n \right) (\mathbf{R}^\top \mathbf{R})^{-1} \mathbf{R}^\top \mathbf{U}}_{H_{2,n}} \\
&+ \underbrace{\frac{1}{\sqrt{\Delta_n}} a^\top \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n \right) (\mathbf{R}^\top \mathbf{R})^{-1} \mathbf{R}^\top \mathbf{J}}_{H_{3,n}} + \frac{1}{\sqrt{\Delta_n}} a^\top \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n \right) (\mathbf{R}^\top \mathbf{R})^{-1} \mathbf{R}^\top \mathbf{Z},
\end{aligned}
\tag{A.105}
$$

where $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^\top$ in the fourth term includes the truncation $\mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}$ on each increment of Y rather than on increments of Z (see Eq. (A.33)). Therefore, we rewrite the fourth term into:

$$
\begin{aligned}
\frac{1}{\sqrt{\Delta_n}} a^\top \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n \right) (\mathbf{R}^\top \mathbf{R})^{-1} \mathbf{R}^\top \mathbf{Z} = &\underbrace{\frac{1}{\sqrt{\Delta_n}} a^\top \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n \right) (\mathbf{R}^\top \mathbf{R})^{-1} \mathbf{R}^\top \widetilde{\mathbf{Z}}}_{H_{4,n}} \\
&+ \underbrace{\frac{1}{\sqrt{\Delta_n}} a^\top \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n \right) (\mathbf{R}^\top \mathbf{R})^{-1} \mathbf{R}^\top (\mathbf{Z} - \widetilde{\mathbf{Z}})}_{H_{5,n}},
\end{aligned}
\tag{A.106}
$$

where $\widetilde{\mathbf{Z}} = (\widetilde{\mathbf{Z}}_1, \dots, \widetilde{\mathbf{Z}}_n)^\top$ with $\widetilde{\mathbf{Z}}_i = \Delta_i^n Z \mathbb{1}_{\{|\Delta_i^n Z| \le u_n\}}$. Next, we verify that the leading term for the stable central limit theorem (CLT) in Eq. (A.104) is $H_{4,n}$, and the other components $H_{1,n}$, $H_{2,n}$, $H_{3,n}$, and $H_{5,n}$ are asymptotically negligible.

(i) Asymptotic distribution of $\boldsymbol{H_{4,n}}$. We define the following vectors:

$$
\boldsymbol{w}_n = (w_{1,n}^{(1)}, \dots, w_{1,n}^{(K_n)}, \dots, w_{p,n}^{(1)}, \dots, w_{p,n}^{(K_n)})^\top = (\mathbf{R}^\top \mathbf{R})^{-1} \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n \right)^\top a \in \mathbb{R}^{pK_n}, \tag{A.107}
$$

$$
\theta_{n,t} = \mathbf{B}_t \boldsymbol{w}_n = \sum_{k=1}^{K_n} B_t^{(k)} \begin{pmatrix} w_{1,n}^{(k)} \\ \vdots \\ w_{p,n}^{(k)} \end{pmatrix} \in \mathbb{R}^p. \tag{A.108}
$$

By the LLN result in Eq. (A.30), a standard Riemann approximation for the deterministic integral, and the continuous mapping theorem, $\theta_{n,t} \xrightarrow{\mathbb{P}} \theta_t = \mathbf{B}_t \boldsymbol{w} \in \mathbb{R}^p$ for any $0 \le t \le T$, where

$$\boldsymbol{w} = (w_1^{(1)}, \ldots, w_p^{(K_n)})^\top = \left(\int_0^T \mathbf{B}_s^\top c_s \mathbf{B}_s ds \right)^{-1} \left(\int_0^T \mathbf{B}_s ds \right)^\top a \in \mathbb{R}^{pK_n}. \tag{A.109}$$

By definition,

$$H_{4,n} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \theta_{n,(i-1)\Delta_n}^\top \Delta_i^n X \Delta_i^n Z \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \mathbb{1}_{\{|\Delta_i^n Z| \le u_n\}}. \tag{A.110}$$

Under Assumption 3, X and Z have zero continuous covariation and no co-jumps. By the standard CLT for truncated realized covariances (Theorem 13.2.1, Jacod and Protter, 2012), and stable Slutsky's theorem, it holds that

$$H_{4,n} \xrightarrow{\mathcal{L}-s} \int_0^T \sqrt{\widetilde{\sigma}_s^2} \theta_s^\top c_s^{1/2} dW_s', \tag{A.111}$$

where W' is a standard p-dimensional Brownian motion defined on an extension of the original probability space. The conditional variance of the limit equals

$$\begin{aligned} \int_0^T \widetilde{\sigma}_s^2 \theta_s^\top c_s \theta_s ds = a^\top & \left(\int_0^T \mathbf{B}_s ds \right) \left(\int_0^T \mathbf{B}_s^\top c_s \mathbf{B}_s ds \right)^{-1} \left(\int_0^T \widetilde{\sigma}_s^2 \mathbf{B}_s^\top c_s \mathbf{B}_s ds \right) \\ & \times \left(\int_0^T \mathbf{B}_s^\top c_s \mathbf{B}_s ds \right)^{-1} \left(\int_0^T \mathbf{B}_s ds \right)^\top a \\ = a^\top \Sigma_T^\beta a. & \end{aligned} \tag{A.112}$$

(ii) $\boldsymbol{H_{1,n} = o_p(1)}$. Decompose $H_{1,n}$ into:

$$\begin{aligned} H_{1,n} &= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \boldsymbol{w}_n^\top \mathbf{R}_i \mathcal{Y}_i \\ &\le \underbrace{\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \boldsymbol{w}_n^\top \mathbf{R}_i M_i^{(1)}}_{H_{1,n}^{(1)}} - \underbrace{\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \boldsymbol{w}_n^\top \mathbf{R}_i M_i^{(2)}}_{H_{1,n}^{(2)}} + \underbrace{\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \boldsymbol{w}_n^\top \mathbf{R}_i \widetilde{\mathbf{Y}}_i^{(2)}}_{H_{1,n}^{(3)}}, \end{aligned} \tag{A.113}$$

where $\mathcal{Y}_i$ represents the i-th element of $\widetilde{\mathbf{Y}} - \mathbf{R}\boldsymbol{\gamma}$, and we employ the decomposition of $\mathcal{Y}_i$ in Eqs. (A.71) to (A.73):

$$\mathcal{Y}_i = M_i^{(1)} - M_i^{(2)} + \widetilde{\mathbf{Y}}_i^{(2)}, \tag{A.114}$$

Next, we verify that all three components on the right-hand side of Eq. (A.113) are $o_p(1)$.

(ii.1) $\boldsymbol{H_{1,n}^{(1)} = o_p(1)}$. It holds for $H_{1,n}^{(1)}$ that, by Hölder's inequality,

$$|H_{1,n}^{(1)}| = \frac{1}{\sqrt{\Delta_n}} \left| \sum_{i=1}^{n} \boldsymbol{w}_n^\top \mathbf{R}_i M_i^{(1)} \right| \leq \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} |\boldsymbol{w}_n^\top \mathbf{R}_i||M_i^{(1)}| \leq \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \|\boldsymbol{w}_n\|_\infty \|\mathbf{R}_i\|_1 |M_i^{(1)}|. \tag{A.115}$$

For any $i = 1, \ldots, n$, by the local support of B-splines (Appendix A.1 (iii)), each $\mathbf{R}_i$ has only a fixed number of nonzero components, and then the equivalence between ℓ_1- and ℓ_2-norms implies that

$$|H_{1,n}^{(1)}| \leq \frac{K}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \sum_{i=1}^{n} \|\mathbf{R}_i\| |M_i^{(1)}|. \tag{A.116}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} |H_{1,n}^{(1)}| &\leq \frac{K}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \sum_{i=1}^{n} \|\mathbf{B}_{(i-1)\Delta_n}\| \|\Delta_i^n X\| |M_i^{(1)}| \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \\ &\leq \frac{K}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \left(K' \max_{1 \leq k \leq K_n} B_{(i-1)\Delta_n}^{(k)} \right) \sum_{i=1}^{n} \|\Delta_i^n X\| |M_i^{(1)}| \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \\ &\leq \frac{K}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \sum_{i=1}^{n} \|\Delta_i^n X\| |M_i^{(1)}| \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \\ &\leq \frac{K u_n}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \sum_{i=1}^{n} |M_i^{(1)}|. \end{aligned} \tag{A.117}$$

Next, we show that $\|\boldsymbol{w}_n\|_\infty$ is bounded on an event E_n with $\mathbb{P}(E_n) \to 1$. Firstly, the local support of B-splines imply that

$$\begin{aligned} \left\| \left(\sum_{i=1}^{n} \mathbf{B}_{(i-1)\Delta_n} \Delta_n \right)^\top a \right\|_\infty &\leq K \|a\|_\infty \max_{1 \leq k \leq K_n} \sum_{i=1}^{n} B_{(i-1)\Delta_n}^{(k)} \Delta_n \\ &\leq K \|a\|_\infty \sup_{t \in [0,T]} B_t^{(k)} \left(\left\lceil \frac{|\mathrm{supp}(B^{(k)})|}{\Delta_n} \right\rceil + 1 \right) \Delta_n \leq \frac{K'}{K_n}, \end{aligned} \tag{A.118}$$

where $|\mathrm{supp}(B^{(k)})| \asymp K_n^{-1}$; see Appendix A.1 (iii). We define a permutation matrix $\mathbf{P}$ that reorders elements in $\mathbf{R}$ (group the p components associated with each spline basis index). Specifically, the

i-th column in the row-permutated $\mathbf{P}\mathbf{R}^\top \in \mathbb{R}^{pK_n \times n}$ is given by

$$\begin{pmatrix} B^{(1)}_{(i-1)\Delta_n} \Delta^n_{1,i} X \mathbb{1}_{\{\|\Delta^n_i X\| \le u_n\}} \\ \vdots \\ B^{(1)}_{(i-1)\Delta_n} \Delta^n_{p,i} X \mathbb{1}_{\{\|\Delta^n_i X\| \le u_n\}} \\ \vdots \\ \vdots \\ B^{(K_n)}_{(i-1)\Delta_n} \Delta^n_{1,i} X \mathbb{1}_{\{\|\Delta^n_i X\| \le u_n\}} \\ \vdots \\ B^{(K_n)}_{(i-1)\Delta_n} \Delta^n_{p,i} X \mathbb{1}_{\{\|\Delta^n_i X\| \le u_n\}} \end{pmatrix} \in \mathbb{R}^{pK_n}. \tag{A.119}$$

Then we consider a rescaled and permutated Gram matrix $\mathbf{G} = K_n \mathbf{P}(\mathbf{R}^\top \mathbf{R})\mathbf{P}^\top \in \mathbb{R}^{pK_n \times pK_n}$. By Eq. (A.32), the eigenvalues of $\mathbf{G}$ are bounded away from 0 and infinity with probability approaching one, i.e.,

$$\mathbb{P}(E_n) \to 1, \qquad \text{where } E_n = \{0 < \underline{\kappa} \le \lambda_{\min}(\mathbf{G}) \le \lambda_{\max}(\mathbf{G}) \le \overline{\kappa} < \infty\}. \tag{A.120}$$

Importantly, the local support property of B-spline basis implies that our rescaled matrix $\mathbf{G}$ is banded with fixed bandwidth, i.e., we have $\mathbf{G}_{ij} = 0$ when $|i-j| > L$ for some $L > 0$.[1] A standard result for symmetric positive definite banded matrices from Demko et al. (1984) is that the inverse has exponentially decaying off-diagonal entries. Specifically in our case, on E_n, $|(\mathbf{G}^{-1})_{ij}| \le K\rho^{|i-j|}$ for some $\rho \in (0,1)$, and therefore

$$\|\mathbf{G}^{-1}\|_\infty \le \max_i \sum_{j=1}^{pK_n} |(\mathbf{G}^{-1})_{ij}| \le K\left(1 + 2\sum_{m\ge 1} \rho^m\right) = K\left(1 + \frac{2\rho}{1-\rho}\right) \le \frac{K'}{1-\rho}. \tag{A.121}$$

Therefore, with

$$\mathbf{P}\boldsymbol{w}_n = \mathbf{P}(\mathbf{R}^\top\mathbf{R})^{-1}\left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n}\Delta_n\right)^\top a = (K_n\mathbf{P}(\mathbf{R}^\top\mathbf{R})\mathbf{P}^\top)^{-1}K_n\mathbf{P}\left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n}\Delta_n\right)^\top a, \tag{A.122}$$

it holds that on E_n, by Eqs. (A.118) and (A.121),

$$\|\boldsymbol{w}_n\|_\infty \le \|\mathbf{P}\boldsymbol{w}_n\|_\infty \le \|\mathbf{G}^{-1}\|_\infty \left\| K_n\mathbf{P}\left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n}\Delta_n\right)^\top a\right\|_\infty \le K. \tag{A.123}$$

[1] More strictly speaking, the matrix $\mathbf{G}$ is block-banded with block size p. When p is fixed, this block-bandedness automatically implies ordinary (scalar) bandedness by taking a half-bandwidth L large enough to cover the finitely many nonzero diagonal blocks.

By Eqs. (A.117) and (A.123), we have

$$\mathbb{E}[|H_{1,n}^{(1)}|^2 \mathbb{1}_{E_n}] \leq \frac{K u_n^2}{\Delta_n} \mathbb{E}\left[\|\boldsymbol{w}_n\|_\infty^2 \mathbb{1}_{E_n} \left(\sum_{i=1}^n |M_i^{(1)}| \right)^2 \right] \leq \frac{K' u_n^2}{\Delta_n} \mathbb{E}\left[\left(\sum_{i=1}^n |M_i^{(1)}| \right)^2 \right], \tag{A.124}$$

so it suffices to show

$$\mathbb{E}\left[\left(\sum_{i=1}^n |M_i^{(1)}| \right)^2 \right] = o\left(\frac{\Delta_n}{u_n^2} \right), \qquad \text{such that} \quad \mathbb{E}[|H_{1,n}^{(1)}|^2 \mathbb{1}_{E_n}] = o(1). \tag{A.125}$$

We define the subset of intervals,

$$\mathcal{I}_n = \{ i : |\Delta_i^n Y| > u_n \text{ or } \|\Delta_i^n X\| > u_n \}, \tag{A.126}$$

and apply the Cauchy-Schwarz inequality on that subset:

$$\left(\sum_{i=1}^n |M_i^{(1)}| \right)^2 = \left(\sum_{i \in \mathcal{I}_n} |M_i^{(1)}| \right)^2 \leq |\mathcal{I}_n| \sum_{i \in \mathcal{I}_n} (M_i^{(1)})^2. \tag{A.127}$$

Therefore, it holds that

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{i=1}^n |M_i^{(1)}| \right)^2 \right] &\leq \mathbb{E}\left[|\mathcal{I}_n| \sum_{i \in \mathcal{I}_n} (M_i^{(1)})^2 \right] \leq \mathbb{E}\left[|\mathcal{I}_n| \left(|\mathcal{I}_n| \sum_{i \in \mathcal{I}_n} (M_i^{(1)})^4 \right)^{1/2} \right] \\ &\leq \mathbb{E}\left[|\mathcal{I}_n|^{3/2} \left(\sum_{i=1}^n (M_i^{(1)})^4 \right)^{1/2} \right] \leq (\mathbb{E}[|\mathcal{I}_n|^3])^{1/2} \left(\sum_{i=1}^n \mathbb{E}[(M_i^{(1)})^4] \right)^{1/2} \\ &\leq K (u_n^{-3r})^{1/2} (\Delta_n^2 u_n^{-r})^{1/2} = K \Delta_n u_n^{-2r} = o(\Delta_n u_n^{-2}), \end{aligned} \tag{A.128}$$

where $\mathbb{E}[|\mathcal{I}_n|^3] = O(u_n^{-3r})$ can be proved in a similar way to Eq. (A.47), and $\mathbb{E}[(M_i^{(1)})^4] = \Delta_n^3 u_n^{-r}$ can be verified by the same steps as in Eq. (A.75) by the BDG inequality and Lemma A.2 (i).

Combining Eqs. (A.124) and (A.128) leads to

$$\mathbb{E}[|H_{1,n}^{(1)}|^2 \mathbb{1}_{E_n}] = o(1). \tag{A.129}$$

By Markov's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}(|H_{1,n}^{(1)}| > \varepsilon) \leq \mathbb{P}(|H_{1,n}^{(1)}| > \varepsilon, E_n) + \mathbb{P}(E_n^\complement) \leq \frac{\mathbb{E}[|H_{1,n}^{(1)}|^2 \mathbb{1}_{E_n}]}{\varepsilon^2} + \mathbb{P}(E_n^\complement) = o(1), \tag{A.130}$$

and therefore $H_{1,n}^{(1)} = o_p(1)$.

(ii.2) $\boldsymbol{H^{(2)}_{1,n} = o_p(1)}$. We follow the same steps as in (ii.1) from Eq. (A.115) to Eq. (A.125), and it remains to show

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} |M_i^{(2)}|\right)^2\right] = o\left(\frac{\Delta_n}{u_n^2}\right), \qquad \text{such that} \quad \mathbb{E}[|H^{(2)}_{1,n}|^2 \mathbb{1}_{E_n}] = o(1). \tag{A.131}$$

We consider

$$|M_i^{(2)}| \leq K \underbrace{\|\Delta_i^n X - \Delta_i^n X^c\| \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}}}_{D_i}, \tag{A.132}$$

so that it suffices to show

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} D_i\right)^2\right] = o\left(\frac{\Delta_n}{u_n^2}\right). \tag{A.133}$$

We consider the same event decomposition as in Eqs. (A.58) to (A.60):

$$\{\|\Delta_i^n X\| \leq u_n\} = \{\|\Delta_i^n X\| \leq u_n, \|\Delta_i^n X^c\| \leq u_n\} \cup \{\|\Delta_i^n X\| \leq u_n, \|\Delta_i^n X^c\| > u_n\}. \tag{A.134}$$

For each i, on the event $\{\|\Delta_i^n X\| \leq u_n, \|\Delta_i^n X^c\| \leq u_n\}$,

$$\|\Delta_i^n X - \Delta_i^n X^c\| \leq \|\Delta_i^n X\| + \|\Delta_i^n X^c\| \leq 2u_n. \tag{A.135}$$

Therefore,

$$D_i \leq \underbrace{\|\Delta_i^n X - \Delta_i^n X^c\| \mathbb{1}_{\{\|\Delta_i^n X - \Delta_i^n X^c\| \leq 2u_n\}}}_{D_i^{(1)}} + \underbrace{\|\Delta_i^n X - \Delta_i^n X^c\| \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \mathbb{1}_{\{\|\Delta_i^n X^c\| > u_n\}}}_{D_i^{(2)}}, \tag{A.136}$$

and also

$$\left(\sum_{i=1}^{n} D_i\right)^2 \leq 2\left(\sum_{i=1}^{n} D_i^{(1)}\right)^2 + 2\left(\sum_{i=1}^{n} D_i^{(2)}\right)^2. \tag{A.137}$$

For $D_i^{(1)}$, we follow the same steps in the proof of Lemma A.2 (ii) (see Eqs. (A.61) to (A.65)):

$$D_i^{(1)} \leq \|\Delta_i^n X - \Delta_i^n X^c\| \wedge 2u_n \leq \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)\mu(ds, dx). \tag{A.138}$$

Then we have

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} D_i^{(1)}\right)^2\right] \le \mathbb{E}\left[\left(\int_0^T \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)\mu(ds,dx)\right)^2\right]$$
$$\le 2\mathbb{E}\left[\left(\int_0^T \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)(\mu-\nu)(ds,dx)\right)^2\right] \tag{A.139}$$
$$+ 2\mathbb{E}\left[\left(\int_0^T \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)\nu(ds,dx)\right)^2\right].$$

For the first term in Eq. (A.139), by the Itô isometry for compensated Poisson integrals,

$$\mathbb{E}\left[\left(\int_0^T \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)(\mu-\nu)(ds,dx)\right)^2\right] = \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)^2\nu(ds,dx)\right]$$
$$\le \mathbb{E}\left[\int_0^T ds \left(\int_{\{\|\delta\|\le 2u_n\}} \|\delta(s,x)\|^2\lambda(dx) + 4u_n^2 \int_{\{\|\delta\|>2u_n\}} \lambda(dx)\right)\right] \le Ku_n^{2-r}, \tag{A.140}$$

by Eqs. (A.38) and (A.39). For the second term in Eq. (A.139), by Jensen's inequality,

$$\mathbb{E}\left[\left(\int_0^T \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)\nu(ds,dx)\right)^2\right] \le T\int_0^T \mathbb{E}\left[\left(\int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)\lambda(dx)\right)^2\right] ds$$
$$\le T\int_0^T \mathbb{E}\left[\left(\int_{\{\|\delta\|\le 2u_n\}} \|\delta(s,x)\|\lambda(dx) + 2u_n \int_{\{\|\delta\|>2u_n\}} \lambda(dx)\right)^2\right] ds \le Ku_n^{2-2r}, \tag{A.141}$$

by Eqs. (A.38) and (A.40). Combining Eqs. (A.139) to (A.141), we obtain

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} D_i^{(1)}\right)^2\right] \le Ku_n^{2-r} + K'u_n^{2-2r} \le Ku_n^{2-2r}. \tag{A.142}$$

For $D_i^{(2)}$, on the event $\{\|\Delta_i^n X\| \le u_n, \|\Delta_i^n X^c\| > u_n\}$,

$$\|\Delta_i^n X - \Delta_i^n X^c\| \le u_n + \|\Delta_i^n X^c\| \le 2\|\Delta_i^n X^c\|, \tag{A.143}$$

and thus it holds that

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} D_i^{(2)}\right)^2\right] \le \mathbb{E}\left[\left(2\sum_{i=1}^{n} \|\Delta_i^n X^c\| \mathbb{1}_{\{\|\Delta_i^n X^c\|>u_n\}}\right)^2\right] \le \mathbb{E}\left[4n\sum_{i=1}^{n} \|\Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X^c\|>u_n\}}\right]$$
$$= 4n\sum_{i=1}^{n} \mathbb{E}\left[\|\Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X^c\|>u_n\}}\right]. \tag{A.144}$$

For any $q > 2$,

$$\mathbb{E}\left[\|\Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X^c\| > u_n\}}\right] \leq \mathbb{E}\left[\|\Delta_i^n X^c\|^2 \left(\frac{\|\Delta_i^n X^c\|}{u_n}\right)^{q-2}\right] \leq \frac{\mathbb{E}[\|\Delta_i^n X^c\|^q]}{u_n^{q-2}} \leq \Delta_n^{q/2} u_n^{2-q}, \tag{A.145}$$

by the BDG inequality. Therefore, Eq. (A.144) turns into

$$\mathbb{E}\left[\left(\sum_{i=1}^n D_i^{(2)}\right)^2\right] \leq \Delta_n^{q/2-2} u_n^{2-q} = o\left(\frac{\Delta_n}{u_n^2}\right), \qquad \text{if } q > \frac{3-4\varpi}{1/2-\varpi}, \tag{A.146}$$

which is always possible because $\varpi < 1/2$. By Eqs. (A.137), (A.142) and (A.146),

$$\mathbb{E}\left[\left(\sum_{i=1}^n D_i\right)^2\right] \leq 2\mathbb{E}\left[\left(\sum_{i=1}^n D_i^{(1)}\right)^2\right] + 2\mathbb{E}\left[\left(\sum_{i=1}^n D_i^{(2)}\right)^2\right] = o\left(\frac{\Delta_n}{u_n^2}\right), \tag{A.147}$$

and therefore

$$\mathbb{E}\left[\left(\sum_{i=1}^n |M_i^{(2)}|\right)^2\right] = o\left(\frac{\Delta_n}{u_n^2}\right), \tag{A.148}$$

so that $\mathbb{E}[|H_{1,n}^{(1)}|^2 \mathbb{1}_{E_n}] = o(1)$. By the same inequality as in Eq. (A.130), we conclude $H_{1,n}^{(2)} = o_p(1)$.

(ii.3) $\boldsymbol{H_{1,n}^{(3)} = o_p(1)}$. Write $H_{1,n}^{(3)}$ into

$$H_{1,n}^{(3)} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \boldsymbol{w}_n^\top \mathbf{R}_i \left(\int_{(i-1)\Delta_n}^{i\Delta_n} (\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n})^\top dX_s^c\right) \mathbb{1}_{\{|\Delta_i^n Y| \leq u_n\}}. \tag{A.149}$$

By the Cauchy-Schwarz inequality,

$$|H_{1,n}^{(3)}| \leq \frac{1}{\sqrt{\Delta_n}} \left(\sum_{i=1}^n (\boldsymbol{w}_n^\top \mathbf{R}_i)^2\right)^{1/2} \left(\sum_{i=1}^n \left(\int_{(i-1)\Delta_n}^{i\Delta_n} (\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n})^\top dX_s^c\right)^2\right)^{1/2}. \tag{A.150}$$

For the first multiplier on the right-hand side of Eq. (A.150),

$$\begin{aligned} \sum_{i=1}^n (\boldsymbol{w}_n^\top \mathbf{R}_i)^2 &= \boldsymbol{w}_n^\top (\mathbf{R}^\top \mathbf{R}) \boldsymbol{w}_n = a^\top \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n\right) (\mathbf{R}^\top \mathbf{R})^{-1} \left(\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n\right)^\top a \\ &\leq K \|a\|^2 \underbrace{\left\|\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n} \Delta_n\right\|^2}_{O(K_n^{-1})} \|(\mathbf{R}^\top \mathbf{R})^{-1}\| = O_p(1), \end{aligned} \tag{A.151}$$

because each basis component $\sum_{i=1}^n B_{(i-1)\Delta_n}^{(k)}\Delta_n \le \sup_{t\in[0,T]} B_t^{(k)}\left(\left\lceil \frac{|\mathrm{supp}(B^{(k)})|}{\Delta_n}\right\rceil + 1\right)\Delta_n = O(K_n^{-1})$, and there are K_n of them in the Euclidean norm. For the other term, by Eq. (A.34),

$$\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} (\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n})^\top dX_s^c\right)^2 \middle| \mathcal{F}_{(i-1)\Delta_n}\right] \le K \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}[\|\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n}\|^2] ds. \quad \text{(A.152)}$$

By the Cauchy-Schwarz inequality,

$$\|\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n}\|^2 \le (s-(i-1)\Delta_n)\int_{(i-1)\Delta_n}^{s} \|\widetilde{\beta}'_u\|^2 du \le \Delta_n \int_{(i-1)\Delta_n}^{i\Delta_n} \|\widetilde{\beta}'_u\|^2 du. \quad \text{(A.153)}$$

Note that here we adopt a different approach from Eq. (A.79), and avoid using the probabilistic order bound for $\|\widetilde{\beta}'\|_\infty$. Then,

$$\int_{(i-1)\Delta_n}^{i\Delta_n} \|\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n}\|^2 ds \le \Delta_n^2 \int_{(i-1)\Delta_n}^{i\Delta_n} \|\widetilde{\beta}'_u\|^2 du. \quad \text{(A.154)}$$

Combining Eqs. (A.152) and (A.154) leads to

$$\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} (\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n})^\top dX_s^c\right)^2 \middle| \mathcal{F}_{(i-1)\Delta_n}\right] \le K\Delta_n^2 \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}[\|\widetilde{\beta}'_u\|^2] du, \quad \text{(A.155)}$$

and thus

$$\mathbb{E}\left[\sum_{i=1}^n \left(\int_{(i-1)\Delta_n}^{i\Delta_n} (\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n})^\top dX_s^c\right)^2\right] \le K\Delta_n^2 \int_0^T \mathbb{E}[\|\widetilde{\beta}'_u\|^2] du = K\Delta_n^2 \mathbb{E}[\|\widetilde{\beta}'\|_{L^2}^2]. \quad \text{(A.156)}$$

Therefore, by Markov's inequality,

$$\sum_{i=1}^n \left(\int_{(i-1)\Delta_n}^{i\Delta_n} (\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n})^\top dX_s^c\right)^2 = O_p(\Delta_n^2 \|\widetilde{\beta}'\|_{L^2}^2). \quad \text{(A.157)}$$

Combining Eqs. (A.151) and (A.157) in Eq. (A.150), we obtain

$$|H_{1,n}^{(3)}| = O_p(\sqrt{\Delta_n}\|\widetilde{\beta}'\|_{L^2}), \quad \text{(A.158)}$$

and it remains to derive the probabilistic order bound for $\|\widetilde{\beta}'\|_{L^2}$.

Lemma A.3. Under Assumption 2, $\|\widetilde{\beta}'\|_{L^2} = O_p(\sqrt{K_n \log K_n})$.

Proof of Lemma A.3. Instead of working directly on $\widetilde{\beta}'$, we consider the quasi-interpolant $\overline{\beta}$ defined in Eq. (A.11) in the proof of Lemma 1, and we have

$$\widetilde{\beta}' = \overline{\beta}' + (\widetilde{\beta} - \overline{\beta})'. \quad \text{(A.159)}$$

with all $\widetilde{\beta}$, $\overline{\beta}$, $\widetilde{\beta}-\overline{\beta} \in \mathbb{G}_n$. By the triangle inequality,

$$\|\widetilde{\beta}'\|_{L^2} \le \|\overline{\beta}'\|_{L^2} + \|(\widetilde{\beta}-\overline{\beta})'\|_{L^2}. \tag{A.160}$$

We start with a general linear combination of B-spline basis functions of degree d: Let $g_t = \sum_{k=1}^{K_n} B_{d,t}^{(k)} r^{(k)} \in \mathbb{G}_n$ for any $r^{(k)} \in \mathbb{R}^p$. By the Cox-de Boor recursion identity in Eq. (A.3),

$$(B_{d,t}^{(k)})' = \frac{d}{\bar{v}_{k+d-1}-\bar{v}_{k-1}} B_{d-1,t}^{(k)} - \frac{d}{\bar{v}_{k+d}-\bar{v}_k} B_{d-1,t}^{(k+1)}. \tag{A.161}$$

Differentiate g_t on t with Eq. (A.161) plugged in, we obtain

$$\begin{aligned}
g_t' &= \sum_{k=1}^{K_n} \frac{d}{\bar{v}_{k+d-1}-\bar{v}_{k-1}} B_{d-1,t}^{(k)} r^{(k)} - \sum_{k=1}^{K_n} \frac{d}{\bar{v}_{k+d}-\bar{v}_k} B_{d-1,t}^{(k+1)} r^{(k)} \\
&= \sum_{k=1}^{K_n} \frac{d}{\bar{v}_{k+d-1}-\bar{v}_{k-1}} B_{d-1,t}^{(k)} r^{(k)} - \sum_{k=2}^{K_n+1} \frac{d}{\bar{v}_{k+d-1}-\bar{v}_{k-1}} B_{d-1,t}^{(k)} r^{(k)} \\
&= \frac{d}{\bar{v}_d-\bar{v}_0} r^{(1)} + \sum_{k=2}^{K_n} \left(\frac{d}{\bar{v}_{k+d-1}-\bar{v}_{k-1}} r^{(k)} - \frac{d}{\bar{v}_{k+d-1}-\bar{v}_{k-1}} r^{(k-1)} \right) B_{d-1,t}^{(k)} - \frac{d}{\bar{v}_{K_n+d}-\bar{v}_{K_n}} r^{(K_n)},
\end{aligned} \tag{A.162}$$

where the first and last terms are both 0 by definition; see Eq. (A.4). Therefore,

$$g_t' = \sum_{k=2}^{K_n} \frac{dB_{d-1,t}^{(k)}}{\bar{v}_{k+d-1}-\bar{v}_{k-1}} (r^{(k)} - r^{(k-1)}) \le K \sum_{k=1}^{K_n-1} (r^{(k+1)} - r^{(k)}), \tag{A.163}$$

and thus

$$\|g'\|_{L^2}^2 = \int_0^T \|g_t'\|^2 dt \le KK_n \sum_{k=1}^{K_n-1} \|r^{(k+1)} - r^{(k)}\|^2 \le K'K_n \sum_{k=1}^{K_n} \|r^{(k)}\|^2 = K'K_n \|\boldsymbol{r}\|^2, \tag{A.164}$$

where $\boldsymbol{r} = ((r^{(1)})^\top, \ldots, (r^{(K_n)})^\top)^\top \in \mathbb{R}^{pK_n}$. On the other hand, by the B-spline norm equivalence in Appendix A.1 (iv),

$$\|g\|_{L^2}^2 \asymp \frac{\|\boldsymbol{r}\|^2}{K_n}. \tag{A.165}$$

Combining Eqs. (A.164) and (A.165) leads to

$$\|g'\|_{L^2} \le KK_n \|g\|_{L^2}. \tag{A.166}$$

Therefore, if $g = \widetilde{\beta} - \overline{\beta}$,

$$\|(\widetilde{\beta}-\overline{\beta})'\|_{L^2} \le KK_n \|\widetilde{\beta}-\overline{\beta}\|_{L^2}. \tag{A.167}$$

By the triangle inequality, Eqs. (A.9) and (A.10),

$$\|\widetilde{\beta}-\overline{\beta}\|_{L^2} \leq \|\beta-\overline{\beta}\|_{L^2}+\|\beta-\widetilde{\beta}\|_{L^2} \leq K\|\beta-\overline{\beta}\|_{L^2} = O_p\left(\sqrt{\frac{\log K_n}{K_n}}\right). \tag{A.168}$$

Therefore, by the above two inequalities,

$$\|(\widetilde{\beta}-\overline{\beta})'\|_{L^2} = O_p(\sqrt{K_n \log K_n}). \tag{A.169}$$

Then it remains to derive the probabilistic order bound for $\|\overline{\beta}'\|_{L^2}$.

Let $g=\overline{\beta}$, by the first inequality in Eq. (A.164),

$$\|\overline{\beta}'\|_{L^2}^2 \leq KK_n \sum_{k=1}^{K_n-1} \|\beta_{\tau^{(k+1)}}-\beta_{\tau^{(k)}}\|^2. \tag{A.170}$$

Consider the same decomposition as in Eq. (A.12). Then we have

$$\sum_{k=1}^{K_n-1} \|\beta_{\tau^{(k+1)}}-\beta_{\tau^{(k)}}\|^2 \leq 2\sum_{k=1}^{K_n-1} \|\beta^c_{\tau^{(k+1)}}-\beta^c_{\tau^{(k)}}\|^2 + 2\sum_{k=1}^{K_n-1} \|\beta^j_{\tau^{(k+1)}}-\beta^j_{\tau^{(k)}}\|^2. \tag{A.171}$$

Under Assumption A.1, the BDG inequality implies that

$$\mathbb{E}\left[\sum_{k=1}^{K_n-1} \|\beta^c_{\tau^{(k+1)}}-\beta^c_{\tau^{(k)}}\|^2\right] \leq K\sum_{k=1}^{K_n-1} (\tau^{(k+1)}-\tau^{(k)}) \leq KT, \tag{A.172}$$

and thus, by Markov's inequality,

$$\sum_{k=1}^{K_n-1} \|\beta^c_{\tau^{(k+1)}}-\beta^c_{\tau^{(k)}}\|^2 = O_p(1). \tag{A.173}$$

Under Assumption 2 with finite-variation jumps,

$$\sum_{k=1}^{K_n-1} \|\beta^j_{\tau^{(k+1)}}-\beta^j_{\tau^{(k)}}\|^2 \leq \left(\sum_{k=1}^{K_n-1} \|\beta^j_{\tau^{(k+1)}}-\beta^j_{\tau^{(k)}}\|\right)^2 \leq V_{\beta^j}([0,T])^2 = O_p(1), \tag{A.174}$$

where $V_{\beta^j}([a,b])$ is defined in Eq. (A.22). Combining Eqs. (A.173) and (A.174) in Eq. (A.171), we obtain

$$\sum_{k=1}^{K_n-1} \|\beta_{\tau^{(k+1)}}-\beta_{\tau^{(k)}}\|^2 = O_p(1), \tag{A.175}$$

and thus, by Eq. (A.170),

$$\|\overline{\beta}'\|_{L^2} = O_p(\sqrt{K_n}). \tag{A.176}$$

Therefore, combining Eqs. (A.169) and (A.176) in Eq. (A.160), we obtain

$$\|\widetilde{\beta}'\|_{L^2} \le O_p(\sqrt{K_n}) + O_p(\sqrt{K_n \log K_n}) = O_p(\sqrt{K_n \log K_n}). \tag{A.177}$$

This completes the proof of Lemma A.3.

□

Therefore, by Eq. (A.158) and Lemma A.3,

$$|H_{1,n}^{(3)}| = O_p(\sqrt{\Delta_n K_n \log K_n}) = o_p(1), \tag{A.178}$$

since $\Delta_n K_n \log K_n \to 0$ as required, and thus $H_{1,n}^{(3)} = o_p(1)$.

By the results in (ii.1), (ii.2), and (ii.3), we conclude that $H_{1,n} = o_p(1)$.

(iii) $\boldsymbol{H_{2,n} = o_p(1)}$. Recall that, with $\theta_{n,t}$ defined in Eq. (A.108),

$$H_{2,n} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \theta_{n,(i-1)\Delta_n}^{\top} \Delta_i^n X \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^{\top} dX_s^c \right) \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}. \tag{A.179}$$

On the high-probability event E_n in Eq. (A.120), it holds that $\|\boldsymbol{w}_n\|_\infty \le K$ by Eq. (A.123). By local support and boundedness of B-splines, on E_n,

$$\sup_{t\in[0,T]} \|\theta_{n,t}\| = \sup_{t\in[0,T]} \left\| \sum_{k=1}^{K_n} B_t^{(k)} \boldsymbol{w}_n \right\| \le K \|\boldsymbol{w}_n\|_\infty \le K', \tag{A.180}$$

and $\|\theta_n'\|_\infty = O_p(K_n)$ by Eq. (A.79). Consider $H_{2,n} = H_{2,n}^{(1)} + H_{2,n}^{(2)}$ with

$$H_{2,n}^{(1)} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \theta_{n,(i-1)\Delta_n}^{\top} \Delta_i^n X^c \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^{\top} dX_s^c \right) \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}, \tag{A.181}$$

$$H_{2,n}^{(2)} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \theta_{n,(i-1)\Delta_n}^{\top} (\Delta_i^n X - \Delta_i^n X^c) \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^{\top} dX_s^c \right) \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}. \tag{A.182}$$

Next, we show both $H_{2,n}^{(1)}$ and $H_{2,n}^{(2)}$ are $o_p(1)$.

(iii.1) $\boldsymbol{H_{2,n}^{(1)} = o_p(1)}$. Instead of working on $H_{2,n}^{(1)}$ directly, we consider $H_{2,n}^{(1)} = \widetilde{H}_{2,n}^{(1)} - \overline{H}_{2,n}^{(1)}$ with

$$\widetilde{H}_{2,n}^{(1)} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \theta_{n,(i-1)\Delta_n}^{\top} \Delta_i^n X^c \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^{\top} dX_s^c \right), \tag{A.183}$$

$$\overline{H}_{2,n}^{(1)} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \theta_{n,(i-1)\Delta_n}^{\top} \Delta_i^n X^c \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^{\top} dX_s^c \right) \left(1 - \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}\right), \tag{A.184}$$

and show that both $\widetilde{H}^{(1)}_{2,n}$ and $\overline{H}^{(1)}_{2,n}$ are $o_p(1)$.

We start with $\widetilde{H}^{(1)}_{2,n}$. Rewrite Eq. (A.183) into $\widetilde{H}^{(1)}_{2,n} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n U_i$ with

$$U_i = \theta^\top_{n,(i-1)\Delta_n} \Delta_i^n X^c \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right). \tag{A.185}$$

By the conditional covariance for continuous martingale increments (with the predictability justified in Remark A.1 via predictable projection),

$$\mathbb{E}[U_i | \mathcal{F}_{(i-1)\Delta_n}] = \theta^\top_{n,(i-1)\Delta_n} \int_{(i-1)\Delta_n}^{i\Delta_n} c_s e_s ds. \tag{A.186}$$

Consider

$$\widetilde{H}^{(1)}_{2,n} = \underbrace{\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \mathbb{E}[U_i | \mathcal{F}_{(i-1)\Delta_n}]}_{A^{(1)}_{2,n}} + \underbrace{\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n (U_i - \mathbb{E}[U_i | \mathcal{F}_{(i-1)\Delta_n}])}_{M^{(1)}_{2,n}}. \tag{A.187}$$

For $A^{(1)}_{2,n}$, we have

$$A^{(1)}_{2,n} = \frac{1}{\sqrt{\Delta_n}} \int_0^T \theta^\top_{n,s} c_s e_s ds - \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} (\theta_{n,s} - \theta_{n,(i-1)\Delta_n})^\top c_s e_s ds, \tag{A.188}$$

where the first term is 0 by $L^2(c)$-orthogonality between e and $\mathbb{G}_n$. For the second term, on E_n,

$$\sup_{(i-1)\Delta_n \le s \le i\Delta_n} \|\theta_{n,s} - \theta_{n,(i-1)\Delta_n}\| \le \Delta_n \|\theta'_n\|_\infty. \tag{A.189}$$

Therefore, by the Cauchy-Schwarz inequality and bounded c under Assumption A.1,

$$\begin{aligned} \mathbb{E}[|A^{(1)}_{2,n}| \mathbb{1}_{E_n}] &\le \frac{K}{\sqrt{\Delta_n}} \mathbb{E}\left[\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \|\theta_{n,s} - \theta_{n,(i-1)\Delta_n}\| \|e_s\| ds \right] \\ &\le K\sqrt{\Delta_n} \mathbb{E}[\|\theta'_n\|_\infty \|e\|_{L^2}] = O(\sqrt{\Delta_n K_n \log K_n}) = o(1), \end{aligned} \tag{A.190}$$

and thus, by Markov's inequality analogous to Eq. (A.130) (but with first-order moment), we obtain

$A_{2,n}^{(1)} = o_p(1)$. Moreover, for $M_{2,n}^{(1)}$, by the same steps as in Eq. (A.75), together with Eq. (A.34),

$$\begin{aligned}
\mathbb{E}[(M_{2,n}^{(1)})^2 \mathbb{1}_{E_n}] &\leq \frac{K}{\Delta_n} \sum_{i=1}^{n} \mathbb{E}[U_i^2] \leq \frac{K}{\Delta_n} \sum_{i=1}^{n} \mathbb{E}\left[\|\Delta_i^n X^c\|^2 \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right)^2 \right] \\
&\leq \frac{K}{\Delta_n} \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}[\|\Delta_i^n X^c\|^2 | \mathcal{F}_{(i-1)\Delta_n}] \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right)^2 \right] \\
&\leq K\mathbb{E}[\|e\|_{L^2}^2] = o(1),
\end{aligned} \tag{A.191}$$

and thus $M_{2,n}^{(1)} = o_p(1)$ by the same Markov's inequality as in Eq. (A.130). Therefore, we conclude that $\widetilde{H}_{2,n}^{(1)} = A_{2,n}^{(1)} + M_{2,n}^{(1)} = o_p(1)$.

For $\overline{H}_{2,n}^{(1)}$, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbb{E}\left[(\overline{H}_{2,n}^{(1)})^2 \mathbb{1}_{E_n} \right] \leq \frac{K}{\Delta_n} &\left(\sum_{i=1}^{n} \mathbb{E}\left[\|\Delta_i^n X^c\|^2 \left(1 - \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \mathbb{1}_{\{|\Delta_i^n Y| \leq u_n\}} \right) \right] \right) \\
\times &\left(\sum_{i=1}^{n} \mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right)^2 \left(1 - \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \mathbb{1}_{\{|\Delta_i^n Y| \leq u_n\}} \right) \right] \right).
\end{aligned} \tag{A.192}$$

By the same steps as in Eq. (A.75), we obtain

$$\sum_{i=1}^{n} \mathbb{E}\left[\|\Delta_i^n X^c\|^2 \left(1 - \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \mathbb{1}_{\{|\Delta_i^n Y| \leq u_n\}} \right) \right] \leq K\Delta_n u_n^{-r}. \tag{A.193}$$

By Eq. (A.34),

$$\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right)^2 \middle| \mathcal{F}_{(i-1)\Delta_n} \right] \leq K \int_{(i-1)\Delta_n}^{i\Delta_n} \|e_s\|^2 ds, \tag{A.194}$$

and thus, by Lemma A.2 (i),

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right)^2 \left(1 - \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \mathbb{1}_{\{|\Delta_i^n Y| \leq u_n\}} \right) \right] \\
&\quad \leq K\mathbb{E}\left[\sum_{i=1}^{n} \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \|e_s\|^2 ds \right) \left(1 - \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}} \mathbb{1}_{\{|\Delta_i^n Y| \leq u_n\}} \right) \right] \\
&\quad \leq K\mathbb{E}[\|e\|_{L^2}^2] \max_{1 \leq i \leq n} \left(\mathbb{P}(|\Delta_i^n Y| > u_n) + \mathbb{P}(\|\Delta_i^n X\| > u_n) \right) \\
&\quad \leq K' \Delta_n u_n^{-r} \mathbb{E}[\|e\|_{L^2}^2].
\end{aligned} \tag{A.195}$$

Combining Eqs. (A.193) and (A.195) in Eq. (A.192) leads to

$$\mathbb{E}\left[(\overline{H}_{2,n}^{(1)})^2 \mathbb{1}_{E_n} \right] \leq K\Delta_n u_n^{-2r} \mathbb{E}[\|e\|_{L^2}^2] = o(1). \tag{A.196}$$

By the same Markov's inequality as in Eq. (A.130), we conclude $\overline{H}^{(1)}_{2,n} = o_p(1)$, and therefore

$$H^{(1)}_{2,n} = \widetilde{H}^{(1)}_{2,n} - \overline{H}^{(1)}_{2,n} = o_p(1). \tag{A.197}$$

(iii.2) $\boldsymbol{H^{(2)}_{2,n} = o_p(1)}$. On E_n, by the Cauchy-Schwarz inequality,

$$\mathbb{E}\left[(H^{(2)}_{2,n})^2 \mathbb{1}_{E_n}\right] \le \frac{K}{\Delta_n} \left(\sum_{i=1}^{n} \mathbb{E}\left[\|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \right] \right) \left(\sum_{i=1}^{n} \mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right)^2 \right] \right). \tag{A.198}$$

By Lemma A.2 (ii),

$$\sum_{i=1}^{n} \mathbb{E}\left[\|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \right] \le K u_n^{2-r}, \tag{A.199}$$

and, by Eq. (A.34),

$$\sum_{i=1}^{n} \mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right)^2 \right] \le K \mathbb{E}[\|e\|^2_{L^2}]. \tag{A.200}$$

Combining Eqs. (A.199) and (A.200) in Eq. (A.198), and by Lemma 1,

$$\mathbb{E}\left[(H^{(2)}_{2,n})^2 \mathbb{1}_{E_n}\right] \le \frac{K}{\Delta_n} u_n^{2-r} \mathbb{E}[\|e\|^2_{L^2}] = o(1), \tag{A.201}$$

which holds for all $u_n \asymp \Delta_n^{\varpi}$ with $\varpi \in \left(\frac{1}{2(2-r)}, \frac{1}{2} \right)$ because

$$\frac{K}{\Delta_n} u_n^{2-r} \mathbb{E}[\|e\|^2_{L^2}] \asymp \Delta_n^{\varpi(2-r)-1} \frac{\log K_n}{K_n} = \Delta_n^{\varpi(2-r)-1/2} \frac{\log K_n}{K_n \sqrt{\Delta_n}} = o(1), \tag{A.202}$$

where $\log K_n / (K_n \sqrt{\Delta_n}) \to 0$ as required, and $\Delta_n^{\varpi(2-r)-1/2} \to 0$ because $\varpi(2-r) - 1/2 > 0$. Therefore, by the same Markov's inequality as in Eq. (A.130), $H^{(2)}_{2,n} = o_p(1)$.

By the results in (iii.1) and (iii.2), we conclude that $H_{2,n} = o_p(1)$.

(iv) $\boldsymbol{H_{3,n} = o_p(1)}$. By the same steps as in Eq. (A.117) and the safe substitution in Eq. (A.82),

$$|H_{3,n}| \le \frac{K u_n}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \sum_{i=1}^{n} |\mathbf{J}_i| \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} = \frac{K u_n}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \sum_{i=1}^{n} |\widetilde{\mathbf{J}}_i|, \tag{A.203}$$

and

$$\mathbb{E}[|H_{3,n}| \mathbb{1}_{E_n}] \le \frac{K u_n}{\sqrt{\Delta_n}} \mathbb{E}\left[\sum_{i=1}^{n} |\widetilde{\mathbf{J}}_i| \right], \tag{A.204}$$

so it suffices to show

$$\mathbb{E}\left[\sum_{i=1}^{n} |\widetilde{\mathbf{J}}_i| \right] = O(u_n^{1-r}) = o\left(\frac{\sqrt{\Delta_n}}{u_n} \right), \qquad \text{such that} \quad \mathbb{E}[|H_{3,n}| \mathbb{1}_{E_n}] = o(1). \tag{A.205}$$

Same as in Eq. (A.84), we have

$$
\begin{aligned}
\sum_{i=1}^{n} |\widetilde{\mathbf{J}}_i| &\le K \sum_{i=1}^{n} \Bigg(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n) \mu(ds,dx) \\
&\qquad + \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\{\|\delta\|>2u_n\}} \|\delta(s,x)\| \mu(ds,dx) \Bigg) \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \\
&= K \int_0^T \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n) \mu(ds,dx) \\
&\qquad + K \sum_{i=1}^{n} \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\{\|\delta\|>2u_n\}} \|\delta(s,x)\| \mu(ds,dx) \right) \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}.
\end{aligned} \tag{A.206}
$$

Following the same steps as in Eq. (A.64),

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^T \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n) \mu(ds,dx) \right] \le K \mathbb{E}\left[\int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n) \lambda(dx) \right] \\
&\quad \le K \mathbb{E}\left[\int_{\{\|\delta\| \le 2u_n\}} \|\delta(s,x)\| \lambda(dx) + 2u_n \int_{\{\|\delta\|>2u_n\}} \lambda(dx) \right] \le K' u_n^{1-r},
\end{aligned} \tag{A.207}
$$

by Eqs. (A.38) and (A.40). For the other term, by Eqs. (A.87), (A.90) and (A.92) to (A.94),

$$
\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\{\|\delta\|>2u_n\}} \|\delta(s,x)\| \mu(ds,dx) \right) \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \right] \\
&\quad \le K \sum_{i=1}^{n} \mathbb{E}\left[N_i^n(2u_n) \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \right] \le K \sum_{i=1}^{n} \mathbb{E}\left[(N_i^n(2u_n))^2 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \right] \\
&\quad \le K \Delta_n u_n^{-2r} = o(u_n^{1-r}).
\end{aligned} \tag{A.208}
$$

Combining Eqs. (A.207) and (A.208) in Eq. (A.206) leads to $\mathbb{E}[\sum_{i=1}^n |\widetilde{\mathbf{J}}_i|] = o(\sqrt{\Delta_n} u_n^{-1})$, and thus $\mathbb{E}[|H_{3,n}| \mathbb{1}_{E_n}] = o(1)$. Therefore, by the same Markov's inequality as in Eq. (A.130), $H_{3,n} = o_p(1)$.

(v) $\boldsymbol{H_{5,n} = o_p(1)}$. By the same steps as in Eq. (A.117),

$$
\begin{aligned}
|H_{5,n}| &\le \frac{K u_n}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \sum_{i=1}^{n} |\Delta_i^n Z| \left| \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} - \mathbb{1}_{\{|\Delta_i^n Z| \le u_n\}} \right| \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \\
&\le \frac{K u_n}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \sum_{i=1}^{n} |\Delta_i^n Z| \left(\mathbb{1}_{\{|\Delta_i^n Y| > u_n, |\Delta_i^n Z| \le u_n\}} + \mathbb{1}_{\{|\Delta_i^n Y| \le u_n, |\Delta_i^n Z| > u_n\}} \right) \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \\
&\le \frac{K u_n}{\sqrt{\Delta_n}} \|\boldsymbol{w}_n\|_\infty \sum_{i=1}^{n} \left(u_n \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}} + |\Delta_i^n Z| \mathbb{1}_{\{|\Delta_i^n Y| \le u_n, |\Delta_i^n Z| > u_n, \|\Delta_i^n X\| \le u_n\}} \right).
\end{aligned} \tag{A.209}
$$

Note that the additional truncation $\mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}$ is harmlessly included because the design matrix $\mathbf{R}$ already includes that for covariate increments. Then,

$$\mathbb{E}[|H_{5,n}|\mathbb{1}_{E_n}] \le \frac{Ku_n^2}{\sqrt{\Delta_n}} \sum_{i=1}^n \mathbb{P}(|\Delta_i^n Y| > u_n) + \frac{K' u_n}{\sqrt{\Delta_n}} \sum_{i=1}^n \mathbb{E}\left[|\Delta_i^n Z| \mathbb{1}_{\{|\Delta_i^n Y| \le u_n,\, |\Delta_i^n Z| > u_n,\, \|\Delta_i^n X\| \le u_n\}}\right]. \tag{A.210}$$

For the first term, by Lemma A.2 (i),

$$\sum_{i=1}^n \mathbb{P}(|\Delta_i^n Y| > u_n) = O(u_n^{-r}), \quad \text{and thus} \quad \frac{Ku_n^2}{\sqrt{\Delta_n}} \sum_{i=1}^n \mathbb{P}(|\Delta_i^n Y| > u_n) = O\left(\frac{u_n^{2-r}}{\sqrt{\Delta_n}}\right) = o(1), \tag{A.211}$$

because $\varpi(2-r) - 1/2 > 0$.

Now we turn to the second term in Eq. (A.210). Intuitively, on the event $\{|\Delta_i^n Y| \le u_n,\, |\Delta_i^n Z| > u_n,\, \|\Delta_i^n X\| \le u_n\}$, the cancellation between $\Delta_i^n Y$, $\Delta_i^n X$ and $\Delta_i^n Z$ imposes a restriction on the size of $\Delta_i^n Z$. To formally establish this, we write $\Delta_i^n Y$ into

$$\Delta_i^n Y = \underbrace{\int_{(i-1)\Delta_n}^{i\Delta_n} \beta_{s-}^\top dX_s^c}_{\Delta_i^n Y^{(1)}} + \underbrace{\sum_{(i-1)\Delta_n \le s \le i\Delta_n} (\beta_{s-}^J)^\top \Delta X_s}_{\Delta_i^n Y^{(2)}} + \Delta_i^n Z. \tag{A.212}$$

On $\{|\Delta_i^n Y| \le u_n,\, |\Delta_i^n Z| > u_n,\, \|\Delta_i^n X\| \le u_n\}$, it holds that

$$\begin{aligned} u_n < |\Delta_i^n Z| &= |\Delta_i^n Y - \Delta_i^n Y^{(1)} - \Delta_i^n Y^{(2)}| \\ &\le |\Delta_i^n Y| + |\Delta_i^n Y^{(1)}| + |\Delta_i^n Y^{(2)}| \\ &= |\Delta_i^n Y| + |\Delta_i^n Y^{(1)}| + |\Delta_i^n Y^{(2)} \mathbb{1}_{\{|\Delta_i^n Y| \le u_n,\, \|\Delta_i^n X\| \le u_n\}}| \\ &= |\Delta_i^n Y| + |\Delta_i^n Y^{(1)}| + |\widetilde{\mathbf{J}}_i| \\ &\le u_n + |\Delta_i^n Y^{(1)}| + |\widetilde{\mathbf{J}}_i|, \end{aligned} \tag{A.213}$$

where $\widetilde{\mathbf{J}}_i$ is defined in Eq. (A.82). Based on the above result, we consider the event decomposition:

$$\begin{aligned} &\{|\Delta_i^n Y| \le u_n,\, |\Delta_i^n Z| > u_n,\, \|\Delta_i^n X\| \le u_n\} \\ &\quad \subseteq \{u_n < |\Delta_i^n Z| \le 3u_n\} \cup \{|\Delta_i^n Y^{(1)}| + |\widetilde{\mathbf{J}}_i| > 2u_n\}. \end{aligned} \tag{A.214}$$

Then, we have

$$\begin{aligned}
&\mathbb{E}\left[|\Delta_i^n Z|\mathbb{1}_{\{|\Delta_i^n Y|\le u_n,\, |\Delta_i^n Z|>u_n,\, \|\Delta_i^n X\|\le u_n\}}\right] \\
&\quad\le \mathbb{E}\left[|\Delta_i^n Z|\mathbb{1}_{\{u_n<|\Delta_i^n Z|\le 3u_n\}}\right] + \mathbb{E}\left[|\Delta_i^n Z|\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|+|\widetilde{\mathbf{J}}_i|>2u_n\}}\right] \\
&\quad\le \underbrace{3u_n\mathbb{P}(|\Delta_i^n Z|>u_n)}_{(1)} + \mathbb{E}\left[|\Delta_i^n Z|\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right] + \mathbb{E}\left[|\Delta_i^n Z|\mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right] \\
&\quad\le (1) + \underbrace{u_n\mathbb{P}\left(|\Delta_i^n Y^{(1)}|>u_n\right)}_{(2)} + \underbrace{\mathbb{E}\left[|\Delta_i^n Y^{(1)}|\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right]}_{(3)} + \underbrace{\mathbb{E}\left[|\widetilde{\mathbf{J}}_i|\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right]}_{(4)} \\
&\qquad + \underbrace{u_n\mathbb{P}\left(|\widetilde{\mathbf{J}}_i|>u_n\right)}_{(5)} + \underbrace{\mathbb{E}\left[|\Delta_i^n Y^{(1)}|\mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right]}_{(6)} + \underbrace{\mathbb{E}\left[|\widetilde{\mathbf{J}}_i|\mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right]}_{(7)}.
\end{aligned} \tag{A.215}$$

Next, we verify that the above terms (1)–(7) are all $O(\Delta_n u_n^{1-r})$.

Term (1): By the same result as in Lemma A.2 (i) for the one-dimensional Itô semimartingale Z,

$$3u_n\mathbb{P}(|\Delta_i^n Z|>u_n) \le K\Delta_n u_n^{1-r}. \tag{A.216}$$

Term (2): By Eq. (A.54),

$$u_n\mathbb{P}\left(|\Delta_i^n Y^{(1)}|>u_n\right) \le K\Delta_n u_n^{1-r}. \tag{A.217}$$

Term (3): By the same steps as in Eq. (A.67) and the BDG inequality, for any $q\ge 2$,

$$\mathbb{E}\left[|\Delta_i^n Y^{(1)}|\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right] \le \mathbb{E}\left[|\Delta_i^n Y^{(1)}|\left(\frac{|\Delta_i^n Y^{(1)}|}{u_n}\right)^{q-1}\right] \le \frac{\mathbb{E}[|\Delta_i^n Y^{(1)}|^q]}{u_n^{q-1}} \le K\Delta_n^{q/2}u_n^{1-q}, \tag{A.218}$$

and it is $O(\Delta_n u_n^{1-r})$ if $q\ge 2(1-\varpi r)/(1-2\varpi)$, which is always possible for $1/2(2-r)<\varpi<1/2$ and $0\le r<1$.

Term (4): By the Cauchy-Schwarz inequality, Eqs. (A.54) and (A.96), for any $q\ge 2$,

$$\begin{aligned}
\mathbb{E}\left[|\widetilde{\mathbf{J}}_i|\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right] &\le \left(\mathbb{E}[|\widetilde{\mathbf{J}}_i|^2]\right)^{1/2}\left(\mathbb{P}\left(|\Delta_i^n Y^{(1)}|>u_n\right)\right)^{1/2} \\
&\le K(\Delta_n u_n^{2-r}\vee \Delta_n^2 u_n^{-2r})^{1/2}(\Delta_n^{q/2}u_n^{-q})^{1/2} \\
&\le K\Delta_n^{1/2+q/4}u_n^{1-r/2-q/2}\vee K\Delta_n^{1+q/4}u_n^{-r-q/2} \\
&= O(\Delta_n u_n^{1-r}),
\end{aligned} \tag{A.219}$$

if we choose some

$$q \ge \frac{2(1-\varpi r)}{1-2\varpi}\bigvee\frac{4\varpi}{1-2\varpi}, \tag{A.220}$$

which is always possible for $1/2(2-r)<\varpi<1/2$ and $0\le r<1$.

Term (5): By Markov's inequality,

$$u_n \mathbb{P}\left(|\widetilde{\mathbf{J}}_i| > u_n\right) \leq u_n \frac{\mathbb{E}[|\widetilde{\mathbf{J}}_i|]}{u_n} \leq \mathbb{E}[|\widetilde{\mathbf{J}}_i|] \leq K\Delta_n u_n^{1-r}, \tag{A.221}$$

which can be calculated following the same steps as in Eqs. (A.207) and (A.208).

Term (6): By Hölder's inequality, for some $q \geq 2$,

$$\begin{aligned} \mathbb{E}\left[|\Delta_i^n Y^{(1)}| \mathbb{1}_{\{|\widetilde{\mathbf{J}}_i| > u_n\}}\right] &\leq \left(\mathbb{E}[|\Delta_i^n Y^{(1)}|^q]\right)^{1/q} \left(\mathbb{P}\left(|\widetilde{\mathbf{J}}_i| > u_n\right)\right)^{1-1/q} \\ &\leq K\sqrt{\Delta_n}(\Delta_n u_n^{-r})^{1-1/q} = K\Delta_n^{3/2-1/q} u_n^{r/q-r}, \end{aligned} \tag{A.222}$$

by the BDG inequality, and it is $O(\Delta_n u_n^{1-r})$ if $q \geq 2(1-\varpi r)/(1-2\varpi)$.

Term (7): Same as in Eq. (A.84), we have

$$|\widetilde{\mathbf{J}}_i| \leq K(A_i^{(1)} + A_i^{(2)})\mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}}, \tag{A.223}$$

where

$$A_i^{(1)} = \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\mathbb{R}^p} (\|\delta(s,x)\| \wedge 2u_n)\mu(ds,dx), \tag{A.224}$$

$$A_i^{(2)} = \int_{(i-1)\Delta_n}^{i\Delta_n} \int_{\{\|\delta\| > 2u_n\}} \|\delta(s,x)\|\mu(ds,dx), \tag{A.225}$$

and thus,

$$\mathbb{E}\left[|\widetilde{\mathbf{J}}_i| \mathbb{1}_{\{|\widetilde{\mathbf{J}}_i| > u_n\}}\right] \leq K\mathbb{E}\left[A_i^{(1)} \mathbb{1}_{\{A_i^{(1)} > \frac{u_n}{2K}\}}\right] + K\mathbb{E}\left[A_i^{(2)} \mathbb{1}_{\{A_i^{(2)} > \frac{u_n}{2K}\}} \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}}\right]. \tag{A.226}$$

For the first expectation on the right-hand side, by Eq. (A.86),

$$\mathbb{E}\left[A_i^{(1)} \mathbb{1}_{\{A_i^{(1)} > \frac{u_n}{2K}\}}\right] \leq \frac{2K}{u_n}\mathbb{E}[(A_i^{(1)})^2] \leq K'\Delta_n u_n^{1-r}. \tag{A.227}$$

For the second expectation, we follow the same steps as in Eq. (A.208),

$$\begin{aligned} \mathbb{E}\left[A_i^{(2)} \mathbb{1}_{\{A_i^{(2)} > \frac{u_n}{2K}\}} \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}}\right] &\leq K'\mathbb{E}\left[N_i^n(2u_n)\mathbb{1}_{\{N_i^n(2u_n) \geq 1\}} \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}}\right] \\ &\leq K'\mathbb{E}\left[(N_i^n(2u_n))^2\mathbb{1}_{\{N_i^n(2u_n) \geq 1\}} \mathbb{1}_{\{\|\Delta_i^n X\| \leq u_n\}}\right] \\ &\leq K''\Delta_n^2 u_n^{-2r} = o(\Delta_n u_n^{1-r}). \end{aligned} \tag{A.228}$$

Therefore, it holds that for the term (7):

$$\mathbb{E}\left[|\widetilde{\mathbf{J}}_i| \mathbb{1}_{\{|\widetilde{\mathbf{J}}_i| > u_n\}}\right] \leq K\Delta_n u_n^{1-r}. \tag{A.229}$$

Therefore, combining all terms (1)–(7) in Eq. (A.215) leads to

$$\mathbb{E}\left[|\Delta_i^n Z|\mathbb{1}_{\{|\Delta_i^n Y|\leq u_n,\, |\Delta_i^n Z|>u_n,\, \|\Delta_i^n X\|\leq u_n\}}\right] \leq K\Delta_n u_n^{1-r}. \tag{A.230}$$

By Eqs. (A.210), (A.211) and (A.230), $\mathbb{E}[|H_{5,n}|\mathbb{1}_{E_n}] = o(1)$. By the same Markov's inequality as in Eq. (A.130), we conclude that $H_{5,n} = o_p(1)$.

By results in (i) to (v), we complete the proof of Eq. (A.104). Next, we show in (vi) and (vii) that the second term (the bias induced by the discretization error) and the third term (the bias induced by the spline approximation error) in Eq. (A.103) are both asymptotically negligible and thus have no impact on the CLT.

(vi) Discretization error. Consider the discretization error term in Eq. (A.103):

$$D_n = \frac{1}{\sqrt{\Delta_n}}\left(\int_0^T \mathbf{B}_s ds - \sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n}\Delta_n\right)\boldsymbol{\gamma} = \frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} (\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n})ds. \tag{A.231}$$

By the Cramér-Wold device, we need to show that, for any fixed linear combination with $a \in \mathbb{R}^p$, it holds that $|a^\top D_n| = o_p(1)$. By the Cauchy-Schwarz inequality, since $\widetilde{\beta}$ is absolutely continuous and has bounded derivative on $[0,T]$ (same as in Eq. (A.78)),

$$\begin{aligned}
|a^\top D_n| &\leq \|a\|\|D_n\| \leq \frac{K}{\sqrt{\Delta_n}}\left\|\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} (\widetilde{\beta}_s - \widetilde{\beta}_{(i-1)\Delta_n})ds\right\| \\
&\leq \frac{K}{\sqrt{\Delta_n}}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n}\left(\int_{(i-1)\Delta_n}^{s} \|\widetilde{\beta}'_u\|du\right)ds \leq \frac{K}{\sqrt{\Delta_n}}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \Delta_n\|\widetilde{\beta}'_u\|du \\
&\leq K\sqrt{\Delta_n}\int_0^T \|\widetilde{\beta}'_u\|du \leq K\sqrt{\Delta_n}\sqrt{T}\|\widetilde{\beta}'\|_{L^2} \leq K'\sqrt{\Delta_n}\|\widetilde{\beta}'\|_{L^2}.
\end{aligned} \tag{A.232}$$

By Lemma A.3,

$$|a^\top D_n| = O_p(\sqrt{\Delta_n K_n \log K_n}) = o_p(1). \tag{A.233}$$

(vii) Approximation error. Define the matrix-valued process $\psi = c^{-1}\mathbf{I}_p$, where $\mathbf{I}_p$ is an identity matrix of dimension p. Under Assumption 1 (iii), consider the same $L^2(c)$-orthogonal projection $\Pi_n\psi \in \mathbb{G}_n$, and then

$$\int_0^T e_s ds = \int_0^T e_s c_s \psi_s ds = \langle e, \psi\rangle_c = \langle e, \psi - \Pi_n\psi\rangle_c. \tag{A.234}$$

By the Cauchy-Schwarz inequality and the norm equivalence in Eq. (A.8),

$$\begin{aligned}\left\|\frac{1}{\sqrt{\Delta_n}}\int_0^T e_s ds\right\| &\leq \frac{1}{\sqrt{\Delta_n}}\|\langle e, \psi - \Pi_n\psi\rangle_c\| \leq \frac{1}{\sqrt{\Delta_n}}\|e\|_{L^2(c)}\|\psi - \Pi_n\psi\|_{L^2(c)} \\ &\leq \frac{K}{\sqrt{\Delta_n}}\|e\|_{L^2}\|\psi - \Pi_n\psi\|_{L^2}.\end{aligned} \tag{A.235}$$

By Lemma 1, $\|e\|_{L^2} = O_p(\sqrt{\log K_n/K_n})$. Moreover, ψ is also an Itô semimartingale with finite-variation jumps (because c is bounded, nonsingular, and an Itô semimartingale under Assumption 1 (iii), so is c^{-1} by Itô's formula for semimartingales with summable jumps; see, e.g., Theorem 3.2.2, Jacod and Protter, 2012), and then the result in Lemma 1 also applies to $\psi - \Pi_n\psi$. Therefore,

$$\|\psi - \Pi_n\psi\|_{L^2} = O_p\left(\sqrt{\frac{\log K_n}{K_n}}\right), \tag{A.236}$$

and then

$$\left\|\frac{1}{\sqrt{\Delta_n}}\int_0^T e_s ds\right\| = O_p\left(\frac{\log K_n}{K_n\sqrt{\Delta_n}}\right) = o_p(1). \tag{A.237}$$

This completes the proof.

A.5 Proof of Theorem 3

By the standard Riemann approximation and the LLN for the empirical Gram matrix in Eq. (A.30),

$$\sum_{i=1}^n \mathbf{B}_{(i-1)\Delta_n}\Delta_n \xrightarrow{\mathbb{P}} \int_0^T \mathbf{B}_s ds, \qquad (\mathbf{R}^\top\mathbf{R})^{-1} \xrightarrow{\mathbb{P}} \left(\int_0^T \mathbf{B}_s^\top c_s \mathbf{B}_s ds\right)^{-1}. \tag{A.238}$$

Hence it suffices to show

$$\mathbf{R}^\top\mathbf{D}\mathbf{R} \xrightarrow{\mathbb{P}} \int_0^T \widetilde{\sigma}_s^2 \mathbf{B}_s^\top c_s \mathbf{B}_s ds \tag{A.239}$$

We define the infeasible and feasible residuals:

$$\varepsilon_i = \Delta_i^n Y - (\Delta_i^n X)^\top \beta_{(i-1)\Delta_n-}, \qquad \widehat{\varepsilon}_i = \Delta_i^n Y - (\Delta_i^n X)^\top \widehat{\beta}_{(i-1)\Delta_n}^{(-i)}, \tag{A.240}$$

and $\widetilde{\mathbf{D}} = \mathrm{diag}(\Delta_n^{-1}\varepsilon_i^2 \mathbb{1}_{\{|\Delta_i^n Y| \leq u_n, \|\Delta_i^n X\| \leq u_n\}})$ as the infeasible counterpart of $\mathbf{D}$. We first show that the difference between $\mathbf{R}^\top\mathbf{D}\mathbf{R}$ and $\mathbf{R}^\top\widetilde{\mathbf{D}}\mathbf{R}$ are asymptotically negligible, i.e.,

$$\|\mathbf{R}^\top(\mathbf{D} - \widetilde{\mathbf{D}})\mathbf{R}\| = o_p(1). \tag{A.241}$$

Firstly, we obtain

$$
\begin{aligned}
&\widehat{\varepsilon}_i^2 - \varepsilon_i^2 = (\Delta_i^n Y - (\Delta_i^n X)^\top \widehat{\beta}^{(-i)}_{(i-1)\Delta_n})^2 - (\Delta_i^n Y - (\Delta_i^n X)^\top \beta_{(i-1)\Delta_n -})^2 \\
&= (2\Delta_i^n Y - (\Delta_i^n X)^\top \widehat{\beta}^{(-i)}_{(i-1)\Delta_n} - (\Delta_i^n Y - (\Delta_i^n X)^\top \beta_{(i-1)\Delta_n -})((\Delta_i^n X)^\top (\beta_{(i-1)\Delta_n -} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n})) \\
&= (2\Delta_i^n Y - (\Delta_i^n X)^\top \beta_{(i-1)\Delta_n -} + (\Delta_i^n X)^\top (\beta_{(i-1)\Delta_n -} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n}))((\Delta_i^n X)^\top (\beta_{(i-1)\Delta_n -} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n})) \\
&= 2\varepsilon_i (\Delta_i^n X)^\top (\beta_{(i-1)\Delta_n -} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n}) + ((\Delta_i^n X)^\top (\beta_{(i-1)\Delta_n -} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n}))^2, \qquad (A.242)
\end{aligned}
$$

and, by the BDG inequality and Eq. (A.39),

$$
\begin{aligned}
&\mathbb{E}\left[(\varepsilon_i)^4 \mathbb{1}_{\{|\Delta_i^n Y| \le u_n,\, \|\Delta_i^n X\| \le u_n\}}\right] \\
&\le K\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} (\beta_s - \beta_{(i-1)\Delta_n -})^\top dX_s^c + \Delta_i^n Z^c\right)^4\right] \\
&\quad + K'\mathbb{E}\left[\left(\sum_{(i-1)\Delta_n \le s \le i\Delta_n} \left(\Delta Y_s + (\beta_s^J - \beta_{(i-1)\Delta_n -})^\top \Delta X_s\right)\right)^4 \mathbb{1}_{\{|\Delta_i^n Y| \le u_n,\, \|\Delta_i^n X\| \le u_n\}}\right] \qquad (A.243) \\
&\le K\Delta_n^2 + K'\Delta_n u_n^{4-r}.
\end{aligned}
$$

By the triangle inequality,

$$
\|\mathbf{R}^\top (\mathbf{D} - \widetilde{\mathbf{D}})\mathbf{R}\| \le S_1 + S_2, \qquad (A.244)
$$

where, by Eq. (A.242) and the boundedness of B-spline basis functions,

$$
S_1 = \frac{K}{\Delta_n} \sum_{i=1}^n |\varepsilon_i| \|\Delta_i^n X\|^3 \|\beta_{(i-1)\Delta_n -} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n}\| \mathbb{1}_{\{|\Delta_i^n Y| \le u_n,\, \|\Delta_i^n X\| \le u_n\}}, \qquad (A.245)
$$

$$
S_2 = \frac{K}{\Delta_n} \sum_{i=1}^n \|\Delta_i^n X\|^4 \|\beta_{(i-1)\Delta_n -} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n}\|^2 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}, \qquad (A.246)
$$

and it suffices to show that both S_1 and S_2 are $o_p(1)$.

Fix i and define the enlarged σ-field

$$
\widetilde{\mathcal{F}}_{(i-1)\Delta_n} = \mathcal{F}_{(i-1)\Delta_n} \vee \sigma\{(\Delta_\ell^n X, \Delta_\ell^n Y) : \ell \ne i\}. \qquad (A.247)
$$

By construction, $\widehat{\beta}^{(-i)}_{(i-1)\Delta_n}$ is $\widetilde{\mathcal{F}}_{(i-1)\Delta_n}$-measurable. By Lemma A.2 (ii),

$$
\begin{aligned}
\mathbb{E}\left[\|\Delta_i^n X\|^4 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}} \,\middle|\, \widetilde{\mathcal{F}}_{(i-1)\Delta_n}\right] &\le K\mathbb{E}\left[\|\Delta_i^n X^c\|^4 + \|\Delta_i^n X - \Delta_i^n X^c\|^4 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}\right] \\
&\le K\Delta_n^2 + K u_n^2 \mathbb{E}\left[\|\Delta_i^n X - \Delta_i^n X^c\|^2 \mathbb{1}_{\{\|\Delta_i^n X\| \le u_n\}}\right] \qquad (A.248) \\
&\le K\Delta_n^2 + K'\Delta_n u_n^{4-r} \le K''\Delta_n^2,
\end{aligned}
$$

and thus

$$\begin{aligned}\mathbb{E}[S_2] &= \frac{K}{\Delta_n}\sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[\|\Delta_i^n X\|^4 \mathbb{1}_{\{\|\Delta_i^n X\|\le u_n\}}\,\middle|\, \widetilde{\mathcal{F}}_{(i-1)\Delta_n}\right]\|\beta_{(i-1)\Delta_n-} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n}\|^2\right] \\ &\le K\Delta_n \sum_{i=1}^n \mathbb{E}\left[\|\beta_{(i-1)\Delta_n-} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n}\|^2\right] \asymp \mathbb{E}\left[\|\widehat{\beta}^- - \beta\|^2_{L^2}\right] = o(1),\end{aligned} \tag{A.249}$$

where $\Delta_n \sum_{i=1}^n \|\beta_{(i-1)\Delta_n} - \widehat{\beta}^{(-i)}_{(i-1)\Delta_n}\|^2$ is a Riemann sum for $\|\widehat{\beta}^- - \beta\|^2_{L^2}$ of the piecewise constant process $\widehat{\beta}^-$ that equals $\widehat{\beta}^{(-i)}_{(i-1)\Delta_n}$ on $((i-1)\Delta_n, i\Delta_n]$, and the proof of Theorem 1 applies to each leave-one-out OLS fit. Therefore, $S_2 = o_p(1)$ by Markov's inequality.

By the Cauchy-Schwarz inequality,

$$\mathbb{E}[S_1] \le K(\mathbb{E}[S_2])^{1/2}\left(\Delta_n^{-1}\sum_{i=1}^n \mathbb{E}\left[\varepsilon_i^2 \|\Delta_i^n X\|^2 \mathbb{1}_{\{|\Delta_i^n Y|\le u_n,\, \|\Delta_i^n X\|\le u_n\}}\right]\right)^{1/2}, \tag{A.250}$$

where, by Eqs. (A.243) and (A.248),

$$\begin{aligned}&\mathbb{E}\left[\varepsilon_i^2 \|\Delta_i^n X\|^2 \mathbb{1}_{\{|\Delta_i^n Y|\le u_n,\, \|\Delta_i^n X\|\le u_n\}}\right] \\ &\quad\le \left(\mathbb{E}\left[\varepsilon_i^4 \mathbb{1}_{\{|\Delta_i^n Y|\le u_n,\, \|\Delta_i^n X\|\le u_n\}}\right]\right)^{1/2}\left(\mathbb{E}\left[\|\Delta_i^n X\|^4 \mathbb{1}_{\{\|\Delta_i^n X\|\le u_n\}}\right]\right)^{1/2} \\ &\quad\le K\Delta_n^2 + K'\Delta_n^{3/2} u_n^{2-r/2} \le K''\Delta_n^2,\end{aligned} \tag{A.251}$$

and thus $S_1 = o_p(1)$ by Markov's inequality. Therefore, it holds that $\|\mathbf{R}^\top(\mathbf{D} - \widetilde{\mathbf{D}})\mathbf{R}\| = o_p(1)$, and thus

$$\mathbf{R}^\top \mathbf{D}\mathbf{R} = \mathbf{R}^\top \widetilde{\mathbf{D}}\mathbf{R} + o_p(1). \tag{A.252}$$

Based on the same arguments on negligible remainders as in the proofs of Theorems 1 and 2, we have $\mathbf{R}^\top \widetilde{\mathbf{D}}\mathbf{R} = \mathbf{R}^\top \mathrm{diag}(\Delta_n^{-1}(\Delta_i^n Z)^2 \mathbb{1}_{\{|\Delta_i^n Z|\le u_n\}})\mathbf{R} + o_p(1)$, and thus, by the standard LLN result of truncated covariances,

$$\mathbf{R}^\top \widetilde{\mathbf{D}}\mathbf{R} \xrightarrow{\mathbb{P}} \int_0^T \widetilde{\sigma}_s^2 \mathbf{B}_s^\top c_s \mathbf{B}_s ds. \tag{A.253}$$

Combining Eqs. (A.238), (A.252) and (A.253) leads to the consistency of $\widehat{\Sigma}^\beta_T$. This completes the proof.

A.6 Proof of Theorem 4

Define the block-wise score $c_j(\boldsymbol{\gamma}) = \frac{\partial Q_n(\boldsymbol{\gamma})}{\partial \gamma_j} = -(\mathbf{R}_j^*)^\top(\mathbf{Y} - \mathbf{R}\boldsymbol{\gamma})$. The standard KKT conditions (e.g., Xue and Qu, 2012, Eqs. (19)–(20)) imply that any local minimizer of $Q_n^*(\boldsymbol{\gamma})$ must satisfy:

$$c_j(\widehat{\boldsymbol{\gamma}}) = 0, \qquad \sqrt{\widehat{\gamma}_j^\top \mathbf{W}_j \widehat{\gamma}_j} > \tau_n, \qquad \text{for } j = 1, \dots, q. \tag{A.254}$$

$$\sqrt{(c_j(\widehat{\boldsymbol{\gamma}}))^\top \mathbf{W}_j c_j(\widehat{\boldsymbol{\gamma}})} \le \frac{\lambda_n}{\tau_n}, \qquad \sqrt{\widehat{\gamma}_j^\top \mathbf{W}_j \widehat{\gamma}_j} = 0, \qquad \text{for } j = q+1, \dots, p. \tag{A.255}$$

Then it suffices to show that $\widehat{\gamma}^\circ$ satisfies both Eqs. (A.254) and (A.255) with probability approaching one.

We denote by $\widehat{\gamma}_q^\circ$ (resp. $\boldsymbol{\gamma}_q^\circ$) the subvector of $\widehat{\boldsymbol{\gamma}}^\circ = ((\widehat{\gamma}_q^\circ)^\top, (\widehat{\gamma}_{-q}^\circ)^\top)^\top$ (resp. $\boldsymbol{\gamma}^\circ$) that includes only the first qK_n elements, and by $\mathbf{R}_q = (\mathbf{R}_1^*, \ldots, \mathbf{R}_q^*)^\top \in \mathbb{R}^{n \times qK_n}$ the submatrix of $\mathbf{R}$ corresponding to the first q regressors. Define the oracle least-squares spline estimator,

$$\widehat{\gamma}_q^\circ = \underset{\boldsymbol{\gamma}:\, \gamma_j \equiv 0,\, \forall q<j\le p}{\operatorname{argmin}} Q_n(\boldsymbol{\gamma}), \qquad \widehat{\gamma}_{-q}^\circ \equiv 0, \qquad \widehat{\gamma}_q^\circ = (\mathbf{R}_q^\top \mathbf{R}_q)^{-1} \mathbf{R}_q^\top \mathbf{Y}. \tag{A.256}$$

Define two standard projection matrices: the orthogonal projection matrix $\mathbf{H}_q = \mathbf{R}_q(\mathbf{R}_q^\top \mathbf{R}_q)^{-1}\mathbf{R}_q^\top \in \mathbb{R}^{n\times n}$ onto $\operatorname{span}(\mathbf{R}_q)$, and the residualizer $\mathbf{M}_q = \mathbf{I}_n - \mathbf{H}_q$, where $\mathbf{I}_n$ is an identity matrix of dimension n. Both $\mathbf{H}_q$ and $\mathbf{M}_q$ are symmetric and idempotent. Then the oracle residual is given by

$$\mathbf{Y} - \mathbf{R}\widehat{\boldsymbol{\gamma}}^\circ = \mathbf{Y} - \mathbf{R}_q \widehat{\gamma}_q^\circ = \mathbf{M}_q \mathbf{Y}. \tag{A.257}$$

Active blocks. By definition, $\widehat{\boldsymbol{\gamma}}^\circ$ minimizes $Q_n(\boldsymbol{\gamma})$ over the restricted model $\{\boldsymbol{\gamma} : \boldsymbol{\gamma}_{-q} \equiv 0\}$. Hence $c_j(\widehat{\boldsymbol{\gamma}}^\circ) = 0$ holds for all $1 \le j \le q$. Same as Eq. (A.29), it holds that

$$(\widehat{\gamma}_j^\circ)^\top \mathbf{W}_j \widehat{\gamma}_j^\circ \xrightarrow{\mathbb{P}} \int_0^T (\widehat{\beta}_{j,s}^\circ)^\top c_s \widehat{\beta}_{j,s}^\circ ds \asymp \|\widehat{\beta}_j^\circ\|_{L^2}^2, \qquad \text{where} \quad \widehat{\beta}_j^\circ = \mathbf{B}_j \widehat{\gamma}_j^\circ. \tag{A.258}$$

By Theorem 1 and the triangle inequality, for any active block $1 \le j \le q$,

$$\Big| \|\widehat{\beta}_j^\circ\|_{L^2} - \|\beta_j\|_{L^2} \Big| \le \|\widehat{\beta}_j^\circ - \beta_j\|_{L^2} \xrightarrow{\mathbb{P}} 0. \tag{A.259}$$

so $\sqrt{(\widehat{\gamma}_j^\circ)^\top \mathbf{W}_j \widehat{\gamma}_j^\circ} \xrightarrow{\mathbb{P}} \|\beta_j\|_{L^2} > 0$, and thus it satisfies $\sqrt{(\widehat{\gamma}_j^\circ)^\top \mathbf{W}_j \widehat{\gamma}_j^\circ} > \tau_n$ for any $\tau_n \to 0$. Hence the active-block KKT conditions hold.

Inactive blocks. For $q+1 \le j \le p$, $\widehat{\gamma}_j^\circ \equiv 0$ by construction. Under the condition

$$\frac{\lambda_n}{\tau_n} \Big/ \left(\sqrt{\frac{\Delta_n \log pK_n}{K_n}} + u_n \log pK_n + u_n^{1-r/2} \right) \to \infty, \tag{A.260}$$

it suffices to show

$$\max_{q+1 \le j \le p} \sqrt{(c_j(\widehat{\boldsymbol{\gamma}}^\circ))^\top \mathbf{W}_j c_j(\widehat{\boldsymbol{\gamma}}^\circ)} = O_p\left(\sqrt{\frac{\Delta_n \log pK_n}{K_n}} + u_n \log pK_n + u_n^{1-r/2} \right), \tag{A.261}$$

which implies Eq. (A.255). By Eq. (A.257), we have the residualized score for any inactive blocks,

$$c_j(\widehat{\boldsymbol{\gamma}}^\circ) = -(\mathbf{R}_j^*)^\top \mathbf{M}_q \mathbf{Y} = -(\widetilde{\mathbf{R}}_j^*)^\top \mathbf{Y}, \qquad \text{where} \quad \widetilde{\mathbf{R}}_j^* = \mathbf{M}_q \mathbf{R}_j^*, \tag{A.262}$$

for any $q+1 \leq j \leq p$. By the Cauchy-Schwarz inequality,

$$\sqrt{(c_j(\widehat{\gamma}^\circ))^\top \mathbf{W}_j c_j(\widehat{\gamma}^\circ)} = \|\mathbf{W}_j^{1/2}(\widetilde{\mathbf{R}}_j^*)^\top \mathbf{Y}\| \leq \|\mathbf{W}_j^{1/2}\|\|(\widetilde{\mathbf{R}}_j^*)^\top \mathbf{Y}\|, \tag{A.263}$$

where $\|\mathbf{W}_j^{1/2}\| = \sqrt{\lambda_{\max}(\mathbf{W}_j)} = O_p(K_n^{-1/2})$ as shown following Eq. (A.32). It suffices to show

$$\max_{q+1\leq j\leq p} \|(\widetilde{\mathbf{R}}_j^*)^\top \mathbf{Y}\| = O_p\left(\sqrt{\Delta_n \log pK_n} + \sqrt{K_n} u_n \log pK_n + \sqrt{K_n} u_n^{1-r/2}\right). \tag{A.264}$$

We consider the decomposition of $\mathbf{Y}$ as in Eq. (A.33) (with the first two components combined):

$$\begin{aligned}\mathbf{Y}_i = &\underbrace{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \beta_{s-}^\top dX_s^c\right) \mathbb{1}_{\{|\Delta_i^n Y|\leq u_n\}}}_{\mathbf{C}_i} + \underbrace{\left(\sum_{(i-1)\Delta_n\leq s\leq i\Delta_n} (\beta_{s-}^J)^\top \Delta X_s\right) \mathbb{1}_{\{|\Delta_i^n Y|\leq u_n\}}}_{\mathbf{J}_i} \\ &+ \underbrace{\Delta_i^n Z \mathbb{1}_{\{|\Delta_i^n Y|\leq u_n\}}}_{\mathbf{Z}_i}.\end{aligned} \tag{A.265}$$

Therefore, it suffices to verify, for each $\Xi \in \{\mathbf{C}, \mathbf{J}, \mathbf{Z}\}$,

$$\max_{q+1\leq j\leq p} \|(\widetilde{\mathbf{R}}_j^*)^\top \Xi\| = O_p\left(\sqrt{\Delta_n \log pK_n} + \sqrt{K_n} u_n \log pK_n + \sqrt{K_n} u_n^{1-r/2}\right). \tag{A.266}$$

We firstly show $\|\widetilde{\mathbf{R}}_j^*\|_\infty$ is bounded for any $q+1 \leq j \leq p$.

(i) Bounded $\|\widetilde{\mathbf{R}}_j^*\|_\infty$. Let $h_{ii} = (\mathbf{H}_q)_{ii}$ and $h_{\max} = \max_{1\leq i\leq n} h_{ii}$. For any $\boldsymbol{r} \in \mathbb{R}^n$, $|(\mathbf{H}_q \boldsymbol{r})_i| \leq \sqrt{h_{ii}}\|\boldsymbol{r}\|$, and thus $\|\mathbf{H}_q \boldsymbol{r}\|_\infty \leq \sqrt{h_{\max}}\|\boldsymbol{r}\|$. By definition,

$$\|\mathbf{M}_q \boldsymbol{r}\|_\infty \leq \|(\mathbf{I}_n - \mathbf{H}_q)\boldsymbol{r}\|_\infty \leq \|\boldsymbol{r}\|_\infty + \sqrt{h_{\max}}\|\boldsymbol{r}\|. \tag{A.267}$$

Moreover, it holds that

$$\begin{aligned} h_{\max} &\leq \left(\max_{1\leq i\leq n} \|\mathbf{R}_i\|^2\right) \|(\mathbf{R}_q^\top \mathbf{R}_q)^{-1}\| \\ &\leq \left(\max_{1\leq i\leq n} \sum_{j=1}^q \sum_{k=1}^{K_n} (B_{(i-1)\Delta_n}^{(k)} \Delta_{j,i}^n X)^2 \mathbb{1}_{\{|\Delta_{j,i}^n X|\leq u_n\}}\right) \|(\mathbf{R}_q^\top \mathbf{R}_q)^{-1}\| \\ &\leq \left(\max_{1\leq i\leq n} q u_n^2 \sum_{k=1}^{K_n} (B_{(i-1)\Delta_n}^{(k)})^2\right) \|(\mathbf{R}_q^\top \mathbf{R}_q)^{-1}\| = O_p(K_n u_n^2), \end{aligned} \tag{A.268}$$

by the boundedness and local support of B-splines, and $\|(\mathbf{R}_q^\top \mathbf{R}_q)^{-1}\| = O_p(K_n)$. Therefore, by Eqs. (A.267) and (A.268),

$$\|\widetilde{\mathbf{R}}_j^*\|_\infty \leq \|\mathbf{R}_j^*\|_\infty + \sqrt{h_{\max}}\|\mathbf{R}_j^*\| \leq K u_n + K' \sqrt{K_n} u_n \frac{1}{\sqrt{K_n}} \leq K'' u_n. \tag{A.269}$$

(ii) Maximal inequality for martingale difference array. We start with a high-dimensional martingale maximal inequality from standard Freedman's inequality, as an analogue to Lemma A.1 in van de Geer (2008) from Bernstein's inequality:

Lemma A.4. Let $(\mathcal{F}_i)_{i=0}^n$ be a filtration and $(M_i^{(\ell)})$ be a triangular array indexed by $1 \le \ell \le m$. Assume there exist a_n, $b_n > 0$ such that for all $1 \le \ell \le m$,

(i) $\mathbb{E}[M_i^{(\ell)}|\mathcal{F}_{i-1}] = 0$ for all i;
(ii) $|M_i^{(\ell)}| \le b_n$ almost surely for all i;
(iii) $\sum_{i=1}^n \mathbb{E}[(M_i^{(\ell)})^2|\mathcal{F}_{i-1}] \le a_n$.

Then,

$$\max_{1\le \ell \le m} \left| \sum_{i=1}^n M_i^{(\ell)} \right| = O_p\left(\sqrt{a_n \log m} + b_n \log m\right). \tag{A.270}$$

Proof of Lemma A.4. By Freedman's inequality (see, e.g., Theorem 1.1, Tropp, 2011), for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left| \sum_{i=1}^n M_i^{(\ell)} \right| \ge \varepsilon\right) \le 2\exp\left(-\frac{\varepsilon^2}{2(a_n + b_n\varepsilon/3)}\right). \tag{A.271}$$

By the union bound over all $1 \le \ell \le m$,

$$\mathbb{P}\left(\max_{1\le \ell \le m}\left| \sum_{i=1}^n M_i^{(\ell)} \right| \ge \varepsilon\right) \le 2m\exp\left(-\frac{\varepsilon^2}{2(a_n + b_n\varepsilon/3)}\right). \tag{A.272}$$

Choose $\varepsilon = C(\sqrt{a_n \log m} + b_n \log m)$ with a sufficiently large constant $C > 0$.

(a) If $\sqrt{a_n \log m} \ge b_n \log m$, we have $\varepsilon \asymp C\sqrt{a_n \log m}$, $\sqrt{a_n} \ge b_n \sqrt{\log m}$, and $a_n + b_n\varepsilon/3 \asymp a_n$, so the right-hand side of Eq. (A.272) satisfies

$$2m\exp\left(-\frac{\varepsilon^2}{2(a_n + b_n\varepsilon/3)}\right) \le 2m\exp(-KC^2\log m) = 2m^{1-KC^2} \to 0, \tag{A.273}$$

for C large enough.

(b) If $\sqrt{a_n \log m} \le b_n \log m$, we have $\varepsilon \asymp Cb_n \log m$, $a_n \le b_n^2 \log m$, and $a_n + b_n\varepsilon/3 \asymp b_n^2 \log m$, so the right-hand side of Eq. (A.272) has the same bound as in Eq. (A.273).

This completes the proof of Lemma A.4.

□

Next, we use Lemma A.4 to prove Eq. (A.266) for each Ξ. We start with some notation: For the j-th block design submatrix $\mathbf{R}_j^*$ in Eq. (30), we write each element as

$$R_{j,k,i} = B_{(i-1)\Delta_n}^{(k)} \Delta_{j,i}^n X \mathbb{1}_{\{|\Delta_{j,i}^n X| \le u_n\}}, \tag{A.274}$$

and the $((j-1)K_n+k)$-column $R_{j,k} = (R_{j,k,i})_{i=1}^n \in \mathbb{R}^n$. For $\widetilde{\mathbf{R}}_j^* = \mathbf{M}_q \mathbf{R}_j^*$, we write the corresponding "residualized" column as

$$\widetilde{R}_{j,k} = \mathbf{M}_q R_{j,k} = (\widetilde{R}_{j,k,i})_{i=1}^n \in \mathbb{R}^n, \tag{A.275}$$

where each element $|\widetilde{R}_{j,k,i}| \le K u_n$ by Eq. (A.269).

(ii.1) $\boldsymbol{\Xi} = \mathbf{Z}$. Define

$$\widetilde{M}_i^{(j,k)} = \widetilde{R}_{j,k,i} \Delta_i^n Z \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}, \tag{A.276}$$

and

$$M_i^{(j,k)} = \widetilde{M}_i^{(j,k)} - \mathbb{E}\left[\widetilde{M}_i^{(j,k)} \middle| \mathcal{F}_{(i-1)\Delta_n} \right], \quad \text{so that} \quad \mathbb{E}\left[M_i^{(j,k)} \middle| \mathcal{F}_{(i-1)\Delta_n} \right] = 0. \tag{A.277}$$

With finite-variation jumps in Z and bounded $\widetilde{\delta}$ under Assumption A.1, it holds that $|\widetilde{M}_i^{(j,k)}| \le K u_n$, and thus $|M_i^{(j,k)}| \le 2K u_n$. For the conditional variance, by Itô's isometry and Markov's inequality, $\mathbb{E}[(\Delta_i^n Z)^2 | \breve{\mathcal{F}}_{(i-1)\Delta_n}] = O_p(\Delta_n)$ with an enlarged σ-field $\breve{\mathcal{F}}_{(i-1)\Delta_n} = \mathcal{F}_{(i-1)\Delta_n} \cup \sigma(\mathbf{M}_q, \Delta_i^n X)$. Then, since $\widetilde{R}_{j,k,i}$ is $\breve{\mathcal{F}}_{(i-1)\Delta_n}$-measurable, by the law of iterated conditional expectations,

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}\left[(M_i^{(j,k)})^2 \middle| \mathcal{F}_{(i-1)\Delta_n} \right] &\le \sum_{i=1}^n \mathbb{E}\left[(\widetilde{M}_i^{(j,k)})^2 \middle| \mathcal{F}_{(i-1)\Delta_n} \right] \\
&\le \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[(\Delta_i^n Z)^2 \middle| \breve{\mathcal{F}}_{(i-1)\Delta_n} \right] \widetilde{R}_{j,k,i}^2 \middle| \mathcal{F}_{(i-1)\Delta_n} \right] \\
&\le K \Delta_n \sum_{i=1}^n \mathbb{E}\left[\widetilde{R}_{j,k,i}^2 \middle| \mathcal{F}_{(i-1)\Delta_n} \right],
\end{aligned} \tag{A.278}$$

where, since $\mathbf{M}_q$ is an orthogonal projection,

$$\mathbb{E}\left[\sum_{i=1}^n \mathbb{E}\left[\widetilde{R}_{j,k,i}^2 \middle| \mathcal{F}_{(i-1)\Delta_n} \right] \right] = \sum_{i=1}^n \mathbb{E}[\widetilde{R}_{j,k,i}^2] = \mathbb{E}[\|\widetilde{R}_{j,k}\|^2] \le \mathbb{E}[\|R_{j,k}\|^2] = O(K_n^{-1}), \tag{A.279}$$

and thus $\sum_{i=1}^n \mathbb{E}[(M_i^{(j,k)})^2 | \mathcal{F}_{(i-1)\Delta_n}] = O_p(K_n^{-1}\Delta_n)$ by Markov's inequality. Therefore, by Lemma A.4,

$$\max_{q+1 \le j \le p} \max_{1 \le k \le K_n} \left| \sum_{i=1}^n M_i^{(j,k)} \right| = O_p\left(\sqrt{\frac{\Delta_n \log pK_n}{K_n}} + u_n \log pK_n \right). \tag{A.280}$$

Next, we bound the predictable compensator uniformly. Write

$$\sum_{i=1}^n \mathbb{E}\left[\widetilde{M}_i^{(j,k)} \middle| \mathcal{F}_{(i-1)\Delta_n} \right] = \sum_{i=1}^n \mathbb{E}\left[\widetilde{R}_{j,k,i} \Delta_i^n Z \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} \middle| \mathcal{F}_{(i-1)\Delta_n} \right] \tag{A.281}$$

By the exogeneity condition under Assumption 3, the untruncated product has zero conditional expectation: $\mathbb{E}[\widetilde{R}_{j,k,i} \Delta_i^n Z | \mathcal{F}_{(i-1)\Delta_n}] = 0$. Hence, we flip the indicator function

$$\mathbb{E}\left[\widetilde{M}_i^{(j,k)} \middle| \mathcal{F}_{(i-1)\Delta_n} \right] = -\mathbb{E}\left[\widetilde{R}_{j,k,i} \Delta_i^n Z \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}} \middle| \mathcal{F}_{(i-1)\Delta_n} \right], \tag{A.282}$$

and then, by the Cauchy-Schwarz inequality, Lemma A.2 (i), and Markov's inequality,

$$
\begin{aligned}
\sum_{i=1}^{n}\left|\mathbb{E}\left[\widetilde{M}_i^{(j,k)}\middle|\mathcal{F}_{(i-1)\Delta_n}\right]\right| &\le K u_n \sum_{i=1}^{n}\mathbb{E}\left[|\Delta_i^n Z|\mathbb{1}_{\{|\Delta_i^n Y|>u_n\}}\middle|\mathcal{F}_{(i-1)\Delta_n}\right] \\
&\le K u_n\left(\sum_{i=1}^{n}\mathbb{E}[(\Delta_i^n Z)^2|\mathcal{F}_{(i-1)\Delta_n}]\right)^{1/2}\left(\sum_{i=1}^{n}\mathbb{P}(|\Delta_i^n Y|>u_n|\mathcal{F}_{(i-1)\Delta_n})\right)^{1/2} \\
&\le K' u_n\left(\sum_{i=1}^{n}O_p(\Delta_n)\right)^{1/2}\left(\sum_{i=1}^{n}O_p(\Delta_n u_n^{-r})\right)^{1/2} = O_p(u_n^{1-r/2}), \quad \text{(A.283)}
\end{aligned}
$$

for any $q+1\le j\le p$ and $1\le k\le K_n$. Combining Eqs. (A.280) and (A.283) in Eq. (A.277), we obtain

$$
\begin{aligned}
\max_{q+1\le j\le p}\|(\widetilde{\mathbf{R}}_j^*)^\top\mathbf{Z}\| &\le \sqrt{K_n}\max_{q+1\le j\le p}\max_{1\le k\le K_n}\left|\sum_{i=1}^{n}\widetilde{M}_i^{(j,k)}\right| \\
&\le \sqrt{K_n}\max_{q+1\le j\le p}\max_{1\le k\le K_n}\left(\left|\sum_{i=1}^{n}M_i^{(j,k)}\right|+\left|\sum_{i=1}^{n}\mathbb{E}\left[\widetilde{M}_i^{(j,k)}\middle|\mathcal{F}_{(i-1)\Delta_n}\right]\right|\right) \qquad \text{(A.284)} \\
&= O_p\left(\sqrt{\Delta_n\log pK_n}+\sqrt{K_n}u_n\log pK_n+\sqrt{K_n}u_n^{1-r/2}\right).
\end{aligned}
$$

(ii.2) $\boldsymbol{\Xi}=\mathbf{C}$. Define

$$
\widetilde{M}_i^{(j,k)} = \widetilde{R}_{j,k,i}\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\beta_{s-}^\top dX_s^c\right)\mathbb{1}_{\{|\Delta_i^n Y|\le u_n\}}, \tag{A.285}
$$

and $M_i^{(j,k)}=\widetilde{M}_i^{(j,k)}-\mathbb{E}[\widetilde{M}_i^{(j,k)}|\mathcal{F}_{(i-1)\Delta_n}]$. By bounded β under Assumption A.1, $|M_i^{(j,k)}|\le Ku_n$. The remaining steps follow exactly as in (ii.1), and we conclude

$$
\max_{q+1\le j\le p}\|(\widetilde{\mathbf{R}}_j^*)^\top\mathbf{C}\| = O_p\left(\sqrt{\Delta_n\log pK_n}+\sqrt{K_n}u_n\log pK_n+\sqrt{K_n}u_n^{1-r/2}\right). \tag{A.286}
$$

(ii.3) $\boldsymbol{\Xi}=\mathbf{J}$. Using the safe replacement $\widetilde{\mathbf{J}}$ for $\mathbf{J}$ in Eq. (A.82),

$$
\begin{aligned}
\widetilde{M}_i^{(j,k)} &= \widetilde{R}_{j,k,i}\left(\sum_{(i-1)\Delta_n\le s\le i\Delta_n}(\beta_{s-}^J)^\top\Delta X_s\right)\mathbb{1}_{\{|\Delta_i^n Y|\le u_n\}} \\
&= \widetilde{R}_{j,k,i}\left(\sum_{(i-1)\Delta_n\le s\le i\Delta_n}(\beta_{s-}^J)^\top\Delta X_s\right)\mathbb{1}_{\{|\Delta_{j,i}^n X|\le u_n,\,|\Delta_i^n Y|\le u_n\}},
\end{aligned} \tag{A.287}
$$

since $\widetilde{R}_{j,k,i}$ already includes the truncation $\mathbb{1}_{\{|\Delta_{j,i}^n X|\le u_n\}}$, and $M_i^{(j,k)}=\widetilde{M}_i^{(j,k)}-\mathbb{E}[\widetilde{M}_i^{(j,k)}|\mathcal{F}_{(i-1)\Delta_n}]$. With bounded β^J and δ under Assumption A.1 and finite-variation jumps in X, $|M_i^{(j,k)}|\le Ku_n$. For the conditional variance, we follow the same steps as in (ii.1). By the result in Eq. (A.96) and

Markov's inequality,

$$\mathbb{E}\left[\left(\sum_{(i-1)\Delta_n\le s\le i\Delta_n}(\beta_{s-}^J)^\top \Delta X_s\right)^2 \mathbb{1}_{\{|\Delta_{j,i}^n X|\le u_n,\,|\Delta_i^n Y|\le u_n\}}\middle|\breve{\mathcal{F}}_{(i-1)\Delta_n}\right] = O_p(\Delta_n u_n^{2-r}+\Delta_n^2 u_n^{-2r}),\tag{A.288}$$

and thus $\sum_{i=1}^n \mathbb{E}[(M_i^{(j,k)})^2|\mathcal{F}_{(i-1)\Delta_n}] = O_p(K_n^{-1}(\Delta_n u_n^{2-r}+\Delta_n^2 u_n^{-2r}))$. Therefore, by Lemma A.4,

$$\begin{aligned}\max_{q+1\le j\le p}\max_{1\le k\le K_n}\left|\sum_{i=1}^n M_i^{(j,k)}\right| &= O_p\left(\sqrt{\frac{(\Delta_n u_n^{2-r}+\Delta_n^2 u_n^{-2r})\log pK_n}{K_n}}+u_n\log pK_n\right)\\ &= O_p\left(\sqrt{\frac{\Delta_n\log pK_n}{K_n}}+u_n\log pK_n\right).\end{aligned}\tag{A.289}$$

We bound the predictable compensator uniformly with the flipped indicator function as in Eq. (A.283),

$$\begin{aligned}&\sum_{i=1}^n\left|\mathbb{E}\left[\widetilde{M}_i^{(j,k)}\middle|\mathcal{F}_{(i-1)\Delta_n}\right]\right|\\ &\le Ku_n\sum_{i=1}^n\mathbb{E}\left[\left|\sum_{(i-1)\Delta_n\le s\le i\Delta_n}(\beta_{s-}^J)^\top\Delta X_s\right|\mathbb{1}_{\{|\Delta_{j,i}^n X|\le u_n,\,|\Delta_i^n Y|>u_n\}}\middle|\mathcal{F}_{(i-1)\Delta_n}\right]\\ &\le K'u_n\left(\sum_{i=1}^n\mathbb{E}\left[\left(\sum_{(i-1)\Delta_n\le s\le i\Delta_n}\Delta X_s\right)^2\middle|\mathcal{F}_{(i-1)\Delta_n}\right]\right)^{1/2}\left(\sum_{i=1}^n\mathbb{P}(|\Delta_i^n Y|>u_n|\mathcal{F}_{(i-1)\Delta_n})\right)^{1/2}\\ &= O_p(u_n^{1-r/2}).\end{aligned}\tag{A.290}$$

Combining Eqs. (A.289) and (A.290), as in Eq. (A.284), leads to

$$\max_{q+1\le j\le p}\|(\widetilde{\mathbf{R}}_j^*)^\top\mathbf{J}\| = O_p\left(\sqrt{\Delta_n\log pK_n}+\sqrt{K_n}u_n\log pK_n+\sqrt{K_n}u_n^{1-r/2}\right).\tag{A.291}$$

Finally, Eq. (A.266) holds by all results in (ii.1), (ii.2) and (ii.3), and this completes the proof.

A.7 Proof of Theorem 5

We define the oracle index set $A^\circ=\{1,\ldots,q\}$. Let $\Gamma_1=\{A: A\subseteq\{1,\ldots,p\}, A^\circ\subset A, A\ne A^\circ\}$ collect all index sets of overfitting models, $\Gamma_2=\{A: A\subseteq\{1,\ldots,p\}, A\subset A^\circ, A\ne A^\circ\}$ collect those of underfitting models, and $\Gamma_3=\{A: A\subseteq\{1,\ldots,p\}, A^\circ\not\subset A, A\not\subset A^\circ, A\ne A^\circ\}$ collect those of other misspecified models. For any $\boldsymbol{\gamma}$, $A(\boldsymbol{\gamma})=\{j:\sqrt{\gamma_j^\top\mathbf{W}_j\gamma_j}\ge\tau_n\}$ must fall into one of Γ_ℓ for $\ell\in\{1,2,3\}$ if $A(\boldsymbol{\gamma})\ne A^\circ$. $|A|$ denotes the cardinality of A. We aim to show that, with probability approaching one,

$$\min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})\ne A^\circ} Q_n^*(\boldsymbol{\gamma}) > Q_n^*(\widehat{\boldsymbol{\gamma}}^\circ),\tag{A.292}$$

which ensures that the global minimizer $\widehat{\boldsymbol{\gamma}}^*$ of $Q_n^*(\boldsymbol{\gamma})$ coincides with the oracle estimator $\widehat{\boldsymbol{\gamma}}^\circ$. It suffices to show that, for $\ell \in \{1,2,3\}$,

$$\sum_{A\in\Gamma_\ell} \mathbb{P}\left(\min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n^*(\boldsymbol{\gamma}) - Q_n^*(\widehat{\boldsymbol{\gamma}}^\circ) \le 0\right) \to 0, \tag{A.293}$$

such that, combining with Theorem 4, it holds that the global minimizer $\widehat{\boldsymbol{\gamma}}^*$ exists and satisfies $\widehat{\boldsymbol{\gamma}}^* = \widehat{\boldsymbol{\gamma}}^\circ$ with probability approaching one. Firstly, we have

$$\begin{aligned}\min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n^*(\boldsymbol{\gamma}) &= \min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A}\left(\frac{1}{2}Q_n(\boldsymbol{\gamma}) + \lambda_n \sum_{j\in A(\boldsymbol{\gamma})} \rho_n\left(\sqrt{\gamma_j^\top \mathbf{W}_j\gamma_j}\right) + \lambda_n \sum_{j\in\{1,\dots,p\}\setminus A(\boldsymbol{\gamma})} \rho_n\left(\sqrt{\gamma_j^\top \mathbf{W}_j\gamma_j}\right)\right)\\ &= \frac{1}{2}\min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n(\boldsymbol{\gamma}) + \lambda_n|A|,\end{aligned} \tag{A.294}$$

and thus comparing any model A to the oracle model A° reduces to comparing unpenalized losses and constant penalties:

$$\min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n^*(\boldsymbol{\gamma}) - Q_n^*(\widehat{\boldsymbol{\gamma}}^\circ) = \frac{1}{2}\left(\min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n(\boldsymbol{\gamma}) - Q_n(\widehat{\boldsymbol{\gamma}}^\circ)\right) + \lambda_n(|A| - q). \tag{A.295}$$

For any index set A, let $\mathbf{R}_A = (\mathbf{R}_j^*)_{j\in A} \in \mathbb{R}^{n\times K_n|A|}$ be the submatrix of $\mathbf{R}$ formed by blocks $j \in A$. We define the orthogonal projection matrix $\mathbf{H}_A = \mathbf{R}_A(\mathbf{R}_A^\top\mathbf{R}_A)^{-1}\mathbf{R}_A^\top \in \mathbb{R}^{n\times n}$ onto $\mathrm{span}(\mathbf{R}_A)$, and the residualizer $\mathbf{M}_A = \mathbf{I}_n - \mathbf{H}_A$, where $\mathbf{I}_n$ is an identity matrix of dimension n. Both $\mathbf{H}_A$ and $\mathbf{M}_A$ are symmetric and idempotent.

Overfitting models. For any overfitting models with $A \in \Gamma_1$, the added blocks in the design matrix $\mathbf{R}_A$ corresponding to the index set $A\backslash A^\circ$ leads to a decrease in the OLS residual sums of squares. We denote the residual vector by $\widehat{\boldsymbol{\varepsilon}} = \mathbf{M}_{A^\circ}\mathbf{Y} = \mathbf{Y} - \mathbf{R}\widehat{\boldsymbol{\gamma}}^\circ = \mathbf{Y} - \mathbf{R}_{A^\circ}\boldsymbol{\gamma}_{A^\circ} \in \mathbb{R}^n$. For any irrelevant group $j > q$, we define the spurious improvement in the unpenalized loss from adding block j as

$$\Delta Q_{n,j} = Q_n(\widehat{\boldsymbol{\gamma}}^\circ) - \min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})\subseteq A^\circ\cup\{j\}} Q_n(\boldsymbol{\gamma}) = \mathbf{Y}^\top(\mathbf{M}_{A^\circ} - \mathbf{M}_{A^\circ\cup\{j\}})\mathbf{Y}. \tag{A.296}$$

The following Lemma A.5 provides a uniform bound for $\Delta Q_{n,j}$:

Lemma A.5. Under the conditions of Theorem 5, there exists a constant $K > 0$ independent of n and p such that, as $\Delta_n \to 0$,

$$\mathbb{P}(E_n) = \mathbb{P}\left(\max_{q+1\le j\le p} \Delta Q_{n,j} \le K\Delta_n(K_n + \log p)\right) \to 1. \tag{A.297}$$

Proof of Lemma A.5. Define the residualized block as $\widetilde{\mathbf{R}}_j^* = \mathbf{M}_{A^\circ}\mathbf{R}_j^* \in \mathbb{R}^{n\times K_n}$, the projection matrix $\mathbf{H}_j^* = \widetilde{\mathbf{R}}_j^*((\widetilde{\mathbf{R}}_j^*)^\top\widetilde{\mathbf{R}}_j^*)^{-1}(\widetilde{\mathbf{R}}_j^*)^\top \in \mathbb{R}^{n\times n}$ onto $\mathrm{span}(\widetilde{\mathbf{R}}_j^*)$, and the corresponding residualizer $\mathbf{M}_j^* = \mathbf{I}_n - \mathbf{H}_j^*$. By the Frisch-Waugh-Lovell theorem (Theorem 2.1, Davidson and MacKinnon, 2004), the OLS residual sums of squares from the overfitted model with the irrelevant block j satisfies

$$\mathbf{Y}^\top \mathbf{M}_{A^\circ \cup \{j\}} \mathbf{Y} = \widehat{\varepsilon}^\top \mathbf{M}_j^* \widehat{\varepsilon}, \tag{A.298}$$

so that

$$\Delta Q_{n,j} = \widehat{\varepsilon}^\top \widehat{\varepsilon} - \widehat{\varepsilon}^\top \mathbf{M}_j^* \widehat{\varepsilon} = \widehat{\varepsilon}^\top \mathbf{H}_j^* \widehat{\varepsilon} = \|(\mathbf{P}_j^*)^\top \widehat{\varepsilon}\|^2, \tag{A.299}$$

where $\mathbf{P}_j^* = \widetilde{\mathbf{R}}_j^* ((\widetilde{\mathbf{R}}_j^*)^\top \widetilde{\mathbf{R}}_j^*)^{-1/2}$, and it suffices to show

$$\max_{q+1 \le j \le p} \|(\mathbf{P}_j^*)^\top \widehat{\varepsilon}\|^2 \le K \Delta_n (K_n + \log p), \tag{A.300}$$

with probability approaching one.

Consider a further decomposition based on Eq. (A.33):

$$\begin{aligned} \mathbf{Y}_i = {} & \underbrace{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \widetilde{\beta}_s^\top dX_s^c \right) \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}}_{\widetilde{\mathbf{Y}}_i} + \underbrace{\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c}_{\mathbf{U}_i^{(1)}} \\ & - \underbrace{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c \right) \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}}}_{\mathbf{U}_i^{(2)}} + \underbrace{\left(\sum_{(i-1)\Delta_n \le s \le i\Delta_n} (\beta_{s-}^J)^\top \Delta X_s \right) \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}}_{\mathbf{J}_i} \\ & + \underbrace{\Delta_i^n Z^c}_{\mathbf{Z}_i^{(1)}} - \underbrace{\Delta_i^n Z^c \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}}}_{\mathbf{Z}_i^{(2)}} + \underbrace{(\Delta_i^n Z - \Delta_i^n Z^c) \mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}}}_{\mathbf{Z}_i^{(3)}}. \end{aligned} \tag{A.301}$$

We denote the leading continuous-martingale term by $\Theta = \mathbf{U}^{(1)} + \mathbf{Z}^{(1)}$, and all remainders by $\Xi \in \{\widetilde{\mathbf{Y}} - \mathbf{R}_{A^\circ} \boldsymbol{\gamma}_{A^\circ}, \mathbf{U}^{(2)}, \mathbf{J}, \mathbf{Z}^{(2)}, \mathbf{Z}^{(3)}\}$. Then, it suffices to show

$$\max_{q+1 \le j \le p} \|(\mathbf{P}_j^*)^\top \mathbf{M}_{A^\circ} \Theta\|^2 \le K \Delta_n (K_n + \log p), \tag{A.302}$$

with probability approaching one, and, for each Ξ,

$$\max_{q+1 \le j \le p} \|(\mathbf{P}_j^*)^\top \mathbf{M}_{A^\circ} \Xi\|^2 = o_p(\Delta_n (K_n + \log p)). \tag{A.303}$$

We firstly verify Eq. (A.302). Since Θ contains only continuous semimartingale increments, $\|(\mathbf{P}_j^*)^\top \mathbf{M}_{A^\circ} \Theta\|^2$ can be written into a quadratic form of (conditional) Gaussian random variables, and Eq. (A.302) can be proved with the Hanson-Wright inequality (Hanson and Wright, 1971; Theorem 6.2.1, Vershynin, 2018). Specifically, we assume $b, \widetilde{b} \equiv 0$ without loss of generality as the drift only contributes a negligible order of Δ_n, and thus, conditional on a σ-field $\mathcal{G} = \sigma(c_s, \widetilde{\sigma}_s, e_s : 0 \le s \le T)$, Θ_i are independent Gaussian variables with mean 0 and variances

$$\mathrm{Var}(\Theta_i | \mathcal{G}) = \int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top c_s e_s ds + \int_{(i-1)\Delta_n}^{i\Delta_n} \widetilde{\sigma}_s^2 ds = \bar{v}_i \Delta_n, \tag{A.304}$$

for all $i = 1, \ldots, n$, with $0 \leq \bar{v}_i \leq \overline{K}$ for some $\overline{K} > 0$ under Assumption A.1. Therefore, we denote the conditional covariance matrix of Θ by

$$\mathbf{\Sigma} = \mathrm{Cov}(\Theta|\mathcal{G}) = \mathrm{diag}(\bar{v}_1\Delta_n, \ldots, \bar{v}_n\Delta_n), \tag{A.305}$$

and thus $\Theta = \mathbf{\Sigma}^{1/2}\Gamma$, where Γ is a $n \times 1$ vector of standard Gaussian random variables and is independent of the conditioning. Then, conditional on the same σ-field, $\|(\mathbf{P}_j^*)^\top \mathbf{M}_{A^\circ}\Theta\|^2$ can be written into a quadratic form of Gaussian random variables: With symmetric and idempotent $\mathbf{M}_{A^\circ}$,

$$\|(\mathbf{P}_j^*)^\top \mathbf{M}_{A^\circ}\Theta\|^2 = \Gamma^\top \mathbf{\Sigma}^{1/2}\mathbf{M}_{A^\circ}\mathbf{H}_j^*\mathbf{M}_{A^\circ}\mathbf{\Sigma}^{1/2}\Gamma = \Gamma^\top \mathbf{\Lambda}_j \Gamma, \quad \text{with} \quad \mathbf{\Lambda}_j = \mathbf{\Sigma}^{1/2}\mathbf{H}_j^*\mathbf{\Sigma}^{1/2}. \tag{A.306}$$

By definition of the block-wise orthogonal projection matrix, $\|\mathbf{H}_j^*\| = 1$ and $\mathrm{rank}(\mathbf{H}_j^*) = K_n$. By the submultiplicativity of operator norms,

$$\|\mathbf{\Lambda}_j\| \leq \|\mathbf{\Sigma}^{1/2}\|\|\mathbf{H}_j^*\|\|\mathbf{\Sigma}^{1/2}\| = \lambda_{\max}(\mathbf{\Sigma}) = \max_{1\leq i\leq n} \bar{v}_i\Delta_n \leq \overline{K}\Delta_n, \tag{A.307}$$

and for symmetric $\mathbf{\Lambda}_j$, with $\mathrm{rank}(\mathbf{\Lambda}_j) = \mathrm{rank}(\mathbf{H}_j^*) = K_n$,

$$\|\mathbf{\Lambda}_j\|_F^2 \leq \mathrm{rank}(\mathbf{\Lambda}_j)\|\mathbf{\Lambda}_j\|^2 \leq \overline{K}^2 K_n \Delta_n^2. \tag{A.308}$$

By the Hanson-Wright inequality, for any $t > 0$ and some constant $k > 0$,

$$\mathbb{P}(|\Gamma^\top \mathbf{\Lambda}_j \Gamma - \mathbb{E}[\Gamma^\top \mathbf{\Lambda}_j \Gamma]| > t \,|\, \mathcal{G}) \leq 2\exp\left(-k \min\left\{\frac{t^2}{\|\mathbf{\Lambda}_j\|_F^2}, \frac{t}{\|\mathbf{\Lambda}_j\|}\right\}\right), \tag{A.309}$$

where

$$\mathbb{E}[\Gamma^\top \mathbf{\Lambda}_j \Gamma] = \mathrm{tr}(\mathbf{\Lambda}_j) = \mathrm{tr}(\mathbf{\Sigma}^{1/2}\mathbf{H}_j^*\mathbf{\Sigma}^{1/2}) = \mathrm{tr}(\mathbf{H}_j^*\mathbf{\Sigma}) \leq \mathrm{tr}(\mathbf{H}_j^*)\|\mathbf{\Sigma}\| \leq K_n\|\mathbf{\Sigma}\| \leq \overline{K}K_n\Delta_n. \tag{A.310}$$

Let $t = \overline{K}\Delta_n(\sqrt{K_n c \log p} + c\log p)$ for some constant $c > 0$ satisfying $kc > 1$. Then,

$$\frac{t^2}{\|\mathbf{\Lambda}_j\|_F^2} \geq \frac{\overline{K}^2\Delta_n^2(\sqrt{K_n c\log p} + c\log p)^2}{\overline{K}^2 K_n \Delta_n^2} \geq \frac{(\sqrt{K_n c\log p} + c\log p)^2}{K_n} \geq c\log p, \tag{A.311}$$

$$\frac{t}{\|\mathbf{\Lambda}_j\|} \geq \frac{\overline{K}\Delta_n(\sqrt{K_n c\log p} + c\log p)}{\overline{K}\Delta_n} \geq c\log p. \tag{A.312}$$

Combining Eqs. (A.310) to (A.312) in Eq. (A.309) leads to

$$\mathbb{P}\left(\Gamma^\top \mathbf{\Lambda}_j \Gamma > \overline{K}\Delta_n(K_n + \sqrt{K_n c\log p} + c\log p)\middle|\, \mathcal{G}\right) \leq 2e^{-kc\log p} = 2p^{-kc}. \tag{A.313}$$

Applying a union bound over all irrelevant blocks, we obtain

$$\mathbb{P}\left(\max_{q+1\leq j\leq p} \Gamma^\top \mathbf{\Lambda}_j \Gamma > \overline{K}\Delta_n(K_n + \sqrt{K_n c\log p} + c\log p)\middle|\, \mathcal{G}\right) \leq 2p^{1-kc} \to 0, \tag{A.314}$$

when $p \to \infty$ and $1 - kc < 0$. Taking expectations on both sides implies the same union bound holds unconditionally. By Eq. (A.306), it holds that with probability approaching one,

$$\max_{q+1 \le j \le p} \|(\mathbf{P}_j^*)^\top \mathbf{M}_{A^\circ} \Theta\|^2 \le \max_{q+1 \le j \le p} \Gamma^\top \mathbf{\Lambda}_j \Gamma \le \overline{K} \Delta_n (K_n + \sqrt{K_n c \log p} + c \log p). \tag{A.315}$$

Furthermore, with $\sqrt{K_n \log p} \le (K_n + \log p)/2$,

$$\max_{q+1 \le j \le p} \|(\mathbf{P}_j^*)^\top \mathbf{M}_{A^\circ} \Theta\|^2 \le K \Delta_n (K_n + \log p), \tag{A.316}$$

with probability approaching one, i.e., Eq. (A.302) holds.

Next, we show Eq. (A.303) holds for each $\Xi \in \{\widetilde{\mathbf{Y}}, \mathbf{U}^{(2)}, \mathbf{J}, \mathbf{Z}^{(2)}, \mathbf{Z}^{(3)}\}$. By definition,

$$\|(\mathbf{P}_j^*)^\top \mathbf{M}_{A^\circ} \Xi\|^2 = \Xi^\top \mathbf{M}_{A^\circ} \mathbf{H}_j^* \mathbf{M}_{A^\circ} \Xi \le \|\mathbf{M}_{A^\circ} \Xi\|^2 \le \|\Xi\|^2. \tag{A.317}$$

Next, we show $\|\Xi\|^2 = o_p(\Delta_n (K_n + \log p))$ for each Ξ.

(i) $\boldsymbol{\Xi = \mathbf{Z}^{(2)}}$. By definition,

$$\|\mathbf{Z}^{(2)}\|^2 = \sum_{i=1}^n (\Delta_i^n Z^c)^2 \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}}. \tag{A.318}$$

By Hölder's and BDG inequalities, as well as Lemma A.2 (i), for any $q \ge 2$,

$$\begin{aligned} \mathbb{E}\left[(\Delta_i^n Z^c)^2 \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}}\right] &\le \left(\mathbb{E}[(\Delta_i^n Z^c)^{2q}]\right)^{1/q} \left(\mathbb{P}(|\Delta_i^n Y| > u_n)\right)^{1-1/q} \\ &\le K \Delta_n (\Delta_n u_n^{-r})^{1-1/q} = K \Delta_n^{2-1/q} u_n^{r/q-r}, \end{aligned} \tag{A.319}$$

and it is $O(\Delta_n u_n^{2-r})$ if we choose $q \ge (1 - \varpi r)/(1 - 2\varpi)$, which is always possible for $1/2(2-r) < \varpi < 1/2$ and $0 \le r < 1$. By Markov's inequality,

$$\|\mathbf{Z}^{(2)}\|^2 = O_p(u_n^{2-r}) = o_p(\Delta_n (K_n + \log p)), \tag{A.320}$$

when $\Delta_n^{1-\varpi(2-r)}(K_n + \log p) \to \infty$ as required.

(ii) $\boldsymbol{\Xi = \mathbf{Z}^{(3)}}$. By definition,

$$\|\mathbf{Z}^{(3)}\|^2 = \sum_{i=1}^n (\Delta_i^n Z - \Delta_i^n Z^c)^2 \left(\mathbb{1}_{\{|\Delta_i^n Z| \le u_n\}} + \left|\mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} - \mathbb{1}_{\{|\Delta_i^n Z| \le u_n\}}\right|\right), \tag{A.321}$$

where, by Lemma A.2 (ii) and Markov's inequality,

$$\sum_{i=1}^n (\Delta_i^n Z - \Delta_i^n Z^c)^2 \mathbb{1}_{\{|\Delta_i^n Z| \le u_n\}} = O_p(u_n^{2-r}), \tag{A.322}$$

and, following the same steps as in Appendix A.4 (v) for truncation mismatch,

$$\sum_{i=1}^{n}(\Delta_i^n Z - \Delta_i^n Z^c)^2 \left|\mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} - \mathbb{1}_{\{|\Delta_i^n Z| \le u_n\}}\right| = O_p(u_n^{2-r} \vee \Delta_n u_n^{-2r}). \tag{A.323}$$

Specifically, we denote $X^\circ = (X_{1,t}, \ldots, X_{q,t})_{t\ge 0}^\top$, and follow Eq. (A.209):

$$\begin{aligned}
&\sum_{i=1}^{n}(\Delta_i^n Z - \Delta_i^n Z^c)^2 \left|\mathbb{1}_{\{|\Delta_i^n Y| \le u_n\}} - \mathbb{1}_{\{|\Delta_i^n Z| \le u_n\}}\right| \\
&\quad \le \sum_{i=1}^{n}(u_n - \Delta_i^n Z^c)^2 \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}} + \sum_{i=1}^{n}(\Delta_i^n Z - \Delta_i^n Z^c)^2 \mathbb{1}_{\{|\Delta_i^n Y| \le u_n, |\Delta_i^n Z| > u_n, \|\Delta_i^n X^\circ\| \le u_n\}}.
\end{aligned} \tag{A.324}$$

Note that X° has only a finite dimension of nonzero components (relevant regressors), so the fixed dimensional result in Appendix A.4 (v) still holds. As in the previous proofs, the OLS regression in Eq. (12) already incorporates the truncation indicator $\mathbb{1}_{\{\|\Delta_i^n X^\circ\| \le u_n\}}$, so it does no harm to include it in any Ξ. For the first sum in Eq. (A.324),

$$\sum_{i=1}^{n}(u_n - \Delta_i^n Z^c)^2 \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}} \le 2\sum_{i=1}^{n} u_n^2 \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}} + 2\underbrace{\sum_{i=1}^{n}(\Delta_i^n Z^c)^2 \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}}}_{\|\mathbf{Z}^{(2)}\|^2}. \tag{A.325}$$

By Lemma A.2 (i) and Markov's inequality,

$$\sum_{i=1}^{n} u_n^2 \mathbb{P}(|\Delta_i^n Y| > u_n) = O(u_n^{2-r}) \quad \Rightarrow \quad \sum_{i=1}^{n} u_n^2 \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}} = O_p(u_n^{2-r}). \tag{A.326}$$

and, with the result for $\|\mathbf{Z}^{(2)}\|^2$ in Eq. (A.320), the first sum in Eq. (A.324) is $O_p(u_n^{2-r})$.

For the second sum, we follow the same steps as from Eqs. (A.212) to (A.215). On $\{|\Delta_i^n Y| \le u_n,\ |\Delta_i^n Z| > u_n,\ \|\Delta_i^n X^\circ\| \le u_n\}$, it holds that, by Eq. (A.213),

$$u_n^2 \le |\Delta_i^n Z|^2 \le (u_n + |\Delta_i^n Y^{(1)}| + |\widetilde{\mathbf{J}}_i|)^2 \le 3(u_n^2 + |\Delta_i^n Y^{(1)}|^2 + |\widetilde{\mathbf{J}}_i|^2), \tag{A.327}$$

and thus

$$|\Delta_i^n Z - \Delta_i^n Z^c|^2 \le K(u_n^2 + |\Delta_i^n Y^{(1)}|^2 + |\widetilde{\mathbf{J}}_i|^2 + |\Delta_i^n Z^c|^2). \tag{A.328}$$

By the same event decomposition as in Eq. (A.214),

$$
\begin{aligned}
&\mathbb{E}\left[|\Delta_i^n Z-\Delta_i^n Z^c|^2 \mathbb{1}_{\{|\Delta_i^n Y|\le u_n,\,|\Delta_i^n Z|>u_n,\,\|\Delta_i^n X^\circ\|\le u_n\}}\right]\\
&\quad\le \mathbb{E}\left[|\Delta_i^n Z-\Delta_i^n Z^c|^2 \mathbb{1}_{\{u_n<|\Delta_i^n Z|\le 3u_n\}}\right]+\mathbb{E}\left[|\Delta_i^n Z-\Delta_i^n Z^c|^2 \mathbb{1}_{\{|\Delta_i^n Y^{(1)}|+|\widetilde{\mathbf{J}}_i|>2u_n\}}\right]\\
&\quad\le 2\underbrace{(3u_n)^2\mathbb{P}(|\Delta_i^n Z|>u_n)}_{(1)}+2\underbrace{\mathbb{E}\left[|\Delta_i^n Z^c|\mathbb{1}_{\{|\Delta_i^n Z|>u_n\}}\right]}_{(2)}\\
&\qquad+\mathbb{E}\left[|\Delta_i^n Z-\Delta_i^n Z^c|^2 \mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right]+\mathbb{E}\left[|\Delta_i^n Z-\Delta_i^n Z^c|^2 \mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right]\\
&\quad\le (1)+(2)\;+\;K\underbrace{u_n^2\mathbb{P}\left(|\Delta_i^n Y^{(1)}|>u_n\right)}_{(3)}+K\underbrace{\mathbb{E}\left[|\Delta_i^n Y^{(1)}|^2\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right]}_{(4)}\\
&\qquad+K\underbrace{\mathbb{E}\left[|\widetilde{\mathbf{J}}_i|^2\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right]}_{(5)}+K\underbrace{\mathbb{E}\left[|\Delta_i^n Z^c|^2\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right]}_{(6)}+K\underbrace{u_n^2\mathbb{P}\left(|\widetilde{\mathbf{J}}_i|>u_n\right)}_{(7)}\\
&\qquad+K\underbrace{\mathbb{E}\left[|\Delta_i^n Y^{(1)}|^2\mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right]}_{(8)}+K\underbrace{\mathbb{E}\left[|\widetilde{\mathbf{J}}_i|^2\mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right]}_{(9)}+K\underbrace{\mathbb{E}\left[|\Delta_i^n Z^c|^2\mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right]}_{(10)}.
\end{aligned}
\tag{A.329}
$$

Next, we verify that the above terms (1)–(10) are all $O(\Delta_n u_n^{2-r}\vee\Delta_n^2 u_n^{-2r})$.

Term (1): By the same result as in Lemma A.2 (i) for the one-dimensional Itô semimartingale Z,

$$u_n^2\mathbb{P}(|\Delta_i^n Z|>u_n)\le K\Delta_n u_n^{2-r}. \tag{A.330}$$

Term (2): Following the same steps as in Eq. (A.319),

$$\mathbb{E}\left[(\Delta_i^n Z^c)^2\mathbb{1}_{\{|\Delta_i^n Z|>u_n\}}\right]=O(\Delta_n u_n^{2-r}). \tag{A.331}$$

Term (3): By Eq. (A.54),

$$u_n^2\mathbb{P}\left(|\Delta_i^n Y^{(1)}|>u_n\right)\le K\Delta_n u_n^{2-r}. \tag{A.332}$$

Term (4): By the same steps as in Eq. (A.67) and the BDG inequality, for any $q\ge 2$,

$$\mathbb{E}\left[|\Delta_i^n Y^{(1)}|^2\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right]\le\mathbb{E}\left[|\Delta_i^n Y^{(1)}|^2\left(\frac{|\Delta_i^n Y^{(1)}|}{u_n}\right)^{q-2}\right]\le\frac{\mathbb{E}[|\Delta_i^n Y^{(1)}|^q]}{u_n^{q-2}}\le K\Delta_n^{q/2}u_n^{2-q}, \tag{A.333}$$

and it is $O(\Delta_n u_n^{2-r})$ if $q\ge 2(1-\varpi r)/(1-2\varpi)$, which is always possible for $1/2(2-r)<\varpi<1/2$ and $0\le r<1$.

Term (5): By Eq. (A.96),

$$\mathbb{E}\left[|\widetilde{\mathbf{J}}_i|^2\mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right]\le\mathbb{E}[|\widetilde{\mathbf{J}}_i|^2]\le K\Delta_n u_n^{2-r}+K'\Delta_n^2 u_n^{-2r}. \tag{A.334}$$

Term (6): Following the same steps as in Eq. (A.319), together with Eq. (A.54),

$$\mathbb{E}\left[|\Delta_i^n Z^c|^2 \mathbb{1}_{\{|\Delta_i^n Y^{(1)}|>u_n\}}\right] = O(\Delta_n u_n^{2-r}). \tag{A.335}$$

Term (7): By Eq. (A.221),

$$u_n^2 \mathbb{P}\left(|\widetilde{\mathbf{J}}_i| > u_n\right) \le K\Delta_n u_n^{2-r}. \tag{A.336}$$

Term (8): By Hölder's and BDG inequalities, together with Eq. (A.221), for some $q \ge 2$,

$$\begin{aligned}\mathbb{E}\left[|\Delta_i^n Y^{(1)}|^2 \mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right] &\le \left(\mathbb{E}[|\Delta_i^n Y^{(1)}|^{2q}]\right)^{1/q} \left(\mathbb{P}\left(|\widetilde{\mathbf{J}}_i| > u_n\right)\right)^{1-1/q} \\ &\le K\Delta_n(\Delta_n u_n^{-r})^{1-1/q} = K\Delta_n^{2-1/q} u_n^{r/q-r},\end{aligned} \tag{A.337}$$

and it is $O(\Delta_n u_n^{2-r})$ if we choose $q \ge (1-\varpi r)/(1-2\varpi)$, which is always possible for $1/2(2-r) < \varpi < 1/2$ and $0 \le r < 1$.

Term (9): Same as term (5), we have

$$\mathbb{E}\left[|\widetilde{\mathbf{J}}_i|^2 \mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right] \le \mathbb{E}[|\widetilde{\mathbf{J}}_i|^2] \le K\Delta_n u_n^{2-r} + K'\Delta_n^2 u_n^{-2r}. \tag{A.338}$$

Term (10): Following the same steps as in Eq. (A.319), together with Eq. (A.221),

$$\mathbb{E}\left[|\Delta_i^n Z^c|^2 \mathbb{1}_{\{|\widetilde{\mathbf{J}}_i|>u_n\}}\right] = O(\Delta_n u_n^{2-r}). \tag{A.339}$$

Combining all terms (1)–(10) in Eq. (A.329) leads to

$$\mathbb{E}\left[|\Delta_i^n Z - \Delta_i^n Z^c|^2 \mathbb{1}_{\{|\Delta_i^n Y|\le u_n,\, |\Delta_i^n Z|>u_n,\, \|\Delta_i^n X^\circ\|\le u_n\}}\right] = O(\Delta_n u_n^{2-r} \vee \Delta_n^2 u_n^{-2r}), \tag{A.340}$$

and thus Eq. (A.323) holds. Combining Eqs. (A.322) and (A.323) in Eqs. (A.321) and (A.317), we obtain

$$\|(\mathbf{P}_j^*)^\top \mathbf{M}_{A^\circ} \mathbf{Z}^{(3)}\|^2 = O_p(u_n^{2-r} \vee \Delta_n u_n^{-2r}) = o_p(\Delta_n(K_n + \log p)), \tag{A.341}$$

when $\Delta_n^{(1-\varpi(2-r))\vee 2\varpi r}(K_n + \log p) \to \infty$ as required.

(iii) $\boldsymbol{\Xi = \widetilde{\mathbf{Y}} - \mathbf{R}_{A^\circ}\gamma_{A^\circ}}$. Consider the same decomposition as in Appendix A.3 (i), i.e., Eqs. (A.71) and (A.72),

$$\widetilde{\mathbf{Y}} - \mathbf{R}_{A^\circ}\boldsymbol{\gamma}_{A^\circ} = (\widetilde{\mathbf{Y}}^{(1)} - \mathbf{R}_{A^\circ}\boldsymbol{\gamma}_{A^\circ}) + \widetilde{\mathbf{Y}}^{(2)}. \tag{A.342}$$

Since the oracle index set A° has fixed cardinality q, the arguments used in the proofs of Theorems 1 and 2 apply verbatim to the restricted oracle regression. By Eqs. (A.74) to (A.76) and Markov's inequality,

$$\|\widetilde{\mathbf{Y}}^{(1)} - \mathbf{R}_{A^\circ}\boldsymbol{\gamma}_{A^\circ}\|^2 = O_p(\Delta_n u_n^{-r} \vee u_n^{2-r}) = O_p(u_n^{2-r}). \tag{A.343}$$

and, by Eq. (A.79),

$$\|\widetilde{\mathbf{Y}}^{(2)}\|^2 = O_p(\Delta_n^2 K_n^2). \tag{A.344}$$

Combining Eqs. (A.343) and (A.344) in Eq. (A.342) leads to

$$\|\widetilde{\mathbf{Y}} - \mathbf{R}_{A^\circ}\boldsymbol{\gamma}_{A^\circ}\|^2 = O_p(u_n^{2-r} \vee \Delta_n^2 K_n^2) = o_p(\Delta_n(K_n + \log p)), \tag{A.345}$$

when $\Delta_n^{1-\varpi(2-r)}(K_n + \log p) \to \infty$, and since $\Delta_n K_n \to 0$,

$$\Delta_n^2 K_n^2 = o(\Delta_n K_n) = o(\Delta_n(K_n + \log p)). \tag{A.346}$$

(iv) $\boldsymbol{\Xi} = \mathbf{U^{(2)}}$. By Hölder's and BDG inequalities, as well as Lemma A.2 (i), for any $q \geq 2$,

$$\begin{aligned}
\mathbb{E}[\mathbf{U}_i^2] &= \mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c\right)^2 \mathbb{1}_{\{|\Delta_i^n Y| > u_n\}}\right] \\
&\leq \left(\mathbb{E}\left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e_s^\top dX_s^c\right)^{2q}\right]\right)^{1/q} (\mathbb{P}(|\Delta_i^n Y| > u_n))^{1-1/q} \\
&\leq K\Delta_n(\Delta_n u_n^{-r})^{1-1/q} = K\Delta_n^{2-1/q} u_n^{r/q-r},
\end{aligned} \tag{A.347}$$

and it is $O(\Delta_n u_n^{2-r})$ if we choose $q \geq (1-\varpi r)/(1-2\varpi)$, which is always possible for $1/2(2-r) < \varpi < 1/2$ and $0 \leq r < 1$. Therefore,

$$\|\mathbf{U}\|^2 = O_p(u_n^{2-r}) = o_p(\Delta_n(K_n + \log p)), \tag{A.348}$$

when $\Delta_n^{1-\varpi(2-r)}(K_n + \log p) \to \infty$.

(v) $\boldsymbol{\Xi} = \mathbf{J}$. Same as in (ii), we employ the safe replacement $\widetilde{\mathbf{J}}$ with the additional truncation indicator $\mathbb{1}_{\{\|\Delta_i^n X^\circ\| \leq u_n\}}$. By Eq. (A.97) and Markov's inequality,

$$\|\widetilde{\mathbf{J}}\|^2 = O_p(u_n^{2-r} \vee \Delta_n u_n^{-2r}) = o_p(\Delta_n(K_n + \log p)), \tag{A.349}$$

when $\Delta_n^{(1-\varpi(2-r))\vee 2\varpi r}(K_n + \log p) \to \infty$.

Combining results in (i)–(v) in Eq. (A.317) completes the proof of Lemma A.5.

□

For any index set A, we define the total loss reduction from adding all blocks in $A \backslash A^\circ$ as

$$\Delta Q_n(A) = Q_n(\widehat{\boldsymbol{\gamma}}^\circ) - \min_{\boldsymbol{\gamma}: A(\boldsymbol{\gamma}) = A} Q_n(\boldsymbol{\gamma}). \tag{A.350}$$

Lemma A.6. Under the conditions of Theorem 5, it holds that for any $A \in \Gamma_1$,

$$\Delta Q_n(A) \leq (|A| - q) \max_{q+1 \leq j \leq p} \Delta Q_{n,j} \tag{A.351}$$

Proof of Lemma A.6. We denote by $Q_n(A) = \min_{\gamma: A(\gamma)=A} Q_n(\gamma)$, where $Q_n(A^\circ) = Q_n(\widehat{\gamma}^\circ)$. Fix $A \in \Gamma_1$, and enumerate the redundant groups as $A \backslash A^\circ = \{j_1, \dots, j_{|A|-q}\}$. Define the nested sequence of index sets

$$\overline{A}_0 = A^\circ, \qquad \overline{A}_k = A^\circ \cup \{j_1, \dots, j_k\}, \qquad \text{where } k = 1, \dots, |A| - q. \tag{A.352}$$

The total improvement from A° to A can be written as a sum of incremental improvements:

$$\Delta Q_n(A) = \mathbf{Y}^\top (\mathbf{M}_{A^\circ} - \mathbf{M}_A)\mathbf{Y} = \sum_{k=1}^{|A|-q} \mathbf{Y}^\top (\mathbf{M}_{\overline{A}_{k-1}} - \mathbf{M}_{\overline{A}_k})\mathbf{Y}. \tag{A.353}$$

We define the residualized block $\breve{\mathbf{R}}^*_{j_k} = \mathbf{M}_{\overline{A}_{k-1}} \mathbf{R}^*_{j_k} \in \mathbb{R}^{n \times K_n}$, the orthogonal projection matrix $\breve{\mathbf{H}}^*_{j_k} = \breve{\mathbf{R}}^*_{j_k}((\breve{\mathbf{R}}^*_{j_k})^\top \breve{\mathbf{R}}^*_{j_k})^{-1}(\breve{\mathbf{R}}^*_{j_k})^\top \in \mathbb{R}^{n \times n}$. We follow the same steps as in Eqs. (A.298) and (A.299), by the Frisch-Waugh-Lovell theorem,

$$\mathbf{Y}^\top (\mathbf{M}_{\overline{A}_{k-1}} - \mathbf{M}_{\overline{A}_k})\mathbf{Y} = (\mathbf{M}_{\overline{A}_{k-1}}\mathbf{Y})^\top \breve{\mathbf{H}}^*_{j_k} (\mathbf{M}_{\overline{A}_{k-1}}\mathbf{Y}) = \mathbf{Y}^\top \breve{\mathbf{H}}^*_{j_k} \mathbf{Y}. \tag{A.354}$$

Since $A^\circ \subseteq \overline{A}_{k-1}$ for all $k = 1, \dots, |A| - q$, $\operatorname{span}(\breve{\mathbf{R}}^*_{j_k}) = \operatorname{span}(\mathbf{M}_{\overline{A}_{k-1}} \mathbf{R}^*_{j_k}) \subseteq \operatorname{span}(\mathbf{M}_{A^\circ} \mathbf{R}^*_{j_k})$, and thus the projector $\breve{\mathbf{H}}^*_{j_k}$ is onto a smaller subspace than $\mathbf{H}^*_{j_k}$, so that

$$\mathbf{Y}^\top \breve{\mathbf{H}}^*_{j_k} \mathbf{Y} \leq \mathbf{Y}^\top \mathbf{H}^*_{j_k} \mathbf{Y}, \qquad \text{where} \quad \mathbf{H}^*_{j_k} = (\mathbf{M}_{A^\circ}\mathbf{R}^*_{j_k})((\mathbf{M}_{A^\circ}\mathbf{R}^*_{j_k})^\top \mathbf{M}_{A^\circ}\mathbf{R}^*_{j_k})^{-1} \mathbf{M}_{A^\circ}\mathbf{R}^*_{j_k}. \tag{A.355}$$

Therefore, by the definition of $\Delta Q_{n,j}$ in Eq. (A.296) and the same steps as in Eq. (A.354),

$$\Delta Q_n(A) \leq \sum_{k=1}^{|A|-q} \mathbf{Y}^\top \mathbf{H}^*_{j_k} \mathbf{Y} \leq (|A| - q) \max_{q+1 \leq j \leq p} \Delta Q_{n,j}. \tag{A.356}$$

This completes the proof of Lemma A.6.

□

By Lemmas A.5 and A.6, it holds for any $A \in \Gamma_1$ on the event E_n in Eq. (A.297) that

$$\Delta Q_n(A) \leq K(|A| - q)\Delta_n(K_n + \log p). \tag{A.357}$$

By Eqs. (A.295) and (A.357), on E_n,

$$\min_{\gamma: A(\gamma)=A} Q^*_n(\gamma) - Q^*_n(\widehat{\gamma}^\circ) \geq (|A| - q)\left(-\frac{K}{2}\Delta_n(K_n + \log p) + \lambda_n\right) > 0, \tag{A.358}$$

since $|A| - q \geq 1$ and $\lambda_n \gg \Delta_n(K_n + \log p)$ as required. Consequently, by Lemma A.5,

$$\mathbb{P}\left(\exists A \in \Gamma_1 : \min_{\boldsymbol{\gamma}: A(\boldsymbol{\gamma})=A} Q_n^*(\boldsymbol{\gamma}) - Q_n^*(\widehat{\boldsymbol{\gamma}}^\circ) \leq 0\right) \leq \mathbb{P}(E_n^\complement) \to 0. \tag{A.359}$$

Underfitting models. For any underfitting models with $A \in \Gamma_2$, the difference between unpenalized losses in Eq. (A.295) satisfies

$$Q_n(\widehat{\boldsymbol{\gamma}}^\circ) - \min_{\boldsymbol{\gamma}: A(\boldsymbol{\gamma})=A} Q_n(\boldsymbol{\gamma}) = \mathbf{Y}^\top(\mathbf{M}_{A^\circ} - \mathbf{M}_A)\mathbf{Y}, \tag{A.360}$$

where the right-hand side corresponds to the decrease of OLS residual sums of squares from adding missed relevant regressors in $A'' = A^\circ \backslash A \neq \varnothing$ to A.

Consider the decomposition of the design matrix $\mathbf{R} = (\mathbf{R}_A^\top, \mathbf{R}_{A''}^\top, \mathbf{R}_{\{1,\ldots,p\}\backslash A^\circ}^\top)^\top$ and the oracle estimator $\widehat{\boldsymbol{\gamma}}^\circ = ((\widehat{\boldsymbol{\gamma}}_A^\circ)^\top, (\widehat{\boldsymbol{\gamma}}_{A''}^\circ)^\top, 0^\top)^\top$, and thus

$$\mathbf{M}_A\mathbf{Y} = \mathbf{M}_A(\mathbf{R}_A\widehat{\boldsymbol{\gamma}}_A^\circ + \mathbf{R}_{A''}\widehat{\boldsymbol{\gamma}}_{A''}^\circ + \widehat{\boldsymbol{\varepsilon}}) = \mathbf{M}_A\mathbf{R}_{A''}\widehat{\boldsymbol{\gamma}}_{A''}^\circ + \widehat{\boldsymbol{\varepsilon}}, \tag{A.361}$$

because $\mathbf{M}_A\mathbf{R}_A = 0$ and $\mathbf{M}_A\widehat{\boldsymbol{\varepsilon}} = \widehat{\boldsymbol{\varepsilon}}$ by definition. With similar applications of the Frisch-Waugh-Lovell theorem as in Eq. (A.298), it holds that

$$\begin{aligned}
\mathbf{Y}^\top(\mathbf{M}_A - \mathbf{M}_{A^\circ})\mathbf{Y} &= \mathbf{Y}^\top(\mathbf{M}_A - \mathbf{M}_{A\cup A''})\mathbf{Y} \\
&= (\mathbf{M}_A\mathbf{Y})^\top\mathbf{M}_A\mathbf{R}_{A''}((\mathbf{M}_A\mathbf{R}_{A''})^\top\mathbf{M}_A\mathbf{R}_{A''})^{-1}(\mathbf{M}_A\mathbf{R}_{A''})^\top\mathbf{M}_A\mathbf{Y} \\
&= (\widehat{\boldsymbol{\gamma}}_{A''}^\circ)^\top\mathbf{R}_{A''}^\top\mathbf{M}_A\mathbf{R}_{A''}\widehat{\boldsymbol{\gamma}}_{A''}^\circ + 2(\widehat{\boldsymbol{\gamma}}_{A''}^\circ)^\top\mathbf{R}_{A''}^\top\widehat{\boldsymbol{\varepsilon}} + \widehat{\boldsymbol{\varepsilon}}^\top\mathbf{R}_{A''}(\mathbf{R}_{A''}^\top\mathbf{M}_A\mathbf{R}_{A''})^{-1}\mathbf{R}_{A''}^\top\widehat{\boldsymbol{\varepsilon}} \\
&\geq (\widehat{\boldsymbol{\gamma}}_{A''}^\circ)^\top\mathbf{R}_{A''}^\top\mathbf{M}_A\mathbf{R}_{A''}\widehat{\boldsymbol{\gamma}}_{A''}^\circ + 2(\widehat{\boldsymbol{\gamma}}_{A''}^\circ)^\top\mathbf{R}_{A''}^\top\widehat{\boldsymbol{\varepsilon}},
\end{aligned} \tag{A.362}$$

where the nonnegative quadratic term of $\widehat{\boldsymbol{\varepsilon}}$ is dropped for a lower bound. Since $\widehat{\boldsymbol{\varepsilon}}$ is orthogonal to the span of $\mathbf{R}_{A^\circ}$, we have $\mathbf{R}_{A''}^\top\widehat{\boldsymbol{\varepsilon}} = \mathbf{0}$ and thus the cross term vanishes. Therefore, it holds that

$$\mathbf{Y}^\top(\mathbf{M}_A - \mathbf{M}_{A^\circ})\mathbf{Y} \geq (\widehat{\boldsymbol{\gamma}}_{A''}^\circ)^\top\mathbf{R}_{A''}^\top\mathbf{M}_A\mathbf{R}_{A''}\widehat{\boldsymbol{\gamma}}_{A''}^\circ \geq K\sum_{j\in A''}(\widehat{\gamma}_j^\circ)^\top\mathbf{W}_j\widehat{\gamma}_j^\circ \geq K(q - |A|)\min_{j\in A''}(\widehat{\gamma}_j^\circ)^\top\mathbf{W}_j\widehat{\gamma}_j^\circ. \tag{A.363}$$

By Eqs. (A.258) and (A.259), $\sqrt{(\widehat{\gamma}_j^\circ)^\top\mathbf{W}_j\widehat{\gamma}_j^\circ} \xrightarrow{\mathbb{P}} \|\beta_j\|_{L^2} > 0$ for all $j \in A^\circ$. Therefore, for some $b_{\min} = K\min_{j\in A^\circ}\|\beta_j\|_{L^2}^2 > 0$, it holds that

$$\mathbb{P}\left(\min_{j\in A^\circ}(\widehat{\gamma}_j^\circ)^\top\mathbf{W}_j\widehat{\gamma}_j^\circ \geq b_{\min}\right) \to 1. \tag{A.364}$$

By Eqs. (A.295) and (A.363),

$$\min_{\boldsymbol{\gamma}: A(\boldsymbol{\gamma})=A} Q_n^*(\boldsymbol{\gamma}) - Q_n^*(\widehat{\boldsymbol{\gamma}}^\circ) \geq \frac{K}{2}|A''|\min_{j\in A''}(\widehat{\gamma}_j^\circ)^\top\mathbf{W}_j\widehat{\gamma}_j^\circ - \lambda_n|A''|. \tag{A.365}$$

With the union bound in Eq. (A.364) and $\lambda_n \to 0$, it holds that

$$\mathbb{P}\left(\exists A \in \Gamma_2 : \min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n^*(\boldsymbol{\gamma}) - Q_n^*(\widehat{\boldsymbol{\gamma}}^\circ) \leq 0\right) \leq 1 - \mathbb{P}\left(\min_{j\in A^\circ} (\widehat{\gamma}_j^\circ)^\top \mathbf{W}_j \widehat{\gamma}_j^\circ \geq b_{\min}\right) \to 0. \tag{A.366}$$

Misspecified models. For any misspecified models with $A \in \Gamma_3$, we denote by $A' = A \cap A^\circ$ the index set of included relevant regressors, by $A'' = A^\circ \backslash A \neq \varnothing$ the index set of missed relevant regressors, and by $A''' = A \backslash A^\circ \neq \varnothing$ the index set of spurious regressors, and thus $A^\circ = A' \cup A''$ and $A = A' \cup A'''$. It holds that

$$\begin{aligned} Q_n(\widehat{\boldsymbol{\gamma}}^\circ) - \min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n(\boldsymbol{\gamma}) &= \mathbf{Y}^\top(\mathbf{M}_{A^\circ} - \mathbf{M}_A)\mathbf{Y} \\ &= \mathbf{Y}^\top(\mathbf{M}_{A'\cup A''} - \mathbf{M}_{A'})\mathbf{Y} - \mathbf{Y}^\top(\mathbf{M}_{A'\cup A'''} - \mathbf{M}_{A'})\mathbf{Y} \\ &= \mathbf{Y}^\top(\mathbf{M}_{A'\cup A''} - \mathbf{M}_{A'})\mathbf{Y} + \mathbf{Y}^\top(\mathbf{M}_{A'} - \mathbf{M}_{A'\cup A'''})\mathbf{Y}, \end{aligned} \tag{A.367}$$

where the first term is the decrease of OLS residual sums of squares from adding missed relevant regressors, and the second term is the increase from excluding spurious regressors. For the second term, we have

$$\begin{aligned} \mathbf{Y}^\top(\mathbf{M}_{A'} - \mathbf{M}_{A'\cup A'''})\mathbf{Y} &= \mathbf{Y}^\top(\mathbf{M}_{A'} - \mathbf{M}_{A'\cup A''} + \mathbf{M}_{A'\cup A''} - \mathbf{M}_{A'\cup A''\cup A'''} + \mathbf{M}_{A'\cup A''\cup A'''} - \mathbf{M}_{A'\cup A'''})\mathbf{Y} \\ &\leq \mathbf{Y}^\top(\mathbf{M}_{A'} - \mathbf{M}_{A'\cup A''})\mathbf{Y} + \mathbf{Y}^\top(\mathbf{M}_{A'\cup A''} - \mathbf{M}_{A'\cup A''\cup A'''})\mathbf{Y}, \end{aligned} \tag{A.368}$$

since $\mathbf{Y}^\top(\mathbf{M}_{A'\cup A''\cup A'''} - \mathbf{M}_{A'\cup A'''})\mathbf{Y} \leq 0$, and therefore

$$Q_n(\widehat{\boldsymbol{\gamma}}^\circ) - \min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n(\boldsymbol{\gamma}) \leq \mathbf{Y}^\top(\mathbf{M}_{A'\cup A''} - \mathbf{M}_{A'\cup A''\cup A'''})\mathbf{Y} = \mathbf{Y}^\top(\mathbf{M}_{A^\circ} - \mathbf{M}_{A^\circ\cup A'''})\mathbf{Y}. \tag{A.369}$$

By Lemmas A.5 and A.6, it holds that, on the event E_n,

$$Q_n(\widehat{\boldsymbol{\gamma}}^\circ) - \min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n(\boldsymbol{\gamma}) \leq K|A'''|\Delta_n(K_n + \log p). \tag{A.370}$$

Bring it back to Eq. (A.295), on the event E_n,

$$\begin{aligned} \min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n^*(\boldsymbol{\gamma}) - Q_n^*(\widehat{\boldsymbol{\gamma}}^\circ) &\geq -\frac{K}{2}|A'''|\Delta_n(K_n + \log p) + \lambda_n(|A'| + |A'''| - q) \\ &\geq |A'''|\left(-\frac{K}{2}\Delta_n(K_n + \log p) + \lambda_n\right) - \lambda_n|A''| > 0, \end{aligned} \tag{A.371}$$

since $|A'''| \geq 1$, $\lambda_n \gg \Delta_n(K_n + \log p)$, and $\lambda_n \to 0$. Consequently, by Lemma A.5,

$$\mathbb{P}\left(\exists A \in \Gamma_3 : \min_{\boldsymbol{\gamma}:A(\boldsymbol{\gamma})=A} Q_n^*(\boldsymbol{\gamma}) - Q_n^*(\widehat{\boldsymbol{\gamma}}^\circ) \leq 0\right) \leq \mathbb{P}(E_n^\complement) \to 0. \tag{A.372}$$

In conclusion, Eqs. (A.359), (A.366) and (A.372) are sufficient for Eq. (A.293). This completes the proof.

Appendix B Supplementary Materials

B.1 Supplementary Simulation Results

In addition to the simulation results in Section 4 (Tables 1 and 2), this appendix reports supplementary Monte Carlo evidence on the sensitivity of the proposed estimation procedure. Unless stated otherwise, the data-generating process and tuning choices follow the baseline design in Section 4.1, and results are reported in Table B.1–B.4.

Alternative truncation threshold. Tables B.1 and B.2 report the model estimation and selection results, respectively, with a less aggressive truncation threshold $4\Delta_n^{0.47}\sqrt{\mathrm{MedRV}_T}$ for the increments of Y and each component of X, with the same simulation design used in Section 4.

Imperfect multicollinearity. To introduce imperfect multicollinearity while preserving the baseline dependence among the relevant factors, we construct the $p \times p$ correlation matrix in block form. We fix the relevant block to have the same Toeplitz structure as in Section 4.1, i.e.,

$$\boldsymbol{\rho}_{11} = \left(0.15^{|i-j|}\right)_{1\leq i,j\leq q}. \tag{B.1}$$

We then generate the redundant-factor block $\boldsymbol{\rho}_{22}$ from a simple latent-factor construction, which is positive definite by construction and can exhibit relatively strong within-block dependence. Specifically, we draw a loading matrix $\mathbf{L} \in \mathbb{R}^{(p-q)\times k}$, where k is the latent-factor dimension, and set

$$\boldsymbol{\rho}_{22} = \mathbf{L}\mathbf{L}^\top + \mathrm{diag}(1 - \|\mathbf{L}_{i\cdot}\|^2), \tag{B.2}$$

where $\|\mathbf{L}_{i\cdot}\|$ is the Euclidean norm of the i-th row of $\mathbf{L}$. Next, we draw a candidate cross-correlation block $\boldsymbol{\rho}_{12}^* \in \mathbb{R}^{q\times(p-q)}$ with entries i.i.d. $U[-1,1]$. Since an arbitrary choice of $\boldsymbol{\rho}_{12}$ may violate positive definiteness of the full matrix, we scale the cross-block by a single factor and set $\boldsymbol{\rho}_{12} = \kappa\boldsymbol{\rho}_{12}^*$, where $\kappa \in (0,1]$ is chosen as large as possible subject to the Schur-complement condition, i.e.,

$$\boldsymbol{\rho}_{22} - \kappa^2(\boldsymbol{\rho}_{12}^*)^\top\boldsymbol{\rho}_{11}^{-1}\boldsymbol{\rho}_{12}^* \tag{B.3}$$

is positive definite. The final correlation matrix

$$\boldsymbol{\rho} = \begin{pmatrix} \boldsymbol{\rho}_{11} & \boldsymbol{\rho}_{12} \\ \boldsymbol{\rho}_{12}^\top & \boldsymbol{\rho}_{22} \end{pmatrix} \tag{B.4}$$

is symmetric with unit diagonal and positive definite. Hence, the relevant-factor dependence remains comparable to the baseline design in the main text, while it allows stronger dependence among the redundant factors and nontrivial correlations between relevant and redundant factors. Fig. B.1 plots the histogram of all off-diagonal entries of $\boldsymbol{\rho}$ for the case $p = 500$. The model estimation and selection results are reported in Tables B.3 and B.4, respectively.

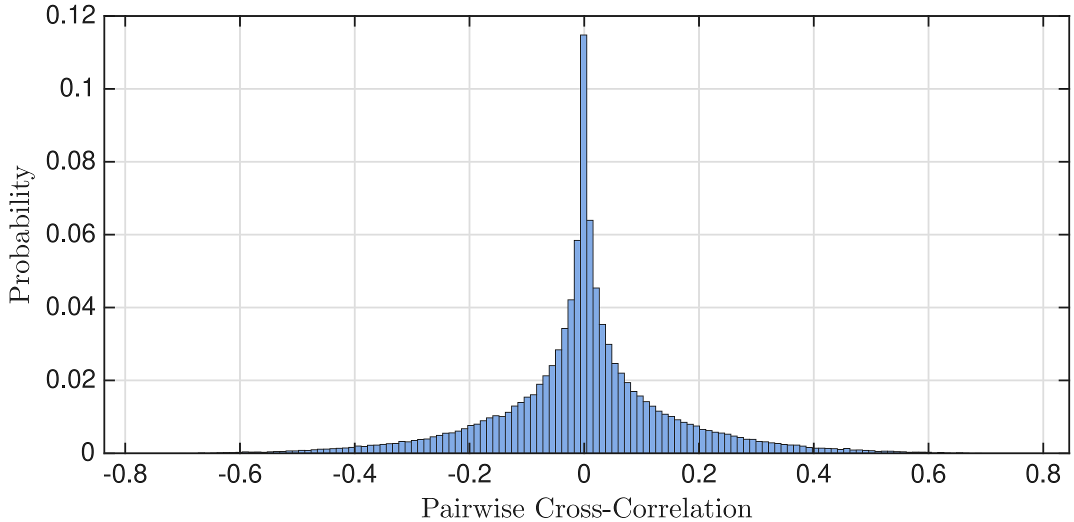

Figure B.1: Histogram of all off-diagonal entries of the correlation matrix $\boldsymbol{\rho}$ for the case $p = 500$.

Table B.1: Simulation results for model estimation with alternative truncation threshold

Estimator	Tuning	$\widehat{I\beta}_{1,T}$ Bias	Stdev	RMSE	$\widehat{I\beta}_{2,T}$ Bias	Stdev	RMSE	$\widehat{I\beta}_{3,T}$ Bias	Stdev	RMSE
Panel A: $p = 3$										
Spline		0.023	2.793	2.792	-0.024	2.537	2.536	-0.022	2.331	2.330
Spline TLP	$\alpha_\tau = 0.05$	0.023	2.793	2.792	-0.024	2.537	2.536	-0.022	2.331	2.330
Spline TLP	$\alpha_\tau = 0.01$	0.023	2.793	2.792	-0.025	2.537	2.536	-0.021	2.331	2.330
AKX	$k_n = 78$	0.000	2.871	2.869	0.000	2.632	2.631	-0.010	2.451	2.450
AKX	$k_n = 91$	-0.018	2.887	2.885	-0.021	2.621	2.619	-0.027	2.475	2.474
AKX	$k_n = 117$	-0.048	2.894	2.893	-0.003	2.640	2.638	0.003	2.495	2.494
Panel B: $p = 10$										
Spline		-0.144	3.032	3.034	-0.026	2.806	2.805	-0.095	2.507	2.508
Spline TLP	$\alpha_\tau = 0.05$	-0.152	3.004	3.006	-0.024	2.776	2.774	-0.085	2.450	2.451
Spline TLP	$\alpha_\tau = 0.01$	-0.152	3.004	3.006	-0.024	2.775	2.774	-0.085	2.451	2.451
AKX	$k_n = 78$	-0.122	3.325	3.326	-0.003	3.022	3.020	-0.165	2.771	2.775
AKX	$k_n = 91$	-0.187	3.221	3.225	0.018	3.102	3.101	-0.094	2.663	2.663
AKX	$k_n = 117$	-0.123	3.233	3.234	-0.004	3.088	3.086	-0.123	2.688	2.690
Panel C: $p = 50$										
Spline		-0.184	2.898	2.902	-0.019	3.158	3.156	-0.070	2.941	2.940
Spline TLP	$\alpha_\tau = 0.05$	-0.199	2.667	2.673	0.014	3.066	3.065	-0.182	2.993	2.997
Spline TLP	$\alpha_\tau = 0.01$	-0.201	2.669	2.675	-0.004	3.053	3.051	-0.092	2.647	2.647
AKX	$k_n = 78$	-0.151	5.135	5.135	0.009	5.095	5.092	-0.314	4.887	4.895
AKX	$k_n = 91$	-0.046	4.671	4.669	-0.030	4.691	4.689	-0.263	4.235	4.241
AKX	$k_n = 117$	-0.127	3.894	3.895	-0.042	4.140	4.138	-0.124	3.866	3.866
Panel D: $p = 100$										
Spline		-0.108	3.351	3.351	0.261	2.964	2.974	-0.241	3.487	3.494
Spline TLP	$\alpha_\tau = 0.05$	-0.082	2.933	2.932	0.280	2.601	2.615	-0.342	3.380	3.396
Spline TLP	$\alpha_\tau = 0.01$	-0.085	2.928	2.928	0.270	2.597	2.610	-0.282	3.348	3.358
AKX	$k_n = 78$	–	–	–	–	–	–	–	–	–
AKX	$k_n = 91$	–	–	–	–	–	–	–	–	–
AKX	$k_n = 117$	-0.112	8.294	8.291	0.530	7.133	7.149	-0.346	9.036	9.038
Panel E: $p = 500$										
Spline		-45.012	1.572	45.039	34.732	1.564	34.768	-20.000	1.538	20.059
Spline TLP	$\alpha_\tau = 0.05$	-0.065	2.545	2.544	0.105	2.626	2.626	-0.058	2.555	2.555
Spline TLP	$\alpha_\tau = 0.01$	-0.065	2.543	2.542	0.103	2.625	2.626	-0.041	2.470	2.469
AKX	$k_n = 78$	–	–	–	–	–	–	–	–	–
AKX	$k_n = 91$	–	–	–	–	–	–	–	–	–
AKX	$k_n = 117$	–	–	–	–	–	–	–	–	–

All these quantities are multiplied by 100. The cutoff level in the TLP function is $\tau_n = \alpha_\tau\sqrt{\mathrm{MedRV}_T}$ for each simulated path of Y. The number of B-spline basis functions K_n and the effective penalty level λ_n/τ_n are selected via 5-fold cross-validation. AKX stands for the estimator of Aït-Sahalia et al. (2020). For both the spline-based and AKX estimators, the truncation thresholds for the increments of Y and each component of X are given by $4\Delta_n^{0.47}\sqrt{\mathrm{MedRV}_T}$. All results are based on 1,000 Monte Carlo replications.

Table B.2: Simulation results for model selection with alternative truncation threshold

	$\alpha_\tau = 0.05$			$\alpha_\tau = 0.01$		
	Relevant	Irrelevant	Correct	Relevant	Irrelevant	Correct
$p = 3$	1.000	–	1.000	1.000	–	1.000
$p = 10$	1.000	0.041	0.745	1.000	0.000	1.000
$p = 50$	1.000	0.011	0.604	1.000	0.000	1.000
$p = 100$	1.000	0.038	0.019	1.000	0.000	1.000
$p = 500$	1.000	0.010	0.014	1.000	0.000	0.788

Average selection frequencies for the relevant and irrelevant factors, and frequencies of correct model specification. The cutoff level in the TLP function is $\tau_n = \alpha_\tau \sqrt{\mathrm{MedRV}_T}$ for each simulated path of Y. The number of B-spline basis functions K_n and the effective penalty level λ_n/τ_n are selected via 5-fold cross-validation. The truncation thresholds for the increments of Y and each component of X are given by $4\Delta_n^{0.47}\sqrt{\mathrm{MedRV}_T}$. All results are based on 1,000 Monte Carlo replications.

Table B.3: Simulation results for model estimation with imperfect multicollinearity

Panel A: $p = 3$		$\widehat{I\beta}_{1,T}$			$\widehat{I\beta}_{2,T}$			$\widehat{I\beta}_{3,T}$		
Estimator	Tuning	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
Spline		-0.249	2.829	2.839	0.161	2.527	2.531	-0.112	2.345	2.347
Spline TLP	$\alpha_\tau = 0.05$	-0.250	2.829	2.839	0.162	2.527	2.531	-0.112	2.345	2.347
Spline TLP	$\alpha_\tau = 0.01$	-0.250	2.829	2.839	0.162	2.527	2.531	-0.112	2.345	2.347
AKX	$k_n = 78$	-0.257	2.900	2.910	0.170	2.622	2.626	-0.103	2.447	2.448
AKX	$k_n = 91$	-0.279	2.920	2.931	0.165	2.619	2.623	-0.127	2.486	2.488
AKX	$k_n = 117$	-0.314	2.929	2.945	0.177	2.631	2.636	-0.091	2.497	2.498
Panel B: $p = 10$		$\widehat{I\beta}_{1,T}$			$\widehat{I\beta}_{2,T}$			$\widehat{I\beta}_{3,T}$		
Estimator	Tuning	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
Spline		-0.337	5.401	5.409	0.288	9.212	9.212	-0.165	52.775	52.748
Spline TLP	$\alpha_\tau = 0.05$	-0.302	3.033	3.046	0.257	2.889	2.899	-0.081	2.552	2.552
Spline TLP	$\alpha_\tau = 0.01$	-0.302	3.032	3.046	0.258	2.889	2.899	-0.089	2.529	2.530
AKX	$k_n = 78$	0.038	11.424	11.418	-0.439	22.124	22.117	4.002	129.714	129.711
AKX	$k_n = 91$	-0.006	10.902	10.897	-0.393	20.901	20.895	3.525	122.475	122.465
AKX	$k_n = 117$	-0.412	10.325	10.328	0.405	19.823	19.817	-0.967	115.436	115.382
Panel C: $p = 50$		$\widehat{I\beta}_{1,T}$			$\widehat{I\beta}_{2,T}$			$\widehat{I\beta}_{3,T}$		
Estimator	Tuning	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
Spline		-1.297	7.651	7.756	1.573	8.529	8.668	-2.245	13.414	13.594
Spline TLP	$\alpha_\tau = 0.05$	-0.226	2.845	2.852	0.225	3.217	3.224	-0.196	2.670	2.676
Spline TLP	$\alpha_\tau = 0.01$	-0.224	2.845	2.852	0.224	3.231	3.237	-0.193	2.757	2.762
AKX	$k_n = 78$	0.124	44.185	44.163	0.216	48.798	48.774	0.928	92.308	92.267
AKX	$k_n = 91$	0.714	37.624	37.612	-0.801	41.825	41.812	2.677	79.908	79.913
AKX	$k_n = 117$	-0.465	27.295	27.285	0.548	30.044	30.034	-0.405	57.546	57.518
Panel D: $p = 100$		$\widehat{I\beta}_{1,T}$			$\widehat{I\beta}_{2,T}$			$\widehat{I\beta}_{3,T}$		
Estimator	Tuning	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
Spline		-6.568	8.022	10.364	5.921	6.179	8.556	-3.756	6.944	7.892
Spline TLP	$\alpha_\tau = 0.05$	-0.531	3.448	3.487	0.599	2.771	2.834	-0.277	3.048	3.059
Spline TLP	$\alpha_\tau = 0.01$	-0.532	3.449	3.488	0.587	2.767	2.828	-0.233	2.921	2.929
AKX	$k_n = 78$	–	–	–	–	–	–	–	–	–
AKX	$k_n = 91$	–	–	–	–	–	–	–	–	–
AKX	$k_n = 117$	3.499	100.218	100.229	-3.375	92.011	92.027	2.155	58.581	58.591
Panel E: $p = 500$		$\widehat{I\beta}_{1,T}$			$\widehat{I\beta}_{2,T}$			$\widehat{I\beta}_{3,T}$		
Estimator	Tuning	Bias	Stdev	RMSE	Bias	Stdev	RMSE	Bias	Stdev	RMSE
Spline		-45.035	1.801	45.071	35.018	1.902	35.069	-20.468	2.029	20.568
Spline TLP	$\alpha_\tau = 0.05$	-0.676	3.072	3.144	0.836	3.398	3.498	-1.109	5.477	5.585
Spline TLP	$\alpha_\tau = 0.01$	-0.668	3.064	3.135	0.756	3.379	3.461	-0.466	4.387	4.410
AKX	$k_n = 78$	–	–	–	–	–	–	–	–	–
AKX	$k_n = 91$	–	–	–	–	–	–	–	–	–
AKX	$k_n = 117$	–	–	–	–	–	–	–	–	–

All these quantities are multiplied by 100. The modified simulation design allows dependence among the redundant factors and also correlations between relevant and redundant factors. The cutoff level in the TLP function is $\tau_n = \alpha_\tau \sqrt{\mathrm{MedRV}_T}$ for each simulated path of Y. The number of B-spline basis functions K_n and the effective penalty level λ_n/τ_n are selected via 5-fold cross-validation. AKX stands for the estimator of Aït-Sahalia et al. (2020). For both the spline-based and AKX estimators, the truncation thresholds for the increments of Y and each component of X are given by $3\Delta_n^{0.47}\sqrt{\mathrm{MedRV}_T}$. All results are based on 1,000 Monte Carlo replications.

Table B.4: Simulation results for model selection with imperfect multicollinearity

	$\alpha_\tau = 0.05$			$\alpha_\tau = 0.01$		
	Relevant	Irrelevant	Correct	Relevant	Irrelevant	Correct
$p = 3$	1.000	–	1.000	1.000	–	1.000
$p = 10$	1.000	0.010	0.931	1.000	0.000	0.997
$p = 50$	1.000	0.003	0.878	1.000	0.000	1.000
$p = 100$	1.000	0.012	0.348	1.000	0.000	1.000
$p = 500$	1.000	0.062	0.000	0.999	0.000	0.984

Average selection frequencies for the relevant and irrelevant factors, and frequencies of correct model specification. The modified simulation design allows dependence among the redundant factors and also correlations between relevant and redundant factors. The cutoff level in the TLP function is $\tau_n = \alpha_\tau \sqrt{\mathrm{MedRV}_T}$ for each simulated path of Y. The number of B-spline basis functions K_n and the effective penalty level λ_n/τ_n are selected via 5-fold cross-validation. The truncation thresholds for the increments of Y and each component of X are given by $3\Delta_n^{0.47}\sqrt{\mathrm{MedRV}_T}$. All results are based on 1,000 Monte Carlo replications.

B.2 Supplementary Empirical Results

Sample information. Table B.5 reports the list of individual stocks in our sample, including their sector and industry group assignments. Sector classifications are based on the GICS sector code from WRDS Compustat. The Fama-French 48 industry group for each firm is assigned based on its four-digit SIC code from Compustat and the SIC-to-industry definitions from Kenneth R. French's data library. Table B.6 reports the fraction of zero 5-minute returns in each year for all 57 stocks.

Table B.5: Descriptive information of selected S&P 100 stocks

Symbol	Security	GICS	Sector	SIC	Fama-French 48 Industry
AAPL	Apple Inc	45	Information Technology	3663	Electronic Equipment (36)
ABBV	AbbVie Inc	35	Health Care	2836	Pharmaceutical Products (13)
ACN	Accenture Plc	45	Information Technology	8742	Business Services (34)
AMGN	Amgen Inc	35	Health Care	2836	Pharmaceutical Products (13)
AMZN	Amazon.com Inc	25	Consumer Discretionary	5961	Retail (42)
AXP	American Express Co	40	Financials	6141	Banking (44)
BA	Boeing Co	20	Industrials	3721	Aircraft (24)
BIIB	Biogen Inc	35	Health Care	2836	Pharmaceutical Products (13)
BLK	BlackRock Inc	40	Financials	6282	Trading (47)
BMY	Bristol-Myers Squibb Co	35	Health Care	2834	Pharmaceutical Products (13)
C	Citigroup Inc	40	Financials	6199	Banking (44)
CAT	Caterpillar Inc	20	Industrials	3531	Machinery (21)
COF	Capital One Financial Corp	40	Financials	6141	Banking (44)
COP	ConocoPhillips	10	Energy	1311	Petroleum and Natural Gas (30)
COST	Costco Wholesale Corp	30	Consumer Staples	5399	Retail (42)
CVS	CVS Health Corp	35	Health Care	8000	Healthcare (11)
CVX	Chevron Corp	10	Energy	2911	Petroleum and Natural Gas (30)
DIS	Walt Disney Co	50	Communication Services	4888	Communication (32)
EMR	Emerson Electric Co	20	Industrials	3823	Measuring and Control Equipment (37)
FDX	FedEx Corp	20	Industrials	4513	Transportation (40)
GD	General Dynamics Corp	20	Industrials	3721	Aircraft (24)
GILD	Gilead Sciences Inc	35	Health Care	2836	Pharmaceutical Products (13)
GOOG	Alphabet Inc	50	Communication Services	7370	Business Services (34)
GS	Goldman Sachs Group Inc	40	Financials	6211	Trading (47)
HD	Home Depot Inc	25	Consumer Discretionary	5211	Retail (42)
HON	Honeywell International Inc	20	Industrials	9997	–
IBM	IBM	45	Information Technology	7370	Business Services (34)
JNJ	Johnson & Johnson	35	Health Care	2834	Pharmaceutical Products (13)
JPM	JPMorgan Chase & Co	40	Financials	6020	Banking (44)
LLY	Eli Lilly & Co	35	Health Care	2834	Pharmaceutical Products (13)
LMT	Lockheed Martin Corp	20	Industrials	3760	Defense (26)
LOW	Lowe's Cos Inc	25	Consumer Discretionary	5211	Retail (42)
MA	Mastercard Inc	40	Financials	7389	Business Services (34)
MCD	McDonald's Corp	25	Consumer Discretionary	5812	Restaurants, Hotels, Motels (43)
MDT	Medtronic Plc	35	Health Care	3845	Medical Equipment (12)
MMM	3M Co	20	Industrials	9997	–
MO	Altria Group Inc	30	Consumer Staples	2111	Tobacco Products (5)
MRK	Merck & Co Inc	35	Health Care	2834	Pharmaceutical Products (13)
MSFT	Microsoft Corp	45	Information Technology	7372	Business Services (34)
NKE	Nike Inc	25	Consumer Discretionary	3021	Apparel (10)
PEP	PepsiCo Inc	30	Consumer Staples	2080	Beer & Liquor (4)

Continued on next page.

Symbol	Security	GICS	Sector	SIC	Fama-French 48 Industry
PG	Procter & Gamble Co	30	Consumer Staples	2840	Consumer Goods (9)
PM	Philip Morris International Inc	30	Consumer Staples	2111	Tobacco Products (5)
QCOM	Qualcomm Inc	45	Information Technology	3674	Electronic Equipment (36)
SBUX	Starbucks Corp	25	Consumer Discretionary	5812	Restaurants, Hotels, Motels (43)
SLB	Schlumberger Ltd	10	Energy	1389	Petroleum and Natural Gas (30)
SPG	Simon Property Group Inc	60	Real Estate	6798	Trading (47)
TGT	Target Corp	30	Consumer Staples	5331	Retail (42)
UNH	UnitedHealth Group Inc	35	Health Care	6324	Insurance (45)
UNP	Union Pacific Corp	20	Industrials	4011	Transportation (40)
UPS	UPS	20	Industrials	4210	Transportation (40)
V	Visa Inc	40	Financials	6099	Banking (44)
VZ	Verizon Communications Inc	50	Communication Services	4812	Communication (32)
WBA	Walgreens Boots Alliance Inc	30	Consumer Staples	5912	Retail (42)
WFC	Wells Fargo & Co	40	Financials	6020	Banking (44)
WMT	Walmart Inc	30	Consumer Staples	5331	Retail (42)
XOM	Exxon Mobil Corp	10	Energy	2911	Petroleum and Natural Gas (30)

Individual stocks used in the empirical analysis. Sector classifications are based on the GICS sector code from WRDS Compustat. The Fama-French 48 industry group for each firm is assigned based on its four-digit SIC code from Compustat and the SIC-to-industry definitions from Kenneth R. French's data library.

Table B.6: Proportions of zero 5-minute returns (%) for selected S&P 100 stocks

Symbol	2016	2017	2018	2019	2020	Symbol	2016	2017	2018	2019	2020
AAPL	1.75	1.59	0.88	1.09	0.39	LLY	3.82	4.71	3.81	2.98	1.45
ABBV	3.35	3.45	1.99	2.86	1.60	LMT	1.24	1.39	0.99	0.92	0.47
ACN	3.05	3.19	2.00	2.04	1.22	LOW	3.69	3.30	2.49	2.70	1.13
AMGN	1.59	1.72	1.17	1.50	0.88	MA	3.49	3.58	1.37	1.44	0.64
AMZN	0.37	0.21	0.16	0.17	0.13	MCD	3.09	2.98	1.87	1.73	1.02
AXP	4.65	4.71	3.16	3.40	1.75	MDT	3.44	4.01	3.12	2.79	1.81
BA	2.05	1.67	0.58	0.61	0.46	MMM	2.14	2.21	1.40	1.69	0.99
BIIB	0.69	0.86	1.01	1.27	0.88	MO	3.90	4.13	3.11	3.06	3.03
BLK	0.63	0.87	0.80	0.85	0.58	MRK	3.93	4.37	3.23	3.23	2.01
BMY	3.40	4.13	3.71	4.17	2.55	MSFT	3.03	3.39	1.63	1.52	0.69
C	3.97	3.51	2.83	3.32	1.91	NKE	3.52	4.39	3.30	2.99	1.78
CAT	3.35	2.93	1.42	2.27	1.24	PEP	3.44	3.55	2.67	2.80	1.57
COF	4.24	3.58	2.97	3.44	2.01	PG	2.95	3.72	2.53	2.70	1.93
COP	3.59	4.16	3.40	3.40	2.43	PM	3.19	3.50	2.73	3.13	2.44
COST	2.06	1.88	1.29	1.12	0.56	QCOM	3.81	3.52	3.08	2.85	1.37
CVS	3.07	3.18	2.92	2.74	1.80	SBUX	3.39	4.09	3.48	3.04	1.66
CVX	2.37	2.80	1.88	2.46	1.34	SLB	3.06	3.42	2.79	4.00	3.43
DIS	2.51	2.62	2.09	1.71	0.83	SPG	1.65	2.03	1.92	1.88	1.19
EMR	4.61	4.63	3.34	4.06	2.54	TGT	3.67	3.57	2.82	2.71	1.32
FDX	1.95	1.75	0.99	1.41	0.78	UNH	2.27	1.91	1.13	1.17	0.58
GD	2.01	1.77	1.48	1.87	1.31	UNP	3.02	2.97	1.88	2.00	1.21
GILD	2.52	3.37	2.77	3.51	2.06	UPS	3.71	3.40	2.07	2.66	1.13
GOOG	0.42	0.51	0.38	0.45	0.26	V	3.22	3.29	1.85	1.86	0.82
GS	1.90	1.40	1.11	1.67	0.83	VZ	3.96	3.73	3.26	3.82	2.58
HD	2.18	2.47	1.43	1.34	0.74	WBA	3.84	4.32	3.32	3.65	2.72
HON	2.85	3.44	2.01	2.22	1.19	WFC	3.96	3.82	3.41	4.22	2.61
IBM	2.30	2.35	1.70	2.27	1.44	WMT	4.26	3.89	2.56	2.88	1.23
JNJ	3.25	2.82	1.88	2.38	1.11	XOM	2.96	3.40	2.59	3.03	1.77
JPM	3.23	2.92	1.83	2.26	1.19		–	–	–	–	–

Factor selection. In addition to the factor selection results for representative stocks in Fig. 5 and the pooled rankings in Tables 3 and 4, we report the full sample results on the average number of selected factors for all selected S&P 100 stocks and Fama-French industry portfolios in Tables B.7 and B.8, respectively.

Table B.7: Average number of selected factors per year for selected S&P 100 stocks

Symbol	2016	2017	2018	2019	2020	Symbol	2016	2017	2018	2019	2020
AAPL	20.67	18.33	16.33	15.50	12.92	LLY	6.50	6.08	6.50	12.00	9.00
ABBV	22.08	17.25	27.17	6.83	14.83	LMT	7.67	4.75	6.83	8.08	7.42
ACN	3.33	1.08	2.50	2.00	1.67	LOW	6.08	3.25	4.50	4.25	6.83
AMGN	12.33	9.75	5.75	8.83	6.83	MA	7.42	13.58	15.75	17.83	18.25
AMZN	24.75	22.33	15.17	16.83	13.58	MCD	9.42	6.83	2.50	10.42	6.08
AXP	2.67	3.08	6.58	9.50	2.83	MDT	1.75	1.08	2.17	2.92	1.67
BA	12.67	12.92	16.42	18.67	22.83	MMM	2.58	2.67	7.33	3.58	3.50
BIIB	3.92	4.67	4.25	1.92	2.92	MO	15.92	12.83	21.17	20.92	9.08
BLK	2.33	2.00	3.33	3.75	2.50	MRK	15.00	9.58	13.67	18.83	15.75
BMY	18.50	7.00	6.83	8.00	18.67	MSFT	21.25	17.67	18.67	12.42	11.42
C	19.25	20.50	17.25	10.33	8.67	NKE	19.00	5.00	9.08	4.25	7.75
CAT	3.58	3.92	7.58	7.83	4.67	PEP	10.58	11.25	15.67	14.75	8.67
COF	4.67	4.92	5.25	5.25	4.08	PG	19.33	19.17	14.00	16.42	16.33
COP	7.67	10.42	12.75	10.75	8.67	PM	18.50	23.08	14.00	22.67	8.25
COST	2.67	0.92	3.17	6.08	7.92	QCOM	14.83	7.17	7.25	13.25	5.25
CVS	10.92	8.58	4.25	8.58	3.92	SBUX	7.17	2.42	1.08	3.58	4.17
CVX	25.58	23.67	23.33	17.42	28.67	SLB	4.58	5.17	9.67	4.50	4.50
DIS	21.75	12.67	16.50	33.00	35.17	SPG	0.92	0.17	1.58	0.58	3.17
EMR	2.08	4.17	4.75	3.75	3.08	TGT	1.83	3.42	5.83	3.33	1.83
FDX	1.17	0.42	2.08	2.92	3.00	UNH	25.50	26.83	13.17	20.67	20.67
GD	1.83	3.42	3.00	2.42	2.75	UNP	6.25	3.75	7.67	14.25	5.10
GILD	25.00	7.75	10.83	5.42	6.75	UPS	1.83	0.92	2.50	4.67	1.42
GOOG	7.00	12.83	15.08	15.50	12.25	V	20.08	21.58	18.75	19.83	17.50
GS	5.92	7.92	11.00	7.92	3.92	VZ	29.75	20.67	21.83	20.50	15.17
HD	22.83	23.92	32.08	24.58	20.42	WBA	13.75	6.58	3.58	2.33	2.67
HON	3.75	3.50	7.92	7.17	2.67	WFC	25.58	26.33	26.75	23.42	14.67
IBM	18.33	7.83	5.67	3.25	4.67	WMT	33.92	27.92	22.58	28.08	17.00
JNJ	15.33	17.00	22.08	26.42	23.33	XOM	22.67	25.50	22.00	20.42	27.33
JPM	18.00	16.17	15.92	20.42	31.75		–	–	–	–	–

Average number of selected factors per year for all selected S&P 100 stocks. The candidate covariate set consists of the 224 high-frequency factors of Aleti (2023). The cutoff level in the TLP function is $\tau_n = 0.01\sqrt{\text{MedRV}_T}$, where MedRV_T denotes the MedRV of the test asset computed over the corresponding estimation window $[0, T]$. The number of B-spline basis functions K_n and the effective penalty level λ_n/τ_n are selected via 5-fold cross-validation using the one-standard-error rule.

Table B.8: Average number of selected factors per year for Fama-French 48 industry portfolios

Ind.	Industry	2016	2017	2018	2019	2020
1	Agriculture	1.00	1.25	2.38	1.00	2.00
2	Food Products	8.08	9.92	6.67	3.25	6.64
3	Candy & Soda	10.42	14.92	15.42	15.50	11.50
4	Beer & Liquor	3.00	9.67	3.00	1.00	1.30
5	Tobacco Products	20.58	26.00	24.42	25.75	11.36
6	Recreation	2.82	1.88	2.27	1.27	4.36
7	Entertainment	9.08	10.83	20.00	33.83	31.83
8	Printing and Publishing	1.00	1.67	1.50	2.44	3.82
9	Consumer Goods	12.42	12.33	9.33	12.00	16.18
10	Apparel	14.18	10.36	9.17	13.58	10.00
11	Healthcare	13.83	5.67	5.50	7.75	6.08
12	Medical Equipment	6.92	4.42	11.58	12.17	11.17
13	Pharmaceutical Products	8.75	6.58	7.17	8.50	7.92
14	Chemicals	13.83	8.58	14.83	5.58	5.50
15	Rubber and Plastic Products	3.83	1.89	2.92	2.33	3.27
16	Textiles	2.73	3.13	2.00	2.00	4.33
17	Construction Materials	2.17	2.73	11.67	6.25	5.58
18	Construction	5.67	2.91	7.50	5.75	6.25
19	Steel Works Etc.	6.92	7.50	6.50	3.92	4.58
20	Fabricated Products	3.25	4.08	3.82	2.33	3.36
21	Machinery	11.17	9.08	12.50	5.75	9.17
22	Electrical Equipment	2.67	4.36	3.50	2.75	4.33
23	Automobiles and Trucks	10.42	9.83	9.50	8.58	8.92
24	Aircraft	4.67	5.92	11.83	10.00	9.08
25	Shipbuilding, Railroad Equipment	2.00	4.00	1.00	1.57	2.92
26	Defense	7.25	4.75	10.33	9.42	6.33
27	Precious Metals	4.80	2.00	2.33	3.00	1.00
28	Non-Metallic and Industrial Metal Mining	7.75	10.50	6.75	2.55	2.08
29	Coal	2.33	3.67	2.44	2.88	3.14
30	Petroleum and Natural Gas	10.33	13.92	15.33	13.83	31.00
31	Utilities	18.50	15.50	13.00	11.58	15.82
32	Communication	15.75	14.75	11.75	14.83	10.92
33	Personal Services	3.00	3.25	3.33	1.58	2.27
34	Business Services	2.83	6.08	5.58	3.92	5.67
35	Computers	12.75	17.00	12.00	10.25	9.83
36	Electronic Equipment	9.92	11.58	13.58	11.42	14.08
37	Measuring and Control Equipment	3.08	3.08	6.42	8.67	3.00
38	Business Supplies	2.67	2.70	2.33	3.42	5.25
39	Shipping Containers	1.83	1.88	2.58	1.09	2.45
40	Transportation	18.25	12.33	15.67	8.33	13.67
41	Wholesale	1.25	1.00	1.58	2.33	3.64
42	Retail	6.42	11.67	8.33	7.42	7.83
43	Restaurants, Hotels, Motels	8.75	2.25	5.67	5.83	21.45
44	Banking	8.50	7.33	6.00	7.67	12.58
45	Insurance	6.92	10.00	5.92	12.50	12.25
46	Real Estate	3.33	2.40	1.67	1.17	2.92
47	Trading	5.17	5.50	5.25	4.42	8.92
48	Other	4.10	1.29	1.42	2.75	2.08

Average number of selected factors per year for Fama-French industry portfolios. The candidate covariate set consists of the 224 high-frequency factors of Aleti (2023). The cutoff level in the TLP function is $\tau_n = 0.01\sqrt{\text{MedRV}_T}$, where MedRV_T denotes the MedRV of the test asset computed over the corresponding estimation window $[0, T]$. The number of B-spline basis functions K_n and the effective penalty level λ_n/τ_n are selected via 5-fold cross-validation using the one-standard-error rule.

References

Aït-Sahalia, Y., Kalnina, I., and Xiu, D. (2020). High-frequency factor models and regressions. *Journal of Econometrics*, 216(1):86–105.

Aleti, S. (2023). The high-frequency factor zoo. Working Paper.

Davidson, R. and MacKinnon, J. G. (2004). *Econometric Theory and Methods*. Oxford University Press.

de Boor, C. (1973). The quasi-interpolant as a tool in elementary polynomial spline theory. In Lorentz, G. G., editor, *Approximation Theory*, pages 269–276. Academic Press.

de Boor, C. (1978). *A Practical Guide to Splines*. Springer.

Demko, S., Moss, W. F., and Smith, P. W. (1984). Decay rates for inverses of band matrices. *Mathematics of Computation*, 43(168):491–499.

Hanson, D. L. and Wright, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Annals of Mathematical Statistics*, 42(3):1079–1083.

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

Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14(3):763–788.

Jacod, J. and Protter, P. (2012). *Discretization of Processes*. Springer.

Jacod, J. and Shiryaev, A. (2013). *Limit Theorems for Stochastic Processes*. Springer.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2):270–296.

Shadrin, A. Y. (2001). The L_∞-norm of the L_2-spline projector is bounded independently of the knot sequence: A proof of de Boor's conjecture. *Acta Mathematica*, 187(1):59–137.

Tropp, J. A. (2011). Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270.

van de Geer, S. A. (2008). High-dimensional generalized linear models and the Lasso. *Annals of Statistics*, 36(2):614–645.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press.

Xue, L. and Qu, A. (2012). Variable selection in high-dimensional varying-coefficient models with global optimality. *Journal of Machine Learning Research*, 13(1):1973–1998.