Supplement to "Estimation and Inference of Change Points in Functional Regression Time Series"

Shivam Kumar[1], Haotian Xu[2], Haeran Cho[3], and Daren Wang[4]

[1]Booth School of Business, University of Chicago, Chicago, USA
[2]Department of Mathematics and Statistics, Auburn University, Auburn, USA
[3]School of Mathematics, University of Bristol, Bristol, UK
[4]Department of Mathematics, University of California San Diego, San Diego, USA

We collect extensive simulation studies, additional real data analysis and all the technical details in the online supplementary material.

A Numerical results

In this section, we perform numerical experiments on simulated datasets to investigate the performance of the proposed change point estimation and inference procedure, which contains three steps: (i) the preliminary estimation of the change points, (ii) the refinement of change point estimators and (iii) the construction of confidence intervals. Throughout, we refer to our combined procedure as 'FRBS'.

A.1 Simulation studies

Settings. We modify the simulation settings of Yuan and Cai (2010) and Cai and Yuan (2012) by introducing temporal dependence in $\{X_j\}_{j=1}^n$ and changes in $\{\beta_j^*\}_{j=1}^n$. Specifically, we simulate data from the model described in (1.1) where the error process $\{\varepsilon_j\}_{j=1}^n$ is a sequence of i.i.d. standard normal random variables, and $\{X_j\}_{j=1}^n$ is a stationary process following

$$X_j = \sum_{m=1}^{50} \zeta_m Z_{m,j} \phi_m, \qquad 1 \le j \le n,$$

with $\phi_1 = 1$, $\phi_{m+1} = \sqrt{2}\cos(m\pi t)$ for $m \ge 1$ and $\zeta_m = (-1)^{m+1} m^{-1}$. For each $m \ge 1$, $\{Z_{m,j}\}_{j=1}^n$ is independently generated as an autoregressive process, i.e. $Z_{m,j} = 0.3 Z_{m,j-1} + \sqrt{1-0.3^2} \cdot e_{m,j}$ with $e_{m,j} \overset{i.i.d.}{\sim} N(0,1)$. Note that $\zeta_m^2 = m^{-2}$ are the eigenvalues of the covariance function of X_j, and ϕ_m are the corresponding eigenfunctions. Let

$$\beta^{(0)} = 4 \sum_{m=1}^{50} (-1)^{m+1} m^{-4} \phi_m \quad \text{and} \quad \beta^{(1)} = (4 - c_\beta) \sum_{m=1}^{50} (-1)^{m+1} m^{-2} \phi_m,$$

where the coefficient $c_\beta \in \{0.5, 1\}$. We consider the slope functions

$$\beta_j^* = \begin{cases} \beta^{(0)} & \text{for } \ j \in \{1, \ldots, \eta_1\}, \\ \beta^{(1)} & \text{for } \ j \in \{\eta_1 + 1, \ldots, \eta_2\}, \\ \vdots & \\ \beta^{(\mathcal{K} \bmod 2)} & \text{for } \ j \in \{\eta_{\mathcal{K}} + 1, \ldots, n\}. \end{cases}$$

The cases with $c_\beta = 0.5$ and $c_\beta = 1$ correspond to the settings with small and large jump sizes, respectively. We further assume that for each j, the random function X_j is observed in an evenly spaced fixed grid with size $p = 200$. We choose the reproducing kernel Hilbert space $\mathcal{H}(K)$ as the Sobolev space

$$\mathcal{W}_2^1 = \{f \in L^2[0,1] : \|f^{(j)}\|_{\mathcal{L}^2} < \infty, \, j = 0, 1\},$$

with the corresponding reproducing kernel

$$K(s,t) = \begin{cases} \frac{\cosh(s)\cosh(1-t)}{\sinh(1)} & 0 \le s \le t \le 1, \\ \frac{\cosh(t)\cosh(1-s)}{\sinh(1)} & 0 \le t \le s \le 1. \end{cases}$$

Note that the reproducing kernel and the covariance function of X_j share a common ordered set of eigenfunctions (see Cai and Yuan, 2012).

Evaluation measurements. Let $\{\eta_k\}_{k=1}^{\mathcal{K}}$ and $\{\widehat{\eta}_k\}_{k=1}^{\widehat{\mathcal{K}}}$ be the set of true change points and a set of estimated change points, respectively. To assess the performance of different methods in localization, we report (i) the proportions (out of 200 repetitions) of over- or under-estimating $\mathcal{K}$, i.e. $\widehat{\mathcal{K}} > \mathcal{K}$ and $\widehat{\mathcal{K}} < \mathcal{K}$, respectively; and (ii) the average and the standard deviation of the scaled Hausdorff distances between $\{\eta_k\}_{k=1}^{\mathcal{K}}$ and $\{\widehat{\eta}_k\}_{k=1}^{\widehat{\mathcal{K}}}$ defined as

$$d_{\mathrm{H}} = \frac{1}{n} \max\left\{ \max_{j=0,\ldots \widehat{\mathcal{K}}+1} \min_{k=0,\ldots,\mathcal{K}+1} |\widehat{\eta}_j - \eta_k|, \max_{k=0,\ldots,\mathcal{K}+1} \min_{j=0,\ldots,\widehat{\mathcal{K}}+1} |\widehat{\eta}_j - \eta_k| \right\},$$

where we set $\widehat{\eta}_0 = 1$ and $\widehat{\eta}_{\widehat{\mathcal{K}}+1} = n+1$. We computed the refined estimators using interpolated refinement intervals, defined in (2.10). Given a confidence level $\alpha \in (0,1)$, we evaluate the performance of the proposed confidence intervals by measuring their coverage of η_k, defined as

$$\mathrm{cover}_k(1-\alpha) = \mathbb{1}\left\{ \eta_k \in \left[\widetilde{\eta}_k + \frac{\widehat{q}_u(\alpha/2)}{\widehat{\kappa}_k^2}, \widetilde{\eta}_k + \frac{\widehat{q}_u(1-\alpha/2)}{\widehat{\kappa}_k^2} \right] \right\},$$

for each $k \in \{1, \ldots, \mathcal{K}\}$. To ensure the validity of the above definition, we compute the averaged coverage among all the repetitions where we obtain $\widehat{\mathcal{K}} = \mathcal{K}$.

Comparison. To the best of our knowledge, no method currently tackles change point problems in the scalar-on-function (functional linear) regression setting we study. To provide meaningful baselines, we construct three competitors. (1) FPCABS replaces the RKHS slope-function estimator in our procedure with the FPCA-based estimator of Yao et al. (2005) and applies exactly the same seeded binary segmentation. (2) HDLR is the high-dimensional linear-regression change-point method of Xu et al. (2024). We include it because densely sampled functional predictors can be viewed as high-dimensional vectors; moreover, HDLR, like our FRBS algorithm, follows a two-stage "preliminary-then-refinement" strategy, allowing a step-by-step performance comparison. Note that among the 3 methods, FRBS and FPCABS belong to the functional linear regression framework, treating covariates explicitly as discretely observed functions, whereas HDLR falls within the high-dimensional linear regression paradigm, viewing covariates simply as high-dimensional vectors in a design matrix.

Selection of tuning parameters and estimation of unknown quantities. Four tuning parameters are involved in the proposed change point localization and inference procedures. These are the number of layers M for the seeded intervals (see Definition 1), ω and τ for the FRBS algorithm (see Algorithm 1) and the block size $2q$ for long-run variance estimation (see Algorithm 2) in the confidence interval construction. We set $M = \lceil \log_2(10) \rceil + 1$. In place of ω, which is used in specifying λ_{e-s}, we propose to select a single $\lambda_{e-s} = \lambda$ along with the threshold τ, adapting the cross-validation method proposed by Rinaldo et al. (2021). Specifically, we first divide $\{(y_j, X_j)\}_{j=1}^n$ into those with odd and even indices, respectively. For each possible combination of $\lambda \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $\tau \in \{1, 1.5, 2, 2.5, 3\} \times n^{2/5}$, we obtain the FRBS outputs ($\widehat{\mathcal{B}}$, $\widehat{\mathcal{K}}$ and $\{\widehat{\beta}_k\}_{k=0}^{\widehat{\mathcal{K}}}$) based on the training set, and compute the least squared prediction error on the test set as the validation loss. We select the combination of λ and τ that minimize the validation loss. Following the discussion after Theorem 3, we set $q = \left\lceil \left(\max_{1\leq k\leq \widehat{\mathcal{K}}} \{e_k - s_k\} \right)^{2/5} \Big/ 2 \right\rceil$ with $\{(s_k, e_k)\}_{k=1}^{\widehat{\mathcal{K}}}$ given in (2.10). We note that the simulation results remain robust against the choices of the tuning parameters M and q.

For the FPCABS, we tune the number of principal components K and the threshold τ using the similar CV approach as for the FRBS. Specifically, for each combination of $K \in \{5, 10, 15, 20\}$ and $\tau \in \{1, 1.5, 2, 2.5, 3\} \times n^{2/5}$, we select the combination of K and τ that minimize the validation loss.

For the HDLR, we use the CV method in Xu et al. (2022) to select the tuning parameters, specifically λ (Lasso penalty) and ξ (L_0 penalty), for the DPDU algorithm therein, with candidate sets $\lambda \in \{0.05, 0.1, 0.5, 1, 2, 3, 4, 5\}$ and $\xi \in \{5, 10, 15, 20, 25, 30, 35, 40\}$, and use the default values of the other tuning parameters.

Scenario I: single change point. Let $\mathcal{K} = 1$ and $\eta = n/2$. We vary $n \in \{200, 400, 600, 800\}$, $c_\beta \in \{0.5, 1\}$ and fix $p = 200$. Tables 1 and 2 summarize the localization and inference performance of FRBS, FPCABS and HDLR. Table 2 excludes the case with $n = 200$ where there are a large number of repetitions with mis-estimated $\mathcal{K}$ for all methods. In Table 1, comparing the Hausdorff distance computed with the preliminary ($d_{\mathrm{H}}^{\mathrm{pre}}$) and the refined estimators ($d_{\mathrm{H}}^{\mathrm{fin}}$), we see that the refinement step improves the performance for all methods in consideration as n increases and/or the jump size increases. The detection power improves with the sample size as evidenced by the decrease in the proportion of under-detection. At the same time, FRBS and FPCABS do not detect more false positives as the sample size increases, unlike HDLR. Overall, the proposed FRBS outperforms both competitors by a large margin, in its detection accuracy as well as localization performance, demonstrating the advantage of adopting a functional approach over the high-dimensional one of HDLR. Although the RKHS and the covariance function of X_j are well-aligned, the dimension reduction-based approach of FPCABS comes short of the RKHS-based FRBS. Table 2 shows that our proposed construction of confidence intervals performs well especially when the jump size is relatively high. In contrast, the intervals constructed based on HDLR perform poorly in capturing the change points, often with the intervals being too narrow. All these observations suggest the benefit of adopting the proposed functional approach.

Scenario II: unequally-spaced two change points. Let $\mathcal{K} = 2$ and the unequally-spaced change points $\{\eta_1, \eta_2\} = \{n/4, 5n/8\}$. We vary $n \in \{400, 600, 800\}$ and fix $p = 200$. Table 3 shows the localization performance of both preliminary and final estimators improve as n increases. The proposed FRBS again significantly outperform the FPCABS and the HDLR in all configurations.

Due to the overall poor detection performance HDLR and FRBS when $c_\beta = 0.5$, we only report the results from the confidence intervals produced by FRBS for the setting with $c_\beta = 1$ in Tables 4 and 5. The comparison between Tables 4 and 5 reveals that our inference procedure performs better when applied to η_2 associated with larger spacing with adjacent change points.

Table 1: In Scenario I, the proportions of under-, over-detection, and the average and standard deviation (in parentheses) of scaled Hausdorff distance over 200 repetitions are reported for FRBS, FPCABS and HDLR. The best performance of each measurement is highlighted in **bold** in each configuration. The single change point is located at $\eta = n/2$.

	$\mathcal{K} = 1$ and $p = 200$							
n	$\widehat{\mathcal{K}} < \mathcal{K}$	$\widehat{\mathcal{K}} > \mathcal{K}$	$d_{\mathrm{H}}^{\mathrm{pre}}$	$d_{\mathrm{H}}^{\mathrm{fin}}$	$\widehat{\mathcal{K}} < \mathcal{K}$	$\widehat{\mathcal{K}} > \mathcal{K}$	$d_{\mathrm{H}}^{\mathrm{pre}}$	$d_{\mathrm{H}}^{\mathrm{fin}}$
	FRBS, $c_\beta = 0.5$ (small jump size)				FRBS, $c_\beta = 1$ (large jump size)			
200	**0.310**	**0.015**	**0.198** (0.212)	**0.190** (0.216)	**0**	**0.025**	**0.033** (0.053)	**0.027** (0.045)
400	**0.105**	**0.045**	**0.091** (0.150)	**0.084** (0.151)	**0**	**0.025**	**0.018** (0.032)	**0.013** (0.026)
600	**0.045**	**0.030**	**0.060** (0.111)	**0.048** (0.108)	**0**	**0.020**	**0.017** (0.037)	**0.012** (0.035)
800	**0.005**	**0.020**	**0.033** (0.053)	**0.020** (0.044)	**0**	**0.010**	**0.013** (0.028)	**0.009** (0.027)
	FPCABS, $c_\beta = 0.5$ (small jump size)				FPCABS, $c_\beta = 1$ (large jump size)			
200	0.385	0.050	0.247 (0.213)	0.245 (0.216)	0.005	0.140	0.069 (0.082)	0.053 (0.079)
400	0.110	0.130	0.114 (0.151)	0.103 (0.153)	**0**	0.105	0.031 (0.048)	0.024 (0.046)
600	**0.045**	0.095	0.068 (0.112)	0.058 (0.112)	**0**	0.060	0.022 (0.046)	0.017 (0.046)
800	0.015	0.085	0.048 (0.079)	0.037 (0.077)	**0**	0.030	0.015 (0.035)	0.010 (0.032)
	HDLR, $c_\beta = 0.5$ (small jump size)				HDLR, $c_\beta = 1$ (large jump size)			
200	0.630	0.050	0.350 (0.196)	0.354 (0.195)	0.080	0.150	0.118 (0.138)	0.115 (0.151)
400	0.275	0.070	0.200 (0.196)	0.201 (0.202)	0.005	0.180	0.066 (0.078)	0.063 (0.104)
600	0.100	0.110	0.127 (0.146)	0.126 (0.159)	**0**	0.100	0.041 (0.061)	0.034 (0.076)
800	0.080	0.105	0.112 (0.137)	0.105 (0.147)	**0**	0.135	0.040 (0.059)	0.033 (0.078)

Table 2: In Scenario 1, the averaged coverage and the average and standard deviation (in parentheses) of the width of the confidence intervals from FRBS and HDLR over 200 repetitions are reported. The single change point is located at $\eta = n/2$.

	$\mathcal{K} = 1$ and $p = 200$			
	$\alpha = 0.01$		$\alpha = 0.05$	
n	cover$(1-\alpha)$	width$(1-\alpha)$	cover$(1-\alpha)$	width$(1-\alpha)$
	FRBS, $c_\beta = 0.5$ (small jump size)			
400	0.982	109.923 (38.551)	0.935	72.441 (24.811)
600	0.973	111.502 (28.749)	0.924	73.076 (19.474)
800	0.974	117.712 (32.737)	0.918	77.015 (20.509)
	FRBS, $c_\beta = 1$ (large jump size)			
400	0.989	42.877 (9.479)	0.923	28.515 (6.469)
600	0.974	43.505 (7.582)	0.969	28.299 (4.811)
800	0.990	43.892 (7.125)	0.949	28.831 (4.435)
	HDLR, $c_\beta = 0.5$ (small jump size)			
400	0.405	19.504 (4.493)	0.321	13.450 (2.944)
600	0.424	23.690 (4.432)	0.367	16.000 (2.849)
800	0.497	26.650 (4.294)	0.387	17.975 (2.685)
	HDLR, $c_\beta = 1$ (large jump size)			
400	0.748	18.153 (3.826)	0.681	12.712 (2.464)
600	0.828	20.722 (3.553)	0.750	14.272 (2.344)
800	0.896	22.879 (3.551)	0.821	15.740 (2.284)

Table 3: In Scenario II, the proportions of under-, over-detection, and the average and standard deviation (in parentheses) of scaled Hausdorff distance over 200 repetitions are reported for FRBS, FPCABS and HDLR. The best performance of each measurement is highlighted in **bold** in each configuration. The single change point is located at $\eta = n/2$. The two change points are located at $\eta_1 = n/4$ and $\eta_2 = 5n/8$.

$\mathcal{K} = 2$ and $p = 200$				
n	$\widehat{\mathcal{K}} < \mathcal{K}$	$\widehat{\mathcal{K}} > \mathcal{K}$	$d_{\mathrm{H}}^{\mathrm{pre}}$	$d_{\mathrm{H}}^{\mathrm{fin}}$
FRBS, $c_\beta = 0.5$ (small jump size)				
400	**0.460**	**0.015**	**0.194** (0.154)	**0.190** (0.157)
600	**0.265**	0.025	**0.120** (0.134)	**0.108** (0.139)
800	**0.150**	**0.015**	**0.081** (0.108)	**0.068** (0.109)
FRBS, $c_\beta = 1$ (large jump size)				
400	**0.010**	0.025	**0.029** (0.042)	**0.024** (0.042)
600	**0**	**0.015**	**0.020** (0.023)	**0.011** (0.018)
800	**0**	0.015	**0.015** (0.021)	**0.011** (0.019)
FPCABS, $c_\beta = 0.5$ (small jump size)				
400	0.555	0.120	0.238 (0.144)	0.235 (0.146)
600	0.465	0.085	0.194 (0.152)	0.187 (0.159)
800	0.240	0.140	0.128 (0.130)	0.119 (0.135)
FPCABS, $c_\beta = 1$ (large jump size)				
400	0.015	0.125	0.049 (0.057)	0.039 (0.057)
600	**0**	0.070	0.027 (0.031)	0.017 (0.028)
800	**0**	0.075	0.023 (0.035)	0.017 (0.035)
HDLR, $c_\beta = 0.5$ (small jump size)				
400	0.930	**0.015**	0.275 (0.073)	0.300 (0.084)
600	0.890	**0.010**	0.252 (0.062)	0.275 (0.080)
800	0.895	**0.015**	0.249 (0.055)	0.273 (0.072)
HDLR, $c_\beta = 1$ (large jump size)				
400	0.820	**0.020**	0.224 (0.045)	0.244 (0.058)
600	0.900	**0.015**	0.229 (0.039)	0.244 (0.043)
800	0.865	**0.005**	0.228 (0.039)	0.239 (0.039)

Table 4: In Scenario II, the averaged coverage and the average and standard deviation (in parentheses) of the width of the confidence intervals from FRBS over 200 repetitions for η_1 are reported.

FRBS, $p = 200$				
	$\alpha = 0.01$		$\alpha = 0.05$	
n	$\text{cover}_1(1-\alpha)$	$\text{width}_1(1-\alpha)$	$\text{cover}_1(1-\alpha)$	$\text{width}_1(1-\alpha)$
$c_\beta = 1$ (large jump size)				
400	0.959	43.591 (12.428)	0.933	28.741 (7.823)
600	0.994	45.492 (12.807)	0.980	29.751 (8.081)
800	0.979	43.477 (9.224)	0.958	28.456 (5.528)

Table 5: In Scenario II, the averaged coverage and the average and standard deviation (in parentheses) of the width of the confidence intervals from FRBS over 200 repetitions for η_2 are reported.

FRBS, $p = 200$				
	$\alpha = 0.01$		$\alpha = 0.05$	
n	$\mathrm{cover}_2(1-\alpha)$	$\mathrm{width}_2(1-\alpha)$	$\mathrm{cover}_2(1-\alpha)$	$\mathrm{width}_2(1-\alpha)$
$c_\beta = 1$ (large jump size)				
400	0.980	45.114 (12.250)	0.938	29.389 (7.501)
600	0.990	44.142 (10.934)	0.980	28.909 (6.844)
800	0.990	44.067 (9.165)	0.958	28.798 (5.734)

B Additional real data analysis

We analyze the relationship between the U.S. Treasury yield curve and monthly inflation using publicly available economic data from January 2000 to December 2024 (300 months). For each month, the functional covariate is constructed as the vector of Treasury yields at eleven standard maturities, ranging from 1 month to 30 years, measured on the last trading day of the month. The scalar response is defined as the month-over-month percentage change in the Consumer Price Index (CPI), representing the inflation rate for that month. This setup provides a time series of paired functional–scalar observations, allowing us to investigate potential structural changes in the predictive association between the yield curve and inflation over the past two decades, a period that includes several well-documented macroeconomic regime shifts.

More specifically, for each month j from January 2000 to December 2024 ($n = 300$), let $X_j(k)$ denote the Treasury yield curve in month j evaluated at maturity k, where k ranges over the set $\{1m, 3m, 6m, 1y, 2y, 3y, 5y, 7y, 10y, 20y, 30y\}$. Thus, $X_j = \{X_j(k)\}_k$ represents a discretely observed yield curve for month j. We use the yield from the last trading day of each month and impute missing values where necessary to ensure a complete functional covariate. The scalar response is the monthly inflation rate, defined as the month-over-month percentage change in the CPI. All data are obtained from publicly available sources at https://fred.stlouisfed.org and are plotted in Figure 1.

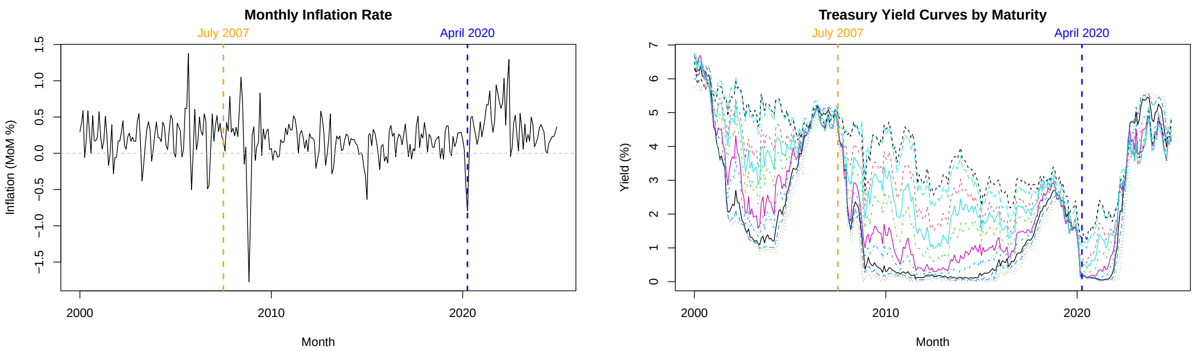

Figure 1: The monthly inflation rate, (y_j, left); the U.S. Treasury yield curves ($X_j(k)$, right). The refined change points estimated by FRBS are marked by vertical dashed lines.

Our empirical analysis of the U.S. Treasury yield curve and inflation relationship reveals that the proposed FRBS method successfully identifies two major structural change points, July 2007 and April 2020, both of which correspond closely to well-documented macro-financial regime shifts: the onset of the global financial crisis and the market turmoil at the onset of the COVID-19 pandemic (see e.g. Guidolin and Tam, 2013; Gorton et al., 2020; He et al., 2022). In contrast, the HDLR detects only a single change point in March 2008, which, while broadly consistent with the financial crisis period, does not align as precisely with established economic milestones. The FPCABS estimates two change points at June 2009 and February 2014, which do not coincide with the main documented policy or market turning points. Overall, these results highlight the superior temporal alignment and interpretability of the change points detected by FRBS, demonstrating its advantage in uncovering regime shifts that are both statistically and economically meaningful.

C Proof of Theorem 1

Proof. For $(s_m, e_m] \in \mathcal{J}$ and for all $t \in (s_m, e_m]$ we define

$$\widetilde{G}_t^{s_m,e_m} = \frac{(t-s_m)(e_m-t)}{e_m-s_m}\Sigma\left[\beta^*_{(s_m,t]}-\beta^*_{(t,e_m]},\beta^*_{(s_m,t]}-\beta^*_{(t,e_m]}\right].$$

For the interval $(s_m, e_m]$, consider the event

$$\begin{aligned}\mathcal{A}(s_m,e_m) = \{&\text{for all } t\in(s_m,e_m],\\ &\left|\widehat{W}_t^{s_m,e_m}-\widetilde{G}_t^{s_m,e_m}\right|-0.5\widetilde{G}_t^{s_m,e_m} \le \left(\frac{n}{\Delta}\right)(\log^{1+2\xi}(n))\left(n^{1/(2r+1)}+0.5\right)\Big\},\end{aligned}$$

and define the event

$$\mathcal{A} = \bigcap_{(s_m,e_m]\in\mathcal{J}} \mathcal{A}(s_m,e_m). \tag{1}$$

We established in Lemma C.2 that

$$P(\mathcal{A}) \longrightarrow 1 \qquad \text{as } n\to\infty.$$

All the analysis in the rest of this proof is under this asymptotically almost sure event $\mathcal{A}$. The strategy here is to use an induction argument. Denote

$$\vartheta_k = C_1\frac{(n/\Delta)\Delta^{1/(2r+1)}\log^{1+2\xi} n}{\kappa_k^2}.$$

Step 1: We show that, FRBS will consistently reject the existence of change points if they are no undetected change points in $(s,e]$. By induction hypothesis, we have

$$|\eta_k - s| \le \vartheta_k, \qquad |e-\eta_{k+1}| \le \vartheta_{k+1}.$$

For each $(s_m, e_m] \in \mathcal{J}$ such that $(s_m, e_m] \subset (s, e]$, there are four possible cases which are outlined below

i) $s_m < \eta_k < \eta_{k+1} < e_m$ with $\eta_k - s_m \le \vartheta_k$ and $\eta_{k+1} - e_m \le \vartheta_{k+1}$

ii) $\eta_k \le s_m < e_m \le \eta_{k+1}$ with $s_m - \eta_k \le \vartheta_k$ and $\eta_{k+1} - e_m \le \vartheta_{k+1}$,

iii) $\eta_{k-1} < s_m \le \eta_k < e_m \le \eta_{k+1}$ with $\eta_k - s_m \le \vartheta_k$,

iv) $\eta_{k-1} \le s_m < \eta_k \le e_m < \eta_{k+1}$ with $e_m - \eta_k \le \vartheta_k$.

We shall consider the first case, all other cases are simpler and could be handled similarly. There are two previously detected change point η_k and η_{k+1} in $(s_m, e_m]$ and we are going to show that FRBS shall not detect any change point in $(s_m, e_m]$. On the event $\mathcal{A}$ we write that

$$\begin{aligned}\forall t \in (s'_m, e'_m] \qquad \widehat{W}_t^{s_m,e_m} &\leq \frac{3}{2}\widetilde{G}_t^{s_m,e_m} + \left(\frac{n}{\Delta}\right)\log^{1+2\xi}(n)\left(n^{1/(2r+1)} + \frac{1}{2}\right) \\ &\leq 3\kappa_k^2(\eta_k - s_m) + 3\kappa_{k+1}^2(e - \eta_{k+1}) + 2\left(\frac{n}{\Delta}\right)n^{1/(2r+1)}\log^{1+2\xi}(n) \\ &\leq (8C_1 + 2)\left(\frac{n}{\Delta}\right)n^{1/(2r+1)}\log^{1+2\xi}(n) < \tau\end{aligned}$$

where the second last line follows from Lemma C.6 and the last line just follows from the definition of τ.

Step 2: We show that FRBS will correctly detect the existence of an undetected change point in $(s, e]$. In this case, there exists some change point, η_k in $(s, e]$, such that

$$\min\{\eta_k - s, e - \eta_k\} > \Delta - \vartheta_k,$$

for some $1 \leq k \leq \mathcal{K}$. Realize that $\Delta - \vartheta_k > 4\Delta/5$, asymptotically. For this step, it is sufficient to show that the set $\mathcal{M}^{s,e}$, form Algorithm 1, is not empty. From the construction of intervals in $\mathcal{J}$ and from Lemma C.1, we can always find an interval $(s_m, e_m] \in \mathcal{J}$ such that $(s_m, e_m] \subset (s, e]$ containing η_k such that

$$e_m - s_m \leq \Delta, \qquad \text{and} \qquad \min\{\eta_k - s_m, e_m - \eta_k\} \geq \Delta/5. \tag{2}$$

On the event $\mathcal{A}$, we have

$$\max_{s_m < t \leq e_m} \widehat{W}_t^{s_m,e_m} \geq \widehat{W}_{\eta_k}^{s_m,e_m} \geq \frac{1}{2}\widetilde{G}_{\eta_k}^{s_m,e_m} - \left(\frac{n}{\Delta}\right)\log^{1+2\xi}(n)\left(n^{1/(2r+1)} + \frac{1}{2}\right). \tag{3}$$

Since η_k is the only change point in $(s_m, e_m]$, using (2), we write that

$$\widetilde{G}_{\eta_k}^{s_m,e_m} = \kappa_k^2\frac{(\eta_k - s_m)(e_m - \eta_k)}{(e_m - s_m)} \geq \frac{1}{2}\kappa_k^2\min\{\eta_k - s_m, e_m - \eta_k\} \geq \frac{1}{10}\kappa_k^2\Delta. \tag{4}$$

We may extend (3) to have

$$\begin{aligned}\max_{s_m < t \leq e_m} \widehat{W}_t^{s_m,e_m} &\geq \frac{1}{2}\widetilde{G}_{\eta_k}^{s_m,e_m} - \left(\frac{n}{\Delta}\right)\log^{1+2\xi}(n)\left(n^{1/(2r+1)} + \frac{1}{2}\right) \\ &\geq \frac{1}{20}\kappa_k^2\Delta - \left(\frac{n}{\Delta}\right)\left(\log^{1+2\xi}(n)\right)\left(n^{1/(2r+1)} + \frac{1}{2}\right) \\ &\geq \frac{1}{20}\kappa_k^2\Delta - o\left(\kappa_k^2\Delta\right) > \tau.\end{aligned}$$

where the second last line follows from (4) and the last line follows from Assumption 3. Therefore $\mathcal{M}^{s,e} \neq \emptyset$.

Step 3: This is the localization step. We have $\mathcal{M}^{s,e} \neq \emptyset$. Let $b = b_{m^*}$ be the chosen point in Algorithm 1. Let $(s_{m^*}, e_{m^*}]$ be the corresponding interval. Since it is the narrowest one, we have $(e_{m^*} - s_{m^*}) \leq (e_m - s_m) \leq \Delta$, where $(s_m, e_m]$ is the interval picked at (2). Therefore, $(s_{m^*}, e_{m^*}]$ can contains exactly one change point η_k.

Without loss of generality, let's assume that $b > \eta_k$. Additionally, we shall assume that $(b-\eta_k) > \frac{3}{\kappa_k^2}$; if not, the localization rate follows directly. Since

$$\widehat{W}_b^{s_{m^*},e_{m^*}} \geq \widehat{W}_{\eta_k}^{s_{m^*},e_{m^*}}.$$

We write that

$$\begin{aligned}
&\sum_{j=s_{m^*}+1}^{b} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(s_{m^*},b]}\right\rangle_{\mathcal{L}^2}\right)^2 + \sum_{j=b+1}^{e_{m^*}} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(b,e_{m^*}]}\right\rangle_{\mathcal{L}^2}\right)^2 \\
\leq &\sum_{j=s_{m^*}+1}^{\eta_k} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(s_{m^*},\eta_k]}\right\rangle_{\mathcal{L}^2}\right)^2 + \sum_{j=\eta_k+1}^{e_{m^*}} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(\eta_k,e_{m^*}]}\right\rangle_{\mathcal{L}^2}\right)^2,
\end{aligned}$$

which is equivalent to

$$\left(\left(\frac{b-\eta_k}{b-s_{m^*}}\right)+1\right)^2 \sum_{j=\eta_k+1}^{b} \left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}^2 \tag{5}$$

$$\leq \left(\sum_{j=s_{m^*}+1}^{\eta_k} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(s_{m^*},\eta_k]}\right\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=s_{m^*}+1}^{\eta_k} \left(Y_j - \langle X_j, \beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2\right) \tag{6}$$

$$+ \left(\sum_{j=\eta_k+1}^{e_{m^*}} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(\eta_k,e_{m^*}]}\right\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=\eta_k+1}^{e_{m^*}} \left(Y_j - \left\langle X_j, \beta^*_{\eta_{k+1}}\right\rangle_{\mathcal{L}^2}\right)^2\right) \tag{7}$$

$$+ \left(\sum_{j=s_{m^*}+1}^{b} \left(Y_j - \left\langle X_j, \beta^*_{(s_{m^*},b]}\right\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=s_{m^*}+1}^{b} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(s_{m^*},b]}\right\rangle_{\mathcal{L}^2}\right)^2\right) \tag{8}$$

$$+ \left(\sum_{j=b+1}^{e_{m^*}} \left(Y_j - \left\langle X_j, \beta^*_{\eta_{k+1}}\right\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=b+1}^{e_{m^*}} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(b,e_{m^*}]}\right\rangle_{\mathcal{L}^2}\right)^2\right) \tag{9}$$

$$+ 2\left(\frac{b-\eta_k}{b-s_{m^*}}\right)\left(\sum_{j=s_{m^*}+1}^{b} \left\langle X_j, \beta^*_{\eta_k} - \beta^*_{(s_{m^*},b]}\right\rangle_{\mathcal{L}^2} \varepsilon_j\right) \tag{10}$$

$$+ 2\left(\sum_{j=\eta_k+1}^{b} \left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2} \varepsilon_j\right). \tag{11}$$

Therefore we have,

$$(5) \leq |(6)| + |(7)| + |(8)| + |(9)| + |(10)| + |(11)|.$$

Step 3A: the order of magnitude of (6), (7), (8) and (9). Following from the Lemma C.3, we have

$$\begin{aligned}
|(6)| &= O_p\left((n/\Delta)\,(\eta_p - s_{m^*})\delta_{\eta_p - s_{m^*}} \log^{1+\xi}(\eta_p - s_{m^*})\right), \\
|(7)| &= O_p\left((n/\Delta)\,(e_{m^*} - \eta_p)\delta_{e_{m^*} - \eta_p} \log^{1+\xi}(e_{m^*} - \eta_p)\right), \\
|(8)| &= O_p\left((n/\Delta)\,(b - s_{m^*})\delta_{b - s_{m^*}} \log^{1+\xi}(b - s_{m^*})\right),
\end{aligned}$$

$$\left|(9)\right| = O_p\left((n/\Delta)\,(e_{m^*}-b)\delta_{e_{m^*}-b}\log^{1+\xi}(e_{m^*}-b)\right),$$

which lead us to

$$\left|(6)\right| + \left|(7)\right| + \left|(8)\right| + \left|(9)\right| = O_p\left((n/\Delta)\,\Delta^{1/(2r+1)}\log^{1+\xi}(\Delta)\right).$$

Step 3B: the order of magnitude of (10) and (11). Observe that from Lemma G.9 we may have

$$\mathbb{E}\left[\frac{1}{\kappa_k^2}\left|\sum_{j=\eta_k+1}^{t'}\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right|^2\right] = O(t'-\eta_k),$$

and using Lemma J.4, we may write

$$\begin{aligned}&\max_{1\le k\le \mathcal{K}}\max_{1/\kappa_k^2<t'<\eta_{k+1}}\left|\frac{1}{\sqrt{(t'-\eta_k)}(\log^{1+\xi}\left((t'-\eta_k)\kappa_k^2\right)+1)}\frac{1}{\kappa_k}\sum_{j=\eta_k+1}^{t'}\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right|^2\\ =&O_p\left(\mathcal{K}\right).\end{aligned} \tag{12}$$

Following this, we have

$$\left|\left(\sum_{j=\eta_k+1}^{b}\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right)\right| = O_p\left(\sqrt{\mathcal{K}}\sqrt{(b-\eta_p)}\kappa_k\left\{\log^{1+\xi}\left((b-\eta_k)\kappa_k^2\right)+1\right\}\right),$$

which is a bound on (11). Similarly using (12), we get

$$\begin{aligned}&\left|\left(\frac{b-\eta_k}{b-s_{m^*}}\right)\left(\sum_{j=s_{m^*}+1}^{b}\left\langle X_j, \beta^*_{\eta_p} - \beta^*_{(s_{m^*},b]}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right)\right|\\ =&O_p\left(\sqrt{\mathcal{K}}\left[\frac{b-\eta_k}{b-s_{m^*}}\right]\sqrt{(b-s_{m^*})}\kappa_k\left\{\log^{1+\xi}\left((b-s_{m^*})\kappa_k^2\right)+1\right\}\right)\\ =&O_p\left(\sqrt{\mathcal{K}}\sqrt{(b-\eta_k)}\kappa_k\log^{1+\xi}\left((b-s_{m^*})\kappa_k^2\right)\right),\end{aligned}$$

where we use $\frac{b-\eta_p}{b-s_{m^*}} \le 1$ and $\log\left((b-s_{m^*})\kappa_k^2\right) > 1$ in the last line. This bound (10). Therefore

$$\left|(10)\right| + \left|(11)\right| = O_p\left(\sqrt{\mathcal{K}}\sqrt{(b-\eta_p)}\kappa_k\left\{\log^{1+\xi}((b-\eta_p)\kappa_k^2)\right\}\right).$$

Step 3C: the lower bound of (5): Observe that from Lemma G.8 we may have

$$\mathbb{E}\left[\frac{1}{\kappa_k^2}\left|\sum_{j=\eta_k+1}^{t'}\left(\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle^2_{\mathcal{L}^2} - \kappa_k^2\right)\right|^2\right] = O(t'-\eta_k),$$

and using Lemma J.4, we may write

$$\max_{1\le k\le \mathcal{K}}\max_{\frac{1}{\kappa_k^2}+\eta_k<t'<\eta_{k+1}}\left|\frac{1}{\sqrt{(t'-\eta_k)}}\frac{1}{(\log^{1+\xi}((t'-\eta_k)\kappa_k^2)+1)}\frac{1}{\kappa_k}\right. \tag{13}$$

$$
\begin{aligned}
&\sum_{j=\eta_k+1}^{t'} \left(\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k} \right\rangle^2_{\mathcal{L}^2} - \kappa_k^2 \right) \Bigg|^2 \\
=&O_p\left(\mathcal{K}\right). \qquad (14)
\end{aligned}
$$

Following (13), we may write

$$
\begin{aligned}
(5) =& \left(\frac{b-\eta_k}{b-s_{m^*}}+1\right)^2 \sum_{j=\eta_k+1}^{b} \left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k} \right\rangle^2_{\mathcal{L}^2} \\
=& \left(\frac{b-\eta_k}{b-s_{m^*}}+1\right)^2 \left[(b-\eta_k)\kappa_k^2 - O_p\left(\sqrt{\mathcal{K}}\sqrt{(b-\eta_k)}\kappa_k \left\{\log^{1+\xi}((b-\eta_k)\kappa_k^2)+1\right\}\right)\right] \\
\geq& (b-\eta_k)\kappa_k^2 - O_p\left(\sqrt{\mathcal{K}}\sqrt{(b-\eta_k)\kappa_k^2}\left(\log^{1+\xi}((b-\eta_k)\kappa_k^2)\right)\right),
\end{aligned}
$$

where we use $\frac{b-\eta_k}{b-s_{m^*}}+1 \geq 1$ and $\log\left((b-s_{m^*})\kappa_k^2\right) > 1$ in the last line.

Following from step 3A, step 3B and step 3C, we get

$$
\begin{aligned}
&(b-\eta_k)\kappa_p^2 - O_p\left(\sqrt{\mathcal{K}}\sqrt{(b-\eta_k)}\kappa_k\left\{\log^{1+\xi}(b-\eta_k)\kappa_k^2)\right\}\right) \\
\leq& O_p\left((n/\Delta)\Delta^{1/(2r+1)}\log^{1+\xi}(\Delta)\right) + O_p\left(\sqrt{\mathcal{K}}\sqrt{(b-\eta_k)}\kappa_k\left\{\log^{1+\xi}(b-\eta_k)\kappa_k^2)\right\}\right),
\end{aligned}
$$

with $\mathcal{K} \leq n/\Delta$, it implies

$$
(b-\eta_k)\kappa_p^2 = O_p\left((n/\Delta)\Delta^{1/(2r+1)}\log^{1+\xi}(\Delta)\right). \qquad (15)
$$

This concludes the induction step when $(s,e]$ contains an undetected change point. □

C.1 Technical results for the proof of Theorem 1

Lemma C.1. *Let $(s,e] \subset (0,n]$ be given. Let η_k be a point in $(s,e]$. Suppose $\min\{\eta_k - s, e-\eta_k\} > 4\Delta/5$. Then there exists an interval $(s_m, e_m] \in \mathcal{J} \cap (s,e]$ containing η_k such that*

$$
e_m - s_m \leq \Delta, \qquad \textit{and} \qquad \min\{\eta_k - s_m, e_m - \eta_k\} \geq \Delta/5.
$$

Proof. There are at most two intervals in each layer $\mathcal{J}_k$, for $1 \leq k \leq M$, that contains any given point. We shall consider the layer with $\mathfrak{l}_k = \Delta$ and $\mathfrak{b}_k = \Delta/2$. Without loss of generality, let $\left(\frac{(i-1)\Delta}{2}, \frac{(i-1)\Delta}{2}+\Delta\right]$ and $\left(\frac{i\Delta}{2}, \frac{i\Delta}{2}+\Delta\right]$ are intervals containing η_k.

Case I: Suppose $\eta_k - i\Delta/2 > (i+1)\Delta/2 - \eta_k$. Observe that $\eta_k - i\Delta/2 \geq \Delta/4$. The interval $(s_m, e_m] = \left(\frac{i\Delta}{2}, \frac{i\Delta}{2}+\Delta\right]$ satisfies the required property because $\eta_k - s_m = \eta_k - i\Delta/2 \geq \Delta/4$ and $e_m - \eta_k > i\Delta/2 + \Delta - ((i-2)\Delta/2 + \Delta) = \Delta/2$.

Case II: Suppose $\eta_k - i\Delta/2 \leq (i+1)\Delta/2 - \eta_k$. Using arguments akin to the previous case, the interval $\left(\frac{(i-1)\Delta}{2}, \frac{(i-1)\Delta}{2}+\Delta\right]$ emerges as the necessary interval. □

C.1.1 Large probability event

Recall for any $a > 0$, $\delta_a \asymp a^{-2r/(2r+1)}$.

Lemma C.2. *Let $\xi > 0$. Then, as $n \to \infty$, we have*

$$P\Bigg(\forall (s_m, e_m] \in \mathcal{J}, \quad \forall t \in (s_m, e_m], \\ \left|\widehat{W}_t^{s_m,e_m} - \widetilde{G}_t^{s_m,e_m}\right| - 0.5\widetilde{G}_t^{s_m,e_m} \le \left(\frac{n}{\Delta}\right) \log^{1+2\xi}(n) \left(n^{1/(2r+1)} + 0.5\right)\Bigg) \to 1.$$

Proof. Let $(s_m, e_m] \in \mathcal{J}$ be fixed. For notational simplicity, denote $s = s_m$ and $e = e_m$. Denote

$$W_t^{*s,e} = \sum_{j=s+1}^{e} \left(Y_j - \langle X_j, \beta^*_{(s,e]}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=s+1}^{t} \left(Y_j - \langle X_j, \beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=t+1}^{e} \left(Y_j - \langle X_j, \beta^*_{(t,e]}\rangle_{\mathcal{L}^2}\right)^2.$$

We show in Step 1 that

$$\max_{s<t\le e} \left|\widehat{W}_t^{s,e} - W_t^{*s,e}\right| = O_p\left((e-s)^{1/(2r+1)} \log^{1+\xi}(e-s)\right). \tag{16}$$

In Step 2, we show that

$$\max_{s<t\le e} \frac{1}{\sqrt{\widetilde{G}_t^{s,e} \log^{1+\xi}(t-s)}} \left|W_t^{*s,e} - \widetilde{G}_t^{s,e}\right| = O_p(1), \tag{17}$$

when $\widetilde{G}_t^{s,e} \neq 0$. It follows from using $4ab \le (a+b)^2$ at (17) that

$$\max_{s<t\le e} \frac{1}{0.5\left(\widetilde{G}_t^{s,e} + \log^{1+\xi}(t-s)\right)} \left|W_t^{*s,e} - \widetilde{G}_t^{s,e}\right| = O_p(1).$$

Therefore,

$$P\left(\forall t \in (s,e], \left|\widehat{W}_t^{s,e} - \widetilde{G}_t^{s,e}\right| - 0.5\widetilde{G}_t^{s,e} \le \left(n^{1/(2r+1)} + 0.5\right) \log^{1+2\xi}(n)\right) \to 1, \tag{18}$$

as $n \to \infty$. The factor $\log^{\xi}(n)$ is to make the event asymptotically almost surely. When $(s,e]$ has no change point, we have $W_t^{*s,e} = \widetilde{G}_t^{s,e} = 0$ and (18) trivially holds. Following the cardinality of $\mathcal{J}$ at (2.8), the main result now follows from the union bound.

Step 1: Using $(a-b)^2 - (a-c)^2 = (b-c)^2 - 2(a-c)(b-c)$, we may write

$$\begin{aligned}
&\widehat{W}_t^{s,e} - W_t^{*s,e} \\
&= \underbrace{\sum_{j=s+1}^{e} \left\langle X_j, \widehat{\beta}_{(s,e]} - \beta^*_{(s,e]}\right\rangle^2_{\mathcal{L}^2}}_{\mathcal{B}_1} - \underbrace{\sum_{j=s+1}^{t} \left\langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\right\rangle^2_{\mathcal{L}^2}}_{\mathcal{B}_2} \\
&\quad - \underbrace{\sum_{j=t+1}^{e} \left\langle X_j, \widehat{\beta}_{(t,e]} - \beta^*_{(t,e]}\right\rangle^2_{\mathcal{L}^2}}_{\mathcal{B}_3} + 2\underbrace{\sum_{j=s+1}^{e} \left\langle X_j, \beta^*_{(s,e]} - \widehat{\beta}_{(s,e]}\right\rangle_{\mathcal{L}^2} \varepsilon_j}_{\mathcal{B}_4} \\
&\quad - 2\underbrace{\sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(s,t]} - \widehat{\beta}_{(s,t]}\right\rangle_{\mathcal{L}^2} \varepsilon_j}_{\mathcal{B}_5} - 2\underbrace{\sum_{j=t+1}^{e} \left\langle X_j, \beta^*_{(t,e]} - \widehat{\beta}_{(t,e]}\right\rangle_{\mathcal{L}^2} \varepsilon_j}_{\mathcal{B}_6}
\end{aligned}$$

$$
\underbrace{+2\sum_{j=s+1}^{e}\left\langle X_j,\beta^*_{(s,e]}-\widehat{\beta}_{(s,e]}\right\rangle_{\mathcal{L}^2}\left\langle X_j,\beta^*_j-\beta^*_{(s,e]}\right\rangle_{\mathcal{L}^2}}_{\mathcal{B}_7}
$$

$$
\underbrace{-2\sum_{j=s+1}^{t}\left\langle X_j,\beta^*_{(s,t]}-\widehat{\beta}_{(s,t]}\right\rangle_{\mathcal{L}^2}\left\langle X_j,\beta^*_j-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}}_{\mathcal{B}_9}
$$

$$
\underbrace{-2\sum_{j=t+1}^{e}\left\langle X_j,\beta^*_{(t,e]}-\widehat{\beta}_{(t,e]}\right\rangle_{\mathcal{L}^2}\left\langle X_j,\beta^*_j-\beta^*_{(t,e]}\right\rangle_{\mathcal{L}^2}}_{\mathcal{B}_9}. \tag{19}
$$

We will show the technique to bound $\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_3$ and the result for $\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6$ and $\mathcal{B}_7,\mathcal{B}_8,\mathcal{B}_9$ follows from the same outlined idea and the corresponding Lemma G.3 and Lemma G.4 respectively.

Observe that for $|\mathcal{B}_2|$

$$
\begin{aligned}
&\max_{s<t\le e}\sum_{j=s+1}^{t}\left\langle X_j,\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right\rangle^2_{\mathcal{L}^2}\\
=&\max_{s<t\le e}(t-s)\widehat{\Sigma}_{(s,t]}\left[\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]},\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right]\\
\le&\max_{s<t\le e}\left((t-s)^{1/(2r+1)}\log^{1+\xi}(t-s)\right)\max_{s<t\le e}\left(\frac{\delta^{-1}_{t-s}}{\log^{1+\xi}(t-s)}\widehat{\Sigma}_{(s,t]}\left[\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]},\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right]\right)\\
=&(s-e)^{1/(2r+1)}\log^{1+\xi}(e-s)O_p(1),
\end{aligned}
$$

where the last line follows from the fact that $z\mapsto z^a\log z$ is strictly increasing for any $a\ge 0$ and the Lemma G.2.

For $|\mathcal{B}_1|$, at $t=e$, we have

$$
\sum_{j=s+1}^{e}\left\langle X_j,\widehat{\beta}_{(s,e]}-\beta^*_{(s,e]}\right\rangle^2_{\mathcal{L}^2}=(e-s)^{1/(2r+1)}\log^{1+\xi}(e-s)O_p(1).
$$

The bound for the term $|\mathcal{B}_3|$ follows by same arguement as $\mathcal{B}_1$. This establish (16).

Step 2: Let $\widetilde{G}^{s,e}_t\neq 0$. Note that

$$
\beta^*_{(s,t]}-\beta^*_{(s,e]}=\left(\frac{e-t}{e-s}\right)\left(\beta^*_{(s,t]}-\beta^*_{(t,e]}\right)
$$

and

$$
\beta^*_{(t,e]}-\beta^*_{(s,e]}=\left(\frac{t-s}{e-s}\right)\left(\beta^*_{(t,e]}-\beta^*_{(s,t]}\right).
$$

Using $(a-b)^2-(a-c)^2=(b-c)^2-2(a-c)(b-c)$, we may write

$$
W^{*s,e}_t=\sum_{j=s+1}^{t}\left\langle X_j,\beta^*_{(s,t]}-\beta^*_{(s,e]}\right\rangle^2_{\mathcal{L}^2}+\sum_{j=t+1}^{e}\left\langle X_j,\beta^*_{(t,e]}-\beta^*_{(s,e]}\right\rangle^2_{\mathcal{L}^2}
$$

$$
\begin{aligned}
&+ 2 \sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(s,e]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \varepsilon_j + 2 \sum_{j=t+1}^{e} \left\langle X_j, \beta^*_{(s,e]} - \beta^*_{(t,e]} \right\rangle_{\mathcal{L}^2} \varepsilon_j \\
&+ 2 \sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(s,e]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \\
&+ 2 \sum_{j=t+1}^{e} \left\langle X_j, \beta^*_{(s,e]} - \beta^*_{(t,e]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \beta^*_{(t,e]} \right\rangle_{\mathcal{L}^2} \\
= &\left(\frac{e-t}{e-s}\right)^2 \sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]} \right\rangle^2_{\mathcal{L}^2} + \left(\frac{t-s}{e-s}\right)^2 \sum_{j=t+1}^{e} \left\langle X_j, \beta^*_{(t,e]} - \beta^*_{(s,t]} \right\rangle^2_{\mathcal{L}^2} \\
&+ 2 \left(\frac{e-t}{e-s}\right) \sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(t,e]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \varepsilon_j + 2 \left(\frac{t-s}{e-s}\right) \sum_{j=t+1}^{e} \left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]} \right\rangle_{\mathcal{L}^2} \varepsilon_j \\
&+ 2 \left(\frac{e-t}{e-s}\right) \sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(t,e]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \\
&+ 2 \left(\frac{t-s}{e-s}\right) \sum_{j=t+1}^{e} \left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \beta^*_{(t,e]} \right\rangle_{\mathcal{L}^2}.
\end{aligned}
$$

Observe that

$$
\widetilde{G}^{s,e}_t = \left(\frac{(t-s)^2(e-t)}{(e-s)^2} + \frac{(t-s)(e-t)^2}{(e-s)^2} \right) \Sigma \left[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]} \right].
$$

Also, $\left(\frac{e-t}{e-s}\right)^2 \leq 1$, $\left(\frac{t-s}{e-s}\right)^2 \leq 1$, $\sum_{j=s+1}^{t} \Sigma[\beta^*_{(t,e]} - \beta^*_{(s,t]}, \beta^*_j - \beta^*_{(s,t]}] = 0$ and $\sum_{j=t+1}^{e} \Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_j - \beta^*_{(t,e]}] = 0$. Using the triangle inequality, we may write

$$
\begin{aligned}
&\left| W^{*s,e}_t - \widetilde{G}^{s,e}_t \right| \\
\leq &\left(\frac{e-t}{e-s}\right) \left| \sum_{j=s+1}^{t} \left(\left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]} \right\rangle^2_{\mathcal{L}^2} - \Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]}] \right) \right| &(20)\\
+ &\left(\frac{t-s}{e-s}\right) \left| \sum_{j=t+1}^{e} \left(\left\langle X_j, \beta^*_{(t,e]} - \beta^*_{(s,t]} \right\rangle^2_{\mathcal{L}^2} - \Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]}] \right) \right| &(21)\\
+2 &\left(\frac{e-t}{e-s}\right) \left| \sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(t,e]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \varepsilon_j \right| + 2 \left(\frac{t-s}{e-s}\right) \left| \sum_{j=t+1}^{e} \left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]} \right\rangle_{\mathcal{L}^2} \varepsilon_j \right| &(22)\\
+2 &\left(\frac{e-t}{e-s}\right) \left| \sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(t,e]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} - \Sigma[\beta^*_{(t,e]} - \beta^*_{(s,t]}, \beta^*_j - \beta^*_{(s,t]}] \right| &(23)\\
+2 &\left(\frac{t-s}{e-s}\right) \left| \sum_{j=t+1}^{e} \left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \beta^*_{(t,e]} \right\rangle_{\mathcal{L}^2} - \Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_j - \beta^*_{(t,e]}] \right|. &(24)
\end{aligned}
$$

Our approach involves bounding each of the six terms through four distinct sub-steps. In Step 2A, we establish the bound for equations (20) and (21). Progressing to Step 2B, we derive the bound for equation (22). Moving on to Step 2C, we obtain the bound for equations (23) and (24). Notably, all these derived bounds are uniform across $t \in (s, e]$. The final step, Step 2D, amalgamates these outcomes into a coherent result.

Step 2A. Using Lemma G.8 we have

$$\begin{aligned}&\mathbb{E}\left[\left|\sum_{j=s+1}^{t}\left(\left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]}\right\rangle^2_{\mathcal{L}^2} - \Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]}]\right)\right|^2\right] \\ =&O(t-s)\Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]}].\end{aligned}$$

Writing $\Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]}] = \frac{(e-s)}{(t-s)(e-t)}\widetilde{G}^{s,e}_t$, we may also write it as

$$\mathbb{E}\left[\frac{(e-t)(t-s)}{(e-s)}\frac{1}{\widetilde{G}^{s,e}_t}\left|\sum_{j=s+1}^{t}\left(\left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]}\right\rangle^2_{\mathcal{L}^2} - \Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]}]\right)\right|^2\right] = O(t-s)$$

Using the Lemma J.3, we may write

$$\begin{aligned}\mathbb{E}\Bigg[&\max_{s<t\le e}\frac{(e-t)}{(e-s)}\frac{1}{\log^{1+\xi}(t-s)}\frac{1}{\widetilde{G}^{s,e}_t} \\ &\left|\sum_{j=s+1}^{t}\left(\left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]}\right\rangle^2_{\mathcal{L}^2} - \Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]}]\right)\right|^2\Bigg] = O(1)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\Bigg[&\max_{s<t\le e}\frac{(t-s)}{(e-s)}\frac{1}{\log^{1+\xi}(t-s)}\frac{1}{\widetilde{G}^{s,e}_t} \\ &\left|\sum_{j=t+1}^{e}\left(\left\langle X_j, \beta^*_{(t,e]} - \beta^*_{(s,t]}\right\rangle^2_{\mathcal{L}^2} - \Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]}]\right)\right|^2\Bigg] = O(1).\end{aligned}$$

This lead us to

$$(20) + (21) = O_p\left(\sqrt{\widetilde{G}^{s,e}_t \log^{1+\xi}(t-s)}\right).$$

Step 2B. Using Lemma G.9, we may have

$$\mathbb{E}\left[\frac{(e-t)(t-s)}{(e-s)}\frac{1}{\widetilde{G}^{s,e}_t}\left|\sum_{j=s+1}^{t}\left\langle X_j, \beta^*_{(t,e]} - \beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right|^2\right] = O(t-s).$$

And again from Lemma J.3, it follows that

$$\begin{aligned}&\mathbb{E}\left[\max_{s<t\le e}\frac{(e-t)}{(e-s)\log^{1+\xi}(t-s)}\frac{1}{\widetilde{G}^{s,e}_t}\left|\sum_{j=s+1}^{t}\left\langle X_j, \beta^*_{(t,e]} - \beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right|^2\right] = O(1) \\ \Longrightarrow\ &\mathbb{E}\left[\max_{s<t\le e}\frac{(t-s)}{(e-s)\log^{1+\xi}(t-s)}\frac{1}{\widetilde{G}^{s,e}_t}\left|\sum_{j=t+1}^{e}\left\langle X_j, \beta^*_{(s,t]} - \beta^*_{(t,e]}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right|^2\right] = O(1),\end{aligned}$$

This lead us to

$$(22) = O_p\left(\sqrt{\widetilde{G}_t^{s,e}\log^{1+\xi}(t-s)}\right).$$

Step 2C. Using Lemma G.9, we may have

$$\mathbb{E}\left[\frac{(e-t)(t-s)}{(e-s)}\frac{1}{\widetilde{G}_t^{s,e}} \quad \left|\sum_{j=s+1}^{t}\left\langle X_j, \beta^*_{(t,e]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\left\langle X_j, \beta^*_j-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\right.\right.$$
$$\left.\left. - \Sigma[\beta^*_{(t,e]}-\beta^*_{(s,t]}, \beta^*_j-\beta^*_{(s,t]}]\right|^2\right] = O(t-s).$$

And again from Lemma J.3, it follows that

$$\mathbb{E}\left[\max_{s<t\le e}\frac{(e-t)}{(e-s)}\frac{(\widetilde{G}_t^{s,e})^{-1}}{\log^{1+\xi}(t-s)}\right.$$
$$\left.\left|\sum_{j=s+1}^{t}\left\langle X_j, \beta^*_{(t,e]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\left\langle X_j, \beta^*_j-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2} - \Sigma[\beta^*_{(t,e]}-\beta^*_{(s,t]}, \beta^*_j-\beta^*_{(s,t]}]\right|^2\right] = O(1),$$

and

$$\mathbb{E}\left[\max_{s<t\le e}\frac{(t-s)}{(e-s)}\frac{(\widetilde{G}_t^{s,e})^{-1}}{\log^{1+\xi}(t-s)}\right.$$
$$\left.\left|\sum_{j=t+1}^{e}\left\langle X_j, \beta^*_{(s,t]}-\beta^*_{(t,e]}\right\rangle_{\mathcal{L}^2}\left\langle X_j, \beta^*_j-\beta^*_{(t,e]}\right\rangle_{\mathcal{L}^2} - \Sigma[\beta^*_{(s,t]}-\beta^*_{(t,e]}, \beta^*_j-\beta^*_{(t,e]}]\right|^2\right] = O(1),$$

This lead us to

$$(23) + (24) = O_p\left(\sqrt{\widetilde{G}_t^{s,e}\log^{1+\xi}(t-s)}\right).$$

Step 2D. Combining the results in Step 2A, Step 2B and Step 2C, we get

$$\max_{s<t\le e}\frac{1}{\sqrt{\widetilde{G}_t^{s,e}\log^{1+\xi}(t-s)}}\left|W_t^{*s,e}-\widetilde{G}_t^{s,e}\right| = O_p(1).$$

□

Lemma C.3. *Let $\xi>0$ and $(s,e]\subset(0,n]$. Suppose $\eta_{k-1}<s<\eta_k<e<\eta_{k+1}$. Then we have uniformly for all $t\in(s,e]$,*

$$\sum_{j=s+1}^{t}\left(Y_j-\left\langle X_j,\widehat{\beta}_{(s,t]}\right\rangle_{\mathcal{L}^2}\right)^2-\sum_{j=s+1}^{t}\left(Y_j-\left\langle X_j,\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\right)^2$$
$$=O_p\left((t-s)\delta_{t-s}\log^{1+\xi}(t-s)\right), \tag{25}$$
$$\sum_{j=t+1}^{e}\left(Y_j-\left\langle X_j,\widehat{\beta}_{(t,e]}\right\rangle_{\mathcal{L}^2}\right)^2-\sum_{j=t+1}^{e}\left(Y_j-\left\langle X_j,\beta^*_{(t,e]}\right\rangle_{\mathcal{L}^2}\right)^2$$
$$=O_p\left((e-t)\delta_{e-t}\log^{1+\xi}(e-t)\right). \tag{26}$$

Consequently, following from the union bound we have uniformly for all $(s_m, e_m] \in \mathcal{J}$ and for all $t \in (s_m, e_m]$

$$\sum_{j=s_m+1}^{t} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(s_m,t]} \right\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=s_m+1}^{t} \left(Y_j - \left\langle X_j, \beta^*_{(s_m,t]} \right\rangle_{\mathcal{L}^2}\right)^2$$
$$= O_p\left((n/\Delta)\,(t-s_m)\delta_{t-s_m} \log^{1+\xi}(t-s_m) \right). \tag{27}$$

Proof. Observe that

$$\sum_{j=s+1}^{t} \left(Y_j - \left\langle X_j, \widehat{\beta}_{(s,t]} \right\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=s+1}^{t} \left(Y_j - \left\langle X_j, \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2}\right)^2$$
$$= \sum_{j=s+1}^{t} \left\langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2}^2 + 2\sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(s,t]} - \widehat{\beta}_{(s,t]} \right\rangle_{\mathcal{L}^2} \varepsilon_j$$
$$+ 2\sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(s,t]} - \widehat{\beta}_{(s,t]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \widehat{\beta}_{(s,t]} \right\rangle_{\mathcal{L}^2}.$$

Following from Lemma G.2, we have uniformly

$$\left| \sum_{j=s+1}^{t} \left\langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2}^2 \right| = O_p\left((t-s)\delta_{t-s} \log^{1+\xi}(t-s) \right).$$

From Lemma G.3

$$\left| \sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(s,t]} - \widehat{\beta}_{(s,t]} \right\rangle_{\mathcal{L}^2} \varepsilon_j \right| = O_p\left((t-s)\delta_{t-s} \log^{1+\xi}(t-s) \right).$$

From Lemma G.4

$$\left| \sum_{j=s+1}^{t} \left\langle X_j, \beta^*_{(s,t]} - \widehat{\beta}_{(s,t]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \widehat{\beta}_{(s,t]} \right\rangle_{\mathcal{L}^2} \right| = O_p\left((t-s)\delta_{t-s} \log^{1+\xi}(t-s) \right).$$

The (25) of this lemma follows from these three bounds. Given the cardinality of $\mathcal{J}$ in (2.8), the expression (27) follows from (25) by the union bound. □

C.1.2 Population CUSUM of functional data

All the notation used in this subsection are specific to this subsection only. We use these general results to prove some results earlier.

Assumption 1. *Let $\{\mathfrak{f}_i\}_{i=1}^{m} \in \mathcal{L}^2$. Assume there are $\{\mathfrak{n}_p\}_{p=0}^{K+1} \subset \{0, 1, \ldots, m\}$ such that $0 = \mathfrak{n}_0 < \mathfrak{n}_1 < \ldots < \mathfrak{n}_K < \mathfrak{n}_{K+1} = m$ and*

$$\mathfrak{f}_t \neq \mathfrak{f}_{t+1} \qquad \textit{if and only if} \qquad t \in \{\mathfrak{n}_1, \ldots, \mathfrak{n}_p\}.$$

Let $\inf_{1\le p\le K}\|\mathfrak{f}_{\mathfrak{n}_p}-\mathfrak{f}_{\mathfrak{n}_{p+1}}\|^2_{\mathcal{L}^2}=\inf_{1\le p\le K}\mathfrak{K}_p^2=\mathfrak{K}^2$.

For $0\le s<t<e\le m$, the CUSUM statistics is

$$\widetilde{\mathfrak{f}}_t^{s,e}=\sqrt{\frac{e-t}{(e-s)(t-s)}}\sum_{i=s+1}^{t}\mathfrak{f}_i-\sqrt{\frac{t-s}{(e-s)(e-t)}}\sum_{i=t+1}^{e}\mathfrak{f}_i. \tag{28}$$

It can be easily shown that the CUSUM statistics at (28) are translational invariant. Consequently assuming $\sum_{i=1}^m\mathfrak{f}_i=0$, we may also write

$$\widetilde{\mathfrak{f}}_t^{s,e}=\left(\sum_{i=s+1}^{t}\mathfrak{f}_i-\frac{t}{e-s}\sum_{i=s+1}^{e}\mathfrak{f}_i\right)/\sqrt{\frac{(t-s)(e-t)}{(e-s)}}=\left(\sum_{i=s+1}^{t}\mathfrak{f}_i\right)/\sqrt{\frac{(t-s)(e-t)}{(e-s)}}. \tag{29}$$

The form at (29) is useful proving many important properties of CUSUM.

The Lemma C.4 below follows directly from the definition of CUSUM statistics.

Lemma C.4. *Suppose $(s,e]$ contains only one change point $\mathfrak{n}_p$, then*

$$\|\widetilde{\mathfrak{f}}_t^{s,e}\|^2_{\mathcal{L}^2}=\begin{cases}\frac{t-s}{(e-s)(e-t)}(e-\mathfrak{n}_p)^2\mathfrak{K}_p^2, & t\le\mathfrak{n}_p\\ \frac{e-t}{(e-s)(t-s)}(\mathfrak{n}_p-s)^2\mathfrak{K}_p^2, & t\ge\mathfrak{n}_p.\end{cases}$$

Consequently, we may write

$$\max_{s<t\le e}\|\widetilde{\mathfrak{f}}_t^{s,e}\|^2_{\mathcal{L}^2}=\frac{(e-\mathfrak{n}_p)(\mathfrak{n}_p-s)}{(e-s)}\mathfrak{K}_p^2.$$

Lemma C.5. *Let $(s,e]$ be such that*

$$\mathfrak{n}_{p-1}\le s<\mathfrak{n}_p<e.$$

Then for any $s<t\le\mathfrak{n}_p$,

$$\|\widetilde{\mathfrak{f}}_t^{s,e}\|^2_{\mathcal{L}^2}=\frac{(t-s)(e-\mathfrak{n}_p)}{(\mathfrak{n}_p-s)(e-t)}\|\widetilde{f}_{\mathfrak{n}_p}^{s,e}\|^2_{\mathcal{L}^2}.$$

Consequently, we may write

$$\max_{s<t\le e}\|\widetilde{\mathfrak{f}}_t^{s,e}\|^2_{\mathcal{L}^2}=\max_{\mathfrak{n}_p\le t\le e}\|\widetilde{\mathfrak{f}}_t^{s,e}\|^2_{\mathcal{L}^2}. \tag{30}$$

Proof. With the form outlined at (29)

$$\begin{aligned}\|\widetilde{\mathfrak{f}}_t^{s,e}\|^2_{\mathcal{L}^2}&=\frac{e-s}{(t-s)(e-t)}\left\|\sum_{i=s+1}^{t}\mathfrak{f}_i\right\|^2_{\mathcal{L}^2}=\frac{(e-s)(t-s)^2}{(t-s)(e-t)}\|\mathfrak{f}_{\mathfrak{n}_p}\|^2_{\mathcal{L}^2}\\&=\frac{(t-s)(e-\mathfrak{n}_p)}{(\mathfrak{n}_p-s)(e-t)}\frac{(e-s)}{(\mathfrak{n}_p-s)(e-\mathfrak{n}_p)}(\mathfrak{n}_p-s)^2\|\mathfrak{f}_{\mathfrak{n}_p}\|^2_{\mathcal{L}^2}=\frac{(t-s)(e-\mathfrak{n}_p)}{(\mathfrak{n}_p-s)(e-t)}\|\widetilde{\mathfrak{f}}_t^{s,e}\|^2_{\mathcal{L}^2}\end{aligned}$$

□

Lemma C.6. *Let $(s, e]$ contains exactly two change points $\mathfrak{n}_p$ and $\mathfrak{n}_{p+1}$. Then*

$$\max_{s<t<e} \|\widetilde{\mathfrak{f}}_t^{s,e}\|_{\mathcal{L}^2}^2 \leq 2(e - \mathfrak{n}_{p+1})\mathfrak{K}_{p+1}^2 + 2(\mathfrak{n}_p - s)\mathfrak{K}_p^2.$$

Proof. Let

$$\mathfrak{g}_t = \begin{cases} \mathfrak{f}_{\mathfrak{n}_{p+1}}, & \text{if} \quad s \leq t \leq \mathfrak{n}_p \\ \mathfrak{f}_t, & \text{if} \quad \mathfrak{n}_p + 1 \leq t \leq \mathfrak{n}_{p+1}. \end{cases}$$

Then $\forall t \geq \mathfrak{n}_r$

$$\begin{aligned} \widetilde{\mathfrak{f}}_t^{s,e} - \widetilde{\mathfrak{g}}_t^{s,e} &= \sqrt{\frac{(e-s)}{(e-t)(t-s)}} \left(\sum_{i=s+1}^{\mathfrak{n}_p} \mathfrak{f}_i - \sum_{i=s+1}^{\mathfrak{n}_p} \mathfrak{g}_i + \sum_{i=\mathfrak{n}_p+1}^{t} \mathfrak{f}_i - \sum_{i=\mathfrak{n}_p+1}^{t} \mathfrak{g}_i \right) \\ &= \sqrt{\frac{(e-s)}{(e-t)(t-s)}} (\mathfrak{n}_p - s)(\mathfrak{f}_{\mathfrak{n}_p} - \mathfrak{f}_{\mathfrak{n}_{p+1}}). \end{aligned}$$

$$\Longrightarrow \|\widetilde{\mathfrak{f}}_t^{s,e} - \widetilde{\mathfrak{g}}_t^{s,e}\|_{\mathcal{L}^2}^2 = \frac{(e-s)(\mathfrak{n}_p - s)}{(e-t)(t-s)} (\mathfrak{n}_p - s)\mathfrak{K}_p^2 \leq (\mathfrak{n}_p - s)\mathfrak{K}_p^2. \tag{31}$$

Observe that

$$\max_{s<t\leq e} \|\widetilde{\mathfrak{g}}_t^{s,e}\|_{\mathcal{L}^2}^2 = \|\widetilde{\mathfrak{g}}_{\mathfrak{n}_{p+1}}^{s,e}\|_{\mathcal{L}^2}^2 = \frac{(e - \mathfrak{n}_{p+1})(\mathfrak{n}_{p+1} - s)}{(e-s)} \mathfrak{K}_{p+1}^2 \leq (e - \mathfrak{n}_{p+1})\mathfrak{K}_{p+1}^2 \tag{32}$$

where the equality follows from the fact that g_t just have one change point and Lemma C.4. Observe that

$$\begin{aligned} \max_{s<t\leq e} \|\widetilde{\mathfrak{f}}_t^{s,e}\|_{\mathcal{L}^2}^2 = \max_{\mathfrak{n}_p\leq t\leq e} \|\widetilde{\mathfrak{f}}_t^{s,e}\|_{\mathcal{L}^2}^2 &\leq 2 \max_{\mathfrak{n}_p\leq t\leq e} \|\widetilde{\mathfrak{f}}_t^{s,e} - \widetilde{\mathfrak{g}}_t^{s,e}\|_{\mathcal{L}^2}^2 + 2 \max_{\mathfrak{n}_p\leq t\leq e} \|\widetilde{\mathfrak{g}}_t^{s,e}\|_{\mathcal{L}^2}^2 \\ &\leq 2(\mathfrak{n}_p - s)\mathfrak{K}_p^2 + 2(e - \mathfrak{n}_{p+1})\mathfrak{K}_{p+1}^2, \end{aligned}$$

where the first line follows from Lemma C.5 and the triangle inequality, and the last line follows from (31) and (32). □

D Proof of Theorem 2

Prior to presenting the proof of the main theorem, we will establish the existence and finiteness of the long-run variance.

Lemma D.1. *Suppose the Assumption 1 hold. For $k \in \{1, \ldots, \mathcal{K}\}$, the long-run variance defined in (3.13) exists and is finite.*

Proof. Denote

$$Z_j = \frac{\langle X_j, \beta_{\eta_k}^* - \beta_{\eta_{k+1}}^* \rangle_{\mathcal{L}^2} \varepsilon_j}{\kappa_k}.$$

Observe that

$$\mathbb{E}\left[|Z_1|^3\right] \leq \sqrt{\mathbb{E}\left[\frac{\langle X_j, \beta_{\eta_k}^* - \beta_{\eta_{k+1}}^* \rangle_{\mathcal{L}^2}^6}{\kappa_k^6}\right]} \sqrt{\mathbb{E}\left[\varepsilon_j^6\right]} = O\left(\mathbb{E}^{3/2}\left[\frac{\langle X_j, \beta_{\eta_k}^* - \beta_{\eta_{k+1}}^* \rangle_{\mathcal{L}^2}^2}{\kappa_k^2}\right]\right) = O(1),$$

where the second last equality follows from Assumption 1. Given that we have $\sum_{k=1}^{\infty} \alpha^{1/3}(k) < \infty$ which is implied by $\sum_{k=1}^{\infty} k^{1/3}\alpha^{1/3}(k) < \infty$ in Assumption 1, all the conditions of Theorem 1.7 of Ibragimov (1962). It follows from therein that $\sigma_\infty^2(k)$ exists and is finite.

□

Sketch of the proof of Theorem 2. We refer to A1 and B1 jointly as uniform tightness. Their proof proceeds in multiple steps where we control diverse errors associated with time series functional linear regression modelling uniformly over the seeded intervals. Let

$$\mathcal{Q}_k^*(t) = \sum_{j=s_k+1}^{t} \left(Y_j - \langle X_j, \beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2 + \sum_{j=t+1}^{e_k} \left(Y_j - \langle X_j, \beta^*_{\eta_{k+1}}\rangle_{\mathcal{L}^2}\right)^2 \tag{33}$$

be the population version of the objective function in (2.9). Observe that $\widetilde{\eta}_k$ is the minimiser of $\mathcal{Q}_k(t)$ and η_k is the minimiser of the $\mathcal{Q}_k^*(t)$. Establishing the limiting distribution in A2, involves understanding the behavior of both $\mathcal{Q}_k^*(\eta_k+t) - \mathcal{Q}_k^*(\eta_k)$ and $\mathcal{Q}_k(\eta_k+t) - \mathcal{Q}_k(\eta_k)$, for fixed t. We show that $\max_t |\mathcal{Q}_k(\eta_k+t) - \mathcal{Q}_k(\eta_k) - \mathcal{Q}_k^*(\eta_k+t) + \mathcal{Q}_k^*(\eta_k)| = o_p(1)$, which in turn hinges on the convergence of $\widehat{\beta}_{(s_k,\widehat{\eta}_k]}$ to $\beta^*_{\eta_k}$ and symmetrically, that of $\widehat{\beta}_{(\widehat{\eta}_k,e_k]}$ to $\beta^*_{\eta_{k+1}}$ in an appropriate norm. This establishes that $\mathcal{Q}_k^*(\eta_k+t) - \mathcal{Q}_k^*(\eta_k)$ and $\mathcal{Q}_k(\eta_k+t) - \mathcal{Q}_k(\eta_k)$ have asymptotically same distribution. We then proceed to show that $\mathcal{Q}_k^*(\eta_k+t) - \mathcal{Q}_k^*(\eta_k)$ converges strongly to $S_k(t)$, and consequently, $\mathcal{Q}_k(\eta_k+t) - \mathcal{Q}_k(\eta_k)$ converges to $S_k(t)$ in distribution.

Finally, we leverage the Argmax continuous mapping theorem (e.g. Theorem 3.2.2 of van der Vaart and Wellner, 1996) to translate the convergence from the functional to the minimizer of the functional, which leads to A2. In this regime, it is noteworthy that t is only taking discrete values, and we are not invoking any central limit theorems.

In the vanishing regime, additional complexities arise. Since κ_k converges to 0, in the light of tightness demonstrated in B1, we invoke the functional CLT and establish that $Q_k^*(\eta_k + t\kappa_k^{-2}) - Q_k^*(\eta_k)$ converges in distribution to a two-sided Brownian motion $\mathbb{W}(t)$, where $1/\kappa_k^2$ acts as a local sample size. The subsequent steps parallels the non-vanishing case but additional intricacies arise due to the convergence behavior as $\kappa_k \to 0$.

Proof of Theorem 2. Let $1 \le k \le \mathcal{K}$ be given. By construction and Lemma D.3, $(s_k, e_k]$ contains only one change point η_k and

$$\eta_k - s_k \ge \Delta/5, \qquad e_k - \eta_k \ge \Delta/5,$$

for large enough n. Recall for any $a > 0$, $\delta_a \asymp a^{-2r/(2r+1)}$.

Let $\widetilde{\eta}_k$ denote the minimiser at (2.9). Without loss of generality assume the minimiser $\widetilde{\eta}_k = \eta_k + \gamma$, with $\gamma > 0$. The results presented here assume that what we establish in Theorem 1 holds.

Uniform tightness: $\kappa_k^2|\widetilde{\eta}_k - \eta_k| = O_p(1)$

Assume $\gamma \ge \max\{1/\kappa_k^2, 2\}$, if not, the uniform tightness follows directly. Let $\mathcal{Q}_k$ be defined as in (2.9). Since $\mathcal{Q}_k(\eta_k+\gamma)$ is a minimum, we may write

$$0 \ge \mathcal{Q}_k(\eta_k+\gamma) - \mathcal{Q}_k(\eta_k) = \sum_{j=\eta_k+1}^{\eta_k+\gamma} \left(Y_j - \langle X_j, \widehat{\beta}_{(s_k,\widehat{\eta}_k]}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=\eta_k+1}^{\eta_k+\gamma} \left(Y_j - \langle X_j, \widehat{\beta}_{(\widehat{\eta}_k,e_k]}\rangle_{\mathcal{L}^2}\right)^2.$$

The preceding inequality is equivalent to

$$0 \ge \left(\sum_{j=\eta_k+1}^{\eta_k+\gamma} \left(Y_j - \langle X_j, \widehat{\beta}_{(s_k,\widehat{\eta}_k]}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=\eta_k+1}^{\eta_k+\gamma} \left(Y_j - \langle X_j, \beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2 \right)$$

$$
\begin{aligned}
&-\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\Big(Y_j-\langle X_j,\widehat{\beta}_{(\widehat{\eta}_k,e_k]}\rangle_{\mathcal{L}^2}\Big)^2-\sum_{j=\eta_k+1}^{\eta_k+\gamma}\Big(Y_j-\langle X_j,\beta^*_{\eta_{k+1}}\rangle_{\mathcal{L}^2}\Big)^2\right)\\
&+\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\big(Y_j-\langle X_j,\beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\big)^2-\sum_{j=\eta_k+1}^{\eta_k+\gamma}\Big(Y_j-\langle X_j,\beta^*_{\eta_{k+1}}\rangle_{\mathcal{L}^2}\Big)^2\right),\\
&=\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\Big(Y_j-\langle X_j,\widehat{\beta}_{(s_k,\widehat{\eta}_k]}\rangle_{\mathcal{L}^2}\Big)^2-\sum_{j=\eta_k+1}^{\eta_k+\gamma}\big(Y_j-\langle X_j,\beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\big)^2\right) &(34)\\
&-\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\Big(Y_j-\langle X_j,\widehat{\beta}_{(\widehat{\eta}_k,e_k]}\rangle_{\mathcal{L}^2}\Big)^2-\sum_{j=\eta_k+1}^{\eta_k+\gamma}\Big(Y_j-\langle X_j,\beta^*_{\eta_{k+1}}\rangle_{\mathcal{L}^2}\Big)^2\right) &(35)\\
&+\left(2\sum_{j=\eta_k+1}^{\eta_k+\gamma}\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\varepsilon_j\right) &(36)\\
&+\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\langle X_j,\beta^*_{\eta_k}-\beta^*_{\eta_{k+1}}\rangle^2_{\mathcal{L}^2}\right). &(37)
\end{aligned}
$$

Therefore, we have

$$(37)\le\big|(36)\big|+\big|(35)\big|+\big|(34)\big|.$$

Recall $\delta_\Delta=\Delta^{-2r/(2r+1)}$. Observe that

$$
\begin{aligned}
\widehat{\eta}_k-s_k\ge\Delta/5 &\implies \delta_{\widehat{\eta}_k-s_k}=O_p\left(\delta_\Delta\right),\\
e_k-\widehat{\eta}_k\ge\Delta/5 &\implies \delta_{e_k-\widehat{\eta}_k}=O_p\left(\delta_\Delta\right).
\end{aligned}
$$

Also using $\gamma\ge 1/\kappa_k^2$ and $r>1$, we have

$$
\begin{aligned}
\delta_\gamma^{3/4}\sqrt{\log^{1+\xi}\gamma}&=O\left(\gamma^{-1/2}\right)=O\left(\kappa_k\right),\\
\delta_\gamma^{1/2}\sqrt{\log^{1+\xi}\gamma}&=O\left(\gamma^{-1/3}\right)=O\left(\kappa_k^{2/3}\right).
\end{aligned}
\tag{38}
$$

From Assumption 3, we get

$$
\begin{aligned}
\delta_{\widehat{\eta}-s}\log^{1+2\xi}(\widehat{\eta}-s)&=O_p\left(\delta_\Delta\log^{1+2\xi}(\Delta)\right)=o_p(\kappa_k^2),\\
\delta_{e-\widehat{\eta}}\log^{1+2\xi}(e-\widehat{\eta})&=O_p\left(\delta_\Delta\log^{1+2\xi}(\Delta)\right)=o_p(\kappa_k^2).
\end{aligned}
\tag{39}
$$

With (38) and (39) we get

$$
\begin{aligned}
\delta_\gamma^{3/4}\sqrt{\log^{1+\xi}\gamma}\left(\delta_{\widehat{\eta}_k-s_k}\right)^{1/4}\sqrt{\log^{1+2\xi}(\widehat{\eta}_k-s_k)}&=o_p(\kappa_k^2)\\
\delta_\gamma^{1/2}\sqrt{\log^{1+\xi}\gamma}\left(\delta_{\widehat{\eta}_k-s_k}\right)^{1/2}\sqrt{\log^{1+2\xi}(\widehat{\eta}_k-s_k)}&=o_p(\kappa_k^2)\\
\delta_\gamma^{3/4}\sqrt{\log^{1+\xi}\gamma}\left(\delta_{e_k-\widehat{\eta}_k}\right)^{1/4}\sqrt{\log^{1+2\xi}(e_k-\widehat{\eta}_k)}&=o_p(\kappa_k^2)\\
\delta_\gamma^{1/2}\sqrt{\log^{1+\xi}\gamma}\left(\delta_{e_k-\widehat{\eta}_k}\right)^{1/2}\sqrt{\log^{1+2\xi}(e_k-\widehat{\eta}_k)}&=o_p(\kappa_k^2).
\end{aligned}
\tag{40}
$$

Also, we have from Theorem 1 that

$$\begin{aligned}\left|\frac{\widehat{\eta}_k-\eta_k}{\widehat{\eta}_k-s}\right| &\lesssim \left|\frac{\widehat{\eta}_k-\eta_k}{\Delta}\right| = o_p(1),\\ \left|\frac{\widehat{\eta}_k-\eta_k}{\widehat{\eta}_k-e}\right| &\lesssim \left|\frac{\widehat{\eta}_k-\eta_k}{\Delta}\right| = o_p(1).\end{aligned} \tag{41}$$

Step 1: the order of magnitude of (34)**.** Following from Lemma D.2, we have

$$\begin{aligned}(34) = & O_p\left(\gamma\delta_\gamma^{1/2}\sqrt{\log^{1+\xi}\gamma}\left(\delta_{\widehat{\eta}_k-s_k}\right)^{1/2}\sqrt{\log^{1+2\xi}(\widehat{\eta}_k-s_k)}\right)\\ &+O_p\left(\gamma\delta_\gamma^{3/4}\sqrt{\log^{1+\xi}\gamma}\left(\delta_{\widehat{\eta}_k-s_k}\right)^{1/4}\sqrt{\log^{1+2\xi}(\widehat{\eta}_k-s_k)}\right)\\ &+O_p\left(\gamma\kappa_k\left(\delta_{\widehat{\eta}_k-s_k}\right)^{1/2}\sqrt{\log^{1+2\xi}(\widehat{\eta}_k-s_k)}\right)+O_p\left(\gamma\delta_{\widehat{\eta}_k-s_k}\log^{1+2\xi}(\widehat{\eta}_k-s_k)\right)\\ &+O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}(\gamma\kappa_k^2)+1\right\}\right)+O_p\left(\gamma\frac{\widehat{\eta}_k-\eta_k}{\widehat{\eta}_k-s_k}\kappa_k^2\right)\\ =& o_p(\gamma\kappa_k^2)+O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}(\gamma\kappa_k^2)+1\right\}\right),\end{aligned} \tag{42}$$

where the last line follows from (40) and (41).

Step 2: the order of magnitude of (35)**.** Following from Lemma D.2, we have

$$\begin{aligned}(35) = & O_p\left(\gamma\delta_\gamma^{1/2}\sqrt{\log^{1+\xi}\gamma}\left(\delta_{e_k-\widehat{\eta}_k}\right)^{1/2}\sqrt{\log^{1+2\xi}(e_k-\widehat{\eta}_k)}\right)\\ &+O_p\left(\gamma\delta_\gamma^{3/4}\sqrt{\log^{1+\xi}\gamma}\left(\delta_{e_k-\widehat{\eta}_k}\right)^{1/4}\sqrt{\log^{1+2\xi}(e_k-\widehat{\eta}_k)}\right)\\ &+O_p\left(\gamma\kappa_k\left(\delta_{e_k-\widehat{\eta}_k}\right)^{1/2}\sqrt{\log^{1+2\xi}(e_k-\widehat{\eta}_k)}\right)+O_p\left(\gamma\delta_{\widehat{\eta}-s}\log^{1+2\xi}(e_k-\widehat{\eta}_k)\right)\\ &+O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}(\gamma\kappa_k^2)+1\right\}\right)+O_p\left(\gamma\frac{\widehat{\eta}_k-\eta_k}{e_k-\widehat{\eta}_k}\kappa_k^2\right)\\ =& o_p(\gamma\kappa_k^2)+O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}(\gamma\kappa_k^2)+1\right\}\right).\end{aligned} \tag{43}$$

where the last line follows from (40) and (41).

Step 3: the order of magnitude of (36)**.** Following from Lemma G.9 and Lemma J.4, we have

$$\max_{1/\kappa_k^2<\gamma<\eta_{k+1}-\eta_k}\left|\frac{1}{\sqrt{\gamma}(\log^{1+\xi}\left((\gamma)\kappa_k^2\right)+1)}\frac{1}{\kappa_k}\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right|^2 = O_p\left(1\right). \tag{44}$$

Using (44), we have

$$(36) = O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}(\gamma\kappa_k^2)+1\right\}\right). \tag{45}$$

Step 4: lower bound of (37)**.** Following from Lemma G.8 and Lemma J.4, we have

$$\max_{1/\kappa_k^2<\gamma<\eta_{k+1}-\eta_k}\left|\frac{1}{\sqrt{\gamma}(\log^{1+\xi}(\gamma\kappa_k^2)+1)}\frac{1}{\kappa_k}\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(\left\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right\rangle^2_{\mathcal{L}^2}-\kappa_k^2\right)\right|^2 = O_p\left(1\right). \tag{46}$$

Using (46), we have

$$(37) \geq \gamma\kappa_k^2 + O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}(\gamma\kappa_k^2)+1\right\}\right). \tag{47}$$

Combining (42), (43), (45) and (47), we have uniformly for all $\gamma \geq \frac{1}{\kappa_k^2}$

$$\gamma\kappa_k^2 + O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}(\gamma\kappa_k^2)+1\right\}\right) \leq O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}(\gamma\kappa_k^2)+1\right\}\right) + o_p(\gamma\kappa_k^2),$$

which gives us

$$\kappa_k^2\left|\widetilde{\eta}_k - \eta_k\right| = O_p(1).$$

Limiting distribution:

Recall the definition of $\mathcal{Q}_k^*(\cdot)$ from (33). For any given $k \in \{1, \ldots, \mathcal{K}\}$, given the end points s_k and e_k and the true coefficients $\beta^*_{\eta_k}$ and $\beta^*_{\eta_{k+1}}$, we have

$$\begin{aligned}(36) + (37) &= \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \beta^*_{\eta_{k+1}}\rangle_{\mathcal{L}^2}\right)^2 \\ &= \mathcal{Q}_k^*(\eta_k+\gamma) - \mathcal{Q}_k^*(\eta_k).\end{aligned}$$

Following from the proof of uniform tightness, we have uniformly in γ, as $n \to \infty$, that

$$\left|\mathcal{Q}_k(\eta_k+\gamma) - \mathcal{Q}_k(\eta_k) - (\mathcal{Q}_k^*(\eta_k+\gamma) - \mathcal{Q}_k^*(\eta_k))\right| \leq \left|(34)\right| + \left|(35)\right| + \left|(36)\right| + \left|(37)\right| \xrightarrow{\mathbb{P}} 0.$$

With Slutsky's theorem, it is sufficient to find the limiting distribution of $\mathcal{Q}_k^*(\eta_k+\gamma) - \mathcal{Q}_k^*(\eta_k)$ when $n \to \infty$.

Non-vanishing regime. For $\gamma > 0$, we have that when $n \to \infty$

$$\begin{aligned}&\mathcal{Q}_k^*(\eta_k+\gamma) - \mathcal{Q}_k^*(\eta_k) = \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \beta^*_{\eta_{k+1}}\rangle_{\mathcal{L}^2}\right)^2 \\ &= \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\{2\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j + \left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle^2_{\mathcal{L}^2}\right\} \\ &\qquad\xrightarrow{\mathcal{D}} \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\{2\varrho_k\left\langle X_j, \Psi_k\right\rangle_{\mathcal{L}^2}\varepsilon_j + \varrho_k^2\left\langle X_j, \Psi_k\right\rangle^2_{\mathcal{L}^2}\right\}.\end{aligned}$$

For $\gamma < 0$, we have when $n \to \infty$

$$\begin{aligned}&\mathcal{Q}_k^*(\eta_k+\gamma) - \mathcal{Q}_k^*(\eta_k) = \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \beta^*_{\eta_{k+1}}\rangle_{\mathcal{L}^2}\right)^2 \\ &= \sum_{j=\eta_k+\gamma+1}^{\eta_k}\left\{2\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j + \left\langle X_j, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right\rangle^2_{\mathcal{L}^2}\right\} \\ &\qquad\xrightarrow{\mathcal{D}} \sum_{j=\eta_k+\gamma+1}^{\eta_k}\left\{-2\varrho_k\left\langle X_j, \Psi_k\right\rangle_{\mathcal{L}^2}\varepsilon_j + \varrho_k^2\left\langle X_j, \Psi_k\right\rangle^2_{\mathcal{L}^2}\right\},\end{aligned}$$

where the last line follows because pointwise convergence implies convergence in $\langle\ ,\ \rangle_{\mathcal{L}^2}$.

From stationarity, the Slutsky's theorem and the Argmax continuous mapping theorem (e.g. Theorem 3.2.2 of van der Vaart and Wellner (1996)), we have

$$\widetilde{\eta}_k - \eta_k \xrightarrow{\mathcal{D}} \arg\min_{\gamma} S_k(\gamma).$$

Vanishing regime. Let $m = \kappa_k^{-2}$, and we have that $m \to \infty$ as $n \to \infty$. For $\gamma > 0$, we have that

$$\mathcal{Q}_k^*(\eta_k + \gamma m) - \mathcal{Q}_k^*(\eta_k) = \sum_{j=\eta_k+1}^{\eta_k+\gamma m} (Y_j - \langle X_j, \beta_{\eta_k}^* \rangle_{\mathcal{L}^2})^2 - \sum_{j=\eta_k+1}^{\eta_k+\gamma m} \left(Y_j - \langle X_j, \beta_{\eta_{k+1}}^* \rangle_{\mathcal{L}^2}\right)^2$$

$$= \sum_{j=\eta_k+1}^{\eta_k+\gamma m} \left\{ 2 \left\langle X_j, \beta_{\eta_{k+1}}^* - \beta_{\eta_k}^* \right\rangle_{\mathcal{L}^2} \varepsilon_j + \left\langle X_j, \beta_{\eta_{k+1}}^* - \beta_{\eta_k}^* \right\rangle_{\mathcal{L}^2}^2 \right\}$$

$$= \frac{2}{\sqrt{m}} \sum_{j=\eta_k+1}^{\eta_k+\gamma m} \left\{ \frac{\left\langle X_j, \beta_{\eta_{k+1}}^* - \beta_{\eta_k}^* \right\rangle_{\mathcal{L}^2}}{\kappa_k} \varepsilon_j \right\} + \frac{1}{m} \sum_{j=\eta_k+1}^{\eta_k+\gamma m} \left\{ \frac{\left\langle X_j, \beta_{\eta_{k+1}}^* - \beta_{\eta_k}^* \right\rangle_{\mathcal{L}^2}^2}{\kappa_k^2} - 1 \right\}$$

$$+ \frac{1}{m} \sum_{j=\eta_k+1}^{\eta_k+\gamma m} 1. \tag{48}$$

Following from the definition of the long-run variance and Theorem I.2, we have

$$\frac{1}{\sigma_\infty(k)} \frac{2}{\sqrt{m}} \sum_{j=\eta_k+1}^{\eta_k+\gamma m} \left\{ \frac{\left\langle X_j, \beta_{\eta_{k+1}}^* - \beta_{\eta_k}^* \right\rangle_{\mathcal{L}^2}}{\kappa_k} \varepsilon_j \right\} \xrightarrow{\mathcal{D}} \mathbb{B}_2(\gamma). \tag{49}$$

We also have

$$\mathbb{E}\left[\left| \frac{1}{m} \sum_{j=\eta_k+1}^{\eta_k+\gamma m} \left\{ \frac{\left\langle X_j, \beta_{\eta_{k+1}}^* - \beta_{\eta_k}^* \right\rangle_{\mathcal{L}^2}^2}{\kappa_k^2} - 1 \right\} \right|^2 \right] = O\left(\frac{\gamma}{m}\right) \to 0, \tag{50}$$

following from (74) in Lemma G.8. Using (50), (49) and $\frac{1}{m}\sum_{j=\eta_k+1}^{\eta_k+\gamma m} 1 \to \gamma$ in (48), we write

$$\mathcal{Q}_k^*(\eta_k + \gamma m) - \mathcal{Q}_k^*(\eta_k) \xrightarrow{\mathcal{D}} \sigma_\infty(k)\mathbb{B}_2(\gamma) + \gamma$$

where $\mathbb{B}_2(\gamma)$ is a standard Brownian motion.

Similarly, for $\gamma < 0$, we may have when $n \to \infty$

$$\mathcal{Q}_k^*(\eta_k + \gamma m) - \mathcal{Q}_k^*(\eta_k) \xrightarrow{\mathcal{D}} -\gamma + \sigma_\infty(k)\mathbb{B}_1(-\gamma),$$

where $\mathbb{B}_1(r)$ is a standard Brownian motion. Let $Z_j^* = \frac{\langle X_j, \beta_{\eta_{k+1}} - \beta_{\eta_k} \rangle_{\mathcal{L}^2} \varepsilon_j}{\kappa_k}$. To see the independence of $\mathbb{B}_1(r)$ and $\mathbb{B}_2(r)$ note that

$$\frac{1}{m}\mathbb{E}\left[\left(\sum_{t=-m\gamma}^{-1} Z_t^* \right) \left(\sum_{t=1}^{m\gamma} Z_t^* \right) \right] = \frac{1}{m} \left\{ \sum_{k=1}^{m\gamma} k\mathbb{E}[Z_1 Z_{1+k}] + \sum_{k=m\gamma+1}^{2m\gamma} (2q-k)\mathbb{E}[Z_1 Z_{1+k}] \right\}$$

$$\leq \frac{1}{m}\sum_{k=1}^{2m\gamma} k\left|\mathbb{E}\left[Z_1 Z_{1+k}\right]\right| \leq \frac{(2m\gamma)^{2/3}}{m}\sum_{k=1}^{2m\gamma} k^{1/3}\|Z_1\|_3^2\alpha^{1/3}(k) = O\left(\frac{1}{m^{1/3}}\right) \to 0,$$

where the second last inequality follows from Lemma I.1 and stationarity and the last inequality follows from $\sum_{k=1}^{\infty} k^{1/3}\alpha^{1/3}(k) < \infty$ and

$$\|Z_1\|_3 = \mathbb{E}^{1/3}\left[\frac{1}{\kappa_k^3}\left\langle X_1, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right\rangle^3 \mathbb{E}\left[\varepsilon_1^3 | X_1\right]\right] \leq O(1)\mathbb{E}^{1/6}\left[\frac{1}{\kappa_k^6}\left\langle X_1, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right\rangle^6\right] = O(1),$$

which follows from Assumption 1.

From the Slutsky's theorem and the Argmax continuous mapping theorem we have

$$\widetilde{\eta}_k - \eta_k \xrightarrow{\mathcal{D}} \arg\min_{\gamma}\left\{|\gamma| + \sigma_\infty(k)\mathbb{W}(\gamma)\right\}.$$

□

D.1 Technical result for the proof of Theorem 2

Recall for any $a > 0$, $\delta_a \asymp a^{-2r/(2r+1)}$.

Lemma D.2. *Let $\eta_{k-1} < s < \eta_k < e < \eta_{k+1}$ be fixed. Let $\xi > 0$. Then,*

$$\max_{\gamma \in \left(1/\kappa_k^2, \eta_{k+1}-\eta_k\right)} \frac{1}{\mathcal{H}(\widehat{\eta}_k - s, \gamma)}\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \widehat{\beta}_{(s,\widehat{\eta}_k]}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2\right) = O_p(1)$$

where for any $t \in (s, e]$

$$\begin{aligned}\mathcal{H}(t-s,\gamma) =& \gamma\left\{\left(\delta_{t-s}^{1/4} + \delta_\gamma^{1/4}\right)\delta_\gamma^{1/2}\sqrt{\log^{1+\xi}\gamma} + \left(\kappa_k + \delta_{t-s}^{1/2}\sqrt{\log^{1+\xi}(t-s)}\right)\delta_{t-s}^{1/4}\right\} \\ & \sqrt{\log^{1+\xi}(t-s)}\delta_{t-s}^{1/4} + \kappa_k\sqrt{\gamma}\left\{\log^{1+\xi}(\gamma\kappa_k^2) + 1\right\} + \left|\frac{t-\eta_k}{t-s}\right|\gamma\kappa_k^2\end{aligned}$$

Proof. Let $t > \eta_k$. The case when $t \leq \eta_k$ follows similar to the proof outlined below. Observe that

$$\beta^*_{(s,t]} - \beta^*_{\eta_k} = \left(\frac{t-\eta_k}{t-s}(\beta^*_{\eta_{k+1}} - \beta^*_{\eta_k})\right).$$

We may write the expression

$$\begin{aligned}&\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \widehat{\beta}_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2\right) \\ =& \left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \widehat{\beta}_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j - \langle X_j, \beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2\right) \qquad (51)\end{aligned}$$

$$+\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2-\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2\right). \tag{52}$$

We show in Step 1 that

$$\max_{\substack{t\in(s,e]\\ \gamma\in\left(1/\kappa_k^2,\eta_{k+1}-\eta_k\right)}}\frac{1}{\mathcal{H}_1(t-s,\gamma)}\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\widehat{\beta}_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2\right.$$
$$\left.-\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2\right)=O_p(1), \tag{53}$$

where

$$\mathcal{H}_1(t-s,\gamma)=\gamma\Bigg\{\left(\delta_{t-s}^{1/4}+\delta_{\gamma}^{1/4}\right)\delta_{\gamma}^{1/2}\sqrt{\log^{1+\xi}(\gamma)}$$
$$+\left(\kappa_k+\delta_{t-s}^{1/2}\sqrt{\log^{1+\xi}(t-s)}\right)\delta_{t-s}^{1/4}\Bigg\}\delta_{t-s}^{1/4}\sqrt{\log^{1+\xi}(t-s)}.$$

We show in Step 2 that

$$\max_{\substack{t\in(s,e]\\ \gamma\in\left(1/\kappa_k^2,\eta_{k+1}-\eta_k\right)}}\frac{1}{\mathcal{H}_2(\gamma)}\left(\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2\right.$$
$$-\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\beta^*_{\eta_k}\rangle_{\mathcal{L}^2}\right)^2 \tag{54}$$
$$\left.-\left\{\left(\frac{\eta_k-s}{t-s}\right)^2-1\right\}\gamma\kappa_k^2\right)=O_p\left(1\right).$$

where

$$\mathcal{H}_2(\gamma)=\kappa_k\sqrt{\gamma}\left\{\log^{1+\xi}\left(\gamma\kappa_k^2\right)+1\right\}.$$

The bound for (51) follows from (53) and the bound for (52) follows from (54) and the realization that $\left\{1-\left(\frac{\eta_k-s}{t-s}\right)^2\right\}\geq\left(1-\frac{\eta_k-s}{t-s}\right)=\left(\frac{t-\eta_k}{t-s}\right)$.

Step 1: Observe that

$$\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\widehat{\beta}_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2-\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2\right|$$
$$=\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(\left\langle X_j,\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}^2\right.\right.$$
$$\left.\left.-2\left\langle X_j,\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\left\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}-2\left\langle X_j,\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right)\right|$$
$$\leq\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\langle X_j,\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}^2\right| \tag{55}$$

$$+\,2\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\langle X_j,\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\left\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\right| \tag{56}$$

$$+\,2\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\langle X_j,\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right|. \tag{57}$$

We are going to bound (55), (56) and (57) in the following three sub-steps. Following from Lemma G.5 we have that

$$(55)=O_p\left(\gamma\left\{\left(\delta_{t-s}^{1/4}+\delta_\gamma^{1/4}\right)\delta_\gamma^{1/2}\sqrt{\log^{1+\xi}(\gamma)}+\delta_{t-s}^{3/4}\sqrt{\log^{1+\xi}(t-s)}\right\}\sqrt{\log^{1+\xi}(t-s)}\delta_{t-s}^{1/4}\right).$$

Following from Lemma G.7, we have that

$$(56)=O_p\left(\gamma\left\{\left(\delta_{t-s}^{1/4}+\delta_\gamma^{1/4}\right)\delta_\gamma^{1/2}\sqrt{\log^{1+\xi}(\gamma)}+\kappa_k\delta_{t-s}^{1/4}\right\}\sqrt{\log^{1+\xi}(t-s)}\delta_{t-s}^{1/4}\right).$$

And following from Lemma G.6, we have that

$$(57)=O_p\left(\left\{1+\left(\frac{\delta_\gamma}{\delta_{t-s}}\right)^{1/4}\right\}\delta_\gamma^{1/2}\sqrt{\log^{1+\xi}(\gamma)}\delta_{t-s}^{1/2}\sqrt{\log^{1+\xi}(t-s)}\right).$$

The stochastic bound (53) now follows directly from these three bounds on (55), (56) and (57).

Step 2: Observe that $\beta^*_{(s,t]}-\beta^*_{\eta_k}=\frac{t-\eta_k}{t-s}(\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k})$. We may write the expansion

$$\begin{aligned}
&\left(\sum_{j=\eta_k}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2-\sum_{j=\eta_k}^{\eta_k+\gamma}(Y_j-\langle X_j,\beta^*_{\eta_k}\rangle_{\mathcal{L}^2})^2-\left\{\left(\frac{\eta_k-s}{t-s}\right)^2-1\right\}\gamma\kappa^2\right)\\
=&\left\{\left(\frac{s-\eta_k}{t-s}\right)^2-1\right\}\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(\left\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}^2-\kappa_k^2\right)\\
&-2\left(\frac{t-\eta_k}{t-s}\right)\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j.
\end{aligned}$$

Consequently, we have that

$$\begin{aligned}
&\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(Y_j-\langle X_j,\beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2-\sum_{j=\eta_k+1}^{\eta_k+\gamma}(Y_j-\langle X_j,\beta^*_{\eta_k+1}\rangle_{\mathcal{L}^2})^2-\left\{\left(\frac{\eta_k-s}{t-s}\right)^2-1\right\}\gamma\kappa_k^2\right|\\
\le&\left|\left\{\left(\frac{\eta_k-s}{t-s}\right)^2-1\right\}\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(\left\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}^2-\kappa_k^2\right)\right|\\
&+2\left|\left(\frac{t-\eta_k}{t-s}\right)\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right|\\
\le&\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(\left\langle X_j,\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}^2-\kappa_k^2\right)\right|
\end{aligned}\tag{58}$$

$$+ 2\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right|, \tag{59}$$

where the last two line follows from $0 \leq \left\{1 - \left(\frac{\eta_k - s}{t-s}\right)^2\right\} \leq 1$ and $\left(\frac{t-\eta_k}{t-s}\right) \leq 1$. For the expression (58), using (46), we get that

$$\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left(\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle^2_{\mathcal{L}^2} - \kappa_k^2\right)\right| = O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}\left(\gamma\kappa_k^2\right) + 1\right\}\right). \tag{60}$$

For the expression (59), we use (44) to have

$$\left|\sum_{j=\eta_k+1}^{\eta_k+\gamma}\left\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right| = O_p\left(\sqrt{\gamma}\kappa_k\left\{\log^{1+\xi}\left(\gamma\kappa_k^2\right) + 1\right\}\right). \tag{61}$$

Bringing (60) and (61) together shall establish (54). □

Lemma D.3. *Let $(s_k, e_k]$ be the refined interval constructed in (2.10). Then, under the event $\mathcal{A}$ defined in (1), η_k is the one and only change point lying in $(s_k, e_k]$. Additionally, under the same event $\mathcal{A}$, we have*

$$\min\left\{e_k - \eta_k, \eta_k - s_k\right\} \geq \Delta/5.$$

Since, event $\mathcal{A}$ is asymptotically almost sure (Lemma C.2). These results holds with probability converging to 1 as $n \to \infty$.

Proof. Observe that, by construction and by the definition of the event $\mathcal{A}$ in (1) as well as Theorem 1, we have

$$e_k - \eta_k = \tfrac{9}{10}\left(\widehat{\eta}_{k+1} - \widehat{\eta}_k\right) + \left(\widehat{\eta}_k - \eta_k\right) \geq \tfrac{9}{10}\Delta + o(\Delta) \geq \frac{\Delta}{5}.$$

Similary,

$$\eta_k - \eta_k = \tfrac{9}{10}\left(\widehat{\eta}_k - \widehat{\eta}_{k-1}\right) + \left(\eta_k - \widehat{\eta}_k\right) \geq \tfrac{9}{10}\Delta + o(\Delta) \geq \frac{\Delta}{5}.$$

□

E Proof of Theorem 3

Proof. Let $\mathcal{P} = \{J_1, J_2, \ldots, J_S\}$. Let $J_1 = \{t_1, t_1+1, \ldots, t_1+(q-1), \ldots, t_1+(2q-1)\}$. Denote $\widetilde{J}_1 = J_1 \setminus (J_1 + q) = \{t_1, t_1, \ldots, t_1 + (q-1)\}$ and $\bar{J}_1 = J_1 \setminus \widetilde{J}_1 = \{t_1+q, \ldots, t_1+(2q-1)\}$ as the two equal partition of the block J_1. Recall that $\delta_a \asymp a^{-2r/2r+1}$ for any $a > 0$. Denote the population version of the process $\{F^*_{J_v}\}_{v=1}^S$ as

$$F^*_{J_1} = \sqrt{\frac{2}{q}}\left\{\sum_{t\in\widetilde{J}_1}\left(Z^*_t - Z^*_{t+q}\right\}\right) = \sqrt{\frac{2}{q}}\left\{\sum_{t\in\widetilde{J}_1} Z^*_t - \sum_{t\in\bar{J}_1} Z^*_t\right\}.$$

where $Z_t^* = \frac{1}{\kappa_k}\left\langle X_t, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right\rangle \varepsilon_t$. Denote

$$\widehat{\beta}_k^{(1)} = \widehat{\beta}_{(s_k, \widehat{\eta}_k]} \qquad \text{and} \qquad \widehat{\beta}_k^{(2)} = \widehat{\beta}_{(\widehat{\eta}_k, e_k]}.$$

This proof is further divided into two steps. Firstly, we establish the consistency of the population version of the estimate. Secondly, we conclude the proof by demonstrating that the deviation of the estimate from the estimator is small in probability. The last redundant step is replacing κ_k with $\widehat{\kappa}_k$ and applying Lemma E.1 along with the Slutsky's theorem.

Step 1a: Note that

$$\mathbb{E}\left[(F^*_{J_1})^2\right] = 2\,\mathbb{E}\left[\left(\frac{1}{\sqrt{q}}\sum_{t\in\widetilde{J}_1} Z_t^*\right)^2\right] + 2\,\mathbb{E}\left[\left(\frac{1}{\sqrt{q}}\sum_{t\in\bar{J}_1} Z_t^*\right)^2\right] - \frac{4}{q}\mathbb{E}\left[\left(\sum_{t\in\widetilde{J}_1} Z_t^*\right)\left(\sum_{t\in\bar{J}_1} Z_t^*\right)\right].$$

Following stationarity, we may write

$$\begin{aligned}
\frac{1}{q}\mathbb{E}\left[\left(\sum_{t\in\widetilde{J}_1} Z_t^*\right)\left(\sum_{t\in\bar{J}_1} Z_t^*\right)\right] &= \frac{1}{q}\left\{\sum_{t=1}^{q} t\mathbb{E}[Z_1 Z_{1+t}] + \sum_{t=q+1}^{2q}(2q-t)\mathbb{E}[Z_1Z_{1+t}]\right\}\\
&\leq \frac{1}{q}\sum_{t=1}^{2q} t\left|\mathbb{E}\left[Z_1Z_{1+t}\right]\right| \leq \frac{(2q)^{2/3}}{q}\sum_{t=1}^{2q} t^{1/3}\|Z_1\|_3^2\alpha^{1/3}(t)\\
&= O\left(\frac{1}{q^{1/3}}\right) \to 0, \qquad (62)
\end{aligned}$$

where the second last inequality follows from Lemma I.1 and stationarity and the last inequality follows from $\sum_{k=1}^{\infty} k^{1/3}\alpha^{1/3}(k) < \infty$ and

$$\begin{aligned}
\|Z_1\|_3 =& \mathbb{E}^{1/3}\left[\frac{1}{\kappa_k^3}\left\langle X_1, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right\rangle^3 \mathbb{E}\left[\varepsilon_1^3|X_1\right]\right]\\
\leq& O\left(\mathbb{E}^{1/6}\left[\frac{1}{\kappa_k^6}\left\langle X_1, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right\rangle^6\right]\right) = O(1), \qquad (63)
\end{aligned}$$

which follows from Assumption 1.

From this, the definition of the long run variance and stationarity at (3.13), we can write

$$\mathbb{E}\left[(F^*_{J_v})^2\right] \to \sigma^2_\infty(k), \qquad \text{as} \qquad q\to\infty, \qquad (64)$$

for all $J_v \in \mathcal{P}$.

Step 1b: We have from Assumption 1 that $\sum_{k=1}^{\infty}(k+1)^{8/3-1}\alpha^{(4/3)/(8/3+4/3)}(k) < \infty$, $\mathbb{E}\left[Z_t^*\right] = 0$ and similar to (63) that

$$\|Z_1^*\|_4 = \mathbb{E}^{1/4}\left[\frac{1}{\kappa_k^4}\left\langle X_1, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right\rangle^4 \mathbb{E}\left[\varepsilon_1^4|X_1\right]\right] \leq O\left(\mathbb{E}^{1/6}\left[\frac{1}{\kappa_k^6}\left\langle X_1, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right\rangle^6\right]\right) < \infty.$$

All of the conditions of Theorem 1 of Yokoyama (1980) are satisfied and therefore

$$\mathbb{E}\left[\left|\sum_{t\in\widetilde{J}_v} Z_t^*\right|^{8/3}\right] = O(q^{4/3}).$$

Following stationarity for all $v \in \{1, \ldots, S\}$, it implies

$$\begin{aligned}\mathbb{E}\left[\left|(F^*_{J_v})^2 - \mathbb{E}\left[(F^*_{J_v})^2\right]\right|^{4/3}\right] \leq& 2^{1/3}\mathbb{E}\left[\left|(F^*_{J_v})^2\right|^{4/3}\right] \\ \leq& 4\left(\mathbb{E}\left[\left|\frac{1}{\sqrt{q}}\sum_{t\in\widetilde{J}_v} Z^*_t\right|^{8/3}\right] + \mathbb{E}\left[\left|\frac{1}{\sqrt{q}}\sum_{t\in\bar{J}_v} Z^*_t\right|^{8/3}\right]\right) < \infty,\end{aligned}$$

where we used $(a+b)^{4/3} \leq 2^{1/3}(a^{4/3} + b^{4/3})$ in the last and the second last inequality. We also have $\alpha(k) = O\left(\frac{1}{k^4}\right)$ which follows from the summability of $\{k^{1/3}\alpha^{1/3}(k)\}_{k=1}^{\infty}$. With $\rho = 8$ and $p = 4/3$, it is what follows that

$$\sum_{v=1}^{S} \frac{\mathbb{E}^{2/\rho}\left[\left|(F^*_{J_v})^2 - \mathbb{E}\left[(F^*_{J_v})^2\right]\right|^p\right]}{v^p} \leq O(1)\sum_{v=1}^{\infty}\frac{1}{v^p} < \infty.$$

We have all the condition of Theorem I.1 satisfied with $\rho = 8$ and $p = 4/3$, therefore,

$$\frac{1}{S}\sum_{v=1}^{S}(F^*_{J_v})^2 - \mathbb{E}\left[(F^*_{J_v})^2\right] \to 0 \qquad \text{a.s.}$$

Combining this with (64) and and the stationarity of $\{F^*_{J_v}\}_{v=1}^S$, we write

$$\frac{1}{S}\sum_{v=1}^{S}(F^*_{J_v})^2 \to \sigma^2_{\infty}(k) \qquad \text{a.s.} \tag{65}$$

Step 2: Let

$$\begin{aligned}\widehat{F}_{J_v} = \sqrt{\frac{2}{q}}\Bigg\{&\sum_{t\in\widetilde{J}_v}\frac{\left\langle X_t, \widehat{\beta}^{(1)}_k - \widehat{\beta}^{(2)}_k\right\rangle_{\mathcal{L}^2}}{\kappa_k}\left(y_t - \left\langle X_t, \widehat{\beta}_{J_v}\right\rangle_{\mathcal{L}^2}\right) \\ &- \sum_{t\in\bar{J}_v}\frac{\left\langle X_t, \widehat{\beta}^{(1)}_k - \widehat{\beta}^{(2)}_k\right\rangle_{\mathcal{L}^2}}{\kappa_k}\left(y_t - \left\langle X_t, \widehat{\beta}_{J_v}\right\rangle_{\mathcal{L}^2}\right)\Bigg\}.\end{aligned}$$

Observe that

$$\left(\widehat{F}_{J_v}\right)^2 = \left(A_v + F^*_{J_v} + B_v\right)^2,$$

where

$$\begin{aligned}A_v = \frac{\sqrt{2}}{\kappa_k\sqrt{q}}\Bigg\{&\sum_{t\in\widetilde{J}_v}\left\langle X_t, \widehat{\beta}^{(1)}_k - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_t + \sum_{t\in\widetilde{J}_v}\left\langle X_t, \beta^*_{\eta_{k+1}} - \widehat{\beta}^{(2)}_k\right\rangle_{\mathcal{L}^2}\varepsilon_t \\ &+ \sum_{t\in\widetilde{J}_v}\left\langle X_t, \widehat{\beta}^{(1)}_k - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\left\langle X_t, \beta^*_{J_v} - \widehat{\beta}_{J_v}\right\rangle_{\mathcal{L}^2} \\ &+ \sum_{t\in\widetilde{J}_v}\left\langle X_t, \beta^*_{\eta_{k+1}} - \widehat{\beta}^{(2)}_k\right\rangle_{\mathcal{L}^2}\left\langle X_t, \beta^*_{J_v} - \widehat{\beta}_{J_v}\right\rangle_{\mathcal{L}^2} \\ &- \sum_{t\in\bar{J}_v}\left\langle X_t, \widehat{\beta}^{(1)}_k - \beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\varepsilon_t - \sum_{t\in\bar{J}_1}\left\langle X_t, \beta^*_{\eta_{k+1}} - \widehat{\beta}^{(2)}_k\right\rangle_{\mathcal{L}^2}\varepsilon_t\end{aligned}$$

$$
\begin{aligned}
&- \sum_{t \in \bar{J}_v} \left\langle X_t, \widehat{\beta}_k^{(1)} - \beta^*_{\eta_k} \right\rangle_{\mathcal{L}^2} \left\langle X_t, \beta^*_{J_v} - \widehat{\beta}_{J_v} \right\rangle_{\mathcal{L}^2} \\
&- \sum_{t \in \bar{J}_v} \left\langle X_t, \beta^*_{\eta_{k+1}} - \widehat{\beta}_k^{(2)} \right\rangle_{\mathcal{L}^2} \left\langle X_t, \beta^*_{J_v} - \widehat{\beta}_{J_v} \right\rangle_{\mathcal{L}^2} \Bigg\}
\end{aligned}
$$

and

$$
\begin{aligned}
B_v = \frac{\sqrt{2}}{\kappa_k \sqrt{q}} \Bigg\{ &\sum_{t \in \widetilde{J}_v} \left\langle X_t, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}} \right\rangle_{\mathcal{L}^2} \left\langle X_t, \beta^*_{J_v} - \widehat{\beta}_{J_v} \right\rangle_{\mathcal{L}^2} \\
&- \sum_{t \in \bar{J}_v} \left\langle X_t, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}} \right\rangle_{\mathcal{L}^2} \left\langle X_t, \beta^*_{J_v} - \widehat{\beta}_{J_v} \right\rangle_{\mathcal{L}^2} \Bigg\}.
\end{aligned}
$$

Such an expansion is possible because, under the event outlined in Lemma C.2, for $1 \le v \le S$, J_v have no change point. This follows from their construction in Algorithm 2 and conditions specified in (4.15). As a consequence, $\beta^*_{\widetilde{J}_v} = \beta^*_{\bar{J}_v} = \beta^*_{J_v}$. Following from Lemma G.2, Lemma G.3, Lemma G.5, Lemma G.6 and the choice of the tuning parameter q detailed in (4.15), we may write

$$
A_v = O_p\left(\frac{1}{\kappa_k} \sqrt{q} \delta_q \log^{1+\xi}(q) \right) = O_p\left(\frac{1}{\kappa_k} q^{\frac{1/2-r}{2r+1}} \log^{1+\xi}(q) \right) = o_p(1).
$$

For the term B_v, we may write

$$
\begin{aligned}
B_v =& \frac{\sqrt{2}}{\kappa_k \sqrt{q}} \Bigg\{ \sum_{t \in \widetilde{J}_v} \left(\left\langle X_t, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}} \right\rangle_{\mathcal{L}^2} \left\langle X_t, \beta^*_{J_v} - \widehat{\beta}_{J_v} \right\rangle_{\mathcal{L}^2} - \Sigma\left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \beta^*_{J_v} - \widehat{\beta}_{J_v} \right] \right) \\
&- \sum_{t \in \bar{J}_v} \left(\left\langle X_t, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}} \right\rangle_{\mathcal{L}^2} \left\langle X_t, \beta^*_{J_v} - \widehat{\beta}_{J_v} \right\rangle_{\mathcal{L}^2} - \Sigma\left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \beta^*_{J_v} - \widehat{\beta}_{J_v} \right] \right) \Bigg\} \\
=& \frac{\sqrt{2q}}{\kappa_k} \left\{ \left(\widehat{\Sigma}_{\widetilde{J}_v} - \Sigma \right) \left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \beta^*_{J_v} - \widehat{\beta}_{J_v} \right] - \left(\widehat{\Sigma}_{\bar{J}_v} - \Sigma \right) \left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \beta^*_{J} - \widehat{\beta}_{J} \right] \right\} \\
=& O_p\left(\frac{\sqrt{q}}{\kappa_k} \sqrt{\frac{\delta_q}{\sqrt{q}}} \log^{1+\xi}(q) \right) = O_p\left(\frac{1}{\kappa_k} \sqrt{q^{\frac{1-2r}{2r+1}}} \log^{1+\xi}(q) \right) = o_p(1),
\end{aligned}
$$

where the first equality in the last line follows from the Holders inequality $\left(\Sigma\,[a,b] \le \sqrt{\Sigma\,[a,a]\,\Sigma\,[b,b]} \right)$, (75) of Lemma G.8, Lemma G.1 and Lemma G.2, the last equality follows from (4.15). Since $A_v = o_p(1), B_v = o_P(1)$ and $F^*_{J_v} = O_p(1)$, we can write

$$
\left(\widehat{F}_{J_v} \right)^2 - \left(F^*_{J_v} \right)^2 = \left(A_v + F^*_{J_v} + B_v \right)^2 - \left(F^*_{J_v} \right)^2 = o_p(1).
$$

Therefore, from (65)

$$
\frac{1}{S} \sum_{v=1}^{S} (\widehat{F}_{J_v})^2 \xrightarrow{\mathbb{P}} \sigma^2_\infty(k).
$$

The main result now follows from the Slutsky's theorem because $F_{J_v} = \frac{\kappa_k}{\widehat{\kappa}_k} \widehat{F}_{J_v}$, and $\frac{\kappa_k}{\widehat{\kappa}_k} \xrightarrow{\mathbb{P}} 1$ by Lemma E.1.

□

E.1 Technical results for the Proof of Theorem 3

Lemma E.1. *Suppose that assumptions of Theorem 1 holds. The estimator $\widehat{\kappa}_k$ defined in (4.14) satisfies*

$$\widehat{\kappa}_k^2 - \kappa_k^2 = O_p\left(\Delta^{-r/(2r+1)} \log^{1+\xi} \Delta\right).$$

Consequently,

$$\frac{\widehat{\kappa}_k^2}{\kappa_k^2} \xrightarrow{\mathbb{P}} 1, \qquad n \to \infty.$$

Proof. WLOG let $\widehat{\eta}_k \geq \eta_k$. Observe that

$$\beta^*_{(s_k,\widehat{\eta}_k]} - \beta^*_{(\widehat{\eta}_k,e_k]} = \left(\frac{\eta_k - s_k}{\widehat{\eta}_k - s_k}\right)\left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right]$$

and because $\Delta/5 \leq \widehat{\eta} - s_k \leq \Delta$ from Lemma D.3, following from Theorem 1 we have that

$$1 - \frac{\widehat{\eta}_k - \eta_k}{\widehat{\eta}_k - s_k} = \left(\frac{\eta_k - s_k}{\widehat{\eta}_k - s_k}\right) \xrightarrow{\mathbb{P}} 1, \qquad n \to \infty.$$

We may write the expansion

$$\begin{aligned}
\widehat{\kappa}_k^2 - \kappa_k^2 =& \widehat{\Sigma}_{(s_k,e_k]}\left[\widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \widehat{\beta}_{(\widehat{\eta}_k,e_k]}, \widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \widehat{\beta}_{(\widehat{\eta}_k,e_k]}\right] - \Sigma\left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right] \\
=& \underbrace{\widehat{\Sigma}_{(s_k,e_k]}\left[\widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \beta^*_{(s_k,\widehat{\eta}_k]}, \widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \beta^*_{(s_k,\widehat{\eta}_k]}\right]}_{A_1} \\
&+ \underbrace{\widehat{\Sigma}_{(s_k,e_k]}\left[\beta^*_{(\widehat{\eta}_k,e_k]} - \widehat{\beta}_{(\widehat{\eta}_k,e_k]}, \beta^*_{(\widehat{\eta}_k,e_k]} - \widehat{\beta}_{(\widehat{\eta}_k,e_k]}\right]}_{A_2} \\
&+ \underbrace{2\widehat{\Sigma}_{(s_k,e_k]}\left[\beta^*_{(s_k,\widehat{\eta}_k]} - \beta^*_{(\widehat{\eta}_k,e_k]}, \widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \beta^*_{(s_k,\widehat{\eta}_k]}\right]}_{A_3} \\
&+ \underbrace{2\widehat{\Sigma}_{(s_k,e_k]}\left[\widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \beta^*_{(s_k,\widehat{\eta}_k]}, \beta^*_{(\widehat{\eta}_k,e_k]} - \widehat{\beta}_{(\widehat{\eta}_k,e_k]}\right]}_{A_4} \\
&+ \underbrace{2\widehat{\Sigma}_{(s_k,e_k]}\left[\beta^*_{(\widehat{\eta}_k,e_k]} - \widehat{\beta}_{(\widehat{\eta}_k,e_k]}, \beta^*_{(s_k,\widehat{\eta}_k]} - \beta^*_{(\widehat{\eta}_k,e_k]}\right]}_{A_5} \\
&+ \underbrace{\left(\widehat{\Sigma}_{(s_k,e_k]} - \Sigma\right)\left[\beta^*_{(s_k,\widehat{\eta}_k]} - \beta^*_{(\widehat{\eta}_k,e_k]}, \beta^*_{(s_k,\widehat{\eta}_k]} - \beta^*_{(\widehat{\eta}_k,e_k]}\right]}_{A_6} \\
&+ \underbrace{\left[\left(\frac{\eta_k - s_k}{\widehat{\eta}_k - s_k}\right)^2 - 1\right]\Sigma\left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right]}_{A_7}.
\end{aligned} \tag{66}$$

Observing $\widehat{\eta}_k - s_k = O(\Delta)$, it follows from Lemma G.5 that for $j = 1, 2, 4$ we have

$$|A_j| = O_p\left(\delta_\Delta \log^{1+\xi} \Delta\right).$$

For the expression A_3, it follows from Lemma G.7 that

$$\begin{aligned}
\left|A_3\right| &= 2\widehat{\Sigma}_{(s_k,e_k]}\left[\beta^*_{(s_k,\widehat{\eta}_k]} - \beta^*_{(\widehat{\eta}_k,e_k]}, \widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \beta^*_{(s_k,\widehat{\eta}_k]}\right] \\
&\leq 2\widehat{\Sigma}_{(s_k,e_k]}\left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \beta^*_{(s_k,\widehat{\eta}_k]}\right] = O_p\left(\sqrt{\delta_\Delta}\log^{1+\xi}\Delta\right),
\end{aligned}$$

and for the fifth expression we have $\left|A_5\right| = O_p\left(\sqrt{\delta_\Delta}\log^{1+\xi}\Delta\right)$ following the same argument. For the expression A_6, we have

$$\begin{aligned}
\left|A_6\right| &= \left(\frac{\eta_k - s_k}{\widehat{\eta}_k - s_k}\right)^2\left(\widehat{\Sigma}_{(s_k,e_k]} - \Sigma\right)\left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right] \\
&\leq \left(\widehat{\Sigma}_{(s_k,e_k]} - \Sigma\right)\left[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\right] = O_p\left(\frac{1}{\sqrt{\Delta}}\kappa_k^2\right),
\end{aligned}$$

where the last equality follows from Lemma G.8 and $e_k - s_k = O(\Delta)$. The deviation for the last expression

$$|A_7| = \left(1 + \frac{\eta_k - s_k}{\widehat{\eta}_k - s_k}\right)\left(\frac{\widehat{\eta}_k - \eta_k}{\widehat{\eta}_k - s_k}\right)\kappa_k^2 \leq 6\frac{5}{\Delta}\kappa_k^2(\widehat{\eta}_k - \eta_k) = O_p\left(\delta_\Delta \log^{1+\xi}(\Delta)\right)$$

follows from the earlier observation $\Delta/5 \leq \eta_k - s_k \leq \widehat{\eta}_k - s_k \leq \Delta$ and (15). The first part this current lemma $\widehat{\kappa}_k^2 - \kappa_k^2 = O_p\left(\Delta^{-r/(2r+1)}\log^{1+\xi}\Delta\right)$ follows by combining this seven deviation bounds.

The deviation from the first part lead us to

$$\left|\frac{\widehat{\kappa}_k^2 - \kappa_k^2}{\kappa_k^2}\right| = O_p\left(\frac{1}{\kappa^2}\Delta^{-r/(2r+1)}\log^{1+\xi}(\Delta)\right) = o_p(1),$$

where the last equality follows from Assumption 3.

□

F Proof of Theorem 4

Proof. Let $k \in [1, \ldots, \widehat{\mathcal{K}}]$ be given. For notational simplicity, we denote $\widehat{u} = \widehat{u}_k^{(b)}$ and $z_j = z_j^{(b)}$.

The proof follows a similar pattern as the proof of Theorem 2. In the first step, we establish the uniform tightness of the minimizer. In the second step, we demonstrate the convergence of the objective function on a compact domain and use the Argmax continuous mapping theorem.

Step 1. Let $\widehat{u}$ be a minimizer. Without loss of generality, assume $\widehat{u} \geq 0$. Since $\widehat{\sigma}^2_\infty(k) = O_p(1)$, we may write

$$\widehat{u} \leq -\widehat{\sigma}^2_\infty(k) \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor n\widehat{u} \rfloor} z_j = O_p\left(\sqrt{\widehat{u} \log^{1+\xi}(\widehat{u})}\right),$$

where the stochastic bound follows from the uniform result Lemma J.3. Therefore, $\widehat{u} = O_p(1)$.

Step 2. Let $M > 0$. We have $\widehat{\sigma}^2_\infty(k) \xrightarrow{\mathbb{P}} \sigma^2_\infty(k)$ from Theorem 3. From functional CLT, we have

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nr \rfloor} z_j \xrightarrow{\mathcal{D}} \mathbb{B}_1(r),$$

uniformly for all $0 \leq r \leq M$. Therefore, with the Argmax continuous mapping theorem (e.g. Theorem 3.2.2 of van der Vaart and Wellner (1996)), we have

$$\widehat{u} \xrightarrow{\mathcal{D}} \arg\min_{r \in \mathbb{R}} \left\{ |r| + \sigma_\infty(k) \mathbb{W}(r) \right\}, \qquad n \to \infty.$$

The main result now follows from the Slutsky's theorem. □

G Deviation bounds in functional linear regression

G.1 Notations

For any $a > 0$, we denote $\delta_a \asymp a^{-2r/(2r+1)}$. Also, $\lambda_a \asymp a^{-2r/(2r+1)}$. This is used in the observation (72) to denote $\widehat{f}_{(s,t]}$ which is estimator of $f^*_{(s,t]}$ from (71). The operator T is defined in (68) and its plug-in estimate $T_{\mathcal{I}}$ is defined in (70). We use $\mathbf{I}$ to denote the identity operator. The expression for $g_{(s,t]}$ and $H_{(s,t]}$ is defined in Proposition G.1 and Lemma G.10 respectively.

G.2 Kernel tools

Following Riesz representation theorem, the norm associated with $\mathcal{H}(K)$ from (1.3) can be equivalently defined through,

$$\langle f, L_K(g)\rangle_{\mathcal{H}(K)} := \langle f, g\rangle_{\mathcal{L}^2}.$$

One may note that

$$\mathbb{E}\left[\langle X, f\rangle^2\right] = \int f(s)\Sigma(s,t)f(t)\, ds\, dt = \langle L_\Sigma(f), f\rangle_{\mathcal{L}^2} = \Sigma[f,f]. \tag{67}$$

Moving forward, the main operator of our interest is the linear operator corresponding to the bi-linear function $K^{1/2}\Sigma K^{1/2}$ and the eigenvalues and eigenfunctions from its expansion. The linear operator on $\mathcal{L}^2$ corresponding to $K^{1/2}\Sigma K^{1/2}$ is given by

$$L_{K^{1/2}\Sigma K^{1/2}}(f)(*) = \langle K^{1/2}\Sigma K^{1/2}(\cdot,*), f(\cdot)\rangle_{\mathcal{L}^2}.$$

We denote the linear operator

$$T = L_{K^{1/2}\Sigma K^{1/2}}, \tag{68}$$

and by Assumption 2

$$T(\phi_l) = \mathfrak{s}_l\phi_l.$$

Following this, for any $a \in \mathbb{R}$, the operator T^a is defined through the operation $T^a(\phi_l) = \mathfrak{s}_l^a\phi_l$. Also for any $\beta \in \mathcal{H}(K)$ such that $f = L_{K^{-1/2}}(\beta)$,

$$\begin{aligned}\Sigma[\beta,\beta] = \Sigma[L_{K^{1/2}}(f), L_{K^{1/2}}(f)] &= \langle L_\Sigma L_{K^{1/2}}(f), L_{K^{1/2}}(f)\rangle_{\mathcal{L}^2} \\ &= \langle L_{K^{1/2}\Sigma K^{1/2}}(f), f\rangle_{\mathcal{L}^2} = \langle T(f), f\rangle_{\mathcal{L}^2} = \|T^{1/2}(f)\|^2_{\mathcal{L}^2}.\end{aligned} \tag{69}$$

The estimator of covariance function based on the sub-sample $\mathcal{I} \subset (0,n]$ is given by

$$\widehat{\Sigma}_{\mathcal{I}}(u,v) = \frac{1}{\mathcal{I}}\sum_{j\in\mathcal{I}} X_j(u)X_j(v).$$

The empirical version of T is $T_{\mathcal{I}} := L_{K^{1/2}\widehat{\Sigma}_{\mathcal{I}}K^{1/2}}$ and its action can be viewed as

$$\begin{aligned}T_{\mathcal{I}}(h) =& L_{K^{1/2}} \circ L_{\widehat{\Sigma}_{\mathcal{I}}} \circ L_{K^{1/2}}(h) = L_{K^{1/2}}\left(\frac{1}{|\mathcal{I}|}\sum_{j\in\mathcal{I}}\langle X_j, L_{K^{1/2}}(h)\rangle X_j\right) \\ =& \frac{1}{|\mathcal{I}|}\sum_{j\in\mathcal{I}}\langle X_j, L_{K^{1/2}}(h)\rangle L_{K^{1/2}}(X_j).\end{aligned} \tag{70}$$

Since, $\mathcal{L}^2$ is bijectively mapped to $\mathcal{H}(K)$, we may have $f^*_{(s,t]}$ and $\widehat{f}_{(s,t]}$ defined as

$$\frac{1}{t-s}\sum_{j=s+1}^{t} f_j^* = f^*_{(s,t]} = L_{K^{-1/2}}\beta^*_{(s,t]} \qquad \text{and} \qquad \widehat{f}_{(s,t]} = L_{K^{-1/2}}(\widehat{\beta}_{(s,t]}). \tag{71}$$

We may also observe that

$$\widehat{f}_{(s,t]} = \arg\min_{f\in\mathcal{L}^2}\left\{\frac{1}{(t-s)}\sum_{j=s+1}^{t}(y_i - \langle X_i, L_{K^{1/2}}(f)\rangle_{\mathcal{L}^2})^2 + \lambda_{(s,t]}\|f\|^2_{\mathcal{L}^2}\right\} \tag{72}$$

Given $(y^\star, X^\star)$ a copy of (y,X) independent of the training data, the excess risk based on $(s,t]$ is defined as

$$\begin{aligned}\mathbb{E}[\langle X^\star, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\rangle^2] &= \int\int(\widehat{\beta}_{(s,t]}(x) - \beta^*_{(s,t]}(x))\Sigma(x,y)(\widehat{\beta}_{(s,t]}(y) - \beta^*_{(s,t]}(y))dxdy \\ &= \Sigma[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}] = \left\|T^{1/2}(\widehat{f}_{(s,t]} - f^*_{(s,t]})\right\|^2_{\mathcal{L}^2}\end{aligned} \tag{73}$$

the last form can be obtained using (67), (69) and (71).

G.3 Roughness regularized estimator and its properties

In order to evaluate the quality of estimation, we rely on the following lemmas. They help us control various deviation terms in the main result presented in this paper. All the proofs of the lemmas stated below are in the next section.

Lemma G.1. *Let $\xi > 0$. Suppose $(s,e] \subset (0,n]$. Then*

$$\max_{s<t\le e}\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}\Sigma\left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\right] = O_p(1).$$

Lemma G.2. *Let $\xi > 0$. Suppose $(s,e] \subset (0,n]$. Then*

$$\max_{s<t\le e}\left(\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}\right)\widehat{\Sigma}_{(s,t]}\left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\right] = O_p(1).$$

Lemma G.3. *Let $\xi > 0$. Suppose $(s,e] \subset (0,n]$. Then*

$$\max_{s<t\le e}\left(\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}\right)\frac{1}{t-s}\sum_{j=s+1}^{t}\langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\varepsilon_j = O_p(1).$$

Lemma G.4. *Let $\xi > 0$. Suppose $(s,e] \subset (0,n]$. Then*

$$\max_{s<t\le e}\left(\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}\right)\frac{1}{t-s}\sum_{j=s+1}^{t}\left\langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\left\langle X_j, \beta_j^* - \beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2} = O_p(1).$$

Lemma G.5. *Let $\xi > 0$. Suppose $(s', e']$ and $(s, e]$ are the subsets of $(0, n]$. Then*

$$\max_{\substack{s<t\leq e\\ s'<t'\leq e'}} \frac{1}{\mathfrak{J}(t-s, t'-s')} \widehat{\Sigma}_{(s',t']} \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\right] = O_p(1),$$

where

$$\mathfrak{J}(t-s, t'-s') = \left(1 + \left(\frac{\delta_{t'-s'}}{\delta_{t-s}}\right)^{1/4}\right) \left\{\delta_{t-s}^{1/2}\delta_{t'-s'}^{1/2}\sqrt{\log^{1+\xi}(t'-s')}\sqrt{\log^{1+\xi}(t-s)}\right\} + \delta_{t-s}\log^{1+\xi}(t-s).$$

Lemma G.6. *Let $\xi > 0$. Suppose $(s', e']$ and $(s, e]$ are the subsets of $(0, n]$. Then*

$$\max_{\substack{s<t\leq e\\ s'<t'\leq e'}} \frac{1}{\mathfrak{H}(t-s, t'-s')} \frac{1}{t'-s'} \sum_{j=s'+1}^{t'} \langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\varepsilon_j = O_p(1).$$

where

$$\mathfrak{H}(t-s, t'-s') = \left(1 + \left(\frac{\delta_{t'-s'}}{\delta_{t-s}}\right)^{1/4}\right) \delta_{t-s}^{1/2}\delta_{t'-s'}^{1/2}\sqrt{\log^{1+\xi}(t'-s')}\sqrt{\log^{1+\xi}(t-s)}.$$

Lemma G.7. *Let $\xi > 0$. Suppose $(s', e']$ and $(s, e]$ are the subsets of $(0, n]$. Then*

$$\max_{\substack{s<t\leq e\\ s'<t'\leq e'}} \frac{1}{\mathfrak{G}(t-s, t'-s')} \frac{1}{t'-s'} \sum_{j=s'+1}^{t'} \langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\rangle_{\mathcal{L}^2}\langle X_j, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\rangle_{\mathcal{L}^2} = O_p(1).$$

where

$$\mathfrak{G}(t-s, t'-s') = \left(1 + \left(\frac{\delta_{t'-s'}}{\delta_{t-s}}\right)^{1/4}\right) \delta_{t-s}^{1/2}\delta_{t'-s'}^{1/2}\sqrt{\log^{1+\xi}(t'-s')}\sqrt{\log^{1+\xi}(t-s)} + \kappa_k\sqrt{\delta_{t-s}\log^{1+\xi}(t-s)}.$$

G.4 Markov type probability bounds

Lemma G.8. *Let $\{\mathfrak{f}_j\}_{j=1}^t$ and h be non-random function in $\mathcal{L}^2$. Suppose $\Sigma[\mathfrak{f}_j, \mathfrak{f}_j] \le M < \infty$, for all $1 \le j \le t$, where M is some absolute constant. Then*

$$\mathbb{E}\left[\left|\sum_{j=1}^{t} \langle X_j, \mathfrak{f}_j\rangle_{\mathcal{L}^2} \langle X_j, h\rangle_{\mathcal{L}^2} - \Sigma[\mathfrak{f}_j, h]\right|^2\right] = O(t)\Sigma[h, h]. \tag{74}$$

When $\mathfrak{f}_1 = \ldots = \mathfrak{f}_t = \mathfrak{f}$, Then

$$\mathbb{E}\left[\left|\sum_{j=1}^{t} \langle X_j, \mathfrak{f}\rangle_{\mathcal{L}^2} \langle X_j, h\rangle_{\mathcal{L}^2} - \Sigma[\mathfrak{f}, h]\right|^2\right] = O(t)\Sigma[h, h]\Sigma[\mathfrak{f}, \mathfrak{f}]. \tag{75}$$

Proof. Given any sequence of stationary random variables $\{W_j\}_{j=1}^t$ with finite second moment it holds

$$\text{Var}\left(\sum_{j=1}^{t} W_j\right)^2 = \sum_{j=1}^{t} Var\left(W_j^2\right) + 2\sum_{j=1}^{t-1}(t-j)Cov\left(W_1, W_{1+j}\right). \tag{76}$$

We are going estblish (74). Let $z_j = \langle X_j, \mathfrak{f}_j\rangle_{\mathcal{L}^2} \langle X_j, h\rangle_{\mathcal{L}^2} - \Sigma[\mathfrak{f}_j, h]$. Then

$$\begin{aligned}
\mathbb{E}[z_j^2] &\le \mathbb{E}\left[\langle X_j, \mathfrak{f}_i\rangle_{\mathcal{L}^2}^2 \langle X_j, h\rangle_{\mathcal{L}^2}^2\right] \\
&\le \sqrt{\mathbb{E}\left[\langle X_j, \mathfrak{f}_j\rangle_{\mathcal{L}^2}^4\right]}\sqrt{\mathbb{E}\left[\langle X_j, h\rangle_{\mathcal{L}^2}^4\right]} \\
&\le c^2\mathbb{E}\left[\langle X_j, \mathfrak{f}_j\rangle_{\mathcal{L}^2}^2\right] c^2\mathbb{E}\left[\langle X_j, h\rangle_{\mathcal{L}^2}^2\right] \\
&= c^4\Sigma[\mathfrak{f}_j, \mathfrak{f}_j]\Sigma[h, h] \le c^4 M\Sigma[h, h].
\end{aligned} \tag{77}$$

The (77) follows from the Assumption 1, where we have the sixth moment bounded by the second moment up to a constant factor c.

We have

$$\mathbb{E}[|z_j z_{j+k}|] = Cov(|z_j|, |z_{j+k}|) \le \|z_j\|_3 \|z_{j+k}\|_3 \alpha^{1/3}(k),$$

following from Lemma I.1. Following from

$$\begin{aligned}
\|z_j\|_3 =& \|\langle X_j, \mathfrak{f}_j\rangle_{\mathcal{L}^2} \langle X_j, h\rangle_{\mathcal{L}^2} - \Sigma[\mathfrak{f}_j, h]\|_3 \\
&\le \|\langle X_j, \mathfrak{f}_j\rangle_{\mathcal{L}^2} \langle X_j, h\rangle_{\mathcal{L}^2}\|_3 + \|\Sigma[\mathfrak{f}_j, h]\|_3 \\
&\le 2\|\langle X_j, \mathfrak{f}_i\rangle_{\mathcal{L}^2} \langle X_j, h\rangle_{\mathcal{L}^2}\|_3,
\end{aligned}$$

one may write

$$\begin{aligned}
\|z_j\|_3 &\le 2\|\langle X_j, \mathfrak{f}_j\rangle_{\mathcal{L}^2} \langle X_j, h\rangle_{\mathcal{L}^2}\|_3 \\
&\le 2\|\langle X_j, \mathfrak{f}_j\rangle_{\mathcal{L}^2}\|_6 \|\langle X_j, h\rangle_{\mathcal{L}^2}\|_6 \\
&\le 2c\|\langle X_j, \mathfrak{f}_j\rangle_{\mathcal{L}^2}\|_2 c\|\langle X_j, h\rangle_{\mathcal{L}^2}\|_2 = 2c^2\sqrt{\Sigma[\mathfrak{f}_i, \mathfrak{f}_j]\Sigma[h, h]} \le 2c^2\sqrt{M}\sqrt{\Sigma[h, h]}.
\end{aligned}$$

The last line here follows the same argument as (77). Similarly

$$\|z_{1+j}\|_3 \le 2c\sqrt{M}\sqrt{\Sigma[h, h]}.$$

Therefore

$$\mathbb{E}[|z_1 z_{1+j}|] \leq 4c^4 M\Sigma[h,h]\alpha^{1/3}(j).$$

Following (76), one may have the expansion

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{j=1}^{t} z_j\right|^2\right] &= \mathbb{E}\left[\sum_{j=1}^{t} z_j^2\right] + 2\sum_{j=1}^{t-1}(t-j)\mathbb{E}\left[z_1 z_{1+j}\right] \\ &\leq \mathbb{E}\left[\sum_{j=1}^{t} z_j^2\right] + 2\sum_{k=1}^{t-1}(t-k)\mathbb{E}\left[|z_i z_{i+k}|\right]. \end{aligned} \tag{78}$$

Using (78), we may write

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{j=1}^{t} z_i\right|^2\right] &\leq \mathbb{E}\left[\sum_{j=1}^{t} z_j^2\right] + 2\sum_{j=1}^{t-1}(t-j)\mathbb{E}\left[|z_1 z_{1+j}|\right] \\ &\leq \sum_{i=1}^{t} c^4 M\Sigma[h,h] + 2\sum_{j=1}^{t-1}(t-k)4c^4 M\Sigma[h,h]\alpha^{1/3}(j) \\ &\leq tc^4 M\Sigma[h,h] + 8c^4 Mt\Sigma[h,h]\sum_{j=1}^{t-1}\alpha^{1/3}(j) \\ &\leq tc^4 M\Sigma[h,h] + 8c^4 Mt\Sigma[h,h]\sum_{j=1}^{\infty}\alpha^{1/3}(j) = (t\Sigma[h,h])\, O(1).\end{aligned}$$

The last line follows from $\sum_{j\geq 1}\alpha^{1/3}(j) < \infty$.

The proof for (75) is very similar and therefore omitted. □

Lemma G.9. *Let h be non-random function in $\mathcal{L}^2$. Then*

$$\mathbb{E}\left[\left|\sum_{j=1}^{t}\langle X_j, h\rangle_{\mathcal{L}^2}\varepsilon_j\right|^2\right] = O\left(t\Sigma[h,h]\right).$$

Proof. The proof here closely follows the proof of the Lemma G.8. Let $z_j = \langle X_j, h\rangle_{\mathcal{L}^2}\varepsilon_j$. We can see $\mathbb{E}\left[z_j\right] = 0$.

Observe that

$$\mathbb{E}\left[z_j^2\right] = \mathbb{E}\left[\langle X_j, h\rangle_{\mathcal{L}^2}^2 \mathbb{E}\left[\varepsilon_j^2|X_i\right]\right] \leq O(1)\|\langle X_j, h\rangle_{\mathcal{L}^2}\|_2^2 = O(1)\Sigma[h,h], \tag{79}$$

here we use the moment assumption outlined Assumption 1.

Following from

$$\mathbb{E}\left[|z_1 z_{1+j}|\right] \leq \|z_1\|_3\|z_{1+j}\|_3\alpha^{1/3}(j),$$

and

$$\|z_j\|_3 \le \left(\mathbb{E}\left[\langle X_j, h\rangle_{\mathcal{L}^2}^3 \mathbb{E}\left[\varepsilon_j^3 | X_j\right]\right]\right)^{1/3} \le O(1)\|\langle X_j, h\rangle_{\mathcal{L}^2}\|_3 = O(1)\sqrt{\Sigma[h,h]},$$

we may have

$$\mathbb{E}\left[|\langle X_j, h\rangle \varepsilon_j \langle X_{j+k}, h\rangle \varepsilon_{j+k}|\right] = \mathbb{E}\left[|z_j z_{j+k}|\right] = O(\Sigma[h,h])\alpha^{1/3}(k). \tag{80}$$

The rest of proof follows from the exactly same arguments as the proof of Lemma G.8 and therefore omitted.

□

G.5 Proofs of Lemmas from Appendix G.3

All the proofs in this section used the notations from Appendix G.1.

G.5.1 Proof of Lemma G.1

The proof of Lemma G.1 follows from Lemma G.13 with $a = 1/2$ and $b = 1$.

G.5.2 Proof of Lemma G.2

Proof. Let $0 < \nu < \frac{1}{2} - \frac{1}{4r}$. Observe that

$$\begin{aligned}
&\left(\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}\right) \widehat{\Sigma}_{(s,t]}\left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\right] \\
=& \left(\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}\right) \left\langle \widehat{f}_{(s,t]} - f^*_{(s,t]}, T_{(s,t]}(\widehat{f}_{(s,t]} - f^*_{(s,t]})\right\rangle_{\mathcal{L}^2} \\
=& \left(\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}\right) \Big\{ \left\langle \widehat{f}_{(s,t]} - f^*_{(s,t]}, (T_{(s,t]} - T)(\widehat{f}_{(s,t]} - f^*_{(s,t]})\right\rangle_{\mathcal{L}^2} \\
&\qquad + \left\langle \widehat{f}_{(s,t]} - f^*_{(s,t]}, T(\widehat{f}_{(s,t]} - f^*_{(s,t]})\right\rangle_{\mathcal{L}^2} \Big\} \\
\le& \left(\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}\right) \Big\{ \left\| T^{\nu}(\widehat{f}_{(s,t]} - f^*_{(s,t]})\right\|_{\mathcal{L}^2} \left\| T^{-\nu}(T_{(s,t]} - T)(\widehat{f}_{(s,t]} - f^*_{(s,t]})\right\|_{\mathcal{L}^2} \\
&\qquad + \left\| T^{1/2}(\widehat{f}_{(s,t]} - f^*_{(s,t]})\right\|_{\mathcal{L}^2} \Big\}.
\end{aligned}$$

The term on the right is bounded by using Lemma G.12 and Lemma G.14. The term on the left is bounded by using Lemma G.1. □

G.5.3 Proof of Lemma G.3

Proof. Observe that

$$\frac{1}{(t-s)} \sum_{j=s+1}^{t} \langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}\rangle_{\mathcal{L}^2} \varepsilon_j$$

$$
\begin{aligned}
&= \left\langle \frac{1}{(t-s)} \sum_{j=s+1}^{t} X_j \varepsilon_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \\
&= \left\langle \frac{1}{(t-s)} \sum_{j=s+1}^{t} L_{K^{1/2}}(X_j) \varepsilon_j, \widehat{f}_{(s,t]} - f^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \\
&= \left\langle g_{(s,t]}, \widehat{f}_{(s,t]} - f^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \\
&= \left\langle T^{-1/4} (T + \lambda_{t-s})^{-1/4} g_{(s,t]}, T^{1/4} (T + \lambda_{t-s})^{1/4} \widehat{f}_{(s,t]} - f^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \\
&\leq \left\| T^{-1/4} (T + \lambda_{t-s})^{-1/4} g_{(s,t]} \right\|_{\mathcal{L}^2} \left\| T^{1/4} (T + \lambda_{t-s})^{1/4} \widehat{f}_{(s,t]} - f^*_{(s,t]} \right\|_{\mathcal{L}^2}
\end{aligned}
$$

where the last line follows from Cauchy-Schwarz inequality.

From Lemma G.15, we have

$$
\max_{s<t\leq e} \sqrt{\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}} \left\| T^{-1/4} (T + \lambda_{t-s})^{-1/4} g_{(s,t]} \right\|_{\mathcal{L}^2} = O_p(1),
$$

and from Lemma G.13

$$
\max_{s<t\leq e} \sqrt{\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}} \left\| T^{1/4} (T + \lambda_{t-s})^{1/4} \widehat{f}_{(s,t]} - f^*_{(s,t]} \right\|_{\mathcal{L}^2} = O_p(1).
$$

The above two bounds establish the result. □

G.5.4 Proof of Lemma G.4

Proof. Observe that

$$
\frac{1}{t-s} \sum_{j=s+1}^{t} \Sigma \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \beta^*_j - \beta^*_{(s,t]} \right] = 0,
$$

because $\beta^*_{(s,t]} = \sum_{j=s+1}^{t} \beta^*_j / (t-s)$.

We may write

$$
\begin{aligned}
&\frac{1}{t-s} \sum_{j=s+1}^{t} \left\langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \\
=& \frac{1}{t-s} \sum_{j=s+1}^{t} \left(\left\langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} - \Sigma \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \beta^*_j - \beta^*_{(s,t]} \right] \right) \\
=& \frac{1}{t-s} \sum_{j=s+1}^{t} \left(\left\langle X_j, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \left\langle X_j, \beta^*_j - \beta^*_{(s,t]} \right\rangle_{\mathcal{L}^2} - \left\langle \widehat{f}_{(s,t]} - f^*_{(s,t]}, T(f^*_j - f^*_{(s,t]}) \right\rangle_{\mathcal{L}^2} \right) \\
=& \left\langle \frac{1}{t-s} \sum_{j=s+1}^{t} \left(\left\langle L_{K^{1/2}}(X_j), f^*_j - f^*_{(s,t]} \right\rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - T(f^*_j - f^*_{(s,t]}) \right), \widehat{f}_{(s,t]} - f^*_{(s,t]} \right\rangle \\
=& \left\langle G_{(s,t]}, \widehat{f}_{(s,t]} - f^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \\
=& \left\langle T^{-1/4} (T + \lambda_{t-s} \mathbf{I})^{-1/4} G_{(s,t]}, T^{1/4} (T + \lambda_{t-s} \mathbf{I})^{1/4} (\widehat{f}_{(s,t]} - f^*_{(s,t]}) \right\rangle_{\mathcal{L}^2}
\end{aligned}
$$

$$\leq \left\| T^{-1/4} \left(T + \lambda_{t-s}\mathbf{I}\right)^{-1/4} G_{(s,t]} \right\|_{\mathcal{L}^2} \left\| T^{1/4} \left(T + \lambda_{t-s}\mathbf{I}\right)^{1/4} (\widehat{f}_{(s,t]} - f^*_{(s,t]}) \right\|_{\mathcal{L}^2},$$

where $G_{(s,t]} = \frac{1}{t-s}\sum_{j=s+1}^{t} \left(\left\langle L_{K^{1/2}}(X_j), f_j^* - f^*_{(s,t]} \right\rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - T(f_j^* - f^*_{(s,t]}) \right)$.

From Lemma G.17, we have

$$\max_{s<t\leq e} \sqrt{\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}} \left\| T^{-1/4} \left(T + \lambda_{t-s}\mathbf{I}\right)^{-1/4} G_{(s,t]} \right\|_{\mathcal{L}^2} = O_p(1),$$

and from Lemma G.13

$$\max_{s<t\leq e} \sqrt{\frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)}} \left\| T^{1/4} \left(T + \lambda_{t-s}\mathbf{I}\right)^{1/4} (\widehat{f}_{(s,t]} - f^*_{(s,t]}) \right\|_{\mathcal{L}^2} = O_p(1).$$

The above two bounds establish the result. □

G.5.5 Proof of Lemma G.5

Proof. Let $\nu < 1/2 - 1/4r$. We may write

$$\begin{aligned}
&\widehat{\Sigma}_{(s',t']} \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right] \\
\leq& \left| \left(\widehat{\Sigma}_{(s',t']} - \Sigma \right) \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right] \right| + \left| \Sigma \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right] \right| \\
=& \left| \left\langle (T_{(s',t']} - T)(\widehat{f}_{(s,t]} - f^*_{(s,t]}), \widehat{f}_{(s,t]} - f^*_{(s,t]} \right\rangle_{\mathcal{L}^2} \right| + \left| \Sigma \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right] \right| \\
=& \left| \left\langle T^{-\nu}(T_{(s',t']} - T)(\widehat{f}_{(s,t]} - f^*_{(s,t]}), T^{\nu}(\widehat{f}_{(s,t]} - f^*_{(s,t]}) \right\rangle_{\mathcal{L}^2} \right| + \left| \Sigma \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right] \right| \\
\leq& \left\| T^{-\nu}(T_{(s',t']} - T)(\widehat{f}_{(s,t]} - f^*_{(s,t]}) \right\|_{\mathcal{L}^2} \left\| T^{\nu}(\widehat{f}_{(s,t]} - f^*_{(s,t]}) \right\|_{\mathcal{L}^2} + \left| \Sigma \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right] \right|,
\end{aligned} \tag{81}$$

where the second line follows from the triangle inequality and the last line follows from the Cauchy-Schwarz inequality. Observe that

$$\begin{aligned}
&\left\| T^{-\nu}(T_{(s',t']} - T)(\widehat{f}_{(s,t]} - f^*_{(s,t]}) \right\|_{\mathcal{L}^2} \\
\leq& \left(1 + \left(\frac{\delta_{t'-s'}}{\delta_{t-s}} \right)^{1/4} \right) \left[\left\| T^{-\nu}(T_{(s',t']} - T)(T + \lambda_{t'-s'}\mathbf{I})^{-1/4} T^{-1/4} \right\|_{\mathrm{op}} \right. \\
& \qquad\qquad \left. \left\| T^{1/4}(T + \lambda_{t-s}\mathbf{I})^{1/4}(\widehat{f}_{(s,t]} - f^*_{(s,t]}) \right\|_{\mathcal{L}^2} \right] \\
=& O_p\left(\left[1 + \left(\frac{\delta_{t'-s'}}{\delta_{t-s}} \right)^{1/4} \right] \delta_{t'-s'}^{1/2} \sqrt{\log^{1+\xi}(t'-s')} \cdot \delta_{t-s}^{1/2} \sqrt{\log^{1+\xi}(t-s)} \right),
\end{aligned} \tag{82}$$

where the first line follows from Lemma G.21 and the last line follows from Lemma G.16 and Lemma G.13. Following from Lemma G.13, we have

$$\left\| T^{-\nu}(\widehat{f}_{(s,t]} - f^*_{(s,t]}) \right\|_{\mathcal{L}^2} = O_p\left(\delta_{t-s}^{\nu} \sqrt{\log^{1+\xi}(t-s)} \right) = O_p(1), \tag{83}$$

and from Lemma G.1 we have

$$\left| \Sigma \left[\widehat{\beta}_{(s,t]} - \beta^*_{(s,t]}, \widehat{\beta}_{(s,t]} - \beta^*_{(s,t]} \right] \right| = O_p\left(\delta_{t-s} \log^{1+\xi}(t-s) \right). \tag{84}$$

The result now follows using (82), (83) and (84) to bound (81).

□

G.5.6 Proof of Lemma G.6

Proof. Observe that

$$
\begin{aligned}
&\left|\frac{1}{t'-s'}\sum_{j=s'+1}^{t'}\left\langle X_j, \widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\varepsilon_j\right| \\
=&\left|\left\langle \frac{1}{t'-s'}\sum_{j=s'+1}^{t'} L_{K^{1/2}}(X_j)\varepsilon_j, \widehat{f}_{(s,t]}-f^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\right| \\
=&\left|\left\langle g_{(s',t']}, \widehat{f}_{(s,t]}-f^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\right| \\
\leq&\left(1+\left(\frac{\delta_{t'-s'}}{\delta_{t-s}}\right)^{1/4}\right)\left\|T^{-1/4}(T+\lambda_{t'-s'}\mathbf{I})^{-1/4}g_{(s',t']}\right\|_{\mathcal{L}^2}\left\|T^{1/4}(T+\lambda_{t-s}\mathbf{I})^{1/4}(\widehat{f}_{(s,t]}-f^*_{(s,t]})\right\|_{\mathcal{L}^2} \\
=&O_p\left(\left[1+\left(\frac{\delta_{t'-s'}}{\delta_{t-s}}\right)^{1/4}\right]\delta_{t'-s'}^{1/2}\sqrt{\log^{1+\xi}(t'-s')}\cdot\delta_{t-s}^{1/2}\sqrt{\log^{1+\xi}(t-s)}\right),
\end{aligned}
$$

where the second last line follows from Lemma G.21 and the last line follows from Lemma G.15 and Lemma G.1.

□

G.5.7 Proof of Lemma G.7

Proof. Observe that

$$
\begin{aligned}
&\left|\frac{1}{t'-s'}\sum_{j=s'+1}^{t'}\left\langle X_j, \widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right\rangle_{\mathcal{L}^2}\left\langle X_j, \beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\right| \\
\leq&\left|\left(\widehat{\Sigma}_{(s',t']}-\Sigma\right)\left[\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}, \beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right]\right|+\left|\Sigma\left[\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}, \beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right]\right| \\
=&\left|\left\langle (T_{(s',t']}-T)(\widehat{f}_{(s,t]}-f^*_{(s,t]}), f^*_{\eta_{k+1}}-f^*_{\eta_k}\right\rangle_{\mathcal{L}^2}\right|+\left|\Sigma\left[\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}, \widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right]\right| \\
=&\left|\left\langle (T_{(s',t']}-T)(\widehat{f}_{(s,t]}-f^*_{(s,t]}), (f^*_{\eta_{k+1}}-f^*_{\eta_k})\right\rangle_{\mathcal{L}^2}\right|+\left|\Sigma\left[\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}, \beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right]\right| \\
\leq&\left\|(T_{(s',t']}-T)(\widehat{f}_{(s,t]}-f^*_{(s,t]})\right\|_{\mathcal{L}^2}\left\|f^*_{\eta_{k+1}}-f^*_{\eta_k}\right\|_{\mathcal{L}^2}+\left|\Sigma\left[\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}, \beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right]\right|,
\end{aligned} \tag{85}
$$

where the second line follows from the triangle inequality and the last line follows from the Cauchy-Schwarz inequality. Observe that

$$
\begin{aligned}
&\left\|(T_{(s',t']}-T)(\widehat{f}_{(s,t]}-f^*_{(s,t]})\right\|_{\mathcal{L}^2} \\
\leq&\left(1+\left(\frac{\delta_{t'-s'}}{\delta_{t-s}}\right)^{1/4}\right)\Bigg[\left\|(T_{(s',t']}-T)(T+\lambda_{t'-s'}\mathbf{I})^{-1/4}T^{-1/4}\right\|_{\text{op}} \\
&\qquad\left\|T^{1/4}(T+\lambda_{t-s}\mathbf{I})^{1/4}(\widehat{f}_{(s,t]}-f^*_{(s,t]})\right\|_{\mathcal{L}^2}\Bigg] \\
=&O_p\left(\left[1+\left(\frac{\delta_{t'-s'}}{\delta_{t-s}}\right)^{1/4}\right]\delta_{t'-s'}^{1/2}\sqrt{\log^{1+\xi}(t'-s')}\cdot\delta_{t-s}^{1/2}\sqrt{\log^{1+\xi}(t-s)}\right),
\end{aligned} \tag{86}
$$

where the first line follows from Lemma G.21 and the last line follows from Lemma G.16 and Lemma G.13. Following from Lemma G.1 we have that

$$
\begin{aligned}
&\left|\Sigma\left[\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]},\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right]\right|\\
\leq&\sqrt{\Sigma\left[\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]},\widehat{\beta}_{(s,t]}-\beta^*_{(s,t]}\right]}\sqrt{\Sigma\left[\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k},\beta^*_{\eta_{k+1}}-\beta^*_{\eta_k}\right]}\\
=&O_p\left(\kappa_k\sqrt{\delta_{t-s}\log^{1+\xi}(t-s)}\right).
\end{aligned}
\tag{87}
$$

The result now follows using (86) and (87) to bound (85). □

G.5.8 Technical results for this section

Proposition G.1. *The analytical expression for the estimator in* (72) *is given by*

$$\widehat{f}_{(s,t]} = (T_{(s,t]} + \lambda_{t-s}\mathbf{I})^{-1} \left(\frac{1}{t-s} \sum_{j=s+1}^{t} \langle L_{K^{1/2}}(X_j), f_j^* \rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) + g_{(s,t]} \right)$$

where $g_{(s,t]} = \frac{1}{(t-s)} \sum_{j=s+1}^{t} \varepsilon_j L_{K^{1/2}}(X_j)$.

Proof. Observe that

$$\frac{\partial \langle f, f \rangle_{\mathcal{L}^2}}{\partial f} = 2f \qquad \text{and} \qquad \frac{\partial \langle f, g \rangle_{\mathcal{L}^2}}{\partial f} = g.$$

Since the objective function is a quadratic form, we just need to differentiate and make it zero to find the minima. We may have

$$\begin{aligned}
0 =& \frac{\partial}{\partial f} \left(\frac{1}{t-s} \sum_{j=s+1}^{t} (y_j - \langle X_j, L_{K^{1/2}}(f) \rangle_{\mathcal{L}^2})^2 + \lambda \|f\|_{\mathcal{L}^2}^2 \right) \Bigg|_{f=\widehat{f}_{(s,t]}} \\
=& \frac{\partial}{\partial f} \left(\frac{1}{t-s} \sum_{j=s+1}^{t} (\langle L_{K^{1/2}}(X_j), f \rangle_{\mathcal{L}^2}^2 + y_j^2 - 2y_j \langle L_{K^{1/2}}(X_j), f \rangle_{\mathcal{L}^2}) + \lambda \|f\|_{\mathcal{L}^2}^2 \right) \Bigg|_{f=\widehat{f}_{(s,t]}} \\
=& \frac{1}{t-s} \sum_{j=s+1}^{t} \Big(2\langle L_{K^{1/2}}(X_j), \widehat{f}_{(s,t]} \rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - 2y_j L_{K^{1/2}}(X_j) \Big) + 2\lambda \widehat{f}_{(s,t]}.
\end{aligned}$$

And it lead us to

$$\begin{aligned}
0 =& \frac{1}{t-s} \sum_{j=s+1}^{t} \Big(\langle L_{K^{1/2}}(X_j), \widehat{f}_{(s,t]} \rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - y_j L_{K^{1/2}}(X_j) \Big) + \lambda \widehat{f}_{(s,t]}, \\
=& \frac{1}{t-s} \sum_{j=s+1}^{t} \langle L_{K^{1/2}}(X_j), \widehat{f}_{(s,t]} \rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - \frac{1}{t-s} \sum_{j=s+1}^{t} \langle L_{K^{1/2}}(X_j), f_j^* \rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) \\
& - \frac{1}{t-s} \sum_{j=s+1}^{t} \varepsilon_j L_{K^{1/2}}(X_j) + \lambda \widehat{f}_{(s,t]} \\
=& T_{(s,t]} \widehat{f}_{(s,t]} - \frac{1}{t-s} \sum_{j=s+1}^{t} \langle L_{K^{1/2}}(X_j), f_j^* \rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - g_{(s,t]} + \lambda \widehat{f}_{(s,t]}.
\end{aligned}$$

The last equation follows from the action of $T_{\mathcal{I}}$ illustrated at (70) in the previous subsection and the result follows.

□

One key component is the expansion of variance term in the error bound. The next lemma is to structure this variance term. Define

$$f_{(s,t]}^{\lambda} = (T + \lambda_{t-s}\mathbf{I})^{-1} T f_{(s,t]}^*. \tag{88}$$

Lemma G.10. *Given* (88) *and the form of the estimator in Proposition G.1, the following holds*

$$\left(\widehat{f}_{(s,t]} - f^{\lambda}_{(s,t]}\right) = (T + \lambda_{t-s}\mathbf{I})^{-1}\left((T - T_{(s,t]})\left(\widehat{f}_{(s,t]} - f^{*}_{(s,t]}\right) + g_{(s,t]} + (T - T_{(s,t]})f^{*}_{(s,t]} + H_{(s,t]}\right)$$

where $H_{(s,t]} = (t-s)^{-1}\sum_{j=s+1}^{t}\left(\langle L_{K^{1/2}}(X_j), f^{*}_j\rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - Tf^{*}_j\right)$ *and* $g_{(s,t]}$ *defined in Proposition G.1.*

Proof.

$$\begin{aligned}
&\widehat{f}_{(s,t]} - f^{\lambda}_{(s,t]}\\
&= (T + \lambda_{t-s}\mathbf{I})^{-1}\left((T - T_{(s,t]})\widehat{f}_{(s,t]} + (T_{(s,t]} + \lambda_{t-s}\mathbf{I})\widehat{f}_{(s,t]} - (T + \lambda_{t-s}\mathbf{I})f^{\lambda}_{(s,t]}\right)\\
&= (T + \lambda_{t-s}\mathbf{I})^{-1}\left((T - T_{(s,t]})\widehat{f}_{(s,t]} + g_{(s,t]} + \frac{1}{t-s}\sum_{j=s+1}^{t}\langle L_{K^{1/2}}(X_j), f^{*}_j\rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - Tf^{*}_{(s,t]}\right)\\
&= (T + \lambda_{t-s}\mathbf{I})^{-1}\Bigg((T - T_{(s,t]})\left(\widehat{f}_{(s,t]} - f^{*}_{(s,t]}\right) + g_{(s,t]}\\
&\qquad\qquad + \frac{1}{t-s}\sum_{j=s+1}^{t}\langle L_{K^{1/2}}(X_j), f^{*}_j\rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - T_{(s,t]}f^{*}_{(s,t]}\Bigg)\\
&= (T + \lambda_{t-s}\mathbf{I})^{-1}\Bigg((T - T_{(s,t]})\left(\widehat{f}_{(s,t]} - f^{*}_{(s,t]}\right) + g_{(s,t]} + (T - T_{(s,t]})f^{*}_{(s,t]}\\
&\qquad\qquad + \frac{1}{t-s}\sum_{j=s+1}^{t}\langle L_{K^{1/2}}(X_j), f^{*}_j\rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - \frac{1}{t-s}\sum_{j=s+1}^{t} Tf^{*}_j\Bigg)
\end{aligned}$$

In the last line, we use the fact that $f^{*}_{(s,t]} = \sum_{j=s+1}^{t} f^{*}_j/(t-s)$ and linearity of the operator T. The changes at the third last line follows from Proposition G.1 and (88). □

Lemma G.11. *Let* $\xi > 0$ *and* $1 \geq b > a > 0$.

$$\max_{s<t\leq e} \delta_{t-s}^{2(b-a-1)} \left\| T^{a}(T + \lambda_{t-s}\mathbf{I})^{1-b}(f^{\lambda}_{(s,t]} - f^{*}_{(s,t]})\right\|^2_{\mathcal{L}^2} = O(1).$$

Proof. Note that because f_j is bounded, the population average $f^{*}_{(s,t]}$ is also bounded. Precisely, if $f^{*}_{(s,t]} = \sum_{l\geq 1} a_l^{s,t}\phi_l$, then $\sum_{l\geq 1}(a_l^{s,t})^2 \leq M < \infty$, for some absolute constant $M > 0$, where $\{\phi_l\}_{l\geq 1}$ is the $\mathcal{L}^2$ basis coming from the spectral decomposition of $K^{1/2}\Sigma K^{1/2}$.

$$\|T^{a}(T + \lambda_{t-s}\mathbf{I})^{1-b}(f^{\lambda}_{(s,t]} - f^{*}_{(s,t]})\|^2_{\mathcal{L}^2} = \|T^{a}(T + \lambda_{t-s}\mathbf{I})^{1-b}\left((T + \lambda_{t-s}\mathbf{I})^{-1}Tf^{*}_{(s,t]} - f^{*}_{(s,t]}\right)\|^2_{\mathcal{L}^2}$$

$$
\begin{aligned}
&= \sum_{l\geq 1} \mathfrak{s}_l^{2a}(\mathfrak{s}_l+\lambda_{t-s})^{2-2b}\left(\frac{\mathfrak{s}_l}{\mathfrak{s}_l+\lambda_{t-s}}-1\right)^2 (a_l^{s,t})^2 \\
&= \sum_{l\geq 1} \mathfrak{s}_l^{2a}\frac{\lambda_{t-s}^2}{(\mathfrak{s}_l+\lambda_{t-s})^{2b}}(a_l^{s,t})^2 \\
&\leq \left\{\max_{l\geq 1} \mathfrak{s}_l^{2a}\frac{\lambda_{t-s}^2}{(\mathfrak{s}_l+\lambda_{t-s})^{2b}}\right\}\sum_{l\geq 1}(a_l^{s,t})^2 \qquad (89) \\
&\leq \left\{\max_{l\geq 1} \mathfrak{s}_l^{2a}\frac{\lambda_{t-s}^2}{\left[(1-a/b)^{-(1-a/b)}\lambda_{t-s}^{1-a/b}(a/b)^{-a/b}\mathfrak{s}_l^{a/b}\right]^{2b}}\right\}\sum_{l\geq 1}(a_l^{s,t})^2 \qquad (90) \\
&= (1-a/b)^{2(1-a/b)}(a/b)^{2a/b}\lambda_{t-s}^{2(1+a-b)}\sum_{l\geq 1}(a_l^{s,t})^2 \\
&= O(1)\lambda_{t-s}^{2(1+a-b)}M = O\left(\lambda_{t-s}^{2(1+a-b)}\right) = O\left(\delta_{t-s}^{2(1+a-b)}\right),
\end{aligned}
$$

where the inequality in (89) is from Holder's inequality and (90) follows from Young's inequality in the following form

$$a+b \leq (pa)^{1/p}(qb)^{1/q},$$

where a, b, p, q are positive real numbers and $p^{-1}+q^{-1}=1$. □

Lemma G.12. *Let $\xi > 0$. Let $0 < \nu < \frac{1}{2}-\frac{1}{4r}$. Then*

$$\max_{s<t\leq e}\left\|T^\nu\left(\widehat{f}_{(s,t]}-f^*_{(s,t]}\right)\right\|\frac{\delta_{t-s}^{-\nu}}{\sqrt{\log^{1+\xi}(t-s)}} = O_p(1).$$

Proof. Following from triangle inequality and the decomposition at Lemma G.10 we have,

$$
\begin{aligned}
\|T^\nu(\widehat{f}_{(s,t]}-f^*_{(s,t]})\|_{\mathcal{L}^2} \leq & \|T^\nu(f^\lambda_{(s,t]}-f^*_{(s,t]})\|_{\mathcal{L}^2}+\|T^\nu(\widehat{f}_{(s,t]}-f^\lambda_{(s,t]})\|_{\mathcal{L}^2} \\
\leq & \quad \|T^\nu(f^\lambda_{(s,t]}-f^*_{(s,t]})\|_{\mathcal{L}^2} \qquad (91) \\
& + \|T^\nu(T+\lambda_{t-s}\mathbf{I})^{-1}(T-T_{(s,t]})T^{-\nu}\|_{\text{op}}\|T^\nu(\widehat{f}_{(s,t]}-f^\lambda_{(s,t]})\|_{\mathcal{L}^2} \qquad (92) \\
& + \|T^\nu(T+\lambda_{t-s}\mathbf{I})^{-1}(T-T_{(s,t]})T^{-\nu}\|_{\text{op}}\|T^\nu(f^\lambda_{(s,t]}-f^*_{(s,t]})\|_{\mathcal{L}^2} \qquad (93) \\
& + \|T^\nu(T+\lambda_{t-s}\mathbf{I})^{-1}g_{(s,t]}\|_{\mathcal{L}^2} \qquad (94) \\
& + \|T^\nu(T+\lambda_{t-s}\mathbf{I})^{-1}(T_{(s,t]}-T)\|_{\text{op}}\|f^*_{(s,t]}\|_{\mathcal{L}^2} \qquad (95) \\
& + \|T^\nu(T+\lambda_{t-s}\mathbf{I})^{-1}H_{(s,t]}\|_{\mathcal{L}^2} \qquad (96)
\end{aligned}
$$

We are going to bound each of the four term uniformly to have result.

For (91), from Lemma G.11, we write that with high probability

$$\forall t \in (s,e], \qquad \|T^\nu(f^\lambda_{(s,t]}-f^*_{(s,t]})\|_{\mathcal{L}^2} \lesssim \delta_{t-s}^\nu.$$

For (92), from Lemma G.16, we write that in high probability

$$\forall t \in (s,e], \qquad \|T^{\nu}(T+\lambda_{t-s}\mathbf{I})^{-1}(T-T_{(s,t]})T^{-\nu}\|_{\mathrm{op}} \lesssim \delta_{t-s}^{\nu}\sqrt{\log^{1+\xi}(t-s)},$$

which would give us

$$(92) \le o(1)\|T^{\nu}(\widehat{f}_{(s,t]} - f^{\lambda}_{(s,t]})\|_{\mathcal{L}^2}$$

in probability, in uniform sense.

Similarly for (93), from Lemma G.16 and Lemma G.11, we write that with high probability

$$\forall t \in (s,e], \qquad (93) \lesssim \delta_{t-s}^{2\nu}\sqrt{\log^{1+\xi}(t-s)}.$$

For (94), from Lemma G.15, we write that with high probability

$$\forall t \in (s,e], \qquad \|T^{\nu}(T+\lambda_{t-s}\mathbf{I})^{-1}g_{(s,t]}\|_{\mathcal{L}^2} \lesssim \delta_{t-s}^{\nu}\sqrt{\log^{1+\xi}(t-s)}.$$

For (95), from Lemma G.16, we write that with high probability

$$\begin{aligned}\forall t \in (s,e], \quad &\|T^{\nu}(T+\lambda_{t-s}\mathbf{I})^{-1}(T_{(s,t]}-T)\|_{\mathrm{op}}\|f^{*}_{(s,t]}\|_{\mathcal{L}^2} \\ &\lesssim \|T^{\nu}(T+\lambda_{t-s}\mathbf{I})^{-1}(T_{(s,t]}-T)\|_{\mathrm{op}} \\ &\lesssim \delta_{t-s}^{\nu}\sqrt{\log^{1+\xi}(t-s)},\end{aligned}$$

where we used the fact that $\|f^{*}_{(s,t]}\|_{\mathcal{L}^2} < \infty$.

For (96), from Lemma G.17, we write that with high probability

$$\forall t \in (s,e], \quad \|T^{\nu}(T+\lambda_{t-s}\mathbf{I})^{-1}H_{(s,t]}\|_{\mathcal{L}^2} \lesssim \delta_{t-s}^{\nu}\sqrt{\log^{1+\xi}(t-s)}.$$

This six individual bounds come together to give us the required result. $\square$

Lemma G.13. *Let $\xi > 0$ and $1 \ge b \ge a + 1/2 > 0$.*

$$\max_{s<t\le e} \frac{\delta_{t-s}^{(b-a-1)}}{\sqrt{\log^{1+\xi}(t-s)}} \left\| T^{a}(T+\lambda_{t-s}\mathbf{I})^{1-b}(\widehat{f}_{(s,t]} - f^{*}_{(s,t]}) \right\|_{\mathcal{L}^2} = O_p(1).$$

Proof. Using triangle inequality we may write,

$$\begin{aligned}\left\| T^{a}(T+\lambda_{t-s}\mathbf{I})^{1-b}(\widehat{f}_{(s,t]} - f^{*}_{(s,t]}) \right\|_{\mathcal{L}^2} \le &\left\| T^{a}(T+\lambda_{t-s}\mathbf{I})^{1-b}(\widehat{f}_{(s,t]} - f^{\lambda}_{(s,t]}) \right\|_{\mathcal{L}^2} \\ &+ \left\| T^{a}(T+\lambda_{t-s}\mathbf{I})^{1-b}(f^{\lambda}_{(s,t]} - f^{*}_{(s,t]}) \right\|_{\mathcal{L}^2} \end{aligned} \tag{97}$$

where $f^{\lambda}_{(s,t]}$ is defined at (88). The second term on the right of (97) can be bounded using Lemma G.11, which gives us

$$\max_{s<t\le e} \frac{\delta_{t-s}^{(b-a-1)}}{\sqrt{\log^{1+\xi}(t-s)}} \left\| T^{a}(T+\lambda_{t-s}\mathbf{I})^{1-b}(f^{\lambda}_{(s,t]} - f^{*}_{(s,t]}) \right\|_{\mathcal{L}^2}$$

$$\leq \max_{s<t\leq e} \delta_{t-s}^{(b-a-1)} \left\| T^a (T+\lambda_{t-s}\mathbf{I})^{1-b} (f^{\lambda}_{(s,t]} - f^*_{(s,t]}) \right\|_{\mathcal{L}^2} = O(1).$$

Now, it is suffice is to bound the first term on the right of (97). Let $0 < \nu < \frac{1}{2} - \frac{1}{4r}$. Following the decomposition at Lemma G.10, we may write

$$\begin{aligned}
&\|T^a (T+\lambda_{t-s}\mathbf{I})^{1-b} (\widehat{f}_{(s,t]} - f^{\lambda}_{(s,t]})\|_{\mathcal{L}^2} \\
\leq &\|T^a (T+\lambda_{t-s}\mathbf{I})^{-b} (T - T_{(s,t]}) T^{-\nu}\|_{\text{op}} \|T^{\nu} (\widehat{f}_{(s,t]} - f^*_{(s,t]})\|_{\mathcal{L}^2} \qquad (98)\\
&+ \|T^a (T+\lambda_{t-s}\mathbf{I})^{-b} g_{(s,t]}\|_{\mathcal{L}^2} \qquad (99)\\
&+ \|T^a (T+\lambda_{t-s}\mathbf{I})^{-b} (T_{(s,t]} - T)\|_{\text{op}} \|f^*_{(s,t]}\|_{\mathcal{L}^2} \qquad (100)\\
&+ \|T^a (T+\lambda_{t-s}\mathbf{I})^{-b} H_{(s,t]}\|_{\mathcal{L}^2} \qquad (101)
\end{aligned}$$

We are going to bound each of the four terms (98), (99), (100) and (101) uniformly over $s < t \leq e$ to have the required result.

For (98), using Lemma G.16 and Lemma G.12, we write that with high probability

$$\begin{aligned}
\forall t \in (s,e], \qquad &\|T^a (T+\lambda_{t-s}\mathbf{I})^{-b} (T - T_{(s,t]}) T^{-\nu}\|_{\text{op}} \|T^{\nu} (\widehat{f}_{(s,t]} - f^*_{(s,t]})\|_{\mathcal{L}^2} \\
&\lesssim \delta_{t-s}^{1+a-b} \sqrt{\log^{1+\xi}(t-s)}\, \delta_{t-s}^{\nu} \sqrt{\log^{1+\xi}(t-s)} \\
&\leq \delta_{t-s}^{1+a-b} \sqrt{\log^{1+\xi}(t-s)}.
\end{aligned}$$

For (99), from Lemma G.15, we write that with high probability

$$\forall t \in (s,e], \qquad \|T^a (T+\lambda_{t-s}\mathbf{I})^{-b} g_{(s,t]}\|_{\mathcal{L}^2} \lesssim \delta_{t-s}^{1+a-b} \sqrt{\log^{1+\xi}(t-s)}.$$

For (100), from Lemma G.16, we write that with high probability

$$\begin{aligned}
\forall t \in (s,e], \qquad &\|T^a (T+\lambda_{t-s}\mathbf{I})^{-b} (T_{(s,t]} - T)\|_{\text{op}} \|f^*_{(s,t]}\|_{\mathcal{L}^2} \\
&\lesssim \|T^a (T+\lambda_{t-s}\mathbf{I})^{-b} (T_{(s,t]} - T)\|_{\text{op}} \\
&\lesssim \delta_{t-s}^{1+a-b} \sqrt{\log^{1+\xi}(t-s)}.
\end{aligned}$$

here we used the fact that $\|f^*_{(s,t]}\|_{\mathcal{L}^2} < \infty$.

For (101), from Lemma G.17, we write that with high probability

$$\forall t \in (s,e], \qquad \|T^a (T+\lambda_{t-s}\mathbf{I})^{-b} H_{(s,t]}\|_{\mathcal{L}^2} \lesssim \delta_{t-s}^{1+a-b} \sqrt{\log^{1+\xi}(t-s)}.$$

This four individual bounds come together to give us the required bound for the first term on the right of (97).

□

Lemma G.14. *Let $0 < \nu < \frac{1}{2} - \frac{1}{4r}$. Let $p \in \{0, \nu\}$ and $\xi > 0$. Then*

$$\max_{s<t\le e} \frac{\delta_{t-s}^{-1}}{\log^{1+\xi}(t-s)} \|T^{-p}(T_{(s,t]} - T)(\widehat{f}_{(s,t]} - f^*_{(s,t]})\|_{\mathcal{L}^2} = O_p(1).$$

Proof. Using the linear operator norm inequality, we may have

$$\begin{aligned}
&\|T^{-p}(T_{(s,t]} - T)(\widehat{f}_{(s,t]} - f^*_{(s,t]})\|_{\mathcal{L}^2} \\
\le &\left\|T^{-p}(T_{(s,t]} - T)(T + \lambda_{t-s}\mathbf{I})^{-1/4}T^{-1/4}\right\|_{\mathrm{op}} \left\|T^{1/4}(T + \lambda_{t-s}\mathbf{I})^{1/4}\left(\widehat{f}_{(s,t]} - f^*_{(s,t]}\right)\right\|_{\mathcal{L}^2}.
\end{aligned}$$

We are going to bound each of the two terms here. For the first one, using Lemma G.16, we write that with high probability

$$\forall t \in (s, e] \qquad \left\|T^{-p}(T_{(s,t]} - T)(T + \lambda_{t-s}\mathbf{I})^{-1/4}T^{-1/4}\right\|_{\mathrm{op}} \lesssim \delta_{t-s}^{1/2}\sqrt{\log^{1+\xi}(t-s)}.$$

And for the second term, we use Lemma G.13 to have

$$\forall t \in (s, e] \qquad \left\|T^{1/4}(T + \lambda_{t-s}\mathbf{I})^{1/4}\left(\widehat{f}_{(s,t]} - f^*_{(s,t]}\right)\right\|_{\mathcal{L}^2} \lesssim \delta_{t-s}^{1/2}\sqrt{\log^{1+\xi}(t-s)}.$$

The two bounds come together to have the required result. $\square$

Lemma G.15. *Let $\xi > 0$ and $1 \ge b \ge a + 1/2 \ge 1/4$. Then we have*

$$\mathbb{E}\left[\max_{s<t\le e} \frac{\delta_{t-s}^{2(b-a-1)}}{\log^{1+\xi}(t-s)} \|T^a(T + \lambda_{t-s}\mathbf{I})^{-b} g_{(s,t]}\|_{\mathcal{L}^2}^2\right] = O(1),$$

where $g_{(s,t]} = \frac{1}{(t-s)}\sum_{j=s+1}^{t} \varepsilon_j L_{K^{1/2}}(X_j)$ defined in Proposition G.1.

Proof. We may write

$$\begin{aligned}
\|T^a(T + \lambda_{t-s}\mathbf{I})^{-b} g_{(s,t]}\|_{\mathcal{L}^2}^2 &= \sum_{l\ge1} \langle T^a(T + \lambda_{t-s}\mathbf{I})^{-b} g_{(s,t]}, \phi_l\rangle_{\mathcal{L}^2}^2 \\
&= \sum_{l\ge1} \langle g_{(s,t]}, T^a(T + \lambda_{t-s}\mathbf{I})^{-b}\phi_l\rangle_{\mathcal{L}^2}^2 \\
&= \sum_{l\ge1} \frac{\mathfrak{s}_l^{2a}}{(\mathfrak{s}_l + \lambda_{t-s})^{2b}} \langle g_{(s,t]}, \phi_l\rangle_{\mathcal{L}^2}^2 \\
&= \sum_{l\ge1} \frac{\mathfrak{s}_l^{2a}}{(\mathfrak{s}_l + \lambda_{t-s})^{2b}} \langle \frac{1}{t-s}\sum_{i=s+1}^{t} \varepsilon_i L_{K^{1/2}}(X_i), \phi_l\rangle_{\mathcal{L}^2}^2
\end{aligned}$$

$$= \sum_{l\geq 1} \frac{\mathfrak{s}_l^{2a}}{(\mathfrak{s}_l+\lambda_{t-s})^{2b}} \left(\frac{1}{t-s} \sum_{j=s+1}^{t} \varepsilon_j \langle X_j, L_{K^{1/2}}\phi_l\rangle_{\mathcal{L}^2} \right)^2$$

By linearity of the expectation it lead us to

$$\mathbb{E}\left[\max_{s<t\leq e} \|T^a(T+\lambda_{t-s}\mathbf{I})^{-b} g_{(s,t]}\|^2 \frac{\delta_t^{2(b-a-1)}}{\log^{1+\xi} t} \right]$$

$$\leq \sum_{l\geq 1} \mathbb{E}\left[\max_{s<t\leq e} \left(\frac{\delta_{t-s}^{(b-a-1)}}{(\log(t-s))^{(1+\xi)/2}} \frac{\mathfrak{s}_l^a}{(t-s)(\mathfrak{s}_l+\lambda_{t-s})^b} \sum_{j=1}^{t} \varepsilon_j \langle X_j, L_{K^{1/2}}\phi_l\rangle_{\mathcal{L}^2} \right)^2 \right]$$

$$= O(1). \tag{102}$$

Here (102) follows from Lemma G.18. □

Lemma G.16. *Let $\xi > 0$ and $1 \geq b \geq a + 1/2 \geq 1/4$. Let $0 < \nu < \frac{1}{2} - \frac{1}{4r}$. Suppose $p \in \{0, \nu\}$. Then*

$$\mathbb{E}\left[\max_{s<t\leq e} \frac{\delta_{t-s}^{2(b-a-1)}}{\log^{1+\xi}(t-s)} \|T^a(T+\lambda_{t-s}\mathbf{I})^{-b}(T_{(s,t]}-T)T^{-p}\|_{\text{op}}^2 \right] = O(1).$$

Proof. Using the definition of operator norm, we may write

$$\|T^a(T+\lambda_{t-s}\mathbf{I})^{-b}(T-T_{(s,t]})T^{-p}\|_{\text{op}} := \sup_{\|h\|_{\mathcal{L}^2}=1} \left| \langle h, T^a(T+\lambda_{t-s}\mathbf{I})^{-b}(T-T_{(s,t]})T^{-p}h\rangle_{\mathcal{L}^2} \right|.$$

Let $h \in \mathcal{L}^2$ such that $\|h\|_{\mathcal{L}^2} = 1$. This means $h = \sum_{j\geq 1} h_j\phi_j$ and $\sum_{j\geq 1} h_j^2 = 1$. Then

$$\begin{aligned}
&\langle h, T^a(T+\lambda_{t-s}\mathbf{I})^{-b}(T-T_{(s,t]})T^{-p}h\rangle_{\mathcal{L}^2} \\
=&\sum_{j\geq 1}\sum_{m\geq 1} h_j h_m \langle \phi_j, T^a(T+\lambda_{t-s}\mathbf{I})^{-b}(T-T_{(s,t]})T^{-p}\phi_m\rangle_{\mathcal{L}^2} \\
=&\sum_{j\geq 1}\sum_{m\geq 1} h_j h_m \langle T^a(T+\lambda_{t-s}\mathbf{I})^{-b}\phi_j, (T-T_{(s,t]})T^{-p}\phi_m\rangle_{\mathcal{L}^2} \\
=&\sum_{j\geq 1}\sum_{m\geq 1} h_j h_m \left\langle \frac{\mathfrak{s}_j^a}{(\mathfrak{s}_j+\lambda_{t-s})^b}\phi_j, (T-T_{(s,t]})\mathfrak{s}_m^{-p}\phi_m \right\rangle_{\mathcal{L}^2} \\
=&\sum_{j\geq 1}\sum_{m\geq 1} h_j h_m \frac{\mathfrak{s}_j^a}{(\mathfrak{s}_j+\lambda_{t-s})^b}\mathfrak{s}_m^{-p} \left\langle \phi_j, (T-T_{(s,t]})\phi_m \right\rangle_{\mathcal{L}^2} \\
\leq&\sqrt{\sum_{j\geq 1}\sum_{m\geq 1} h_j^2 h_m^2}\sqrt{\sum_{j\geq 1}\sum_{m\geq 1} \frac{\mathfrak{s}_j^{2a}}{(\mathfrak{s}_j+\lambda_{t-s})^{2b}}\mathfrak{s}_m^{-2p} \left\langle \phi_j, (T-T_{(s,t]})\phi_m \right\rangle_{\mathcal{L}^2}^2} &\quad (103) \\
=&\sqrt{\sum_{j\geq 1}\sum_{m\geq 1} \frac{\mathfrak{s}_j^{2a}}{(\mathfrak{s}_j+\lambda_{t-s})^{2b}}\mathfrak{s}_m^{-2\nu} \left\langle \phi_j, (T-T_{(s,t]})\phi_m \right\rangle_{\mathcal{L}^2}^2} &\quad (104)
\end{aligned}$$

The second last inequality (103) follows from Cauchy-Schwarz, where one may think $\langle A, B\rangle = \sum_{j\geq 1}\sum_{m\geq 1} A_{jm}B_{jm}$. The last equality (104) follows from

$$\sum_{j\geq 1}\sum_{m\geq 1} h_j^2 h_m^2 = \sum_{j\geq 1} h_j^2 \sum_{m\geq 1} h_m^2 = 1,$$

by definition of h.

We have,

$$\|T^a(T+\lambda_{t-s}\mathbf{I})^{-b}(T-T_{(s,t]})T^{-p}\|_{\mathrm{op}}^2 \leq \sum_{j\geq 1}\sum_{m\geq 1} \frac{\mathfrak{s}_j^{2a}}{(\mathfrak{s}_j+\lambda_{t-s})^{2b}} \mathfrak{s}_m^{-2p} \left\langle \phi_j, (T-T_{(s,t]})\phi_m \right\rangle_{\mathcal{L}^2}^2 \tag{105}$$

By linearity of expectation

$$\begin{aligned}
&\mathbb{E}\left[\max_{s<t\leq e} \|T^a(T+\lambda_{t-s}\mathbf{I})^{-b}(T_{(s,t]}-T)T^{-p}\|_{\mathrm{op}}^2 \frac{\delta_{t-s}^{2(b-a-1)}}{\log^{1+\xi} t}\right] \\
&\leq \sum_{m\geq 1} \mathfrak{s}_m^{-2p} \mathbb{E}\left[\max_{s<t\leq e} \sum_{j\geq 1} \frac{\mathfrak{s}_j^{2a}}{(\mathfrak{s}_j+\lambda_{t-s})^{2b}} \frac{\delta_t^{2(b-a-1)}}{\log^{1+\xi} t} \left|\langle \phi_k, (T_{(s,t]}-T)\phi_j\rangle_{\mathcal{L}^2}\right|^2\right] \\
&\lesssim \sum_{m\geq 1} \mathfrak{s}_m^{1-2p} < \infty,
\end{aligned}$$

where the last line follows from Lemma G.19, and $\sum_{m\geq 1} \mathfrak{s}_m^{1-2p} \asymp \sum_{m\geq 1} m^{-(1-2p)2r}$ is summable given we have $(1-2p)2r > 1$. $\square$

Lemma G.17. *Let $\xi > 0$ and $1 \geq b \geq a + 1/2 \geq 1/4$. Suppose $\{h_j\}$ be some $\mathcal{L}^2$ sequence that satisfies $\Sigma\left[L_{K^{1/2}}(h_j), L_{K^{1/2}}(h_j)\right] \leq M < \infty$. Then we have*

$$\begin{aligned}
&\mathbb{E}\left[\max_{s<t\leq e} \frac{\delta_{t-s}^{2(b-a-1)}}{\log^{1+\xi}(t-s)} \left\| T^a \left(T+\lambda_{t-s}\mathbf{I}\right)^{-b} \right.\right. \\
&\left.\left.\left(\frac{1}{t-s}\sum_{j=s+1}^{t} \left(\langle L_{K^{1/2}}(X_j), h_j\rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - Th_j\right)\right)\right\|_{\mathrm{op}}^2\right] = O(1).
\end{aligned} \tag{106}$$

Consequently, it holds that

$$\mathbb{E}\left[\max_{s<t\leq e} \frac{\delta_{t-s}^{2(b-a-1)}}{\log^{1+\xi}(t-s)} \|T^a \left(T+\lambda_{t-s}\mathbf{I}\right)^{-b} H_{(s,t]}\|_{\mathrm{op}}^2\right] = O(1), \tag{107}$$

where $H_{(s,t]} = \frac{1}{t-s}\sum_{j=s+1}^{t}\left(\langle L_{K^{1/2}}(X_j), f_j^*\rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - Tf_j^*\right)$ defined in Lemma G.10, and that

$$\mathbb{E}\left[\max_{s<t\leq e} \frac{\delta_{t-s}^{2(b-a-1)}}{\log^{1+\xi}(t-s)} \|T^a \left(T+\lambda_{t-s}\mathbf{I}\right)^{-b} G_{(s,t]}\|_{\mathrm{op}}^2\right] = O(1), \tag{108}$$

where $G_{(s,t]} = \frac{1}{t-s}\sum_{j=s+1}^{t}\left(\langle L_{K^{1/2}}(X_j), f_j^* - f_{(s,t]}^*\rangle_{\mathcal{L}^2} L_{K^{1/2}}(X_j) - T(f_j^* - f_{(s,t]}^*)\right)$.

Proof. We may write

$$
\begin{aligned}
&\left\| T^a\left(T+\lambda_{t-s}\mathbf{I}\right)^{-b}\left(\frac{1}{t-s}\sum_{j=s+1}^{t}\left(\langle L_{K^{1/2}}(X_j),h_j\rangle_{\mathcal{L}^2}L_{K^{1/2}}(X_j)-Th_j\right)\right)\right\|_{\mathrm{op}}^2\\
&=\sum_{m\geq 1}\left\langle T^a(T+\lambda_{t-s}\mathbf{I})^{-b}\left(\frac{1}{t-s}\sum_{j=s+1}^{t}\left(\langle L_{K^{1/2}}(X_j),h_j\rangle_{\mathcal{L}^2}L_{K^{1/2}}(X_j)-Th_j\right)\right),\phi_m\right\rangle_{\mathcal{L}^2}^2\\
&=\sum_{m\geq 1}\left\langle \left(\frac{1}{t-s}\sum_{j=s+1}^{t}\left(\langle L_{K^{1/2}}(X_j),h_j\rangle_{\mathcal{L}^2}L_{K^{1/2}}(X_j)-Th_j\right)\right),T^a(T+\lambda_{t-s}\mathbf{I})^{-b}\phi_m\right\rangle_{\mathcal{L}^2}^2\\
&=\sum_{m\geq 1}\frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m+\lambda_{t-s})^{2b}}\left\langle \left(\frac{1}{t-s}\sum_{j=s+1}^{t}\left(\langle L_{K^{1/2}}(X_j),h_j\rangle_{\mathcal{L}^2}L_{K^{1/2}}(X_j)-Th_j\right)\right),\phi_m\right\rangle_{\mathcal{L}^2}^2\\
&=\sum_{m\geq 1}\frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m+\lambda_{t-s})^{2b}}\left(\frac{1}{t-s}\sum_{j=s+1}^{t}\left\{\langle X_i,L_{K^{1/2}}(h_j)\rangle_{\mathcal{L}^2}\langle X_i,L_{K^{1/2}}\phi_m\rangle_{\mathcal{L}^2}-\langle h_j,T\phi_m\rangle_{\mathcal{L}^2}\right\}\right)^2
\end{aligned}
$$

By linearity of the expectation it lead us to

$$
\begin{aligned}
&\mathbb{E}\left[\max_{1<t\leq n}\left\|T^a(T+\lambda_{t-s}\mathbf{I})^{-b}\left(\frac{1}{t-s}\sum_{j=s+1}^{t}\left(\langle L_{K^{1/2}}(X_j),h_j\rangle_{\mathcal{L}^2}L_{K^{1/2}}(X_j)-Th_j\right)\right)\right\|^2\frac{\delta_t^{2(b-a-1)}}{\log^{1+\xi}t}\right]\\
&\leq\sum_{m\geq 1}\mathbb{E}\Bigg[\max_{1<t\leq n}\Bigg(\frac{\delta_{t-s}^{(b-a-1)}\mathfrak{s}_m^{a}(\mathfrak{s}_m+\lambda_{t-s})^{-b}}{(t-s)(\log(t-s))^{(1+\xi)/2}}\\
&\qquad\sum_{j=s+1}^{t}\left\{\langle X_i,L_{K^{1/2}}(h_j)\rangle_{\mathcal{L}^2}\langle X_i,L_{K^{1/2}}\phi_m\rangle_{\mathcal{L}^2}-\langle h_j,T\phi_m\rangle_{\mathcal{L}^2}\right\}\Bigg)^2\Bigg]=O(1).
\end{aligned}
\tag{109}
$$

Here (109) follows from Lemma G.20 because we have $\Sigma\left[L_{K^{1/2}}(h_j),L_{K^{1/2}}(h_j)\right]<\infty$.

The result (107) follows from (106) because $\|\beta_j^*\|_{\mathcal{H}(K)}<\infty$. For (108), we can again use (106) as we have $\Sigma\left[\beta_j^*-\beta_{(s,t]}^*,\beta_j^*-\beta_{(s,t]}^*\right]\leq O(1)\max_{1\leq k\leq\mathcal{K}}\|\beta_{\eta_k}^*\|_{\mathcal{H}(K)}<\infty$. □

Lemma G.18. *Let $\xi>0$ and $1\geq b\geq a+1/2\geq 1/4$. Then we have*

$$
\mathbb{E}\left[\max_{s<t\leq e}\sum_{m\geq 1}\frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m+\lambda_{t-s})^{2b}}\frac{\delta_{t-s}^{2(b-a-1)}}{\log^{1+\xi}(t-s)}\left|\frac{1}{t-s}\sum_{j=s+1}^{t}\langle X_j,L_{K^{1/2}}(\phi_m)\rangle_{\mathcal{L}^2}\varepsilon_j\right|^2\right]=O(1).
$$

Proof. For simplicity, denote $Y_{j,m} = \langle X_j, L_{K^{1/2}}(\phi_m)\rangle_{\mathcal{L}^2}\varepsilon_j$. Observe that

$$\Sigma\left[L_{K^{1/2}}(\phi_m), L_{K^{1/2}}(\phi_m)\right] = \langle T\phi_m, \phi_m\rangle_{\mathcal{L}^2} = \mathfrak{s}_m.$$

We are going to prove the result for a general interval $\{1, \dots, T\}$, the result for the $(s, e]$ follows from translation and stationarity.

Using Lemma G.9, we may write

$$\mathbb{E}\left[\left|\sum_{j=1}^{t} Y_{j,m}\right|^2\right] \le O(t)\mathfrak{s}_m.$$

We apply Lemma J.3 and to have this result: for any non-decreasing sequence $\{\gamma_t\}_{t=1}^T$

$$\mathbb{E}\left[\max_{1<t\le T}\left|\frac{1}{\gamma_t}\sum_{j=1}^{t} Y_{j,m}\right|^2\right] = C\sum_{t=1}^{T}\frac{1}{\gamma_t^2}\mathfrak{s}_m, \tag{110}$$

for some constant $C > 0$.

Observe that

$$\frac{\delta_t^{2(b-a-1)}}{(\mathfrak{s}_m + \lambda_t)^{2b-(1+2a)+(1+2a)}} \le \frac{\delta_t^{2(b-a-1)}}{\lambda_t^{2b-(1+2a)}(\mathfrak{s}_m + \lambda_t)^{(1+2a)}} \lesssim \frac{\delta_t^{-1}}{(\mathfrak{s}_m + \lambda_t)^{1+2a}}.$$

It led us to

$$\frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m + \lambda_t)^{2b}}\frac{\delta_t^{2(b-a-1)}}{\log^{1+\xi} t}\left|\frac{1}{t}\sum_{j=1}^{t} Y_{j,m}\right|^2 \lesssim \frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m + \lambda_t)^{1+2a}}\frac{\delta_t^{-1}}{\log^{1+\xi} t}\left|\frac{1}{t}\sum_{j=1}^{t} Y_{j,m}\right|^2$$
$$\le \left\{\frac{\mathfrak{s}_m^{2a}}{\mathfrak{s}_m^{1+2a}} \wedge \frac{\mathfrak{s}_m^{2a}}{\lambda_t^{1+2a}}\right\}\frac{\delta_t^{-1}}{\log^{1+\xi} t}\left|\frac{1}{t}\sum_{j=1}^{t} Y_{j,m}\right|^2,$$

which further implies that

$$\sum_{m\ge 1}\frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m + \lambda_t)^{2b}}\frac{\delta_t^{2(b-a-1)}}{\log^{1+\xi} t}\left|\frac{1}{t}\sum_{j=1}^{t} Y_{j,m}\right|^2$$
$$\lesssim \sum_{m\ge 1}\left\{\frac{\mathfrak{s}_m^{2a}}{\mathfrak{s}_m^{1+2a}} \wedge \frac{\mathfrak{s}_m^{2a}}{\lambda_t^{1+2a}}\right\}\frac{\delta_t^{-1}}{\log^{1+\xi} t}\left|\frac{1}{t}\sum_{j=1}^{t} Y_{j,m}\right|^2. \tag{111}$$

Case I: $a \le 0$

Let $f_m = \lfloor m^{(2r+1)} \rfloor \wedge T$. Using (111), we write

$$\begin{aligned}
&\sum_{m\geq 1} \frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m+\lambda_t)^{2b}} \frac{\delta_t^{2(b-a-1)}}{\log^{1+\xi} t} \left| \frac{1}{t}\sum_{j=1}^{t} Y_{j,m} \right|^2 \\
&\leq \sum_{m\geq 1} \mathbf{I}\left\{t \leq f_m\right\} \mathfrak{s}_m^{2a} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{\lambda_t^{a+1/2} t} \sum_{j=1}^{t} Y_{j,m} \right|^2 \\
&\quad + \sum_{m\geq 1} \mathbf{I}\left\{t > f_m\right\} \mathfrak{s}_m^{-1} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{t} \sum_{j=1}^{t} Y_{j,m} \right|^2 .
\end{aligned} \tag{112}$$

Observe that for $2 \leq t \leq T$,

$$\frac{d}{dt}\left(t\delta_t^{1/2}(\log t)^{(1+\xi)/2}\right) = t^{(1+r)/(2r+1)}(\log t)^{(\xi-1)/2}\left(\frac{1+r}{2r+1} + \frac{1+\xi}{2}\log t\right) > 0$$

and

$$\frac{d}{dt}\left(t\lambda_t^{a+1/2}\delta_t^{1/2}(\log t)^{(1+\xi)/2}\right) = \lambda_t^{a+1/2}\delta_t^{1/2}(\log t)^{(\xi-1)/2}\left(\frac{1+r-(a+1/2)2r}{2r+1}\log t + \frac{1+\xi}{2}\right) > 0.$$

This says that $\left\{t\delta_t^{1/2}(\log t)^{(1+\xi)/2}\right\}$ and $\left\{t\lambda_t^{a+1/2}\delta_t^{1/2}(\log t)^{(1+\xi)/2}\right\}$ satisfies the criteria for $\{\gamma_t\}$ in (110).

This observation on derivatives and (112) helps us to write

$$\begin{aligned}
&\mathbb{E}\left[\max_{1<t\leq T} \sum_{m\geq 1} \frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m+\lambda_t)^{2b}} \frac{\delta_t^{2(b-a-1)}}{\log^{1+\xi} t} \left| \frac{1}{t}\sum_{j=1}^{t} Y_{j,m} \right|^2\right] \\
&\leq \sum_{m\geq 1} \mathfrak{s}_m^{2a} \mathbb{E}\left[\max_{1<t\leq f_m} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{\lambda_t^{a+1/2} t} \sum_{j=1}^{t} Y_{j,m} \right|^2\right] \\
&\quad + \sum_{m\geq 1} \mathfrak{s}_m^{-1} \mathbb{E}\left[\max_{f_m<t\leq T} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{t} \sum_{j=1}^{t} Y_{j,m} \right|^2\right] \\
&\leq \sum_{m\geq 1} \mathfrak{s}_m^{2a} c \sum_{t\leq f_m} \frac{\delta_t^{-1}}{\log^{1+\xi} t \lambda_t^{1+2a} t^2} \mathfrak{s}_m + \sum_{m\geq 1} \mathfrak{s}_m^{-1} c \sum_{t>f_m} \frac{\delta_t^{-1}}{t^2 \log^{1+\xi} t} \mathfrak{s}_m && (113) \\
&= c \sum_{1<t\leq T} \frac{\delta_t^{-1}}{t^2 \log^{1+\xi} t} \sum_{m\geq \delta_t^{-1/2r}} \frac{\mathfrak{s}_m^{1+2a}}{\lambda_t^{1+2a}} + c \sum_{1<t\leq T} \frac{\delta_t^{-1}}{t^2 \log^{1+\xi} t} \sum_{m<\delta_t^{-1/2r}} 1 \\
&= c \sum_{1<t\leq T} \frac{\delta_t^{-1}}{t^2 \log^{1+\xi} t} O(\delta_t^{-1/2r}) + c \sum_{1<t\leq T} \frac{\delta_t^{-1}}{t^2 \log^{1+\xi} t} O(\delta_t^{-1/2r}) && (114) \\
&= \sum_{1<t\leq T} \frac{1}{t \log^{1+\xi} t} O\left(\frac{\delta_t^{-1-1/2r}}{t}\right) = \sum_{1<t\leq T} \frac{1}{t \log^{1+\xi} t} O(1) < \infty.
\end{aligned}$$

The (113) follows from (110) and (114) follows from (126) with the observation

$$\sum_{m\geq \delta_t^{-1/2r}} \frac{\mathfrak{s}_m^{1+2a}}{\lambda_t^{1+2a}} \lesssim \sum_{m\geq \delta_t^{-1/2r}} \frac{1}{(m^{2r}\delta_t)^{1+2a}} \leq O(1) \int_{\delta_t^{-1/2r}}^{\infty} \frac{1}{(x^{2r}\delta_t)^{1+2a}} dx.$$

Case II: $a > 0$

Let $f_m = \lfloor m^{(2r+1)} \rfloor \wedge T$. Using (111), we write

$$
\begin{aligned}
&\sum_{m\geq 1} \frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m+\lambda_t)^{2b}} \frac{\delta_t^{2(b-a-1)}}{\log^{1+\xi} t} \left| \frac{1}{t} \sum_{j=1}^{t} Y_{j,m} \right|^2 \\
&\leq \sum_{m\geq 1} \mathbf{I}\left\{t < f_m\right\} \mathfrak{s}_m^{2a} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{\lambda_t^{a+1/2} t} \sum_{j=1}^{t} Y_{j,m} \right|^2 \\
&\quad + \sum_{m\geq 1} \mathbf{I}\left\{t \geq f_m\right\} \mathfrak{s}_m^{-1} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{t} \sum_{j=1}^{t} Y_{j,m} \right|^2 \\
&\leq \sum_{m\geq 1} \mathbf{I}\left\{t < f_m\right\} \frac{\mathfrak{s}_m^{2a}}{\lambda_{f_m}^{2a}} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{\lambda_t^{1/2} t} \sum_{j=1}^{t} Y_{j,m} \right|^2 \qquad (115) \\
&\quad + \sum_{m\geq 1} \mathbf{I}\left\{t \geq f_m\right\} \mathfrak{s}_m^{-1} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{t} \sum_{j=1}^{t} Y_{j,m} \right|^2 \qquad (116)
\end{aligned}
$$

We have $t < f_m \Rightarrow \lambda_t > \lambda_{f_m}$ which gives us (115) and (116).
Observe that for $2 \leq t \leq n$,

$$
\frac{d}{dt}\left(t\delta_t^{1/2} (\log t)^{(1+\xi)/2} \right) = t^{(1+r)/(2r+1)} (\log t)^{(\xi-1)/2} \left(\frac{1+r}{2r+1} + \frac{1+\xi}{2} \log t \right) > 0
$$

and

$$
\frac{d}{dt}\left(t\lambda_t^{1/2}\delta_t^{1/2} (\log t)^{(1+\xi)/2} \right) = \lambda_t^{1/2}\delta_t^{1/2} (\log t)^{(\xi-1)/2} \left(\frac{1}{2r+1} \log t + \frac{1+\xi}{2} \right) > 0.
$$

This says that $\left\{ t\delta_t^{1/2} (\log t)^{(1+\xi)/2} \right\}$ and $\left\{ t\lambda_t^{1/2}\delta_t^{1/2} (\log t)^{(1+\xi)/2} \right\}$ satisfies the criteria for $\{\gamma_t\}$ in (110).

This observation on derivatives and (115) helps us to write

$$
\begin{aligned}
&\mathbb{E}\left[\max_{1<t\leq T} \sum_{m\geq 1} \frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m+\lambda_t)^{2b}} \frac{\delta_t^{2(b-a-1)}}{\log^{1+\xi} t} \left| \frac{1}{t} \sum_{j=1}^{t} Y_{j,m} \right|^2 \right] \\
&\leq \sum_{m\geq 1} \frac{\mathfrak{s}_m^{2a}}{\lambda_{f_m}^{2a}} \mathbb{E}\left[\max_{1<t<f_m} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{\lambda_t^{1/2} t} \sum_{j=1}^{t} Y_{j,m} \right|^2 \right] \\
&\quad + \sum_{m\geq 1} \mathfrak{s}_m^{-1} \mathbb{E}\left[\max_{f_m\leq t\leq T} \left| \frac{\delta_t^{-1/2}}{(\log t)^{(1+\xi)/2}} \frac{1}{t} \sum_{j=1}^{t} Y_{j,m} \right|^2 \right] \\
&\leq \sum_{m\geq 1} \frac{\mathfrak{s}_m^{2a}}{\lambda_{f_m}^{2a}} c \sum_{t<f_m} \frac{\delta_t^{-1}}{\log^{1+\xi} t \lambda_t t^2} \mathfrak{s}_m + \sum_{m\geq 1} \mathfrak{s}_m^{-1} c \sum_{t\geq f_m} \frac{\delta_t^{-1}}{t^2 \log^{1+\xi} t} \mathfrak{s}_m \qquad (117)
\end{aligned}
$$

$$
\begin{aligned}
&= c\sum_{1<t\le T}\frac{\delta_t^{-1}}{t^2\log^{1+\xi}t}\sum_{m>\delta_t^{-1/2r}}\frac{\mathfrak{s}_m^{1+2a}}{\lambda_t\lambda_{f_m}^{2a}} + c\sum_{1<t\le T}\frac{\delta_t^{-1}}{t^2\log^{1+\xi}t}\sum_{m\le\delta_t^{-1/2r}} 1\\
&= c\sum_{1<t\le T}\frac{\delta_t^{-1}}{t^2\log^{1+\xi}t}O(\delta_t^{-1/2r}) + c\sum_{1<t\le T}\frac{\delta_t^{-1}}{t^2\log^{1+\xi}t}O(\delta_t^{-1/2r}) \qquad (118)\\
&= \sum_{1<t\le T}\frac{1}{t\log^{1+\xi}t}O\left(\frac{\delta_t^{-1-1/2r}}{t}\right) = \sum_{1<t\le T}\frac{1}{t\log^{1+\xi}t}O(1) < \infty.
\end{aligned}
$$

The (117) follows from (110). For (118), with the realization $\lambda_{f_m} \asymp \mathfrak{s}_m$ we may write

$$
\frac{\mathfrak{s}_m^{1+2a}}{\lambda_t\lambda_{f_m}^{2a}} \le c_2\frac{1}{m^{2r}\lambda_t} \implies \sum_{m>\delta_r^{-1/2r}}\frac{\mathfrak{s}_m^{1+2a}}{\lambda_t\lambda_{f_m}^{2a}} \le c_2\sum_{m>\delta_t^{-1/2r}}\frac{1}{m^{2r}\lambda_t} \le c_2\int_{\delta_t^{-1/2r}}^{\infty}\frac{1}{x^{2r}\lambda_t}dx = O(\delta_t^{-1/2r}),
$$

similar idea is outlined at (126) which comes as a consequence from Lemma J.1.

□

Lemma G.19. *Let $\xi > 0$ and $1 \ge b \ge a + 1/2 \ge 1/4$. Then for any $k \ge 1$, we have*

$$
\mathbb{E}\left[\max_{s<t\le e}\sum_{j\ge1}\frac{\mathfrak{s}_j^{2a}}{(\mathfrak{s}_j+\lambda_{t-s})^{2b}}\frac{\delta_{t-s}^{2(b-a-1)}}{\log^{1+\xi}(t-s)}\left|\langle\phi_k,(T_{(s,t]}-T)\phi_j\rangle_{\mathcal{L}^2}\right|^2\right] = O(\mathfrak{s}_k).
$$

Proof. Denote $u_{j,k} = \langle X_j, L_{K^{1/2}}(\phi_k)\rangle_{\mathcal{L}^2}$ and $u_{j,m} = \langle X_j, L_{K^{1/2}}(\phi_m)\rangle_{\mathcal{L}^2}$. Let $Y_{j,m}^k = u_{j,k}u_{j,m} - \mathbb{E}[u_{j,k}u_{j,m}] = \langle X_j, L_{K^{1/2}}(\phi_k)\rangle_{\mathcal{L}^2}\langle X_j, L_{K^{1/2}}(\phi_m)\rangle_{\mathcal{L}^2} - \langle\phi_k, T\phi_m\rangle_{\mathcal{L}^2}$. Observe that

$$
\langle\phi_k,(T_{(s,t]}-T)\phi_j\rangle_{\mathcal{L}^2} = \frac{1}{t-s}\sum_{i=s+1}^{t}Y_{j,m}^k.
$$

Again, We are going to prove the result for a general interval $\{1,\dots,T\}$, the result for the $(s,e]$ follows from translation and stationarity.

Using Lemma G.8, we may write

$$
\mathbb{E}\left[\left|\sum_{j=1}^{t}Y_{j,m}^k\right|^2\right] = O(t)\mathfrak{s}_k\mathfrak{s}_m.
$$

We use Lemma J.3 to establish for any non-decreasing sequence $\{\gamma_t\}$

$$
\mathbb{E}\left[\max_{1<t\le T}\left|\frac{1}{\gamma_t}\sum_{j=1}^{t}Y_{j,m}^k\right|^2\right] = C\sum_{t=1}^{T}\frac{1}{\gamma_t^2}\mathfrak{s}_k\mathfrak{s}_m, \qquad (119)
$$

for some constant $C > 0$.

The rest of proof follows exactly as the proof of Lemma G.18, just by replacing $Y_{j,m}^k$ with $Y_{j,m}$ and therefore omitted.

□

Lemma G.20. *Let $\xi > 0$ and $1 \geq b \geq a + 1/2 \geq 1/4$. Let $\{h_i\}$ be sequence of $\mathcal{L}^2$ functions such that $\Sigma[L_{K^{1/2}}(h_i), L_{K^{1/2}}(h_i)] \leq M < \infty$. Then we have*

$$\mathbb{E}\Bigg[\max_{s<t\leq e}\sum_{m\geq 1}\frac{\mathfrak{s}_m^{2a}}{(\mathfrak{s}_m+\lambda_{t-s})^{2b}}\frac{\delta_{t-s}^{2(b-a-1)}}{(t-s)\log^{1+\xi}(t-s)}$$
$$\left|\sum_{j=s+1}^{t}\left(\langle X_j, L_{K^{1/2}}(h_j)\rangle_{\mathcal{L}^2}\langle X_j, L_{K^{1/2}}(\phi_m)\rangle_{\mathcal{L}^2} - \langle h_j, T\phi_m\rangle_{\mathcal{L}^2}\right)\right|^2\Bigg] = O(1).$$

Proof. Let
$$Y_{j,m} = \langle X_j, L_{K^{1/2}}(h_j)\rangle_{\mathcal{L}^2}\langle X_j, L_{K^{1/2}}(\phi_m)\rangle_{\mathcal{L}^2} - \langle h_j, T\phi_m\rangle_{\mathcal{L}^2}$$

Similar to the last two proofs, we are going to establish the result on a generic interval $\{1, \dots, T\}$, the case in the lemma follows from translation and stationarity.

Observe that
$$\mathbb{E}\left[\langle X_j, L_{K^{1/2}}(h_j)\rangle_{\mathcal{L}^2}\langle X_j, L_{K^{1/2}}(\phi_m)\rangle_{\mathcal{L}^2}\right] = \langle h_j, T\phi_m\rangle_{\mathcal{L}^2},$$
and
$$\Sigma[L_{K^{1/2}}(\phi_m), L_{K^{1/2}}(\phi_m)] = \langle T\phi_m, \phi_m\rangle_{\mathcal{L}^2} = \mathfrak{s}_m.$$
Using this, from Lemma G.8, we may establish
$$\mathbb{E}\left[\left|\sum_{j=1}^{t} Y_{j,m}\right|^2\right] \leq O(1)\sum_{j=1}^{t}\mathfrak{s}_m.$$

Now, similar to (110), we apply Lemma J.3 to have: for any non-decreasing sequence $\{\gamma_t\}_{t=1}^T$
$$\mathbb{E}\left[\max_{1<t\leq T}\left|\frac{1}{\gamma_t}\sum_{j=1}^{t} Y_{j,m}\right|^2\right] = C\sum_{t=1}^{T}\frac{1}{\gamma_t^2}\mathfrak{s}_m, \tag{120}$$
for some constant $C > 0$.

The rest of proof follows exact same steps as the proof of Lemma G.18 and therefore omitted. □

Lemma G.21. *Let $a, b, q > 0$. Let p, r be some constant. Suppose $D : \mathcal{L}^2 \to \mathcal{L}^2$ be some linear operator. Suppose $f, h \in \mathcal{L}^2$. Then we have*

$$\|T^p D f\|_{\mathcal{L}^2} \leq \left(1 + \left(\frac{\lambda_b}{\lambda_a}\right)^q\right)\|T^p D(T+\lambda_b\mathbf{I})^{-q}T^{-r}\|_{\text{op}}\|T^r(T+\lambda_a\mathbf{I})^q f\|_{\mathcal{L}^2} \tag{121}$$
$$|\langle h, f\rangle_{\mathcal{L}^2}| \leq \left(1 + \left(\frac{\lambda_b}{\lambda_a}\right)^q\right)\|T^{-p}(T+\lambda_b\mathbf{I})^{-q}h\|_{\mathcal{L}^2}\|T^p(T+\lambda_a\mathbf{I})^q h\|_{\mathcal{L}^2}. \tag{122}$$

Proof. We are going to establish (121) and the proof for (122) follows similarly. The proof is divided in three steps. We establish some necessary result in Step 1 and Step 2 and complete the proof in Step 3 by using them.

Step 1: For $d \geq c$, we are going to establish the following in this step.

$$\|T^p(T+\lambda_d\mathbf{I})^q f\|_{\mathcal{L}^2} \leq \|T^p(T+\lambda_c\mathbf{I})^q f\|_{\mathcal{L}^2} \leq \left(\frac{\lambda_c}{\lambda_d}\right)^q \|T^p(T+\lambda_d\mathbf{I})^q f\|_{\mathcal{L}^2}, \tag{123}$$

Let $f = \sum_{j\geq 1} a_j\phi_j$. Observe that $j^{-2r} \asymp \mathfrak{s}_j > 0$ and

$$d \geq c \iff (\mathfrak{s}_j+\lambda_c) \geq (\mathfrak{s}_j+\lambda_d) \iff \frac{1}{(\mathfrak{s}_j+\lambda_d)} \geq \frac{1}{(\mathfrak{s}_j+\lambda_c)} \iff \frac{\lambda_c}{\lambda_d} \geq \frac{\mathfrak{s}_j+\lambda_c}{\mathfrak{s}_j+\lambda_d}.$$

It lead us to

$$\|T^p(T+\lambda_d\mathbf{I})^q f\|_{\mathcal{L}^2}^2 = \sum_{j\geq 1} \mathfrak{s}_j^{2p}(\mathfrak{s}_j+\lambda_d)^{2q}a_j^2 \leq \sum_{j\geq 1} \mathfrak{s}_j^{2p}(\mathfrak{s}_j+\lambda_c)^{2q}a_j^2 = \|T^p(T+\lambda_c\mathbf{I})^q f\|_{\mathcal{L}^2}^2$$

and

$$\begin{aligned}\|T^p(T+\lambda_c\mathbf{I})^q f\|_{\mathcal{L}^2}^2 &= \sum_{j\geq 1} \mathfrak{s}_j^{2p}(\mathfrak{s}_j+\lambda_c)^{2q}a_j^2 = \sum_{j\geq 1}\left(\frac{\mathfrak{s}_j+\lambda_c}{\mathfrak{s}_j+\lambda_d}\right)^{2b} \mathfrak{s}_j^{2p}(\mathfrak{s}_j+\lambda_d)^{2q}a_j^2 \\ &\leq \left(\frac{\lambda_c}{\lambda_d}\right)^{2b} \sum_{j\geq 1} \frac{\mathfrak{s}_j^{2p}}{(\mathfrak{s}_j+\lambda_d)^{2q}}a_j^2 = \left(\frac{\lambda_c}{\lambda_d}\right)^{2b} \|T^p(T+\lambda_d\mathbf{I})^{-q} f\|_{\mathcal{L}^2}^2.\end{aligned}$$

Step 2: For $d \geq c$, we are going to establish the following in this step.

$$\|T^pD(T+\lambda_c\mathbf{I})^{-q}T^r\|_{\text{op}} \leq \|T^pD(T+\lambda_d\mathbf{I})^{-q}T^r\|_{\text{op}}. \tag{124}$$

Observe that

$$b \geq a \iff (\mathfrak{s}_j+\lambda_c) \geq (\mathfrak{s}_j+\lambda_b) \iff \frac{1}{(\mathfrak{s}_j+\lambda_b)} \geq \frac{1}{(\mathfrak{s}_j+\lambda_c)}.$$

It lead us to

$$\begin{aligned}&\|T^pD(T+\lambda_c\mathbf{I})^{-q}T^r\|_{\text{op}} \\ &= \sup_{\substack{h\in\mathcal{L}^2 \\ \|h\|_{\mathcal{L}^2}=1}} \left|\left\langle h, T^pD(T+\lambda_c\mathbf{I})^{-q}T^r h\right\rangle_{\mathcal{L}^2}\right| \\ &= \sup_{\substack{h\in\mathcal{L}^2 \\ \|h\|_{\mathcal{L}^2}=1}} \left|\sum_{j\geq 1}\sum_{m\geq 1} h_jh_m \left\langle \phi_j, T^pD(T+\lambda_c\mathbf{I})^{-q}T^r\phi_m\right\rangle_{\mathcal{L}^2}\right| \\ &= \sup_{\substack{h\in\mathcal{L}^2 \\ \|h\|_{\mathcal{L}^2}=1}} \left|\sum_{j\geq 1}\sum_{m\geq 1} h_jh_m \left\langle T^p\phi_j, D(T+\lambda_c\mathbf{I})^{-q}T^r\phi_m\right\rangle_{\mathcal{L}^2}\right|\end{aligned}$$

$$
\begin{aligned}
&= \sup_{\substack{h\in\mathcal{L}^2\\ \|h\|_{\mathcal{L}^2}=1}} \left| \sum_{j\geq 1}\sum_{m\geq 1} h_j h_m \frac{\mathfrak{s}_j^{2p}\mathfrak{s}_m^{2r}}{(\mathfrak{s}_m+\lambda_c)^{2q}} \left\langle \phi_j, D\phi_m \right\rangle_{\mathcal{L}^2} \right| \\
&\leq \sup_{\substack{h\in\mathcal{L}^2\\ \|h\|_{\mathcal{L}^2}=1}} \left| \sum_{j\geq 1}\sum_{m\geq 1} h_j h_m \frac{\mathfrak{s}_j^{2p}\mathfrak{s}_m^{2r}}{(\mathfrak{s}_m+\lambda_d)^{2q}} \left\langle \phi_j, D\phi_m \right\rangle_{\mathcal{L}^2} \right| \\
&= \sup_{\substack{h\in\mathcal{L}^2\\ \|h\|_{\mathcal{L}^2}=1}} \left| \sum_{j\geq 1}\sum_{m\geq 1} h_j h_m \left\langle T^p\phi_j, D(T+\lambda_d\mathbf{I})^{-q}T^r\phi_m \right\rangle_{\mathcal{L}^2} \right| \\
&= \sup_{\substack{h\in\mathcal{L}^2\\ \|h\|_{\mathcal{L}^2}=1}} \left| \sum_{j\geq 1}\sum_{m\geq 1} h_j h_m \left\langle \phi_j, T^pD(T+\lambda_d\mathbf{I})^{-q}T^r\phi_m \right\rangle_{\mathcal{L}^2} \right| \\
&= \sup_{\substack{h\in\mathcal{L}^2\\ \|h\|_{\mathcal{L}^2}=1}} \left| \left\langle h, T^pD(T+\lambda_d\mathbf{I})^{-q}T^r h \right\rangle_{\mathcal{L}^2} \right| \\
&= \|T^pD(T+\lambda_d\mathbf{I})^{-q}T^r\|_{\mathrm{op}}.
\end{aligned}
$$

Step 3: Using (123) and (124), we may write

$$
\begin{aligned}
&\|T^pDf\|_{\mathcal{L}^2} \\
=&\mathbf{I}\{b\geq a\}\|T^pD(T+\lambda_b\mathbf{I})^{-q}T^{-r}T^r(T+\lambda_b\mathbf{I})^q f\|_{\mathcal{L}^2} + \mathbf{I}\{b\leq a\}\|T^pD(T+\lambda_b\mathbf{I})^{-q}T^{-r}T^r(T+\lambda_b\mathbf{I})^q f\|_{\mathcal{L}^2} \\
\leq&\mathbf{I}\{b\geq a\}\|T^pD(T+\lambda_b\mathbf{I})^{-q}T^{-r}\|_{\mathrm{op}}\|T^r(T+\lambda_a\mathbf{I})^q f\|_{\mathcal{L}^2} \\
&\qquad + \mathbf{I}\{b\leq a\}\left(\frac{\lambda_b}{\lambda_a}\right)^q \|T^pD(T+\lambda_b\mathbf{I})^{-q}T^{-r}\|_{\mathrm{op}}\|T^r(T+\lambda_a\mathbf{I})^q f\|_{\mathcal{L}^2} \\
\leq&\left(1+\left(\frac{\lambda_b}{\lambda_a}\right)^q\right)\|T^pD(T+\lambda_b\mathbf{I})^{-q}T^{-r}\|_{\mathrm{op}}\|T^r(T+\lambda_a\mathbf{I})^q f\|_{\mathcal{L}^2}.
\end{aligned}
$$

□

H Lower bound

Proof of Lemma 1. We prove a more general result and the required result follows as a special case.

For $Z_j = (Y_j, X_j)$, let P_0^n be the joint distribution of $\{Z_j\}_{j=1}^n$ following

$$\begin{aligned} y_j &= \langle X_j, \beta\rangle_{\mathcal{L}^2} + \varepsilon_j, && \text{for } 1 \le j \le \Delta, \\ y_j &= \varepsilon_j, && \text{for } \Delta < j \le n, \end{aligned}$$

where $\{X_j\}_{j=1}^n$ is independent standard Brownian motion and $\{\varepsilon_j\}_{j=1}^n \overset{iid}{\sim} N(0,1)$. Let P_1^n be the joint distribution of $\{Z_j'\}_{j=1}^n$ with $Z_j' = (Y_j', X_j')$ which follows

$$\begin{aligned} y_j' &= \langle X_j', \beta\rangle_{\mathcal{L}^2} + \varepsilon_j', && \text{for } 1 \le j \le \Delta + \delta, \\ y_j' &= \varepsilon_j', && \text{for } \Delta + \delta < j \le n. \end{aligned}$$

where $\{X_j'\}_{j=1}^n$ is independent standard Brownian motion and $\{\varepsilon_j'\}_{j=1}^n \overset{iid}{\sim} N(0,1)$. We assume that the two datasets are independent. Observe that

$$KL\left(P_0^n; P_1^n\right) = \sum_{j=\Delta+1}^{\Delta+\delta} KL\left(P_0^{j,n}; P_1^{j,n}\right),$$

where $P_0^{j,n}(y,x)$ and $P_1^{j,n}(y,x)$ are distributions of (y_j, X_j) and (y_j', X_j') respectively. For $\Delta < j \le \Delta + \delta$, one may write

$$\begin{aligned} KL\left(P_0^{j,n}; P_1^{j,n}\right) &= \iint \log\left\{\frac{p_0^{j,n}(y|x)}{p_1^{j,n}(y|x)}\right\} p_0^{j,n}(y|x)p(x)\, dy\, dx \\ &= \iint \frac{1}{2}\left(\langle x, \beta\rangle_{\mathcal{L}^2}^2 - 2y\langle x, \beta\rangle_{\mathcal{L}^2}\right) p_0^{j,n}(y|x)p(x)\, dy\, dx \\ &= \frac{1}{2}\int \langle x, \beta\rangle_{\mathcal{L}^2}^2 p(x) dx = \frac{\kappa^2}{2}, \end{aligned}$$

where in the first line we used the conditional density $p_0^{j,n}(y|x)$, $p_1^{j,n}(y|x)$, and $p(x)$ as the density of X_j; in the second and the last line we use the fact that $y_j|X_j \sim N(0,1)$ under $p_0^{j,n}(y|x)$. This lead us to $KL\left(P_0^n; P_1^n\right) = \delta\kappa^2/2$ and we already have $\eta(P_1^n) - \eta(P_0^n) = \delta$. Following from LeCam's lemma (see e.g. Yu, 1997 and Theorem 2.2 of Tsybakov, 2004), we may write

$$\inf_{\widehat{\eta}} \sup_{P \in \mathfrak{P}} \mathbb{E}\left[|\widehat{\eta} - \eta(P)|\right] \ge \frac{\delta}{4} e^{-\delta\kappa^2/2}.$$

The result now follows by putting $\delta = \frac{4}{\kappa^2}$ with the realization that, for large n, $\frac{4}{\kappa^2} \ll \Delta < n/2$. $\square$

I α-mixing

The strong mixing or α-mixing coefficient between two $\sigma-$fields $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$\alpha(\mathcal{A},\mathcal{B}) = \sup_{A\in\mathcal{A},B\in\mathcal{B}} |P(A\cap B) - P(A)P(B)|.$$

Lemma I.1. *Let X and Y be random variables. Then for any positive numbers p,q,r satisfying $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1$, we have*

$$|Cov(X,Y)| \leq 4\|X\|_p\|Y\|_q \{\alpha(\sigma(X),\sigma(Y))\}^{1/r}.$$

I.1 Strong law of large numbers

Theorem I.1. *Let $\{Z_t\}$ be centered alpha mixing time series such that $\alpha(k) = O\left(\frac{1}{(kL_k^2)^{\rho/(\rho-2)}}\right)$ for some $\rho > 2$, where L_k is non-decreasing sequence satisfying*

$$\sum_{k=1}^{\infty}\frac{1}{kL_k} < \infty \qquad \textit{and} \qquad L_k - L_{k-1} = O\left(L_k/k\right).$$

Suppose for some $1 \leq p \leq \rho < \infty$ one has

$$\sum_{t=1}^{\infty}\frac{\mathbb{E}^{2/\rho}\left[|Z_t|^p\right]}{t^p} < \infty.$$

Then $\sum_{t=1}^{n} Z_t/n$ converges a.s to 0.

Proof. Using L_n in the Definition 1.4 of McLeish (1975), $\{\alpha(n)\}$ is sequence of size$-\rho/(\rho-2)$. Following their Remark 2.6b, the results directly follows from their Lemma 2.9 with $g_t(x) = x^p$, $d_t = 1$ and $X_t = Z_t/t$. □

I.2 Central limit theorem

Below is the central limit theorem for α-mixing random variable. For a proof, one may see Doukhan (2012).

Theorem I.2. *Let $\{Z_t\}$ be a centred α-mixing stationary time series. Suppose that it holds for some $\delta > 0$,*

$$\sum_{k=1}^{\infty}\alpha(k)^{\delta/(2+\delta)} < \infty \quad \textit{and} \quad \mathbb{E}(|Z_1|^{2+\delta}) < \infty.$$

Denote $S_n = \sum_{t=1}^{n} Z_t$ and $\sigma_n^2 = \mathbb{E}\left[|S_n|^2\right]$. Then

$$\frac{S_{\lfloor nt\rfloor}}{\sigma_n} \to W(t),$$

where the convergence is in Skorohod topology and $W(t)$ is the standard Brownian motion on $[0,1]$.

J Inequalities

Lemma J.1. *Let $f : [0, \infty] \to [0, \infty]$ be monotonically decreasing continuous function such that $\int_1^\infty f(x)dx < \infty$. Then*

$$\int_1^\infty f(x)dx \le \sum_{k\ge 1} f(k) \le f(1) + \int_1^\infty f(x)dx \le f(0) + \int_1^\infty f(x)dx.$$

Lemma J.2. *Let $r > 1$ be a constant. For any positive sequence $\mathfrak{s}_j \asymp j^{-2r}$ and $\varphi \ge 1/2$ we have*

$$\sum_{j\ge 1} \frac{\mathfrak{s}_j^\varphi}{(\alpha + \mathfrak{s}_j)^\varphi} \le c_1 \alpha^{-1/2r} \tag{125}$$

given any $\alpha > 0$. Here $c_1 > 0$ is some constant.

Proof. Given $\mathfrak{s}_j \asymp j^{-2r}$, we have some positive constants c_2 and c_3 such that $c_2 j^{-2r} \le \mathfrak{s}_j \le c_3 j^{-2r}$,

$$\implies \sum_{j\ge 1} \frac{\mathfrak{s}_j^\varphi}{(\alpha + \mathfrak{s}_j)^\varphi} \le \sum_{j\ge 1} \frac{(c_3 j^{-2r})^\varphi}{(\alpha + c_2 j^{-2r})^\varphi} = c_3 \sum_{j\ge 1} \frac{1}{(\alpha j^{2r} + c_2)^\varphi}.$$

Now, we shall upper bound the quantity on the right of above equation using Lemma J.1. Observe that the function defined by $x \mapsto \frac{1}{(\alpha x^{2r} + c_2)^\varphi}$ satisfy the conditions of Lemma J.1 and therefore

$$\sum_{j\ge 1} \frac{1}{(\alpha j^{2r} + c_2)^\varphi} \le c_4 + \int_1^\infty \frac{1}{(\alpha x^{2r} + c_2)^\varphi} dx$$

Observe that

$$\begin{aligned} \int_1^{(c_2/\alpha)^{1/2r}} \frac{1}{(\alpha x^{2r} + c_2)^\varphi} dx &\le \int_1^{(c_2/\alpha)^{1/2r}} \frac{1}{(c_2)^\varphi} dx = \frac{c_5}{\alpha^{1/2r}} - c_4 \le \frac{c_5}{\alpha^{1/2r}} \\ \int_{(c_2/\alpha)^{1/2r}}^\infty \frac{1}{(\alpha x^{2r} + c_2)^\varphi} dx &\le \int_{(c_2/\alpha)^{1/2r}}^\infty \frac{1}{(\alpha x^{2r})^\varphi} dx = \frac{c_6}{\alpha^{1/2r}}. \end{aligned} \tag{126}$$

Using (126) we may write,

$$\int_1^\infty \frac{1}{(\alpha x^{2r} + c_2)^\varphi} dx \le \frac{c_5 + c_6}{\alpha^{1/2r}}$$

which lead us to the required result.

□

Corollary J.1. *Let $\{\alpha_n\}_{n\ge 1}$ be positive sequence converging to 0. Under assumptions of Lemma J.2, we have*

$$\sum_{j\ge 1} \frac{\mathfrak{s}_j^{1+2t}}{(\alpha_n + \mathfrak{s}_j)^{1+2t}} = O\left(\alpha_n^{-1/2r}\right) \tag{127}$$

Lemma J.3. *Let $\{Z_i\}$ be a sequence of random variable. Let $\xi > 0$. Suppose*

$$\mathbb{E}\left[\left|\sum_{m=1}^t Z_j\right|^2\right] = O(t).$$

Then for any positive non-increasing sequence $\{\gamma_t\}$, we have

$$\mathbb{E}\left[\max_{1<t\leq n}\left|\gamma_t\sum_{m=1}^{t}Z_j\right|^2\right] = O(1)\sum_{t=2}^{n}\gamma_t^2. \tag{128}$$

Consequently we have

$$\mathbb{E}\left[\max_{1<t\leq n}\left|\frac{1}{\sqrt{t\log^{1+\xi}t}}\sum_{m=1}^{t}Z_j\right|^2\right] = O(1). \tag{129}$$

Proof. Observe that (128) follows directly from Theorem B.3 of Kirch (2006).

Note that $\left\{\frac{1}{\sqrt{t\log^{1+\xi}t}}\right\}_{t=2}^{n}$ is a non-increasing sequence and from (128)

$$\mathbb{E}\left[\max_{1<t\leq n}\left|\frac{1}{\sqrt{t\log^{1+\xi}t}}\sum_{m=1}^{t}Z_j\right|^2\right] = O(1)\sum_{t=2}^{n}\frac{1}{t\log^{1+\xi}t},$$

and the (129) follows from the fact that $\sum_{t=2}^{\infty}\frac{1}{t\log^{1+\xi}t} < \infty$. □

Lemma J.4. *Let $\nu > 0$. Let $\{Z_i\}$ be a sequence of random variable. Suppose we have*

$$\mathbb{E}\left[|S_i^j|^2\right] \leq c'(j-i),$$

where $S_i^j = \sum_{k=i}^{j} Z_k$ and $c' > 0$ is some absolute constant. Then for any given $\varepsilon > 0$

$$P\left(\forall r > 1/\nu, \qquad |S_r| \leq \frac{C_1}{\sqrt{\varepsilon}}\sqrt{r}(\log r\nu + 1)\right) > 1-\varepsilon,$$

where $C_1 = \frac{\pi}{\log 2}\sqrt{\frac{c}{6}}$.

In other words

$$\max_{r>1/\nu}\frac{1}{\sqrt{r}\,(\log(r\nu)+1)}\left|\sum_{j=1}^{r}Z_j\right| = O_p(1).$$

Proof. Observe that using Lemma J.3 with $\gamma_t = 1$, we can get

$$\mathbb{E}\left[\max_{i<t\leq j}|S_i^t|^2\right] \leq c(j-i)$$

for some absolute constant $c > 0$.

We are going to use the peeling argument for the proof. With

$$\mathbb{E}\left[\max_{m\leq k\leq 2m}\left|\frac{S_k}{\sqrt{k}}\right|^2\right] \leq \frac{1}{m}\mathbb{E}\left[\max_{m\leq k\leq 2m}|S_k|^2\right] \leq c.$$

Let's define $A_j = \left[2^{j-1}/\nu, 2^j/\nu\right]$. Using Markov's inequality we may write,

$$\begin{aligned}
&P\left(\max_{m\le k\le 2m}\left|\frac{S_k}{\sqrt{k}}\right| \ge x\right) \le \frac{1}{x^2}\mathbb{E}\left[\max_{m\le k\le 2m}\left|\frac{S_k}{\sqrt{k}}\right|^2\right] \le cx^{-2}\\
\implies &P\left(\bigcup_{k\in A_j}\left\{\left|\frac{S_k}{\sqrt{k}}\right| \ge \alpha j\right\}\right) \le \frac{c}{\alpha^2 j^2}\\
\implies &P\left(\bigcup_{k\in A_j}\left\{\left|\frac{S_k}{\sqrt{k}}\right| \ge \alpha(\log_2 k\nu + 1)\right\}\right) \le \frac{c}{\alpha^2 j^2}.
\end{aligned}$$

The last equation follows from

$$2^{j-1}/\nu \le k \le 2^j/\nu \implies j \le \log_2 k\nu + 1 \le j+1.$$

And

$$\begin{aligned}
&P\left(\bigcup_{r\ge 1/\nu}\left\{\left|\frac{S_r}{\sqrt{r}}\right| \ge \alpha(\log_2 r\nu + 1)\right\}\right)\\
=&P\left(\bigcup_{j\ge 1}\bigcup_{k\in A_j}\left\{\left|\frac{S_k}{\sqrt{k}}\right| \ge \alpha(\log_2 k\nu + 1)\right\}\right)\\
\le&\sum_{j=1}^{\infty} P\left(\bigcup_{k\in A_j}\left\{\left|\frac{S_k}{\sqrt{k}}\right| \ge \alpha(\log_2 k\nu + 1)\right\}\right)\\
\le&\frac{c}{\alpha^2}\sum_{j=1}^{\infty}\frac{1}{j^2} = \frac{c\pi^2}{6\alpha^2}.
\end{aligned}$$

With $\alpha^2 = \frac{c\pi^2}{6\varepsilon \log^2 2}$ and $\log 2 < 1$, we can have

$$P\left(\max_{r\ge 1/\nu}|S_r| \ge \frac{C_1}{\sqrt{\varepsilon}}\sqrt{r}(\log r\nu + 1)\right) \le \varepsilon.$$

$\square$

References

Cai, T. T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association*, 107(499):1201–1216.

Doukhan, P. (2012). *Mixing: properties and examples*, volume 85. Springer Science & Business Media.

Gorton, G. B., Metrick, A., and Xie, L. (2020). An econometric chronology of the financial crisis of 2007-2008. *Available at SSRN 2615029*.

Guidolin, M. and Tam, Y. M. (2013). A yield spread perspective on the great financial crisis: Break-point test evidence. *International Review of Financial Analysis*, 26:18–39.

He, Z., Nagel, S., and Song, Z. (2022). Treasury inconvenience yields during the covid-19 crisis. *Journal of Financial Economics*, 143(1):57–79.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382.

Kirch, C. (2006). *Resampling methods for the change analysis of dependent data*. PhD thesis, Universität zu Köln.

McLeish, D. L. (1975). A Maximal Inequality and Dependent Strong Laws. *The Annals of Probability*, 3(5):829 – 839.

Rinaldo, A., Wang, D., Wen, Q., Willett, R., and Yu, Y. (2021). Localizing changes in high-dimensional regression models. In *International Conference on Artificial Intelligence and Statistics*, pages 2089–2097. PMLR.

Tsybakov, A. B. (2004). *Introduction to nonparametric estimation, 2009*, volume 9.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence*. Springer New York, New York, NY.

Xu, H., Padilla, O. M., Wang, D., and Li, M. (2022). *changepoints: A Collection of Change-Point Detection Methods*. R package version 1.1.0.

Xu, H., Wang, D., Zhao, Z., and Yu, Y. (2024). Change-point inference in high-dimensional regression models under temporal dependence. *The Annals of Statistics*, 52(3):999–1026.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 52(1):45–57.

Yu, B. (1997). Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam: research papers in probability and statistics*, pages 423–435. Springer.

Yuan, M. and Cai, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *The Annals of Statistics*, 38(6):3412 – 3444.