Contents

Supplementary Materials to "Change Point Detection for High-dimensional Linear Models: A General Tail-adaptive Approach"

Bin Liu[1], Zhengling Qi[2], Xinsheng Zhang[3] and Yufeng Liu[4]

[1,3]*Department of Statistics and Data Science, School of Management at Fudan University*

[2]*Department of Decision Sciences, George Washington University*

[4]*Department of Statistics and Operations Research, Department of Genetics, Department of Biostatistics, Carolina Center for Genome Sciences, Linberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, U.S.A*

This document provides detailed proofs of the main theoretical results as well as full numerical studies. In Section S1, we demonstrate how to combine our proposed tail-adaptive methods with the wild binary segmentation technique to detect multiple change points. In Section S2, we provide detailed numerical experiments. In Section S3, we apply our proposed new method to the S&P100 data to detect multiple change points. In Section S4, we introduce some additional notations. In Section S5, we provide the detailed model assumptions for the theory developed in the main paper. In Section S6, some useful lemmas are provided. In Section S7, we give the detailed proofs of theoretical results in the main paper. In Section S8, we provide the proof of lemmas used in Section S7. In Sections S9 and S10, we prove the useful lemmas in Section S6 as well as some additional lemmas.

S1 Extensions to multiple change point detection

In practical applications, it may exist multiple change points in describing the relationship between $\boldsymbol{X}$ and Y. Therefore, it is essential to perform estimation of multiple change points if $\mathbf{H}_0$ is rejected by our powerful tail-adaptive test. In this section, we extend our single change point detection method by the idea of WBS proposed in Fryzlewicz (2014) to estimate the locations of all possible multiple change points.

Consider a single change point detection task in any interval $[s, e]$, where $0 \leq q_0 \leq s < e \leq 1 - q_0$. Following Section 2.4, we can compute the corresponding adaptive test statistics as $\widehat{P}_{\text{ad}}(s, e)$ using the subset of our data, i.e., $\{\boldsymbol{X}_{\lfloor ns \rfloor}, \boldsymbol{X}_{\lfloor ns \rfloor+1}, \cdots, \boldsymbol{X}_{\lfloor ne \rfloor}\}$ and $\{Y_{\lfloor ns \rfloor}, Y_{\lfloor ns \rfloor+1}, \cdots, Y_{\lfloor ne \rfloor}\}$. Following the idea of WBS, we first independently generate a series of random intervals by the uniform distribution. Denote the number of these random intervals as V. For each random interval $[s_\nu, e_\nu]$ among $\nu = 1, 2, \cdots, V$, we compute $\widehat{P}_{\text{ad}}(s_\nu, e_\nu)$ as long as $0 \leq q_0 \leq s_\nu < e_\nu \leq 1 - q_0$ and $e_\nu - s_\nu \geq v_0$, where v_0 is the minimum length for implementing Section 2.4. The threshold v_0

Algorithm S1.1 : A WBS-typed tail-adaptive test for multiple change point detection

Input: Given the data $(\mathcal{X}, \mathcal{Y}) = \{(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)\}$, set the values for $\widetilde{\boldsymbol{\tau}}$, the significance level γ, s_0, q_0, the bootstrap replication number B, the candidate subset $\mathcal{A} \subset [0, 1]$, and a set of random intervals $\{(s_\nu, e_\nu\}_{\nu=1}^{V}$ with thresholds v_0 and v_1. Initialize an empty set $\mathcal{C}$.

Step 1: For each $\nu = 1, \cdots, V$, compute $\widehat{P}_{\mathrm{ad}}(s_\nu, e_\nu)$ following Section 2.4.

Step 2: Perform the following function with $S = q_0$ and $E = 1 - q_0$.

Function(S, E): S and E are the starting and ending points for the change point detection.

(a) RETURN if $E - S \leq v_1$.

(b) Define $\mathcal{M} = \{1 \leq \nu \leq V \mid [s_\nu, e_\nu] \subset [E, S]\}$.

(c) Compute the test statistics as $\overline{P}_{\mathrm{ad}} = \min_{\nu \in \mathcal{M}, v_0 \leq e_\nu - s_\nu} \widehat{P}_{\mathrm{ad}}(s_\nu, e_\nu)$ and the corresponding optimal solution ν^*.

(d) If $\overline{P}_{\mathrm{ad}} \geq \gamma / V$, RETURN. Otherwise, add the corresponding change point estimator $\widehat{t}_{\nu^*}$ to $\mathcal{C}$, and perform Function(S, ν^*) and Function(ν^*, E).

Output: The set of multiple change points $\mathcal{C}$.

is used to reduce the variability of our algorithm for multiple change point detection. Based on the test statistics computed from the random intervals, we consider the final test statistics as $\overline{P}_{\mathrm{ad}} = \min_{1 \leq \nu \leq V, v_0 \leq e_\nu - s_\nu} \widehat{P}_{\mathrm{ad}}(s_\nu, e_\nu)$, based on which we make decisions if there exists at least one change point among these intervals. We stop the algorithm if $\overline{P}_{\text{ad}} \geq \bar{c}$, otherwise we report the change point estimation in $[s_{\nu^*}, e_{\nu^*}]$, where $\nu^* \in \underset{1 \leq \nu \leq V, v_0 \leq e_\nu - s_\nu}{\arg\min} \widehat{P}_{\text{ad}}(s_\nu, e_\nu)$, and continue our algorithm. Given the first change point estimator denoted by $\widehat{t}_{\nu^*}$, we split our data into two folds, i.e., before and after the estimated change point. Then we apply the previous procedure on each fold of the data using the same set of the random intervals as long as it satisfies the constraints. We repeat this step until the algorithm stops returning the change point estimation. For each step, we choose $\bar{c} = \gamma/V$, where γ is the significance level used in each single change point detection algorithm. While we do not have the theoretical guarantee of using $\bar{c}$ in the proposed algorithm for controlling the size, the selection of this constant is based on the idea of Bonferroni correction, which is conservative. The numerical experiments in the appendix demonstrate the superiority of our proposed method in detecting multiple change points. Nevertheless, it is interesting to study the asymptotic property of $\overline{P}_{\text{ad}}$, which we leave for the future work. The full algorithm of the multiple change point detection can be found in Algorithm S1.1.

S2 Numerical experiments

In this section, we investigate the numerical performance of our proposed method and compare with the existing techniques in terms of change point detection and identification. In Sections S2.1 - S2.3, we consider single change point testing and estimation. In Section S2.5, we investigate multiple change point detection.

S2.1 Single change point testing

We consider the performance of singe change point testing for the following model:

$$Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(1)} \mathbf{1}\{i \leq k_1\} + \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(2)} \mathbf{1}\{i > k_1\} + \epsilon_i, \quad i = 1, \ldots, n, \tag{S2.1}$$

where $k_1 = \lfloor nt_1 \rfloor$. To show the broad applicability of our method, we generate data from various model settings. Specifically, for the design matrix $\mathbf{X}$, we generate $\boldsymbol{X}_i$ (i.i.d) from $N(\mathbf{0}, \boldsymbol{\Sigma})$ under two different models:

- **Model 1:** We generate $\boldsymbol{X}_i$ with banded $\boldsymbol{\Sigma}$. Specifically, we set $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'$, where $\boldsymbol{\Sigma}' = (\sigma'_{ij}) \in \mathbb{R}^{p \times p}$ with $\sigma'_{ij} = 0.8^{|i-j|}$ for $1 \leq i, j \leq p$.

- **Model 2:** We generate $\boldsymbol{X}_i$ with blocked $\boldsymbol{\Sigma}$. Specifically, we set $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^\star$, where $\boldsymbol{\Sigma}^\star = (\sigma^\star_{ij}) \in \mathbb{R}^{p \times p}$ with $\sigma^\star_{ii} \overset{\text{i.i.d.}}{\sim} \mathrm{U}(1, 2)$, $\sigma^\star_{ij} = 0.6$ for $5(k-1) + 1 \leq i \neq j \leq 5k$ $(k = 1, \ldots, \lfloor p/5 \rfloor)$, and $\sigma^\star_{ij} = 0$ otherwise.

Moreover, to show the tail-adaptivity of our new testing method, we generate the error term ϵ_i from various types of distributions including both lighted-tailed and heavy-tailed distributions. In particular, we generate ϵ_i from the Gaussian distribution $N(0,1)$ and the Student's t_v distribution with a degree of freedom $v \in \{1,2,3,4\}$. Note that t_v with $v = 2$ and $v = 1$ correspond to the error without second moments and first moments, respectively. For the regression coefficient $\boldsymbol{\beta}^{(1)}$, for each replication, we generate $\boldsymbol{\beta}^{(1)} = (1,1,1,1,1,0,\cdots,0)^\top \in \mathbb{R}^p$. In other words, only the first five elements in $\boldsymbol{\beta}^{(1)}$ are non-zero with magnitudes of ones, which are called the active set. Under $\mathbf{H}_0$, we set $\boldsymbol{\beta}^{(2)} = \boldsymbol{\beta}^{(1)} := \boldsymbol{\beta}^{(0)}$. Under $\mathbf{H}_1$, we set $\boldsymbol{\beta}^{(2)} = \boldsymbol{\beta}^{(1)} + \boldsymbol{\delta}$, where $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)^\top \in \mathbb{R}^p$ is the signal jump with

$$\delta_s = \begin{cases} c\sqrt{\log(p)/n}, & \text{for } s \in \{1,2,3,4,5\}, \\ 0, & \text{for } s \in \{6,\dots,p\}. \end{cases}$$

In other words, we add a signal jump with a magnitude of $c\sqrt{\log(p)/n}$ on the first five elements of $\boldsymbol{\beta}^{(1)}$. To avoid the trivial power performance (too low or high powers), we set $c = 1$ and $c = 1.5$ for the normal and the Student's t distributions, respectively.

Throughout the simulations, we fix the sample size at $n = 200$ and the dimension at $p = 400$. The number of bootstrap replications is $B = 200$. Without additional specifications, all numerical results are based on 1000 replications. In addition, we consider the $L_1 - L_2$ composite loss by setting $\widetilde{\tau} = 0.5$ and $K = 1$ in (2.6), which is of special interest in high dimensional data analysis. Note that our proposed method involves the optimization problem in (2.10). We use the coordinate descent algorithm for obtaining the corresponding LASSO estimators. As for the tuning parameters λ_α, for $\alpha = 1$, we use the cross-validation technique to select the "best" λ_1; for $\alpha = 0$, we adopt the method recommended in Belloni and Chernozhukov (2011) (see Section 2.3 therein) to set λ_0; for $\alpha \in (0, 1)$, we use an idea of weighted combination and let $\lambda_\alpha = (1 - \alpha)\lambda_0 + \alpha\lambda_1$.

S2.2 Empirical sizes

We consider the size performance with a significance level $\gamma = 5\%$. Tables S2.1 provides the size results for the individual tests T_α with $\alpha \in \mathcal{A} = \{0, 0.1, 0.5, 0.9, 1\}$ and the tail-adaptive test T_{ad} under **Models 1** and **2** with various error distributions. Note that the construction of our testing statistic involves a selection of $s_0 \in \{1, \ldots, p\}$. To show the effect of different s_0, we consider various $s_0 \in \{1, 3, 5, 7\}$. Note that $s_0 = 1$ corresponds to the ℓ_∞-norm based individual test and $s_0 = 5$ corresponds to the test that aggregates the active set of variables in $\boldsymbol{\beta}^{(0)}$. As shown in Table S2.1, for a given s_0, our individual test T_α and tail-adaptive test

Table S2.1: Empirical sizes of the individual and tail-adaptive tests for **Models 1-2** with banded and blocked covariance matrices for $s_0 \in \{1, 3, 5, 7\}$. The results are based on 1000 replications with $B = 200$ for each replication.

Empirical sizes for Model 1 with $p = 400$							
Dist	$\mathbf{s_0}$	$\boldsymbol{\alpha = 0}$	$\boldsymbol{\alpha = 0.1}$	$\boldsymbol{\alpha = 0.5}$	$\boldsymbol{\alpha = 0.9}$	$\boldsymbol{\alpha = 1}$	**Adaptive**
$N(0,1)$	$s_0 = 1$	0.062	0.041	0.036	0.033	0.038	0.040
	$s_0 = 3$	0.052	0.056	0.041	0.032	0.034	0.045
	$s_0 = 5$	0.051	0.053	0.040	0.032	0.027	0.040
	$s_0 = 7$	0.050	0.048	0.041	0.035	0.027	0.046
t_4	$s_0 = 1$	0.049	0.056	0.052	0.048	0.040	0.062
	$s_0 = 3$	0.058	0.057	0.052	0.050	0.040	0.051
	$s_0 = 5$	0.049	0.035	0.041	0.038	0.035	0.041
	$s_0 = 7$	0.064	0.043	0.045	0.050	0.048	0.048
t_3	$s_0 = 1$	0.058	0.052	0.046	0.038	0.048	0.064
	$s_0 = 3$	0.053	0.053	0.045	0.050	0.052	0.058
	$s_0 = 5$	0.062	0.060	0.055	0.053	0.063	0.074
	$s_0 = 7$	0.051	0.051	0.053	0.053	0.055	0.066
Empirical sizes Model 2 with $p = 400$							
Dist	$\mathbf{s_0}$	$\boldsymbol{\alpha = 0}$	$\boldsymbol{\alpha = 0.1}$	$\boldsymbol{\alpha = 0.5}$	$\boldsymbol{\alpha = 0.9}$	$\boldsymbol{\alpha = 1}$	**Adaptive**
$N(0,1)$	$s_0 = 1$	0.052	0.049	0.043	0.030	0.028	0.042
	$s_0 = 3$	0.068	0.062	0.042	0.025	0.025	0.048
	$s_0 = 5$	0.068	0.058	0.028	0.018	0.017	0.043
	$s_0 = 7$	0.043	0.043	0.022	0.015	0.011	0.031
t_4	$s_0 = 1$	0.059	0.050	0.047	0.041	0.040	0.053
	$s_0 = 3$	0.048	0.046	0.044	0.041	0.036	0.044
	$s_0 = 5$	0.073	0.059	0.034	0.036	0.042	0.060
	$s_0 = 7$	0.064	0.051	0.030	0.038	0.038	0.055
t_3	$s_0 = 1$	0.059	0.063	0.044	0.036	0.041	0.058
	$s_0 = 3$	0.070	0.055	0.042	0.041	0.043	0.057
	$s_0 = 5$	0.052	0.055	0.047	0.042	0.042	0.049
	$s_0 = 7$	0.056	0.048	0.044	0.042	0.035	0.054

Table S2.2: Empirical sizes of the individual and tail-adaptive tests for **Models 1-2** for the error term being Student's t_2 and t_1 distributed. The results are based on 1000 replications with $B = 200$ for each replication.

p	Dist	s_0	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 1$	Adaptive
Empirical sizes for Mode 1 with heavy tails								
400	t_2	$s_0 = 1$	0.041	0.057	0.079	0.087	0.085	0.077
	t_2	$s_0 = 3$	0.057	0.060	0.092	0.088	0.090	0.090
	t_2	$s_0 = 5$	0.068	0.074	0.110	0.107	0.129	0.128
	t_2	$s_0 = 7$	0.062	0.065	0.116	0.115	0.113	0.126
400	t_1	$s_0 = 1$	0.057	0.207	0.222	0.217	0.217	0.192
	t_1	$s_0 = 3$	0.043	0.228	0.244	0.240	0.232	0.208
	t_1	$s_0 = 5$	0.058	0.299	0.315	0.310	0.300	0.266
	t_1	$s_0 = 7$	0.057	0.276	0.306	0.308	0.300	0.267
Empirical sizes for Mode 2 with heavy tails								
p	**Dist**	s_0	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 1$	**Adaptive**
400	t_2	$s_0 = 1$	0.066	0.082	0.095	0.096	0.094	0.106
	t_2	$s_0 = 3$	0.072	0.086	0.115	0.113	0.118	0.127
	t_2	$s_0 = 5$	0.058	0.074	0.138	0.151	0.152	0.147
	t_2	$s_0 = 7$	0.056	0.085	0.113	0.127	0.118	0.136
400	t_1	$s_0 = 1$	0.070	0.220	0.249	0.251	0.247	0.223
	t_1	$s_0 = 3$	0.046	0.402	0.440	0.432	0.430	0.406
	t_1	$s_0 = 5$	0.055	0.455	0.500	0.487	0.489	0.462
	t_1	$s_0 = 7$	0.057	0.479	0.526	0.511	0.501	0.488

T_{ad} can have a size that is very close to the nominal level. This strongly suggests that our bootstrap-based procedure in Algorithms 1 and 2 can approximate the theoretical distributions very well. Interestingly, it can be seen that under a specific error distribution, the individual test T_α may have different size performance in the sense that the corresponding size can be slightly above or below the nominal level. In contrast, after the combination, the size of the tail-adaptive test T_{ad} is near the nominal level as compared to its individual test. This indicates that in practice, the tail-adaptive test is more reliable in terms of size control.

Table S2.2 provides additional size performance under Student's t_2 and t_1 distributions. Note that these two distributions are known as seriously heavy-tailed. It is also well known that controlling the size for these two distributions is a challenging task, especially for high-dimensional change point analysis. As can be seen from Table S2.2, in these cases, the individual test T_α except $\alpha = 0$ suffers from serious size distortion. In particular, as α increases from 0.1 to 1, it is more difficult to control the size. Moreover, when the error is Cauchy distributed, the size is completely out of control for $\alpha \in \{0.1, \ldots, 1\}$. As a result, the corresponding tail-adaptive method becomes oversized. As an exception, we can see that the individual test T_α with $\alpha = 0$ enjoys satisfactory size performance for both t_2 and t_1

distributions. A reasonable explanation is that for $\alpha = 0$, our individual test reduces to the median regression based method which does not require any moment constraints on the error terms. Hence, our proposed individual test with $\alpha = 0$ contributes to the literature for handling the extremely heavy-tailed case. In practice, if the practitioners strongly believe that the data are seriously heavy-tailed, we can just set $\mathcal{A} = \{0\}$.

S2.3 Empirical powers

We next consider the power performance, where various error distributions, data dimensions as well as change point locations are investigated. The results are summarized in Tables S2.3 and S2.4. Note that according to our model setups, there are five coordinates in $\boldsymbol{\beta}^{(1)}$ having a change point. It can be seen that for light-tailed error distributions such as $N(0, 1)$, the individual tests with $\alpha = 0.5, 0.9, 1$ have the best power performance and those with $\alpha = 0$ have the worst performance. This indicates that for a light-tailed error distribution, using median regression can lose power efficiency, and using the moment information with a larger weight α can increase the signal to noise ratio. Interestingly, in this empirical study, the individual test with $\alpha = 0.5$ generally has slightly higher powers than that with $\alpha = 1$, even though the latter one is expected to have the best power performance (see Figure 1). As for the tail-adaptive test, in the light-tailed case, it has very close powers to the best individual tests.

We next turn to the heavy-tailed case, where the individual tests have power performance that is very different from the light-tailed case. Specifically, for t_3 distributions, the individual test with $\alpha = 0$ and $\alpha = 0.1$ have higher powers than the remaining ones. This indicates that for data with heavy tails, it is beneficial to use more rank information instead of using only moments. More specifically, we see that T_α with $\alpha = 0.1$ has the highest powers and that with $\alpha = 1$ has the lowest powers. This result is consistent with the theoretical SNR in Figure 1. In this case, using a non-trivial weight ($\alpha = 0.1$) can significantly enhance the power efficiency via increasing the SNR. As for the tail-adaptive method, it still has very close powers to the best individual test, i.e. $\alpha = 0.1$ when the data are heavy-tailed. In addition to $N(0, 1)$ and t_3 distributions, we can observe that for t_4 distributions, even though the individual tests may present various power performances, the tail-adaptive method consistently has powers close to that of the corresponding best individual test. The above results suggest that our proposed tail-adaptive method can sufficiently account for the unknown tail-structures, and enjoy satisfactory power performance under various data generating mechanisms. Lastly, we remark that when the change point location gets closer to the boundary of data observations, e.g. from $t_1 = 0.5$ to $t_1 = 0.3$, it becomes more difficult to detect a change point, which is also consistent with our theoretical result.

Next, we consider the effect of different s_0 on the power performance. We find that for any given s_0, the performance of the individual and the tail-adaptive tests are similar to our above findings. This suggests that the tail-adaptivity of our testing method is robust to the choice of s_0. Moreover, for each case with a specific error distribution and data dimension, both the individual and tail-adaptive tests with $s_0 = 3, 5, 7$ have higher powers than those with $s_0 = 1$. More specifically, tests with $s_0 = 5$ generally have the best performance and those with $s_0 = 3$ and $s_0 = 7$ have close powers to $s_0 = 5$. This indicates that for high dimensional sparse linear models, instead of using the ℓ_∞-norm, it is more efficient to detect a change point via aggregating the CUSUM statistics using the first $s_0 > 1$ order statistics.

Table S2.3: Empirical powers of the individual and tail-adaptive tests for **Model 1** with banded covariance matrix under various distributions with $s_0 \in \{1, 3, 5, 7\}$ and $t_1 \in \{0.3, 0.5\}$. The dimension is $p = 400$. The results are based on 1000 replications with $B = 200$ for each replication.

Empirical powers for N(0,1)								
Dist	$\mathbf{t_1}$	$\mathbf{s_0}$	$\boldsymbol{\alpha = 0}$	$\boldsymbol{\alpha = 0.1}$	$\boldsymbol{\alpha = 0.5}$	$\boldsymbol{\alpha = 0.9}$	$\boldsymbol{\alpha = 1}$	**Adaptive**
		$s_0 = 1$	0.482	0.612	0.768	0.732	0.722	0.733
$N(0,1)$	0.5	$s_0 = 3$	0.529	0.655	0.787	0.759	0.739	0.759
		$s_0 = 5$	0.546	0.641	0.783	0.759	0.749	0.760
		$s_0 = 7$	0.525	0.634	0.802	0.778	0.773	0.765
		$s_0 = 1$	0.295	0.398	0.546	0.506	0.489	0.495
$N(0,1)$	0.3	$s_0 = 3$	0.318	0.415	0.573	0.534	0.516	0.518
		$s_0 = 5$	0.286	0.418	0.568	0.543	0.514	0.505
		$s_0 = 7$	0.315	0.418	0.560	0.522	0.505	0.522
Empirical powers for Student's t_4								
Dist	$\mathbf{t_1}$	$\mathbf{s_0}$	$\boldsymbol{\alpha = 0}$	$\boldsymbol{\alpha = 0.1}$	$\boldsymbol{\alpha = 0.5}$	$\boldsymbol{\alpha = 0.9}$	$\boldsymbol{\alpha = 1}$	**Adaptive**
		$s_0 = 1$	0.792	0.880	0.835	0.769	0.756	0.873
t_4	0.5	$s_0 = 3$	0.836	0.903	0.852	0.795	0.782	0.904
		$s_0 = 5$	0.847	0.914	0.884	0.813	0.787	0.915
		$s_0 = 7$	0.831	0.895	0.861	0.813	0.807	0.896
		$s_0 = 1$	0.595	0.724	0.687	0.588	0.555	0.722
t_4	0.3	$s_0 = 3$	0.644	0.762	0.737	0.627	0.591	0.765
		$s_0 = 5$	0.646	0.773	0.745	0.618	0.606	0.765
		$s_0 = 7$	0.603	0.765	0.712	0.582	0.564	0.743
Empirical powers for Student's t_3								
Dist	$\mathbf{t_1}$	$\mathbf{s_0}$	$\boldsymbol{\alpha = 0}$	$\boldsymbol{\alpha = 0.1}$	$\boldsymbol{\alpha = 0.5}$	$\boldsymbol{\alpha = 0.9}$	$\boldsymbol{\alpha = 1}$	**Adaptive**
		$s_0 = 1$	0.773	0.819	0.663	0.580	0.572	0.802
t_3	0.5	$s_0 = 3$	0.777	0.826	0.693	0.612	0.583	0.827
		$s_0 = 5$	0.791	0.840	0.685	0.604	0.594	0.822
		$s_0 = 7$	0.789	0.847	0.713	0.623	0.602	0.829
		$s_0 = 1$	0.554	0.629	0.475	0.380	0.362	0.599
t_3	0.3	$s_0 = 3$	0.599	0.692	0.527	0.422	0.403	0.656
		$s_0 = 5$	0.587	0.697	0.525	0.404	0.390	0.640
		$s_0 = 7$	0.549	0.650	0.487	0.362	0.348	0.613

Table S2.4: Empirical powers of the individual and data-adaptive tests for **Model 2** with blocked covariance matrix under various distributions with $s_0 \in \{1, 3, 5, 7\}$ and $t_1 \in \{0.3, 0.5\}$. The dimension p is 400. The results are based on 1000 replications with $B = 200$ for each replication.

			Empirical powers for N(0,1)					
Dist	$\mathbf{t_1}$	$\mathbf{s_0}$	$\boldsymbol{\alpha = 0}$	$\boldsymbol{\alpha = 0.1}$	$\boldsymbol{\alpha = 0.5}$	$\boldsymbol{\alpha = 0.9}$	$\boldsymbol{\alpha = 1}$	**Adaptive**
		$s_0 = 1$	0.328	0.441	0.651	0.626	0.626	0.594
$N(0,1)$	0.5	$s_0 = 3$	0.428	0.547	0.733	0.714	0.704	0.695
		$s_0 = 5$	0.462	0.585	0.761	0.712	0.702	0.714
		$s_0 = 7$	0.476	0.591	0.760	0.718	0.703	0.712
		$s_0 = 1$	0.175	0.232	0.361	0.326	0.321	0.301
$N(0,1)$	0.3	$s_0 = 3$	0.245	0.334	0.486	0.453	0.437	0.458
		$s_0 = 5$	0.244	0.356	0.483	0.428	0.412	0.428
		$s_0 = 7$	0.246	0.338	0.470	0.409	0.389	0.419
			Empirical powers for Student's t_4					
Dist	$\mathbf{t_1}$	$\mathbf{s_0}$	$\boldsymbol{\alpha = 0}$	$\boldsymbol{\alpha = 0.1}$	$\boldsymbol{\alpha = 0.5}$	$\boldsymbol{\alpha = 0.9}$	$\boldsymbol{\alpha = 1}$	**Adaptive**
		$s_0 = 1$	0.618	0.722	0.749	0.654	0.659	0.742
t_4	0.5	$s_0 = 3$	0.791	0.862	0.849	0.784	0.763	0.873
		$s_0 = 5$	0.780	0.866	0.855	0.778	0.769	0.874
		$s_0 = 7$	0.802	0.879	0.868	0.806	0.782	0.889
		$s_0 = 1$	0.398	0.518	0.547	0.427	0.404	0.511
t_4	0.3	$s_0 = 3$	0.511	0.661	0.645	0.535	0.514	0.665
		$s_0 = 5$	0.522	0.663	0.631	0.505	0.483	0.651
		$s_0 = 7$	0.531	0.661	0.637	0.509	0.482	0.655
			Empirical powers for Student's t_3					
Dist	$\mathbf{t_1}$	$\mathbf{s_0}$	$\boldsymbol{\alpha = 0}$	$\boldsymbol{\alpha = 0.1}$	$\boldsymbol{\alpha = 0.5}$	$\boldsymbol{\alpha = 0.9}$	$\boldsymbol{\alpha = 1}$	**Adaptive**
		$s_0 = 1$	0.571	0.640	0.571	0.460	0.455	0.621
t_3	0.5	$s_0 = 3$	0.726	0.790	0.683	0.588	0.579	0.782
		$s_0 = 5$	0.761	0.794	0.674	0.583	0.576	0.808
		$s_0 = 7$	0.753	0.803	0.717	0.619	0.600	0.807
		$s_0 = 1$	0.359	0.455	0.360	0.293	0.261	0.422
t_3	0.3	$s_0 = 3$	0.470	0.574	0.485	0.370	0.349	0.560
		$s_0 = 5$	0.490	0.581	0.451	0.339	0.328	0.567
		$s_0 = 7$	0.498	0.601	0.440	0.330	0.308	0.581

S2.4 The choices of $\mathcal{A}$ and s_0

Table S2.5: Empirical powers of the tail-adaptive tests for **Model 1** with banded covariance matrix under various choices of s_0 and $\mathcal{A}$. The dimension is $p = 400$. The results are based on 1000 replications with $B = 200$ for each replication.

	$N(0,1)$							t_3						
s_0	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	$\mathcal{A}_6$	$\mathcal{A}_7$
1	0.456	0.746	0.725	0.546	0.682	0.698	0.700	0.760	0.634	0.524	0.804	0.486	0.786	0.748
2	0.479	0.800	0.748	0.584	0.718	0.706	0.708	0.792	0.666	0.544	0.828	0.520	0.786	0.778
4	0.521	0.773	0.752	0.608	0.738	0.704	0.662	0.766	0.728	0.584	0.840	0.534	0.812	0.788
$[\log(p)]$	0.498	0.780	0.746	0.586	0.682	0.742	0.682	0.822	0.676	0.560	0.820	0.538	0.814	0.758
8	0.488	0.798	0.724	0.586	0.650	0.710	0.678	0.788	0.688	0.592	0.860	0.536	0.826	0.762
16	0.442	0.738	0.716	0.554	0.680	0.714	0.626	0.722	0.616	0.480	0.806	0.516	0.750	0.676
32	0.426	0.692	0.644	0.470	0.590	0.586	0.526	0.646	0.616	0.416	0.720	0.448	0.736	0.608
64	0.346	0.586	0.512	0.398	0.476	0.524	0.508	0.584	0.516	0.350	0.630	0.384	0.574	0.508
128	0.264	0.556	0.436	0.320	0.426	0.396	0.354	0.468	0.476	0.322	0.550	0.336	0.540	0.396

Note that our approach involves the selection of the candidate set $\mathcal{A}$ and the parameter s_0, both of which can be regarded as tuning parameters. Intuitively, $\mathcal{A}$ determines the weight between the quantile loss and the least squared losses, whereas s_0 indicates how much information on change points among regression coefficient components should be integrated into the CUSUM statistic. Therefore, we conducted numerical simulations to investigate how different choices of $\mathcal{A}$ and s_0 affect the efficacy of change point detection. We selected seven different subsets for $\mathcal{A}$ including $\mathcal{A}_1 = \{0\}$, $\mathcal{A}_2 = \{0.5\}$, $\mathcal{A}_3 = \{1\}$, $\mathcal{A}_4 = \{0, 0.1\}$, $\mathcal{A}_5 = \{0.9, 1\}$, $\mathcal{A}_6 = \{0, 0.1, 0.5, 0.9, 1\}$ and $\mathcal{A}_7 = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$. Additionally, we selected different values for s_0, including $s_0 = 2^0, 2^1, \ldots, 2^{\lfloor \log_2(p)+1 \rfloor}$. Table S2.5 displays the performance of the adaptive change point detection method under $N(0,1)$ and t_3 distributions for these selections of $\mathcal{A}$ and s_0. The model settings are the same as in Section S4.3. We observed that for any given $\mathcal{A}$, when s_0 increases from small to large, the efficacy of the adaptive detection method initially increases and then decreases, indicating that as s_0 increases, the statistic extracts more change point information from the regression components, enhancing the efficacy of change point detection. However, once s_0 becomes larger, additional noise accumulates, leading to a decrease in the detection efficacy. Considering the sparsity assumptions for regression coefficients and the requirements of Gaussian approximation theory, which requires $s_0^3 \log(pn) = O(n^{\xi_1})$ for some $0 < \xi_1 < 1/7$ and $s_0^4 \log(pn) = O(n^{\xi_2})$ for some $0 < \xi_2 < \frac{1}{6}$, we recommend the use of $s_0 = \lfloor \log(p) \rfloor$ in practice.

Regarding the selection of $\mathcal{A}$, we note that for data with light-tailed distributions, sets with larger values such as $\mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_5, \mathcal{A}_6, \mathcal{A}_7$ exhibit higher efficacy. Conversely, for data with heavy-tailed distributions, sets with smaller values such as $\mathcal{A}_1, \mathcal{A}_4, \mathcal{A}_6, \mathcal{A}_7$ perform satisfactorily. Therefore, if the tail structure of the data is unknown in practical applications, we might consider a candidate set that includes both larger and smaller values. Interestingly, we find that adding too many weights, such as in $\mathcal{A}_7$ does not yield much additional benefit. Considering the balance between detection efficacy and computational efficiency, we recommend $\mathcal{A} = \{0, 0.1, 0.5, 0.9, 1\}$ for practical use.

S2.5 Multiple change point detection

In this section, we consider the performance of multiple change point detection and compare our method with the existing techniques. In this numerical study, we set $n = 1000$ and $p = 100$ with three change points ($m = 3$) at $k_1 = 300$, $k_2 = 500$, and $k_3 = 700$, respectively. The above three change points divide the data into four segments with piecewise constant regression coefficients $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(2)}$, $\boldsymbol{\beta}^{(3)}$ and $\boldsymbol{\beta}^{(4)}$ as follows:

$$\begin{cases} Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(1)} + \epsilon_i, & \text{for } i = 1, \ldots, k_1, \\ Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(2)} + \epsilon_i, & \text{for } i = k_1 + 1, \ldots, k_2, \\ Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(3)} + \epsilon_i, & \text{for } i = k_2 + 1, \ldots, k_3, \\ Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(4)} + \epsilon_i, & \text{for } i = k_3 + 1, \ldots, n. \end{cases}$$

Table S2.6: Multiple change point estimation results with $(n, p) = (1000, 100)$. The results are based on 100 replications with $B = 100$ for each replication.

	N(0,1) ($c = 3$)		N(0,1) ($c = 6$)		t3 ($c = 4$)		t3 ($c = 6$)	
Methods	Haus	(Sd)	Haus	(Sd)	Haus	(Sd)	Haus	(Sd)
$\alpha = 0$ (BS)	0.186	(0.1310)	0.028	(0.0374)	0.144	(0.1300)	0.040	(0.0603)
$\alpha = 0.1$ (BS)	0.138	(0.1286)	0.033	(0.0442)	0.111	(0.1229)	0.032	(0.0580)
$\alpha = 0.5$ (BS)	0.072	(0.1018)	0.028	(0.0432)	0.145	(0.1353)	0.051	(0.0740)
$\alpha = 0.9$ (BS)	0.088	(0.1092)	0.030	(0.0443)	0.180	(0.1371)	0.063	(0.0877)
$\alpha = 1$ (BS)	0.095	(0.1146)	0.024	(0.0394)	0.187	(0.1405)	0.065	(0.0963)
Adaptive (BS)	0.092	(0.1085)	0.032	(0.0428)	0.113	(0.1260)	0.039	(0.0638)
$\alpha = 0$ (WBS)	0.087	(0.0926)	0.018	(0.0381)	0.046	(0.0690)	0.018	(0.0381)
$\alpha = 0.1$ (WBS)	0.060	(0.0846)	0.012	(0.0277)	0.049	(0.0710)	0.012	(0.0277)
$\alpha = 0.5$ (WBS)	0.036	(0.0599)	0.012	(0.0279)	0.077	(0.0908)	0.012	(0.0279)
$\alpha = 0.9$ (WBS)	0.041	(0.0662)	0.011	(0.0216)	0.095	(0.0999)	0.011	(0.0216)
$\alpha = 1$ (WBS)	0.043	(0.0659)	0.012	(0.0216)	0.101	(0.1014)	0.012	(0.0216)
Adaptive (WBS)	0.031	(0.0478)	0.014	(0.0292)	0.033	(0.0511)	0.014	(0.0292)
VPWBS	0.135	(0.1004)	0.038	(0.0474)	0.138	(0.0782)	0.085	(0.0636)
DPDU	0.082	(0.1097)	0.009	(0.0087)	0.118	(0.0830)	0.045	(0.0598)

The covariates $\boldsymbol{X}_i$ are generated from $N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ being banded which is introduced in Model 1. For each replication, we first randomly select five covariates (denoted by $\mathcal{S}_1$) from $\{1, \ldots, 10\}$. For generating $\boldsymbol{\beta}^{(1)}$, we set $\beta_s^{(1)} = 1$ if $s \in \mathcal{S}_1$ and $\beta_s^{(1)} = 0$ if $s \notin \mathcal{S}_1$. For $\boldsymbol{\beta}^{(2)}$, we set $\beta_s^{(2)} = \beta_s^{(1)} + c\sqrt{\log(p)/n}$ if $s \in \mathcal{S}_1$ and $\beta_s^{(2)} = 0$ if $s \notin \mathcal{S}_1$. Then, we set $\boldsymbol{\beta}^{(3)} = \boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(4)} = \boldsymbol{\beta}^{(2)}$. We compare our proposed method with the Variance-Projected Wild Binary Segmentation (VPWBS) method in Wang et al. (2021) and the dynamic programming with dynamic update method in Xu et al. (2022). As compared in Wang et al. (2021), VPWBS has better performance than the binary segmentation based technique in Leonardi and Bühlmann (2016) and the sparse graphical LASSO based method in Zhang et al. (2015). Hence, we do not compare with Leonardi and Bühlmann (2016) and Zhang et al. (2015). For VPWBS, we use the R codes published by the authors on GitHub (https://github.com/darenwang/VPBS) and employ a cross-validation method to select tuning parameters for estimating change points. For DPDU, we utilize the DPDU.regression.R function from the R package named "changepoints" to estimate multiple change points. As for our methods, we combine the individual and tail-adaptive procedures with the Binary Segmentation and Wild Binary Segmentation techniques. For WBS, we use Algorithm S1.1 with parameters as $\gamma = 0.05$, $s_0 = 5$,

$q_0 = 0.1$, $B = 100$, $V = 150$, and $v_0 = 0.1$. In this numerical study, we set the replication number as 100.

To evaluate the performance in identifying the change point, we use the scaled Hausdorff distance to evaluate the performance in change point estimation, which is defined as:

$$d(\mathcal{S}_1, \mathcal{S}_2) = \frac{\max(\max_{s_1 \in \mathcal{S}_1} \min_{s_2 \in \mathcal{S}_2} |s_1 - s_2|, \max_{s_2 \in \mathcal{S}_2} \min_{s_1 \in \mathcal{S}_1} |s_1 - s_2|)}{1000},$$

where $\mathcal{S}_1 = \{300, 500, 7000, 1000\}$ are the true change points and $\mathcal{S}_2$ are the estimated change points. Note that scaled Hausdorff distance is a number between 0 and 1, and a smaller one indicates better change point estimation. Table S2.6 provides the results for $N(0, 1)$ and t_3 distributions with various signal strength $c \in \{3, 4, 6\}$. For light-tailed error distributions, the individual methods with a larger α generally have better performance than those with a smaller one for identifying the change point number and locations. This can be seen by smaller Hausdorff. On the contrary, in the heavy-tailed case, the individual methods with a smaller α are more preferred. As for the tail-adaptive method, it has comparable performance to the best individual one under both light and heavy-tailed errors.

Additionally, we note that for both individual and tail-adaptive testing methods, those based on Wild Binary Segmentation (WBS) generally outperform those based on Binary Segmentation (BS). Therefore, we recommend combining our proposed method with the WBS algorithm for multiple change point estimation in practical applications. For VPWBS and DPDU, these methods show satisfactory performance under light-tailed distributions such as the normal distribution. Particularly, DPDU, which employs a dynamic programming algorithm for multiple change point estimation, achieves the lowest estimation errors when data follow a normal distribution with strong signals. Our adaptive method performs comparably to these two methods under light-tailed distributions. However, their detection capabilities decrease when the data follow heavy-tailed distributions, such as the Student's t_3 distribution. This indicates that methods based on the least squared loss are not robust for heavy-tailed data.

Lastly, we report the computational complexity of the algorithm. For our individual and tail-adaptive testing methods, when combined with the WBS algorithm, the complexities are $O(M\text{Lasso}(n,p))$ and $O(M|\mathcal{A}|\text{Lasso}(n,p))$, respectively, where M represents the number of small intervals in WBS, and $\text{Lasso}(n,p)$ denotes the computational cost for calculating lasso with sample size n and data dimension p. For the DPDU algorithm, it uses a backward iterative dynamic programming approach, and its complexity is $O(n^2p^2 + n^2\text{Lasso(p)})$. Figure S2.1 shows the computational time of our method and the DPDU algorithm under various $n \in \{200, 300, 400, 500\}$

and $p \in \{200, 300, 400, 500\}$, where the model setup is the same as in Section S2.3. We set the number of intervals in WBS to $\log^2(n)$. We can observe that the computational costs of both our method and the DPDU method increase with n and p. Our individual testing method has comparable computational time to that of the DPDU. The computational cost for the tail-adaptive testing method is the highest. This is not surprising, as we aim to construct a testing method that is adaptive to the tail structure of the error terms. To that end, we need to calculate lasso estimates with different weights α to obtain the best individual testing method.

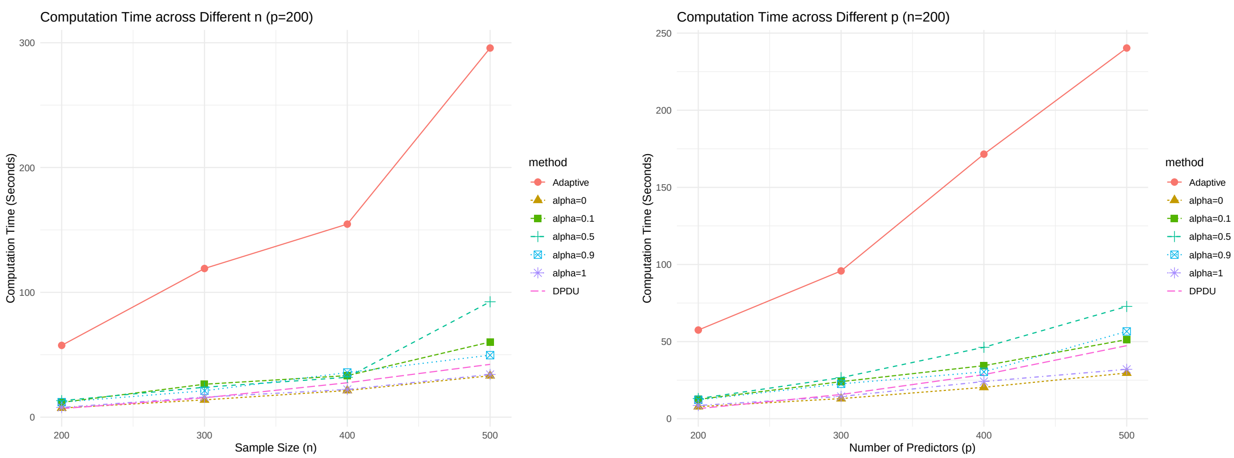

Figure S2.1: Computational time for our proposed method and the DPDU algorithm with $n \in \{200, 300, 400, 500\}$ and $p \in \{200, 300, 400, 500\}$.

S3 An application to the S&P 100 dataset

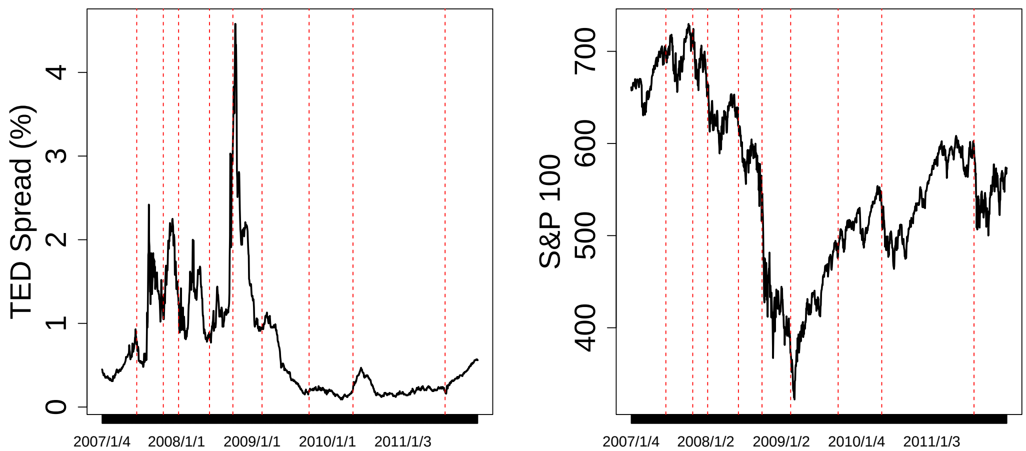

Figure S3.1: Plots of the Ted spread (left) and the S&P 100 index (right) with the estimated change points (vertical lines) marked by # in Table S3.1.

In this section, we apply our proposed method to the S&P 100 dataset to find multiple change points. We obtain the S&P 100 index as well as the associated stocks from Yahoo! Finance (https://finance.yahoo.com/) including the largest and most established 100 companies in the S&P 100. For this dataset, we collect the daily prices of 76 stocks that have remained in the S&P 100 index consistently from January 3, 2007 to December 30, 2011. This covers the recent financial crisis beginning in 2008 and some other important events, resulting in a sample size $n = 1259$.

In financial marketing, it is of great interest to predict the S&P 100 index since it reveals the direction of the entire financial system. To this end, we use the daily prices of the 76 stocks to predict the S&P 100 index. Specifically, let $Y_t \in \mathbb{R}^1$ be the S&P 100 index for the t-th day and $\boldsymbol{X}_t \in \mathbb{R}^{76\times 2}$ be the stock prices with lag-1 and lag-3 for the t-th day. Our goal is to predict Y_t using $\boldsymbol{X}_t$ under the high dimensional linear regression models and detect multiple change points for the linear relationships between the S&P 100 index and the 76 stocks' prices. Note that we have calculated differences of the data to remove the temporal trend. It is well known that the financial data are typically heavy-tailed and we have no prior-knowledge about the tail structure of the data. Hence, for this real data analysis, it seems very suitable to use our proposed tail-adaptive method. We combine our proposed tail-adaptive test with the WBS method (Fryzlewicz (2014)) to detect multiple change points, which is demonstrated in Algorithm S1.1. To implement this algorithm, we set $\mathcal{A} = \{0, 0.1, 0.5, 0.9, 1\}$, $s_0 = 5$, $B = 100$, and $V = 500$ (number of random intervals). Moreover, we consider the $L_1 - L_2$ weighted loss by setting $\widetilde{\boldsymbol{\tau}} = 0.5$ in (2.6). The data are scaled to have mean zeros and variance ones before the change point detection. There are 14 change points detected which are reported in Table S3.1.

To further justify the meaningful findings of our proposed new methods, we refer to the T-bills and ED (TED) spread, which is short for the difference between the 3-month of London Inter-Bank Offer Rate (LIBOR), and the 3-month short-term U.S. government debt (T-bills). It is well-known that TED spread is an indicator of perceived risk in the general economy and an increased TED spread during the financial crisis reflects an increase in credit risk. Figure S3.1 shows the plot of TED where the red dotted lines correspond to the estimated change points. We can see that during the financial crisis from 2007 to 2009, the TED spread has experienced very dramatic fluctuations and the estimated change points can capture some big changes in the TED spread. In addition, the S&P 100 index obtains its highest level during the financial crisis in October 2007 and then has a huge drop. Our method identifies October 29, 2007 as a change point. Moreover, the third detected change point is January 10th, 2008. The National Bureau of Economic Research (NBER) identifies December of 2007 as the beginning of the great recession which is captured by our method. In addition, it is well known that affected by the 2008 financial crisis, Europe experienced a debt crisis from 2009 to 2012, with the Greek government debt crisis in October 2009 serving as the starting point. Our method identifies October 5, 2009 as a change point after which S&P 100 index began to experience a significant decline. Moreover, it is known that countries such as Italy and Spain were facing severe debt issues in July 2011, raising fears about the stability of the Eurozone and the potential impact on global financial markets. As a result, there exists another huge drop for the S&P 100 index in July 26, 2011, which can be successfully detected by our method.

Table S3.1: Multiple change point detection for the S&P 100 dataset.

Change points	Date	Events
117	2007/06/21	TED Spread#
207	2007/10/29	TED Spread#
257	2008/01/10	Global Financial Crisis (TED Spread)#
360	2008/06/09	TED Spread#
439	2008/09/30	TED Spread#
535	2009/02/18	Nadir of the crisis#
632	2009/07/08	
694	2009/10/05	Greek debt crisis#
840	2010/05/05	Global stock markets fell due to fears of contagion of the European sovereign debt crisis#
890	2010/07/16	
992	2010/12/09	
1074	2011/04/07	
1149	2011/07/26	Spread of the European debt crisis to Spain and Italy#
1199	2011/10/05	

S4 Some notations

Before the proofs, we give some notations. Under $\mathbf{H}_0$, we set $\boldsymbol{\beta}^{(0)} := \boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$ and $s^{(0)} := s^{(1)} = s^{(2)}$. Under $\mathbf{H}_1$, For the regression vectors $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$, define $\mathcal{S}^{(1)} = \{1 \leq j \leq p : \beta_j^{(1)} \neq 0\}$ and $\mathcal{S}^{(2)} = \{1 \leq j \leq p : \beta_j^{(2)} \neq 0\}$ as the active sets of variables. Denote $s^{(1)} = |\mathcal{S}^{(1)}|$ and $s^{(2)} = |\mathcal{S}^{(2)}|$ as the cardinalities of $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$, respectively. We set $\mathcal{S} = \mathcal{S}^{(1)} \cup \mathcal{S}^{(2)}$ and $s = |\mathcal{S}|$. For a vector $\boldsymbol{v} \in \mathbb{R}^p$, we denote $J(\boldsymbol{v}) = \{1 \leq j \leq p : v_j \neq 0\}$ as the set of non-zero elements of $\boldsymbol{v}$ and set $\mathcal{M}(\boldsymbol{v}) := |J(\boldsymbol{v})|$ as the number of non-zero elements of $\boldsymbol{v}$. For a set J and $\boldsymbol{v} \in \mathbb{R}^p$, denote $\boldsymbol{v}_J$ as the vector in $\mathbb{R}^p$ that has the same coordinates as $\boldsymbol{v}$ on J and zero coordinates on the complement J^c of J. For any vector $\boldsymbol{x} \in \mathbb{R}^p$ and a matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, define $\|\boldsymbol{x}\|_{\mathbf{A}}^2 = \boldsymbol{x}^\top \mathbf{A} \boldsymbol{x}$. Denote $\mathcal{X} = \{\mathbf{X}, \boldsymbol{Y}\}$. We use $C_1, C_2, \ldots$ to denote constants that may vary from line to line. We use w.p.a.1 for the abbreviation of with probability approaching to one. For $\beta > 0$, we define the function $\psi_\beta : [0, \infty) \to [0, \infty)$ as $\psi_\beta(x) := \exp(x^\beta) - 1$. Then, for any random variable X, we define

$$\|X\|_{\psi_\beta} := \inf\big\{C > 0 : \mathbb{E}\psi_\beta(|X|/C)|) \leq 1\big\}.$$

For any $0 \leq s < t \leq 1$, we denote

$$\widehat{\boldsymbol{\Sigma}}(s:t) = \frac{1}{\lfloor nt \rfloor - \lfloor ns \rfloor + 1} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \boldsymbol{X}_i^\top. \tag{S4.2}$$

S5 Basic assumptions

We introduce some basic assumptions for deriving our main theorems. Before that, we introduce some notations. Let $e_i(\widetilde{\boldsymbol{\tau}}) := K^{-1}\sum_{k=1}^{K}\big(\mathbf{1}\{\epsilon_i \leq b_k^{(0)}\} - \tau_k\big) := K^{-1}\sum_{k=1}^{K} e_i(\tau_k)$. We set $\mathcal{V}_{s_0} := \{\boldsymbol{v} \in \mathbb{S}^p : \|\boldsymbol{v}\|_0 \leq s_0\}$, where $\mathbb{S}^p := \{\boldsymbol{v} \in \mathbb{R}^p : \|\boldsymbol{v}\| = 1\}$. For each $\alpha \in [0,1]$, we introduce $\underset{\sim}{\boldsymbol{\beta}}^* = ((\boldsymbol{\beta}^*)^\top, (\boldsymbol{b}^*)^\top)^\top \in \mathbb{R}^{p+K}$ with $\boldsymbol{\beta}^* \in \mathbb{R}^p$, $\boldsymbol{b}^* = (b_1^*, \ldots, b_K^*)^\top \in \mathbb{R}^K$, where

$$\underset{\sim}{\boldsymbol{\beta}}^* := \underset{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{b}\in\mathbb{R}^K}{\arg\min}\ \mathbb{E}\Big[(1-\alpha)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\rho_{\tau_k}(Y_i - b_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}) + \frac{\alpha}{2n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta})^2\Big]. \quad \text{(S5.3)}$$

Note that by definition, we can regard $\underset{\sim}{\boldsymbol{\beta}}^*$ as the true parameters under the population level with pooled samples. We can prove that under $\mathbf{H}_0$, $\underset{\sim}{\boldsymbol{\beta}}^* = ((\boldsymbol{\beta}^{(0)})^\top, (\boldsymbol{b}^{(0)})^\top)^\top$ with $\boldsymbol{b}^{(0)} = (b_1^{(0)}, \ldots, b_K^{(0)})^\top$. Under $\mathbf{H}_1$, $\underset{\sim}{\boldsymbol{\beta}}^*$ is generally a weighted combination of the parameters before the change point and those after the change point. For example, when $\alpha = 1$, it has the explicit form of $\underset{\sim}{\boldsymbol{\beta}}^* = ((t_1\boldsymbol{\beta}^{(1)} + t_2\boldsymbol{\beta}^{(2)})^\top, (\boldsymbol{b}^{(0)})^\top)^\top$. With the above notations, we are ready to introduce our assumptions as follows:

Assumption A (Design matrix): The design matrix $\mathbf{X}$ has i.i.d rows $\{\boldsymbol{X}_i\}_{i=1}^n$. (A.1) Assume that there are positive constants κ_1 and κ_2 such that $\lambda_{\min}(\boldsymbol{\Sigma}) \geq \kappa_1 > 0$ and $\lambda_{\max}(\boldsymbol{\Sigma}) \leq \kappa_2 < \infty$ hold. (A.2) There exists some constant $M \geq 1$ such that $\max_{1\leq i\leq n}\max_{1\leq j\leq p}|X_{ij}| \leq M$ almost surely for every n and p.

Assumption B (Error distribution): The error terms $\{\epsilon_i\}_{i=1}^n$ are i.i.d. with mean zero and finite variance σ_ϵ^2. There exist positive constants c_ϵ and C_ϵ such that $c_\epsilon^2 \leq \text{Var}(\epsilon_i) \leq C_\epsilon^2$ hold. In addition, ϵ_i is independent with $\boldsymbol{X}_i$ for $i = 1, \ldots, n$.

Assumption C (Moment constraints): (C.1) There exists some constant $b > 0$ such that $\mathbb{E}(\boldsymbol{v}^\top \boldsymbol{X}_i \epsilon_i)^2 \geq b$ and $\mathbb{E}(\boldsymbol{v}^\top \boldsymbol{X}_i e_i(\widetilde{\boldsymbol{\tau}}))^2 \geq b$, for $\boldsymbol{v} \in \mathcal{V}_{s_0}$ and all $i = 1, \ldots, n$. Moreover, assume that $\inf_{i,j} \mathbb{E}[X_{ij}^2] \geq b$ holds. (C.2) There exists a constant $K > 0$ such that $\mathbb{E}|\epsilon_i|^{2+\ell} \leq K^\ell$, for $\ell = 1, 2$.

Assumption D (Underlying distribution): The distribution function ϵ has a continuously differentiable density function $f_\epsilon(t)$ whose derivative is denoted by $f'_\epsilon(t)$. Furthermore, suppose there exist some constants C_+, C_- and C'_+ such that

$$(\text{D.1}) \sup_{t \in \mathbb{R}} f_\epsilon(t) \leq C_+; (\text{D.2}) \inf_{j=1,2} \inf_{1 \leq k \leq K} f_\epsilon(\boldsymbol{x}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(j)}) + b_k^*) \geq C_-;$$
$$(\text{D.3}) \sup_{t \in \mathbb{R}} |f'_\epsilon(t)| \leq C'_+.$$

Assumption E (Parameter space):

(E.1) We require $s_0^3 \log(pn) = O(n^{\xi_1})$ for some $0 < \xi_1 < 1/7$ and $s_0^4 \log(pn) = O(n^{\xi_2})$ for some $0 < \xi_2 < \frac{1}{6}$.

(E.2) Assume that $\dfrac{s_0^2 s^3 \log^3(pn)}{n} \to 0$ as $(n, p) \to \infty$, where s is the overall sparsity of $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$.

(E.3) We require $\max(\|\boldsymbol{\beta}^{(1)}\|_\infty, \|\boldsymbol{\beta}^{(2)}\|_\infty) < C_{\boldsymbol{\beta}}$ for some $C_{\boldsymbol{\beta}} > 0$. Moreover, we require $\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_1 \leq C_{\boldsymbol{\Delta}}$ for some constant $C_{\boldsymbol{\Delta}} > 0$.

(E.4) For the tuning parameters λ_α in (2.10), we require $\lambda_\alpha = C_\lambda \sqrt{\log(pn)/n}$ for some $C_\lambda > 0$.

Assumption A gives some conditions for the design matrix, requiring $\boldsymbol{X}$ has a non-degenerate covariance matrix $\boldsymbol{\Sigma}$ in terms of its eigenvalues. This is important for deriving the high-dimensional LASSO property with $\alpha \in [0, 1]$ under both $\mathbf{H}_0$ and $\mathbf{H}_1$. Assumption B mainly requires the underlying error term ϵ_i has non-degenerate variance. Assumption C imposes some restrictions on the moments of the error terms as well as the design matrix. In particular, Assumption C.1 requires that $\boldsymbol{v}^\top \boldsymbol{X}\epsilon$, $\boldsymbol{v}^\top \boldsymbol{X} e(\widetilde{\boldsymbol{\tau}})$, as well as X_{ij} have non-degenerate variances. Moreover, Assumption C.2 requires that the errors have at most fourth moments, which is much weaker than the commonly used Gaussian or sub-Gaussian assumptions. Both Assumptions C.1 and C.2 are basic moment conditions for bootstrap approximations for the individual-based tests. See Lemma C.6 in the proof. Assumptions D.1 - D.3 are some regular conditions for the underlying distribution of the errors, requiring ϵ has a bounded density function as well as bounded derivatives. Assumption D.2 also requires the density function at $\boldsymbol{x}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(j)}) + b_k^*$ to be strictly bounded away from zero. Lastly, Assumption E imposes some conditions for the parameter spaces in terms of

$(s_0, n, p, s, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)})$. Specifically, Assumption E.1 scales the relationship between s_0, n, and p, which allows s_0 can grow with the sample size n. This condition is mainly used to establish the high-dimensional Gaussian approximation for our individual tests. Assumption E.2 also gives some restrictions on (s_0, s, n, p). Note that both Assumptions E.1 and E.2 allow the data dimension p to be much larger than the sample size n as long as the required conditions hold. Assumption E.3 requires that the regression coefficients as well as signal jump in terms of its ℓ_1-norm are bounded. Assumption E.4 imposes the regularization parameter $\lambda_\alpha = O(\sqrt{\log(pn)/n})$, which is important for deriving the desired error bound for the LASSO estimators under both $\mathbf{H}_0$ and $\mathbf{H}_1$ using our weighted composite loss function. See Lemmas C.9 - C.11 in the proof.

Remark 1. Assumption C.2 with the finite fourth moment is mainly for the individual test with $\alpha = 1$, while Assumption D is for that with $\alpha = 0$. Note that Assumption D only imposes some conditions on the density functions of the errors instead of the moments, which can be statisfied for the errors with heavy tails. Hence, in both cases, our proposed individual-based change point method extends the high-dimensional linear models with sub-Gaussian distributed errors to those with only finite moments or without any moments, covering a wide range of errors with different tails.

S6 Useful lemmas

Lemma 1 (Lemma E.1 in Chernozhukov et al. (2017)). *Let* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^p$ *with* $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^\top$ *be independet and centered random vectors. Define* $Z = \max_{1 \le j \le p} |\sum_{i=1}^n X_{ij}|$, $M = \max_{1 \le i \le n} \max_{1 \le j \le p} |X_{ij}|$ *and* $\sigma^2 = \max_j \sum_i \mathbb{E}[X_{ij}^2]$. *Then,*

$$\mathbb{E}[Z] \le C(\sigma\sqrt{\log p} + \sqrt{\mathbb{E}[M^2]} \log p),$$

where C *is some universal constant.*

Lemma 2 (Lemma E.2 in Chernozhukov et al. (2017)). *(a) Assume the setting of Lemma 1 holds. For every* $\eta > 0, \beta \in (0, 1]$ *and* $t > 0$*, we have*

$$\mathbb{P}(Z \ge (1+\eta)\mathbb{E}[Z] + t) \le \exp(-\frac{t^2}{3\sigma^2}) + 3\exp\Big(-\big(\frac{t}{K\|M\|_{\psi_\beta}}\big)^\beta\Big),$$

where $K = K(\eta, \beta)$ *is a constant only depending on* η *and* β.

(b) Assume the setting of Lemma 1 holds. For every $\eta > 0, s \ge 1$ *and* $t > 0$*, we have*

$$\mathbb{P}(Z \ge (1+\eta)\mathbb{E}[Z] + t) \le \exp(-\frac{t^2}{3\sigma^2}) + K'\frac{\mathbb{E}[M^s]}{t^s},$$

where $K' = K(\eta, s)$ *is a constant only depending on* η *and* s.

Lemma 3 (Hoeffding's inequality). *Suppose* $X_1, \ldots, X_n \in \mathbb{R}^1$ *be independent random variables with* $|X_i| \le K$ *for some* $K > 0$. *Let* $\bar{X}$ *be the sample* *mean. Then, for any $x > 0$, we have*

$$\mathbb{P}(|\bar{X} - \mathbb{E}\bar{X}| \geq x) \leq 2\exp\Big(-\frac{nx^2}{2K^2}\Big). \tag{S6.4}$$

Lemma 4 (Nazarovs inequality in Nazarov (2003))**.** *Let $\boldsymbol{W} = (W_1, W_2, \cdots, W_d)^\top \in \mathbb{R}^p$ be centered Gaussian random vector with $\inf_{1\leq k\leq p}\mathbb{E}(W_k)^2 \geq b > 0$. Then for any $\boldsymbol{x} \in \mathbb{R}^p$ and $a > 0$, we have*

$$\mathbb{P}(\boldsymbol{W} \leq \boldsymbol{x} + a) - \mathbb{P}(\boldsymbol{W} \leq \boldsymbol{x}) \leq Ca\sqrt{\log p},$$

where C is a constant only depending on b.

Before introducing Lemma 5, we need some definitions for an m-generated convex set A^m. We say a set A^m is m-generated if it is generated by intersecting m half spaces. In other words, the set A^m is a convex polytope with at most m facets. Moreover, for any $\epsilon > 0$ and an m-generated convex set A^m, we define

$$A^{m,\epsilon} = \bigcap_{\boldsymbol{v}\in\mathcal{V}(A^m)} \{\boldsymbol{w} \in \mathbb{R}^d : \boldsymbol{w}^\top\boldsymbol{v} \leq S_{A^m}(\boldsymbol{v}) + \epsilon\}, \tag{S6.5}$$

where $\mathcal{V}(A^m)$ consists m unit vectors that are outward normal to the facets of A^m, and $S_{A^m}(\boldsymbol{v})$ is the support function for A^m (see Chernozhukov et al. (2017)).

Let $\boldsymbol{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$. For any $1 \leq s_0 \leq p$, define $\|\boldsymbol{x}\|_{(s_0,2)} = (\sum_{j=1}^{s_0}|x_{(j)}|^2)^{1/2}$, where $|x_{(1)} \geq |x_{(2)}| \cdots \geq |x_{(p)}|$ be the order statistics of $\boldsymbol{x}$.

The following lemma shows that the set $V_{(s_0,2)}^{z,p} := \{\boldsymbol{x} \in \mathbb{R}^p : \|\boldsymbol{x}\|_{(s_0,2)} \leq z\}$ can be approximated by m-generated convex set.

Lemma 5 (Zhou et al. (2018))**.** *Let $\mathcal{E}^{R,p} = \{\boldsymbol{x} \in \mathbb{R}^p : \|\boldsymbol{x}\| \leq R\}$ and $V_{(s_0,2)}^{z,p} = \{\boldsymbol{x} \in \mathbb{R}^p : \|\boldsymbol{x}\|_{(s_0,2)} \leq z\}$. For any $\gamma > e/4\sqrt{2}$, there is a m-generated convex set $A^m \in \mathbb{R}^p$ and a constant ϵ_γ such that for any $0 < \epsilon < \epsilon_\gamma$, we have*

$$A^m \subset \mathcal{E}^{R,p} \cap V_{(s_0,2)}^{z,p} \subset A^{m,R\epsilon} \quad \textit{and} \quad m \leq p^{s_0}\Big(\frac{\gamma}{\sqrt{\epsilon}}\ln(\frac{1}{\epsilon})\Big)^{s_0^2}.$$

The following Lemma 6 shows the Gaussian approximation theory for the testing statistic, which is very important for the size control. To show that, we need some notations and assumptions. In particular, let $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n \sim (\mathbf{0}, \boldsymbol{\Sigma})$ be independent and centered random vectors in $\mathbb{R}^p$ with $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{ip})^\top$ for $i = 1, \ldots, n$. Let $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n$ be independent centered Gaussian random vectors in $\mathbb{R}^p$ such that each $\boldsymbol{G}_i$ has the same covariance matrix as $\boldsymbol{Z}_i$. Let $\mathcal{V}_{s_0} := \{\boldsymbol{v} \in \mathbb{S}^{q-1} : \|\boldsymbol{v}\|_0 \leq s_0\}$, where $\mathbb{S}^{q-1} := \{\boldsymbol{v} \in \mathbb{R}^p : \|\boldsymbol{v}\| = 1\}$. We require the following conditions:

(M1) There is a constant $b > 0$ such that $\inf_{\boldsymbol{v} \in \mathcal{V}_{s_0}} \dfrac{1}{n}\sum_{i=1}^{n} \mathbb{E}(\boldsymbol{v}^\top \boldsymbol{Z}_i)^2 \geq b$ for $i = 1, \ldots, n$.

(M2) There exists some constant $K > 0$ such that $\max\limits_{1\leq j\leq p} \dfrac{1}{n}\sum_{i=1}^{n} \mathbb{E}|Z_{ij}|^{2+\ell} \leq K^\ell$ for $\ell = 1, 2$.

(M3) There exists a constant $K > 0$ and $q > 0$ such that $\mathbb{E}((\max_{1\le j\le p}|Z_{ij}|/K)^q) \le 2$ holds for all $i = 1, \ldots, n$.

Lemma 6. *Assume that that* $s_0^3 K^{2/7}\log(pn) = O(n^{\xi_1})$ *for some* $0 < \xi_1 < 1/7$ *and* $s_0^4 K^{2/3}\log(pn) = O(n^{\xi_2})$ *for some* $0 < \xi_2 < \frac{1}{3}(1-2/q)$*. Let*

$$\boldsymbol{S}^{\boldsymbol{Z}}(k) = \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{k}\boldsymbol{Z}_i - \frac{k}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big), \quad \boldsymbol{S}^{\boldsymbol{G}}(k) = \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{k}\boldsymbol{G}_i - \frac{k}{n}\sum_{i=1}^{n}\boldsymbol{G}_i\Big), \tag{S6.6}$$

be the partial sum processes for $(\boldsymbol{Z}_i)_{i\ge 1}$ *and* $(\boldsymbol{G}_i)_{i\ge 1}$*, respectively. If* $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ *satisfy* (**M1**), (**M2**) *and* (**M3**)*, then there is a constant* $\zeta_0 > 0$ *such that*

$$\sup_{z\in(0,\infty)}\Big|\mathbb{P}(\max_{k_0\le k\le n-k_0}\|\boldsymbol{S}^{\boldsymbol{Z}}(k)\|_{(s_0,2)} \le z\big) - \mathbb{P}(\sup_{k_0\le k\le n-k_0}\|\boldsymbol{S}^{\boldsymbol{G}}(k)\|_{(s_0,2)} \le z\big)\Big| \le Cn^{-\zeta_0}, \tag{S6.7}$$

where C *is a constant only depending on* b, q*,* K *and* $k_0 := \lfloor nq_0 \rfloor$ *for some* $0 < q_0 < 0.5$.

The following Lemmas 7 and 8 present the orders for the partial sum process of $\{\boldsymbol{X}\epsilon\}_{i=1}^n$ as well as the ℓ_∞-norm based uniform large deviation bound for $\widehat{\boldsymbol{\Sigma}}(0:t)$ and $\widehat{\boldsymbol{\Sigma}}(t:1)$, which will be frequently used throughout the proofs.

Lemma 7. *Let* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ *be independent centered random vectors in* $\mathbb{R}^p$ *and* $\epsilon_1, \ldots, \epsilon_n$ *be independent centered random vectors in* $\mathbb{R}^1$*. Suppose further that* $\{\boldsymbol{X}_i\}_{i=1}^n$ *and* $\{\epsilon_i\}_{i=1}^n$ *satisfy* **Assumptions A − C** *in the* *main paper. Then, for any sequence $a_n \in (0,1)$ and $b_n \in (0,1)$ satisfying $\lfloor na_n \rfloor \to \infty$ and $\lfloor nb_n \rfloor \to \infty$ as $n \to \infty$, we have*

$$\begin{aligned}
&\max_{t\in[a_n,1-b_n]} \max_{1\le j\le p} \Big| \frac{1}{\sqrt{n}} \Big(\sum_{i=1}^{\lfloor nt \rfloor} X_{ij}\epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} X_{ij}\epsilon_i \Big) \Big| \\
&\quad = \max_{t\in[a_n,1-b_n]} \Big\| \frac{1}{\sqrt{n}} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i\epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i\epsilon_i \Big) \Big\|_\infty \\
&\quad = O_p(M\sqrt{\log(p(n - \underline{k}_n - \overline{k}_n))}),
\end{aligned} \tag{S6.8}$$

where $\underline{k}_n := \lfloor na_n \rfloor$ and $\overline{k}_n := \lfloor nb_n \rfloor$. Moreover, we can also have the following results:

$$\begin{aligned}
&\max_{t\in[a_n,1-b_n]} \max_{1\le j\le p} \Big| \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} X_{ij}\epsilon_i \Big| = \max_{t\in[a_n,1-b_n]} \Big\| \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i\epsilon_i \Big\|_\infty \\
&= O_p\Big(M\sqrt{\frac{\log(pn)}{\underline{k}_n}} \max\big\{1, n^{1/4}\sqrt{\frac{\log(pn)}{\underline{k}_n}}\big\} \Big).
\end{aligned} \tag{S6.9}$$

Lemma 8. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent centered random vectors in $\mathbb{R}^p$ satisfying* **Assumption A**. *Let $\boldsymbol{\Sigma} = Cov(\boldsymbol{X}_1)$. Recall $\widehat{\boldsymbol{\Sigma}}(s:t)$ defined in (S4.2). Then, for any sequence $a_n \in (0,1)$ and $b_n \in (0,1)$ satisfying $\lfloor na_n \rfloor \to \infty$ and $\lfloor nb_n \rfloor \to \infty$ as $n \to \infty$, with probability at least $1-(np)^{-C_1}$, we have:*

$$\begin{aligned}
\max_{a_n\le t\le 1-b_n} \|\widehat{\boldsymbol{\Sigma}}(0:t) - \boldsymbol{\Sigma}\|_\infty &\le C_2 M^2 \sqrt{\frac{\log(pn)}{\lfloor na_n \rfloor}}, \\
\max_{a_n\le t\le 1-b_n} \|\widehat{\boldsymbol{\Sigma}}(t:1) - \boldsymbol{\Sigma}\|_\infty &\le C_3 M^2 \sqrt{\frac{\log(pn)}{\lfloor nb_n \rfloor}}.
\end{aligned}$$

Moreover, if we take $a_n = b_n = q_0 \in (0, 0.5)$, we have

$$\max_{q_0 \le t \le 1-q_0} \|\widehat{\mathbf{\Sigma}}(0:t) - \mathbf{\Sigma}\|_\infty \le C_4 M^2 \sqrt{\frac{\log(pn)}{n}},$$
$$\max_{q_0 \le t \le 1-q_0} \|\widehat{\mathbf{\Sigma}}(t:1) - \mathbf{\Sigma}\|_\infty \le C_5 M^2 \sqrt{\frac{\log(pn)}{n}},$$

where $C_1, \ldots, C_5$ are some universal constants. Note that Lemma 8 is a direct consequence of Lemma 3. The proof is ommitted.

Note that for proving our results, we need some theoretical analysis for the lasso estimator defined in (2.10). The following Lemmas 9 - 11 show the lasso property for $\alpha \in [0, 1]$ under both $\mathbf{H}_0$ and $\mathbf{H}_1$, which is very important for deriving the theoretical results for the individual test. Before presenting the details, for each $\alpha \in [0, 1]$, we introduce $\underset{\sim}{\boldsymbol{\beta}}^* = ((\boldsymbol{\beta}^*)^\top, (\boldsymbol{b}^*)^\top)^\top \in \mathbb{R}^{p+K}$ with $\boldsymbol{b}^* = (b_1^*, \ldots, b_K^*)^\top \in \mathbb{R}^K$, where

$$\underset{\sim}{\boldsymbol{\beta}}^* := \underset{\boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{b} \in \mathbb{R}^K}{\arg\min}\, \mathbb{E}\Big[(1-\alpha)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\rho_{\tau_k}(Y_i - b_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}) + \frac{\alpha}{2n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta})^2\Big]. \tag{S6.10}$$

Note that by definition, we can regard $\underset{\sim}{\boldsymbol{\beta}}^*$ as the true parameters under the population level. In this paper, we assume $\boldsymbol{\beta}^*$ enjoys some sparsity property in the sense that $\mathcal{M}(\boldsymbol{\beta}^*) = O(s)$. Moreover, the properties of $\underset{\sim}{\boldsymbol{\beta}}^*$ are discussed in Sections S9.3 - S9.5, respectively.

The following Lemma 9 shows the lasso property with $\alpha = 1$. The proof of Lemma 9 is given in Section S9.3.

Lemma 9 (Lasso property with $\alpha = 1$)**.** *Let $\widehat{\boldsymbol{\beta}}$ be the lasso estimator with $\alpha = 1$ defined in (2.10). Let $\lambda_\alpha = C_\lambda M^2 \sqrt{\log(pn)/n}$ for some big enough constant $C_\lambda > 0$. Assume* **Assumptions A, B, C.2, E.2 - E.3** *hold. Then, with probability tending to one, we have*

$$\frac{1}{2n}\left\|\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right\|^2 \leq C_2\lambda^2 s, \quad \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_q \leq C_3\lambda s^{1/q}, \ \textit{and} \ \ \mathcal{M}(\widehat{\boldsymbol{\beta}}) \leq C_4 s, \quad \textit{for } q = 1, 2. \tag{S6.11}$$

The following Lemma 10 shows the lasso property with $\alpha = 0$ under both $\mathbf{H}_0$ and $\mathbf{H}_1$. The proof of Lemma 10 is given in Section S9.4.

Lemma 10 (Lasso property with $\alpha = 0$)**.** *Let $\widehat{\boldsymbol{\beta}}$ be the lasso estimator with $\alpha = 0$ defined in (2.10). Let $\lambda = C_\lambda M \sqrt{\log(pn)/n}$ for some big enough constant $C_\lambda > 0$. Assume* **Assumptions A, D, E.2 - E.3** *hold. Then, with probability tending to one, we have*

$$\frac{1}{n}\left\|\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right\|^2 \leq C_2\lambda^2 s, \quad \|\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^*\|_q \leq C_3\lambda s^{1/q}, \ \textit{and} \ \ \mathcal{M}(\widehat{\boldsymbol{\beta}}) \leq C_4 s, \quad \textit{for } q = 1, 2. \tag{S6.12}$$

The following Lemma 11 shows the lasso property with $\alpha \in (0, 1)$ under both $\mathbf{H}_0$ and $\mathbf{H}_1$. The proof of Lemma 11 is given in Section S9.5.

Lemma 11 (Lasso property with $\alpha \in (0, 1)$)**.** *Let $\widehat{\boldsymbol{\beta}}$ be the lasso estimator with $\alpha \in (0, 1)$ defined in (2.10). Let $\lambda = C_\lambda M^2 \sqrt{\log(pn)/n}$ for some big enough constant $C_\lambda > 0$. Assume* **Assumptions A, B, C.2, D, E.2 -**

E.3 *hold. Then, with probability tending to one, we have*

$$\frac{1}{n}\big\|\mathbf{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\big\|^2 \le C_2\lambda^2 s, \quad \|\underset{\sim}{\widehat{\boldsymbol{\beta}}}-\underset{\sim}{\boldsymbol{\beta}}^*\|_q \le C_3\lambda s^{1/q}, \ \textit{and} \ \mathcal{M}(\widehat{\boldsymbol{\beta}}) \le C_4 s, \quad \textit{for } q=1,2. \tag{S6.13}$$

S7 Proof of main results

S7.1 Proof of Theorem 1

In this section, we prove the variance estimation results under $\mathbf{H}_0$, which are given in Sections S7.1.1 - S7.1.2, respectively. For simplicity, we omit the subscript α whenever needed.

S7.1.1 Proof of Theorem 1 with $\alpha = 1$

Note that for $\alpha = 1$, the variance estimators $\widehat{\sigma}_-^2(1,\widetilde{\boldsymbol{\tau}})$ and $\widehat{\sigma}_+^2(1,\widetilde{\boldsymbol{\tau}})$ reduce to

$$\widehat{\sigma}_-^2(1,\widetilde{\boldsymbol{\tau}}) := \frac{1}{|n_-|}\sum_{i\in n_-}\big[\widehat{\epsilon}_i\big]^2, \quad \widehat{\sigma}_+^2(1,\widetilde{\boldsymbol{\tau}}) := \frac{1}{|n_+|}\sum_{i\in n_+}\big[\widehat{\epsilon}_i\big]^2,$$

where $\widehat{\epsilon}_i$ is defined in (2.20). Moreover, under $\mathbf{H}_0$, the change point estimator $\widehat{t}_1$ can be an arbitrary number which satisfies $\widehat{t}_1 \in [q_0, 1-q_0]$. We aim to prove both $\widehat{\sigma}_-^2(1,\widetilde{\boldsymbol{\tau}})$ and $\widehat{\sigma}_+^2(1,\widetilde{\boldsymbol{\tau}})$ are consistent. We first consider

$\widehat{\sigma}^2_{-}(1,\widetilde{\boldsymbol{\tau}})$. In fact, by the definition of $Y_i = \epsilon_i + \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(0)}$, we have:

$$
\begin{aligned}
&\widehat{\sigma}^2_{-}(1,\widetilde{\boldsymbol{\tau}})\\
&= \underbrace{\frac{1}{|n_-|}\sum_{i\in n_-}\epsilon_i^2}_{I} + \underbrace{(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(0)})^\top \frac{1}{|n_-|}\sum_{i\in n_-}\boldsymbol{X}_i\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(0)})}_{II}\\
&\qquad + \underbrace{\frac{2}{|n_-|}\sum_{i\in n_-}\epsilon_i\boldsymbol{X}_i^\top(\boldsymbol{\beta}^{(0)} - \widehat{\boldsymbol{\beta}}^{(1)})}_{III}.
\end{aligned}
$$

For I, by Assumption C.2 and according to the law of large numbers, we have $I - \sigma^2 = O_p(\frac{1}{\sqrt{n}})$. For II, similar to the proof of Lemma 9, under $\mathbf{H}_0$ and **Assumptions A, B, C.2, E.2 - E.4**, one can prove $II = O_p(s\frac{\log(pn)}{n})$. For III, using the Cauchy-Swartz inequality, we have:

$$
III \le 2\sqrt{\frac{1}{|n_-|}\sum_{i\in n_-}\epsilon_i^2}\times\sqrt{(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(0)})^\top \frac{1}{|n_-|}\sum_{i\in n_-}\boldsymbol{X}_i\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(0)})} = O_p(\sqrt{s\frac{\log(pn)}{n}}).
$$

Combining the above results, by Assumption E.2, we have:

$$
\widehat{\sigma}^2_{-}(1,\widetilde{\boldsymbol{\tau}}) - \sigma^2 = O_p(\sqrt{s\frac{\log(pn)}{n}}).
$$

With a similar analysis, we can prove that the same bound applies to $\widehat{\sigma}^2_{+}(1,\widetilde{\boldsymbol{\tau}}) - \sigma^2$, which yields:

$$
|\widehat{\sigma}^2(1,\widetilde{\boldsymbol{\tau}}) - \sigma^2| = |\widehat{t}_1\times(\widehat{\sigma}^2_{-}(1,\widetilde{\boldsymbol{\tau}}) - \sigma^2) + (1-\widehat{t}_1)\times(\widehat{\sigma}^2_{+}(1,\widetilde{\boldsymbol{\tau}}) - \sigma^2)| = O_p(\sqrt{s\frac{\log(pn)}{n}}).
$$

S7.1.2 Proof of Theorem 1 with $\alpha = 0$

Note that for $\alpha = 0$, the true variance has the following explicit form:

$$\sigma^2(0, \widetilde{\boldsymbol{\tau}}) := \mathrm{Var}[e_i(\widetilde{\boldsymbol{\tau}})] = \frac{1}{K^2} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \gamma_{k_1 k_2}, \text{ with } \gamma_{k_1 k_2} := \min(\tau_{k_1}, \tau_{k_2}) - \tau_{k_1} \tau_{k_2}.$$

In this case, the variance estimators $\widehat{\sigma}^2_-(0, \widetilde{\boldsymbol{\tau}})$ and $\widehat{\sigma}^2_+(0, \widetilde{\boldsymbol{\tau}})$ reduce to

$$\widehat{\sigma}^2_-(0, \widetilde{\boldsymbol{\tau}}) := \frac{1}{|n_-|} \sum_{i \in n_-} \left[\widehat{e}_i(\widetilde{\boldsymbol{\tau}})\right]^2, \quad \widehat{\sigma}^2_+(0, \widetilde{\boldsymbol{\tau}}) := \frac{1}{|n_+|} \sum_{i \in n_+} \left[\widehat{e}_i(\widetilde{\boldsymbol{\tau}})\right]^2,$$

where $\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) = K^{-1} \sum_{k=1}^{K} \widehat{e}_i(\tau_k)$ with $\widehat{e}_i(\tau_k)$ being defined in (2.21). Let $\underset{\sim}{\boldsymbol{\beta}}^{(0)} := ((\boldsymbol{\beta}^{(0)})^\top, (\boldsymbol{b}^{(0)})^\top)^\top$ be the true parameters under $\mathbf{H}_0$ and $\underset{\sim}{\widehat{\boldsymbol{\beta}}}^{(1)} := ((\widehat{\boldsymbol{\beta}}^{(1)})^\top, \widehat{\boldsymbol{b}}^{(1)})^\top)^\top$ and $\underset{\sim}{\widehat{\boldsymbol{\beta}}}^{(2)} := ((\widehat{\boldsymbol{\beta}}^{(2)})^\top, \widehat{\boldsymbol{b}}^{(1)})^\top)^\top$ be the estimators using samples in n_- and n_+, respectively. Similar to the proof of Lemma 10, under $\mathbf{H}_0$ and **Assumptions A, D, E.2 - E.4**, we can prove that:

$$\|\underset{\sim}{\widehat{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1 = O_p(s\sqrt{\frac{\log(pn)}{n}}), \quad \|\underset{\sim}{\widehat{\boldsymbol{\beta}}}^{(2)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1 = O_p(s\sqrt{\frac{\log(pn)}{n}}). \tag{S7.14}$$

We first prove the consistency of $\widehat{\sigma}^2_-(0, \widetilde{\boldsymbol{\tau}})$. For $\widehat{e}_i(\widetilde{\boldsymbol{\tau}})$ with $i \in n_-$, it has the following decomposition:

$$\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) = e_i(\widetilde{\boldsymbol{\tau}}) + \mathbb{E}[\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}})] + \underbrace{\{\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}}) - \mathbb{E}[\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}})]\}}_{V_i(\widetilde{\boldsymbol{\tau}})}, \tag{S7.15}$$

where

$$\begin{aligned} e_i(\widetilde{\boldsymbol{\tau}}) &:= \frac{1}{K}\sum_{k=1}^{K} e_i(\tau_k), \text{ with } e_i(\tau_k) = \mathbf{1}\{\epsilon_i \le b_k^{(0)}\} - \tau_k, \\ \mathbb{E}[\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}})] &:= \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\big[\widehat{e}_i(\tau_k) - e_i(\tau_k)\big], \\ V_i(\widetilde{\boldsymbol{\tau}}) &:= \frac{1}{K}\sum_{k=1}^{K} V_i(\tau_k), \end{aligned} \tag{S7.16}$$

and

$$\begin{aligned} V_i(\tau_k) = {} & [\mathbf{1}\{Y_i - \widehat{b}_k^{(1)} - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}^{(1)} \le 0\} - \mathbf{1}\{\epsilon_i \le b_k^{(0)}\}] \\ & - \mathbb{E}[\mathbf{1}\{Y_i - \widehat{b}_k^{(1)} - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}^{(1)} \le 0\} - \mathbf{1}\{\epsilon_i \le b_k^{(0)}\}]. \end{aligned}$$

By the Taylor's expansion, for $\mathbb{E}[\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}})]$, we can further decompose it into two terms:

$$\mathbb{E}[\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}})] = \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\big[\widehat{e}_i(\tau_k) - e_i(\tau_k)\big] = \underbrace{\frac{1}{K}\sum_{k=1}^{K} M_i^{(1)}(\tau_k)}_{M_i^{(1)}(\widetilde{\boldsymbol{\tau}})} + \underbrace{\frac{1}{K}\sum_{k=1}^{K} M_i^{(2)}(\tau_k)}_{M_i^{(2)}(\widetilde{\boldsymbol{\tau}})}, \tag{S7.17}$$

where

$$\begin{aligned} M_i^{(1)}(\tau_k) &:= f_\epsilon(b_k^{(0)})\big(\widehat{b}_k^{(1)} - b_k^{(0)} + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(0)})\big), \\ M_i^{(2)}(\tau_k) &:= \frac{1}{2} f_\epsilon'(\xi_{ik})\big(\widehat{b}_k^{(1)} - b_k^{(0)} + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(0)})\big)^2, \end{aligned} \tag{S7.18}$$

with ξ_{ik} being some constant that between $b_k^{(0)}$ and $\widehat{b}_k^{(1)} + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(0)})$. Hence, based on the above decomposition, for $\widehat{\sigma}_-^2(0, \widetilde{\boldsymbol{\tau}}) - \sigma^2(0, \widetilde{\boldsymbol{\tau}})$, it can be decomposed into ten terms:

$$\widehat{\sigma}_-^2(0, \widetilde{\boldsymbol{\tau}}) - \sigma^2(0, \widetilde{\boldsymbol{\tau}}) = A_1 + \cdots + A_{10},$$

where $A_1, \cdots, A_{10}$ are defined as:

$$A_1 := \frac{1}{|n_-|}\sum_{i\in n_-}\left[e_i(\widetilde{\boldsymbol{\tau}})\right]^2 - \sigma^2(0,\widetilde{\boldsymbol{\tau}}), \quad A_2 := \frac{1}{|n_-|}\sum_{i\in n_-}\left[M_i^{(1)}(\widetilde{\boldsymbol{\tau}})\right]^2$$
$$A_3 = \frac{1}{|n_-|}\sum_{i\in n_-}\left[M_i^{(2)}(\widetilde{\boldsymbol{\tau}})\right]^2, \quad A_4 = \frac{1}{|n_-|}\sum_{i\in n_-}\left[V_i(\widetilde{\boldsymbol{\tau}})\right]^2$$
$$A_5 := \frac{2}{|n_-|}\sum_{i\in n_-}\left[e_i(\widetilde{\boldsymbol{\tau}})M_i^{(1)}(\widetilde{\boldsymbol{\tau}})\right], \quad A_6 := \frac{2}{|n_-|}\sum_{i\in n_-}\left[e_i(\widetilde{\boldsymbol{\tau}})M_i^{(2)}(\widetilde{\boldsymbol{\tau}})\right]$$
$$A_7 := \frac{2}{|n_-|}\sum_{i\in n_-}\left[e_i(\widetilde{\boldsymbol{\tau}})V_i(\widetilde{\boldsymbol{\tau}})\right], \quad A_8 := \frac{2}{|n_-|}\sum_{i\in n_-}\left[M_i^{(1)}(\widetilde{\boldsymbol{\tau}})M_i^{(2)}(\widetilde{\boldsymbol{\tau}})\right]$$
$$A_9 := \frac{2}{|n_-|}\sum_{i\in n_-}\left[M_i^{(1)}(\widetilde{\boldsymbol{\tau}})V_i(\widetilde{\boldsymbol{\tau}})\right], \quad A_{10} := \frac{2}{|n_-|}\sum_{i\in n_-}\left[M_i^{(2)}(\widetilde{\boldsymbol{\tau}})V_i(\widetilde{\boldsymbol{\tau}})\right].$$

Next, we consider the above ten terms, respectively. For A_1, by the law of large numbers, we have $A_1 = O_p(n^{-1/2})$. For A_2 and A_3, by Assumption A.2 and the bounds in (S7.14), we can prove that

$$|A_2| = O_p(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^2), \quad |A_3| = O_p(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^4).$$

For A_4, similar to the proof in Lemma 17 but using very tedious modifications, we can prove

$$|A_4| = O_p\Big(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^{1/2}\sqrt{s\frac{\log(pn)}{n}}\Big) = O_p\Big(s(\frac{\log(pn)}{n})^{\frac{3}{4}}\Big).$$

For $A_5 - A_{10}$, using the obtained bounds and the Cauchy-Swartz inequality, we have:

$$
\begin{aligned}
&|A_5| = O_p(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1), \quad |A_6| = O_p(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^2),\\
&|A_7| = O_p\Big(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^{1/4}\big(s\frac{\log(pn)}{n}\big)^{1/4}\Big) = O_p\Big(s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}\Big),\\
&|A_8| = O_p(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^3), \quad |A_9| = O_p\Big(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^{5/4}\big(s\frac{\log(pn)}{n}\big)^{1/4}\Big),\\
&|A_{10}| = O_p\Big(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^{9/4}\big(s\frac{\log(pn)}{n}\big)^{1/4}\Big).
\end{aligned}
$$

By Assumption E.2, we can see that $|A_5|$ and $|A_7|$ dominate the other terms. Hence, we have:

$$
\widehat{\sigma}_-^2(0, \widetilde{\boldsymbol{\tau}}) - \sigma^2(0, \widetilde{\boldsymbol{\tau}}) = O_p\Big(s\sqrt{\frac{\log(pn)}{n}} \vee s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}\Big).
$$

With a similar analysis, we can prove that the same bound applies to $\widehat{\sigma}_+^2(0, \widetilde{\boldsymbol{\tau}}) - \sigma^2(0, \widetilde{\boldsymbol{\tau}})$, which yields:

$$
\begin{aligned}
&|\widehat{\sigma}^2(0, \widetilde{\boldsymbol{\tau}}) - \sigma^2(0, \widetilde{\boldsymbol{\tau}})|\\
&= |\widehat{t}_0 \times (\widehat{\sigma}_-^2(0, \widetilde{\boldsymbol{\tau}}) - \sigma^2(0, \widetilde{\boldsymbol{\tau}})) + (1 - \widehat{t}_0) \times (\widehat{\sigma}_+^2(0, \widetilde{\boldsymbol{\tau}}) - \sigma^2(0, \widetilde{\boldsymbol{\tau}}))|\\
&= O_p\Big(s\sqrt{\frac{\log(pn)}{n}} \vee s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}\Big).
\end{aligned}
$$

S7.1.3 Proof of Theorem 1 with $\alpha \in (0, 1)$

Note that for $\alpha \in (0, 1)$, the true variance has the following explicit form:

$$
\sigma^2(\alpha, \widetilde{\boldsymbol{\tau}}) = (1-\alpha)^2 \mathbb{E}[e_i^2(\widetilde{\boldsymbol{\tau}})] + \alpha^2\sigma^2 - 2\alpha(1-\alpha)\mathbb{E}[e_i(\widetilde{\boldsymbol{\tau}})\epsilon_i].
$$

In this case, the variance estimators $\widehat{\sigma}^2_-(\alpha,\widetilde{\boldsymbol{\tau}})$ and $\widehat{\sigma}^2_+(\alpha,\widetilde{\boldsymbol{\tau}})$ reduce to

$$\widehat{\sigma}^2_-(\alpha,\widetilde{\boldsymbol{\tau}}) := \frac{1}{|n_-|}\sum_{i\in n_-}\left[(1-\alpha)\widehat{e}_i(\widetilde{\boldsymbol{\tau}})-\alpha\widehat{\epsilon}_i\right]^2,\ \ \widehat{\sigma}^2_+(\alpha,\widetilde{\boldsymbol{\tau}}) := \frac{1}{|n_+|}\sum_{i\in n_+}\left[(1-\alpha)\widehat{e}_i(\widetilde{\boldsymbol{\tau}})-\alpha\widehat{\epsilon}_i\right]^2,$$

where $\widehat{\epsilon}_i$ is defined in (2.20) and $\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) = K^{-1}\sum_{k=1}^{K}\widehat{e}_i(\tau_k)$ with $\widehat{e}_i(\tau_k)$ being defined in (2.21). Recall $\underset{\sim}{\boldsymbol{\beta}}^{(0)} := ((\boldsymbol{\beta}^{(0)})^\top,(\boldsymbol{b}^{(0)})^\top)^\top$ are the true parameters under $\mathbf{H}_0$, and $\underset{\sim}{\widehat{\boldsymbol{\beta}}}^{(1)} := (\widehat{\boldsymbol{\beta}}^{(1)})^\top,\widehat{\boldsymbol{b}}^{(1)})^\top)^\top$ as well as $\underset{\sim}{\widehat{\boldsymbol{\beta}}}^{(2)} := (\widehat{\boldsymbol{\beta}}^{(2)})^\top,\widehat{\boldsymbol{b}}^{(1)})^\top)^\top$ are the estimators using samples in n_- and n_+, respectively. Similar to the proof of Lemma 11, under $\mathbf{H}_0$ and **Assumptions A, B, C.2, D, E.2 - E.4** , one can prove that:

$$\begin{aligned}
&\|\underset{\sim}{\widehat{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1 = O_p\Big(s\sqrt{\frac{\log(pn)}{n}}\Big),\quad \|\underset{\sim}{\widehat{\boldsymbol{\beta}}}^{(2)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1 = O_p\Big(s\sqrt{\frac{\log(pn)}{n}}\Big),\\
&(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(0)})^\top\frac{1}{|n_-|}\sum_{i\in n_-}\boldsymbol{X}_i\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(0)}) = O_p\Big(s\frac{\log(pn)}{n}\Big),\\
&(\widehat{\boldsymbol{\beta}}^{(2)} - \boldsymbol{\beta}^{(0)})^\top\frac{1}{|n_+|}\sum_{i\in n_+}\boldsymbol{X}_i\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}}^{(2)} - \boldsymbol{\beta}^{(0)}) = O_p\Big(s\frac{\log(pn)}{n}\Big).
\end{aligned} \tag{S7.19}$$

We first prove the consistency of $\widehat{\sigma}^2_-(\alpha,\widetilde{\boldsymbol{\tau}})$. For $\widehat{\sigma}^2_-(\alpha,\widetilde{\boldsymbol{\tau}}) - \sigma^2(\alpha,\widetilde{\boldsymbol{\tau}})$, it can be decomposed into three terms:

$$\begin{aligned}
&\widehat{\sigma}^2_-(\alpha,\widetilde{\boldsymbol{\tau}}) - \sigma^2(\alpha,\widetilde{\boldsymbol{\tau}})\\
&= (1-\alpha)^2\underbrace{\frac{1}{|n_-|}\sum_{i\in n_-}\left(\widehat{e}_i(\widetilde{\boldsymbol{\tau}})^2 - \mathbb{E}[e_i^2(\widetilde{\boldsymbol{\tau}})]\right)}_{A} + \alpha^2\underbrace{\frac{1}{|n_-|}\sum_{i\in n_-}(\widehat{\epsilon}_i^2 - \sigma^2)}_{B}\\
&-2\alpha(1-\alpha)\underbrace{\frac{1}{|n_-|}\sum_{i\in n_-}\left(\widehat{e}_i(\widetilde{\boldsymbol{\tau}})\widehat{\epsilon}_i - \mathbb{E}[e_i(\widetilde{\boldsymbol{\tau}})\epsilon_i]\right)}_{C}.
\end{aligned} \tag{S7.20}$$

Next, we consider the three terms A, B, and C, respectively. For B, using the bounds obtained in Section S7.1.1, we have: $B = O_p(\sqrt{s\frac{\log(pn)}{n}})$. For A, using the bounds in Section S7.1.2, we have:

$$A = O_p\Big(s\sqrt{\frac{\log(pn)}{n}} \vee s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}\Big).$$

Next, we consider C. By the decomposition of $e_i(\widetilde{\boldsymbol{\tau}})$ in (S7.16) and the fact that $\widehat{\epsilon_i} = \epsilon_i + \widehat{\epsilon_i} - \epsilon_i$, we can decompose C into eight terms:

$$\frac{1}{|n_-|}\sum_{i\in n_-}(\widehat{e}_i(\widetilde{\boldsymbol{\tau}})\widehat{\epsilon}_i - \mathbb{E}[e_i(\widetilde{\boldsymbol{\tau}})\epsilon_i]) = C_1 + \cdots + C_8,$$

where

$$\begin{aligned}
C_1 &= \frac{1}{|n_-|}\sum_{i\in n_-}(e_i(\widetilde{\boldsymbol{\tau}})\epsilon_i - \mathbb{E}[e_i(\widetilde{\boldsymbol{\tau}})\epsilon_i]) \quad C_2 = \frac{1}{|n_-|}\sum_{i\in n_-} M_i^{(1)}(\widetilde{\boldsymbol{\tau}})\epsilon_i,\\
C_3 &= \frac{1}{|n_-|}\sum_{i\in n_-} M_i^{(2)}(\widetilde{\boldsymbol{\tau}})\epsilon_i, \quad C_4 = \frac{1}{|n_-|}\sum_{i\in n_-} V_i(\widetilde{\boldsymbol{\tau}})\epsilon_i,\\
C_5 &= \frac{1}{|n_-|}\sum_{i\in n_-} e_i(\widetilde{\boldsymbol{\tau}})(\widehat{\epsilon_i} - \epsilon_i), \quad C_6 = \frac{1}{|n_-|}\sum_{i\in n_-} M_i^{(1)}(\widetilde{\boldsymbol{\tau}})(\widehat{\epsilon_i} - \epsilon_i),\\
C_7 &= \frac{1}{|n_-|}\sum_{i\in n_-} M_i^{(2)}(\widetilde{\boldsymbol{\tau}})(\widehat{\epsilon_i} - \epsilon_i), \quad C_8 = \frac{1}{|n_-|}\sum_{i\in n_-} V_i(\widetilde{\boldsymbol{\tau}})(\widehat{\epsilon_i} - \epsilon_i).
\end{aligned}$$

For C_1, by the law of large numbers, we have $C_1 = O_p(n^{-1/2})$. Note that using the bounds in (S7.19), with similar proof techniques as in Sections S7.1.1 and S7.1.2, we can prove:

$$\frac{1}{|n_-|}\sum_{i\in n_-}\big[M_i^{(1)}(\widetilde{\boldsymbol{\tau}})\big]^2 = O_p(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^2), \quad \frac{1}{|n_-|}\sum_{i\in n_-}\big[M_i^{(2)}(\widetilde{\boldsymbol{\tau}})\big]^2 = O_p(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^4),$$

$$\frac{1}{|n_-|}\sum_{i\in n_-}\big[V_i(\widetilde{\boldsymbol{\tau}})\big]^2 = O_p\Big(\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(0)}\|_1^{1/2}\sqrt{s\frac{\log(pn)}{n}}\Big) = O_p\Big(s(\frac{\log(pn)}{n})^{\frac{3}{4}}\Big),$$

$$\frac{1}{|n_-|}\sum_{i\in n_-}\big[\widehat{\epsilon_i} - \epsilon_i\big]^2 = O_p(s\frac{\log(pn)}{n}).$$

Hence, for $C_2 - C_8$ using the above bounds and the Cauchy-Swartz inequality, one can see that C_2 and C_4 dominate the other terms. Specifically, we have:

$$|C_2| \leq \sqrt{\frac{1}{|n_-|}\sum_{i\in n_-}\epsilon_i^2} \times \sqrt{\frac{1}{|n_-|}\sum_{i\in n_-}\big[M_i^{(1)}(\widetilde{\boldsymbol{\tau}})\big]^2} = O_p\Big(s\sqrt{\frac{\log(pn)}{n}}\Big),$$

and

$$|C_4| \leq \sqrt{\frac{1}{|n_-|}\sum_{i\in n_-}\epsilon_i^2} \times \sqrt{\frac{1}{|n_-|}\sum_{i\in n_-}\big[V_i(\widetilde{\boldsymbol{\tau}})\big]^2} = O_p\Big(s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}\Big),$$

which implies

$$C = O_p\Big(s\sqrt{\frac{\log(pn)}{n}} \vee s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}\Big).$$

Lastly, combining (S7.20) and the obtained upper bounds for A, B and C, we have

$$\widehat{\sigma}_-^2(\alpha, \widetilde{\boldsymbol{\tau}}) - \sigma^2(\alpha, \widetilde{\boldsymbol{\tau}}) = O_p\Big(s\sqrt{\frac{\log(pn)}{n}} \vee s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}\Big).$$

With a similar analysis, we can prove the same bound applies to $\widehat{\sigma}_+^2(\alpha, \widetilde{\boldsymbol{\tau}}) - \sigma^2(\alpha, \widetilde{\boldsymbol{\tau}})$, which yields:

$$|\widehat{\sigma}^2(\alpha, \widetilde{\boldsymbol{\tau}}) - \sigma^2(\alpha, \widetilde{\boldsymbol{\tau}})| = O_p\Big(s\sqrt{\frac{\log(pn)}{n}} \vee s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}\Big).$$

S7.2 Proof of Theorem 2

In this section, we prove the Gaussian approximation results under $\mathbf{H}_0$, which are given in Sections S7.2.1 - S7.2.3, respectively. For simplicity, we omit the subscript α whenever needed.

S7.2.1 Gaussian approximation for $\alpha = 1$

Proof. In this section, we give the size results for $\alpha = 1$. Note that in this case, our individual based test statistic T_1 reduces to the least squared based score type test statistic. Let $\boldsymbol{z}(\boldsymbol{x}, y; \boldsymbol{\beta}) = \boldsymbol{x}(y - \boldsymbol{x}^\top \boldsymbol{\beta}) := -S_1(y, \boldsymbol{x}; \widetilde{\boldsymbol{\tau}}, \boldsymbol{b}, \boldsymbol{\beta})$ be the negative score for the ℓ_2-loss. Let $\boldsymbol{Z}_i(\boldsymbol{X}_i, Y_i; \boldsymbol{\beta}) = \boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta})$ be the sample version. In this section, we aim to prove:

$$\sup_{z \in (0,\infty)} \left| \mathbb{P}(T_1 \le z) - \mathbb{P}(T_1^b \le z | \mathcal{X}) \right| = o_p(1), \text{ as } n, p \to \infty. \tag{S7.21}$$

The proof proceeds into three steps.

Step 1: Decomposition of T_1. Note that for $\alpha = 1$, the score based CUSUM process reduces to:

$$\boldsymbol{C}_1(t) = \frac{1}{\sqrt{n}\widehat{\sigma}(1, \widetilde{\boldsymbol{\tau}})} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{Z}_i(\boldsymbol{X}_i, Y_i; \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{Z}_i(\boldsymbol{X}_i, Y_i; \widehat{\boldsymbol{\beta}}) \Big), \tag{S7.22}$$

where $\widehat{\boldsymbol{\beta}}$ is the lasso estimator defined in (2.10) and $\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})$ is the variance estimator defined in (2.22). By definition, we have $T_1 = \max_{q_0 \le t \le 1 - q_0} \|\boldsymbol{C}_1(t)\|_{(s_0, 2)}$. Replacing $\widehat{\boldsymbol{\beta}}$ by $\boldsymbol{\beta}^{(0)}$ in $\boldsymbol{C}_1(t)$, we have:

$$\boldsymbol{C}_1(t) = \boldsymbol{C}_1^I(t) + \boldsymbol{C}_1^{II}(t),$$

where $\boldsymbol{C}_1^I(t)$ and $\boldsymbol{C}_{I1}^I(t)$ are defined as

$$\begin{aligned} \boldsymbol{C}_1^I(t) &= \frac{1}{\sqrt{n}\widehat{\sigma}(1, \widetilde{\boldsymbol{\tau}})} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \epsilon_i \Big), \\ \boldsymbol{C}_1^{II}(t) &= \frac{1}{\sqrt{n}\widehat{\sigma}(1, \widetilde{\boldsymbol{\tau}})} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \boldsymbol{X}_i^\top (\boldsymbol{\beta}^{(0)} - \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top (\boldsymbol{\beta}^{(0)} - \widehat{\boldsymbol{\beta}}) \Big). \end{aligned} \tag{S7.23}$$

Note that we can regard $\boldsymbol{C}_1^I(t)$ as the leading term of $\boldsymbol{C}_1(t)$ and $\boldsymbol{C}_1^{II}(t)$ as the residual term. Moreover, replacing $\widehat{\sigma}(1, \widetilde{\boldsymbol{\tau}})$ by $\sigma^2 := \mathrm{Var}(\epsilon)$ in $\boldsymbol{C}_1^I(t)$, we can define the oracle leading term as:

$$\widetilde{\boldsymbol{C}}_1^I(t) = \frac{1}{\sqrt{n}\sigma}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \epsilon_i\Big). \tag{S7.24}$$

The following Lemma 12 shows that we can approximate T_1 by $\widetilde{\boldsymbol{C}}_1^I(t)$ in terms of the $(s_0, 2)$-norm. The proof of Lemma 12 is provided in Section S8.1.

Lemma 12. *Assume* **Asssumptions A, B, C.2, E.2-E.4** *hold. Under* $\mathbf{H}_0$*, we have*

$$\mathbb{P}\Big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_1(t) - \widetilde{\boldsymbol{C}}_1^I(t)\big\|_{(s_0,2)} \ge \epsilon\Big) = o(1), \tag{S7.25}$$

where $\epsilon := C s_0^{1/2} s M^2 \log(p)/\sqrt{n}$ *for some big enough universal constant* $C > 0$.

Step 2: Gaussian approximation for the oracle leading term. By Lemma 12, we only need to consider Gaussian approximation for the process $\{\widetilde{\boldsymbol{C}}_1^I(t), q_0 \le t \le 1 - q_0\}$. Recall the bootstrap based CUSUM process for $\alpha = 1$ as:

$$\boldsymbol{C}_1^b(t) = \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i e_i^b - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i e_i^b\Big), \tag{S7.26}$$

where $e_i^b \sim N(0,1)$. By definition, the bootstrap based testing statistic is

$$T_1^b = \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_1^b(t)\|_{(s_0,2)}.$$

Let $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{ip})^\top$ with $Z_{ij} := X_{ij}\epsilon_i/\sigma$ and $\boldsymbol{G}_i = (G_{i1}, \ldots, G_{ip})^\top$ with $\boldsymbol{G}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}_1)$. One can see that $\boldsymbol{G}_i$ has the same covariance matrix as $\boldsymbol{Z}_i$. Define

$$\boldsymbol{C}_1^{\boldsymbol{G}}(t) = \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{G}_i - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n} \boldsymbol{G}_i\Big), \quad \text{and} \quad T_1^{\boldsymbol{G}} = \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\|_{(s_0,2)}.$$

By **Assumptions A, C, E.1**, we can verify that the Conditions **(M1) - (M3)** in Lemma 6 hold. Hence, by Lemma 6, we can prove

$$\sup_{z\in(0,\infty)} \Big|\mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \|\widetilde{\boldsymbol{C}}_1^{\boldsymbol{I}}\|_{(s_0,2)} \le z\big) - \mathbb{P}(T_1^{\boldsymbol{G}} \le z)\Big| \le n^{-\xi_0}, \text{ for some } \xi_0 > 0. \tag{S7.27}$$

Next, we aim to approximate $T_1^b|\mathcal{X}$ by $T_1^{\boldsymbol{G}}$. The result is based on the following Lemma 13.

Lemma 13. *Suppose* **Assumptions A, E.1** *are satisfied. Then, under* $\mathbf{H}_0$*, we have*

$$\sup_{z\in(0,\infty)} \Big|\mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\|_{(s_0,2)} > z\big) - \mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_1^b(t)\|_{(s_0,2)} > z|\mathcal{X}\big)\Big| = o_p(1).$$

Hence, based on Lemma 13, we show that the two Gaussian processes $\boldsymbol{C}_1^{\boldsymbol{G}}(t)$ and $\boldsymbol{C}^b(t)|\mathcal{X}$ with $q_0 \le t \le 1-q_0$ can be uniformly close to each other with the $(s_0,2)$-norm. The proof of Lemma 13 is provided in Section

S8.2.

Step 3: Combining the previous results. In this step, we aim to combine the previous two steps for proving:

$$\sup_{z\in(0,\infty)} \left|\mathbb{P}(T_1 \le z) - \mathbb{P}(T_1^b \le z|\mathcal{X})\right| = o_p(1), \text{ as } n, p \to \infty. \tag{S7.28}$$

In particular, we need to obtain the upper and lower bounds of ρ_0, where

$$\rho_0 := \mathbb{P}(T_1 > z) - \mathbb{P}(T_1^b > z|\mathcal{X}). \tag{S7.29}$$

We first consider the upper bound. Note that $T_1 = \max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_1(t)\|_{(s_0,2)}$. By plugging $\widetilde{\boldsymbol{C}}_1^I(t)$ in T_1 and using the triangle inequality of $\|\cdot\|_{(s_0,2)}$, we have

$$\mathbb{P}(T_1 > z) \le \mathbb{P}(\max_{t\in[q_0,1-q_0]} \|\widetilde{\boldsymbol{C}}_1^I(t)\|_{(s_0,2)} > z - \epsilon) + \rho_1, \tag{S7.30}$$

where $\rho_1 := \mathbb{P}(\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_1(t) - \widetilde{\boldsymbol{C}}_1^I(t)\|_{(s_0,2)} > \epsilon)$. By Lemma 12, we have $\rho_1 = o(1)$. For $\mathbb{P}(\max_{q_0\le t\le 1-q_0} \|\widetilde{\boldsymbol{C}}_1^{\mathrm{I}}(t)\|_{(s_0,2)} \ge z - \epsilon)$, by the triangle inequality, we have

$$\mathbb{P}\Big(\max_{q_0\le t\le 1-q_0} \big\|\widetilde{\boldsymbol{C}}_1^{\mathrm{I}}(t)\big\|_{(s_0,2)} \ge z - \epsilon\Big) \le \mathbb{P}\Big(\max_{q_0\le t\le 1-q_0} \big\|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\big\|_{(s_0,2)} \ge z - \epsilon\Big) + \rho_2, \tag{S7.31}$$

where

$$\rho_2 = \max_{x>0} \left|\mathbb{P}(\max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\|_{(s_0,2)} > x\big) - \mathbb{P}(\max_{q_0\le t\le 1-q_0} \|\widetilde{\boldsymbol{C}}^{\mathrm{I}}(t)\|_{(s_0,2)} > x\big)\right|.$$

By Lemma 6, we have $\rho_2 \leq Cn^{-\zeta_0}$. Therefore, by (S7.30) and (S7.31), we have proved that

$$\mathbb{P}\big(\max_{q_0\leq t\leq 1-q_0}\big\|\boldsymbol{C}_1(t)\big\|_{(s_0,2)} \geq z\big) \leq \underbrace{\mathbb{P}\big(\max_{q_0\leq t\leq 1-q_0}\big\|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\big\|_{(s_0,2)} \geq z-\epsilon\big)}_{\rho_3} + o(1). \tag{S7.32}$$

We next consider ρ_3. We decompose ρ_3 as $\rho_3 = \rho_4 + \rho_5$, where ρ_4 and ρ_5 are defined as

$$\rho_4 = \mathbb{P}\big(z-\epsilon \leq \max_{q_0\leq t\leq 1-q_0}\big\|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\big\|_{(s_0,2)} \leq z\big),\ \ \rho_5 = \mathbb{P}(\max_{q_0\leq t\leq 1-q_0}\|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\|_{(s_0,2)} \geq z).$$

By Lemmas 4 and 5, we can show that $\rho_4 = o(1)$. For ρ_5, we have

$$\mathbb{P}\big(\max_{q_0\leq t\leq 1-q_0}\big\|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\big\|_{(s_0,2)} \geq z\big) \leq \mathbb{P}\big(\max_{q_0\leq t\leq 1-q_0}\big\|\boldsymbol{C}_1^{b}(t)\big\|_{(s_0,2)} \geq z|\mathcal{X}\big)\big) + \rho_6, \tag{S7.33}$$

where

$$\rho_6 = \sup_{z\in(0,\infty)}|\mathbb{P}(\max_{q_0\leq t\leq 1-q_0}\|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\|_{(s_0,2)} > z) - \mathbb{P}(\max_{q_0\leq t\leq 1-q_0}\|\boldsymbol{C}_1^{b}(t)\|_{(s_0,2)} > z|\mathcal{X})|.$$

By Lemma 13, we have $\rho_6 = o_p(1)$. Therefore, by (S7.30) – (S7.33), we have proved

$$\mathbb{P}(T_1 \geq z) - \mathbb{P}(T_1^b \geq z|\mathcal{X}) = o_p(1),$$

uniformly for $z > 0$. Similarly, we can obtain the lower bound and prove that

$$\sup_{z\in(0,\infty)}\big|\mathbb{P}(T_1 \geq z) - \mathbb{P}(T_1^b \geq z|\mathcal{X})\big| = o_p(1),$$

which finishes the proof of Theorem 2 for the individual test with $\alpha = 1$. $\square$

S7.2.2 Gaussian approximation for $\alpha = 0$

Proof. In this section, we give the size results for $\alpha = 0$. Recall $0 < \tau_1 < \cdots < \tau_K < 1$ are user-specified K quantile levels. Let $\widetilde{\boldsymbol{\tau}} := (\tau_1, \ldots, \tau_K)^\top$ and $\boldsymbol{b} = (b_1, \ldots, b_K)^\top$. Note that in this case, our individual based test statistic T_0 reduces to composite quantile loss based score type test statistic. Define the score function as:

$$z(\boldsymbol{x}, y; \widetilde{\boldsymbol{\tau}}, \boldsymbol{b}, \boldsymbol{\beta}) := \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{x}\big(\mathbf{1}\{y - b_k - \boldsymbol{x}^\top \boldsymbol{\beta} \leq 0\} - \tau_k\big), \tag{S7.34}$$

and $Z(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \boldsymbol{b}, \boldsymbol{\beta})$ as its sample version. For $\alpha = 0$, we aim to prove:

$$\sup_{z \in (0,\infty)} \big|\mathbb{P}(T_0 \leq z) - \mathbb{P}(T_0^b \leq z|\mathcal{X})\big| = o_p(1), \text{ as } n, p \to \infty. \tag{S7.35}$$

The proof proceeds into three steps.

Note that for $\alpha = 0$, the score based CUSUM process reduces to:

$$\boldsymbol{C}_0(t) = \frac{1}{\sqrt{n}\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})} \big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}})\big), \tag{S7.36}$$

where $\widehat{\boldsymbol{b}}$ and $\widehat{\boldsymbol{\beta}}$ are the lasso estimators defined in (2.10), and $\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})$ is the variance estimator defined in (2.22). By definition of T_0, we have $T_0 = \max_{q_0 \leq t \leq 1-q_0} \|\boldsymbol{C}_0(t)\|_{(s_0,2)}$. Before the proof, we need some notations. Let $\boldsymbol{\Delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^{(0)} \in \mathbb{R}^p$, $\boldsymbol{\delta} = \boldsymbol{b} - \boldsymbol{b}^{(0)} \in \mathbb{R}^K$, $\delta_k = b_k - b_k^{(0)} \in \mathbb{R}^1$, $\underset{\sim}{\boldsymbol{\Delta}_k} = (\boldsymbol{\Delta}^\top, \delta_k)^\top \in$

$\mathbb{R}^{p+1}$, and $\underset{\sim}{\boldsymbol{\Delta}} = (\boldsymbol{\Delta}^\top, \boldsymbol{\delta}^\top)^\top \in \mathbb{R}^{p+K}$. Accordingly, we define $\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\delta}}$, $\widehat{\delta}_k$, $\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}$, $\underset{\sim}{\widehat{\boldsymbol{\Delta}}}$ by using the corresponding estimators. Moreover, we define $\underset{\sim}{\boldsymbol{X}_i} = (\boldsymbol{X}_i^\top, 1)^\top \in \mathbb{R}^{p+1}$ or $\underset{\sim}{\boldsymbol{X}_i} = (\boldsymbol{X}_i^\top, \mathbf{1}_K)^\top \in \mathbb{R}^{p+K}$ whenever it is used, where $\mathbf{1}_K$ is an $\mathbb{R}^K$ dimensional vector with elements being 1s. By the definition of $Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(0)} + \epsilon_i$, we have $Y_i \leq b_k + \boldsymbol{X}_i^\top \boldsymbol{\beta}$ which is equal to $\epsilon_i \leq \underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\boldsymbol{\Delta}_k} + b_k^{(0)}$. Hence, by replacing $\widehat{\boldsymbol{\beta}}$ by $\boldsymbol{\beta}^{(0)}$ and $\widehat{\boldsymbol{b}}$ by $\boldsymbol{b}^{(0)}$ in $\boldsymbol{C}_0(t)$, we have the following decomposition:

$$\boldsymbol{C}_0(t) = \boldsymbol{C}_0^I(t) + \boldsymbol{C}_0^{II}(t), \tag{S7.37}$$

where $\boldsymbol{C}_0^I(t)$ and $\boldsymbol{C}_0^{II}(t)$ are defined as

$$\begin{aligned}
\boldsymbol{C}_0^I(t) &= \frac{1}{\sqrt{n}\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i e_i(\widetilde{\boldsymbol{\tau}}) - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i e_i(\widetilde{\boldsymbol{\tau}})\Big), \\
\boldsymbol{C}_0^{II}(t) &= \frac{1}{\sqrt{n}\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{X}_i \big(\mathbf{1}\{\epsilon_i \leq \underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k} + b_k^{(0)}\} - \mathbf{1}\{\epsilon_i \leq b_k^{(0)}\}\big) \\
&\quad - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{X}_i \big(\mathbf{1}\{\epsilon_i \leq \underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k} + b_k^{(0)}\} - \mathbf{1}\{\epsilon_i \leq b_k^{(0)}\}\big)\Big),
\end{aligned} \tag{S7.38}$$

where $e_i(\widetilde{\boldsymbol{\tau}}) := \frac{1}{K} \sum_{k=1}^{K} \big(\mathbf{1}\{\epsilon_i \leq b_k^{(0)}\} - \tau_k\big) := \frac{1}{K} \sum_{k=1}^{K} e_i(\tau_k)$ be a random sample satisfying

$$\mathbb{E}[e_i(\widetilde{\boldsymbol{\tau}})] = 0 \quad \text{and} \quad \mathrm{Var}[e_i(\widetilde{\boldsymbol{\tau}})] = \frac{1}{K^2} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \gamma_{k_1 k_2} \tag{S7.39}$$

with $\gamma_{k_1 k_2} := \min(\tau_{k_1}, \tau_{k_2}) - \tau_{k_1}\tau_{k_2}$ for $\tau_{k_1}, \tau_{k_2} \in (0, 1)$. Under this decomposition, we can regard $\boldsymbol{C}_0^I(t)$ as the leading term of $\boldsymbol{C}_0(t)$ and $\boldsymbol{C}_0^{II}(t)$ as the residual term. Moreover, replacing $\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})$ by $\sigma^2 := \mathrm{Var}(e_i(\widetilde{\boldsymbol{\tau}}))$ in $\boldsymbol{C}_0^I(t)$, we can define the oracle leading term as:

$$\widetilde{\boldsymbol{C}}_0^I(t) = \frac{1}{\sqrt{n}\sigma}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i e_i(\widetilde{\boldsymbol{\tau}}) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n} \boldsymbol{X}_i e_i(\widetilde{\boldsymbol{\tau}})\Big). \tag{S7.40}$$

The following Lemma 14 shows that we can approximate T_0 by $\widetilde{\boldsymbol{C}}_1^0(t)$ in terms of $(s_0, 2)$-norm. The proof of Lemma 14 is provided in Section S8.3.

Lemma 14. *Assume* **Assumptions A, D, E.2-E.4** *hold. Under* $\mathbf{H}_0$*, we have*

$$\mathbb{P}\Big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_0(t) - \widetilde{\boldsymbol{C}}_0^I(t)\big\|_{(s_0,2)} \ge \epsilon\Big) = o(1), \tag{S7.41}$$

where $\epsilon := CM^2 s_0^{1/2}(s\log(pn))^{3/4}/n^{1/4}$ *for some big enough universal constant* $C > 0$.

Note that for the case of $\alpha = 0$, the error term $\epsilon_i(\widetilde{\boldsymbol{\tau}})$ is a bounded random variable, which satisfies the assumptions in Lemma 6 trivially. Hence, by Lemma 14, Lemma 6, and using similar arguments of Steps 2 and 3 in Section S7.2.1, we can finish the proof of Theorem 2. □

S7.2.3 Gaussian approximation for $\alpha \in (0, 1)$

Proof. In this section, we give the size results for $\alpha \in (0, 1)$. Recall $0 < \tau_1 < \cdots < \tau_K < 1$ are user-specified K quantile levels. Let $\widetilde{\boldsymbol{\tau}} := (\tau_1, \ldots, \tau_K)^\top$

and $\boldsymbol{b} = (b_1, \dots, b_K)^\top$. For $\alpha \in (0,1)$, define the score function as:

$$z(\boldsymbol{x}, y; \widetilde{\boldsymbol{\tau}}, \boldsymbol{b}, \boldsymbol{\beta}) := (1-\alpha)\frac{1}{K}\sum_{k=1}^{K} \boldsymbol{x}\big(\mathbf{1}\{y - b_k - \boldsymbol{x}^\top \boldsymbol{\beta} \le 0\} - \tau_k\big) - \alpha \boldsymbol{x}(y - \boldsymbol{x}^\top \boldsymbol{\beta}), \tag{S7.42}$$

and $Z(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \boldsymbol{b}, \boldsymbol{\beta})$ as its sample version. For $\alpha \in (0,1)$, we aim to prove:

$$\sup_{z \in (0,\infty)} \left| \mathbb{P}(T_\alpha \le z) - \mathbb{P}(T_\alpha^b \le z | \mathcal{X}) \right| = o_p(1), \text{ as } n, p \to \infty. \tag{S7.43}$$

Note that for $\alpha \in (0,1)$, the score based CUSUM process reduces to:

$$\boldsymbol{C}_\alpha(t) = \frac{1}{\sqrt{n}\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})}\big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}})\big), \tag{S7.44}$$

where $\widehat{\boldsymbol{b}}$ and $\widehat{\boldsymbol{\beta}}$ are the lasso estimators defined in (2.10), and $\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})$ is the variance estimator defined in (2.22). By definition of T_α, we have $T_\alpha = \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_\alpha(t)\|_{(s_0,2)}$. Recall $\boldsymbol{\Delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^{(0)} \in \mathbb{R}^p$, $\boldsymbol{\delta} = \boldsymbol{b} - \boldsymbol{b}^{(0)} \in \mathbb{R}^K$, $\delta_k = b_k - b_k^{(0)} \in \mathbb{R}^1$, $\underset{\sim}{\boldsymbol{\Delta}_k} = (\boldsymbol{\Delta}^\top, \delta_k)^\top \in \mathbb{R}^{p+1}$, and $\underset{\sim}{\boldsymbol{\Delta}} = (\boldsymbol{\Delta}^\top, \boldsymbol{\delta}^\top)^\top \in \mathbb{R}^{p+K}$. Accordingly, recall $\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\delta}}$, $\widehat{\delta}_k$, $\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}$, $\underset{\sim}{\widehat{\boldsymbol{\Delta}}}$ by using the corresponding lasso estimators. Moreover, we define $\underset{\sim}{\boldsymbol{X}_i} = (\boldsymbol{X}_i^\top, 1)^\top \in \mathbb{R}^{p+1}$ or $\underset{\sim}{\boldsymbol{X}_i} = (\boldsymbol{X}_i^\top, \mathbf{1}_K)^\top \in \mathbb{R}^{p+K}$ whenever it is used, where $\mathbf{1}_K$ is an $\mathbb{R}^K$ dimensional vector with elements being 1s. Under $\mathbf{H}_0$, by the definition of $Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(0)} + \epsilon_i$, we have $Y_i \le b_k + \boldsymbol{X}_i^\top \boldsymbol{\beta}$ which is equal to $\epsilon_i \le \underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\boldsymbol{\Delta}_k} + b_k^{(0)}$. Hence, by replacing $\widehat{\boldsymbol{\beta}}$ by $\boldsymbol{\beta}^{(0)}$ and $\widehat{\boldsymbol{b}}$ by $\boldsymbol{b}^{(0)}$ in $\boldsymbol{C}_\alpha(t)$, under $\mathbf{H}_0$, we have the following decomposition:

$$\boldsymbol{C}_\alpha(t) = \boldsymbol{C}_\alpha^I(t) + \boldsymbol{C}_\alpha^{II}(t), \tag{S7.45}$$

where $\boldsymbol{C}_\alpha^I(t)$ and $\boldsymbol{C}_\alpha^{II}(t)$ are defined as

$$\boldsymbol{C}_\alpha^I(t) = \frac{1}{\sqrt{n}\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})}\big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i((1-\alpha)e_i(\widetilde{\boldsymbol{\tau}}) - \alpha\epsilon_i) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n} \boldsymbol{X}_i((1-\alpha)e_i(\widetilde{\boldsymbol{\tau}}) - \alpha\epsilon_i)\big),$$
$$\text{and } \boldsymbol{C}_\alpha^{II}(t) = (1-\alpha)\boldsymbol{C}_0^{II}(t) + \alpha\boldsymbol{C}_1^{II}(t), \tag{S7.46}$$

where $e_i(\widetilde{\boldsymbol{\tau}}) := K^{-1}\sum_{k=1}^{K}\big(\mathbf{1}\{\epsilon_i \leq b_k^{(0)}\} - \tau_k\big) := K^{-1}\sum_{k=1}^{K} e_i(\tau_k)$, $\boldsymbol{C}_1^{II}(t)$ is defined in (S7.23), and $\boldsymbol{C}_0^{II}(t)$ is defined in (S7.38). Under this decomposition, we can regard $\boldsymbol{C}_\alpha^I(t)$ as the leading term of $\boldsymbol{C}_\alpha(t)$ and $\boldsymbol{C}_\alpha^{II}(t)$ as the residual term. Moreover, replacing $\widehat{\sigma}(\alpha, \widetilde{\boldsymbol{\tau}})$ by $\sigma^2 := \mathrm{Var}[(1-\alpha)e_i(\widetilde{\boldsymbol{\tau}}) - \alpha\epsilon_i]$ in $\boldsymbol{C}_\alpha^I(t)$, we can define the oracle leading term as:

$$\widetilde{\boldsymbol{C}}_\alpha^I(t) = \frac{1}{\sqrt{n}\sigma}\big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i((1-\alpha)e_i(\widetilde{\boldsymbol{\tau}}) - \alpha\epsilon_i) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n} \boldsymbol{X}_i((1-\alpha)e_i(\widetilde{\boldsymbol{\tau}}) - \alpha\epsilon_i)\big). \tag{S7.47}$$

The following Lemma 15 shows that we can approximate T_α by $\widetilde{\boldsymbol{C}}_1^\alpha(t)$ in terms of the $(s_0, 2)$-norm. The proof of Lemma 15 is provided in Section S8.4.

Lemma 15. *Assume* **Assumptions A, B, C.2, D, E.2 - E.4** *hold. Under* $\mathbf{H}_0$, *we have*

$$\mathbb{P}\big(\max_{q_0 \leq t \leq 1-q_0} \big\|\boldsymbol{C}_\alpha(t) - \widetilde{\boldsymbol{C}}_\alpha^I(t)\big\|_{(s_0,2)} \geq \epsilon\big) = o(1), \tag{S7.48}$$

where $\epsilon := CM^2 s_0^{1/2}(s\log(pn))^{3/4}/n^{1/4}$ *for some big enough constant* $C > 0$ *and* C *is a universal constant not depending on* n *or* p.

Note that for the case of $\alpha \in (0,1)$, the error term $(1-\alpha)e_i(\widetilde{\boldsymbol{\tau}}) - \alpha\epsilon_i$ is a combination of a bounded random variable $e_i(\widetilde{\boldsymbol{\tau}})$ and ϵ_i, which can be proved to satisfy the assumptions in Lemma 6. Hence, by Lemma 14, Lemma 6, and using similar arguments of Steps 2 and 3 as in Section S7.2.1, we finish the proof of Theorem 2 with $\alpha \in (0,1)$. □

S7.3 Proof of Theorem 3

In this section, we give the change point estimation results for $\alpha = 1$, $\alpha = 0$ and $\alpha \in (0,1)$, respectively. Before the proof, we need some notations. Note that by **Assumption A**, we have $\|\boldsymbol{x}\|_{(s_0,2)} \approx \|\boldsymbol{\Sigma}\boldsymbol{x}\|_{(s_0,2)}$ for any $\boldsymbol{x} \in \mathbb{R}^p$. Hence, for simplicity, we assume $\boldsymbol{\Sigma} = \mathbf{I}$. Moreover, to make a clear result, we assume s_0 is fixed with $s_0 \le s := |\mathcal{S}^{(1)}| \vee |\mathcal{S}^{(2)}|$. Recall $\mathcal{M} = \{j : \beta_j^{(1)} \neq \beta_j^{(2)}\} \subset \{1,\ldots,p\}$ as the set of coordinates having a change point. For any $\boldsymbol{x} \in \mathbb{R}^p$ and the subset $J \subset \{1,\ldots,p\}$, define the projection operator $\Pi_J\boldsymbol{x} \in \mathbb{R}^{|J|}$ being the sub-vector of $\boldsymbol{x}$ with the same coordinates of $\boldsymbol{x}$ on J, e.g., $\Pi_J\boldsymbol{x} := (x'_1,\ldots,x'_{|J|})$ with $x'_j = x_j$ for $j \in J$. Based on the definition of $\|\boldsymbol{x}\|_{(s_0,2)}$, we have $\|\boldsymbol{x}\|_{(s_0,2)} = \max_{J\subset\{1,\ldots,p\},|J|=s_0} \|\Pi_J\boldsymbol{x}\|_2$. In addition, for notational simplicity, we also assume $\lfloor nt \rfloor = nt$ for any

$t \in (0,1)$.

Throughout the following Sections S7.3.1 - S7.3.3, we assume

$$\|\boldsymbol{\Delta}\|_{(s_0,2)} \gg C^* s_0^{1/2} M^2 \sqrt{\frac{\log(pn)}{n}} \text{ and } s_0^{1/2} M^2 \sqrt{\frac{\log(pn)}{n}} = o(1) \quad \text{(S7.49)}$$

for some big enough constant $C^* > 0$.

S7.3.1 Change point estimation for $\alpha = 1$

Proof. Recall $\boldsymbol{Z}_i(\boldsymbol{X}_i, Y_i; \boldsymbol{\beta}) = \boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta})$ is the negative score for $\alpha = 1$. For each $t \in [q_0, 1-q_0]$, define $\widetilde{\boldsymbol{C}}_1(t) = (\widetilde{C}_{11}(t), \ldots, \widetilde{C}_{1p}(t))^\top$ with

$$\widetilde{\boldsymbol{C}}_1(t) = \frac{1}{\sqrt{n}} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{Z}_i(\boldsymbol{X}_i, Y_i; \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{Z}_i(\boldsymbol{X}_i, Y_i; \widehat{\boldsymbol{\beta}}) \Big). \quad \text{(S7.50)}$$

Note that there is no variance estimator in $\widetilde{\boldsymbol{C}}_1(t)$. By definition, we have

$$\widehat{t}_1 := \operatorname*{arg\,max}_{t \in [q_0, 1-q_0]} \|\widetilde{\boldsymbol{C}}_1(t)\|_{(s_0,2)}.$$

Let $\boldsymbol{\Delta} = \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}$ be the signal difference. Moreover, define the estimation error ϵ_n as:

$$\epsilon_n = C(s_0, M, q_0) \frac{\log(pn)}{n \|\boldsymbol{\Delta}\|_{(s_0,2)}^2}. \quad \text{(S7.51)}$$

To prove Theorem 3 with $\alpha = 1$, we need to prove that as $n, p \to \infty$, by choosing a large enough constant $C(s_0, M, q_0)$ in ϵ_n, we have

$$\mathbb{P}\big(|\widehat{t}_1 - t_1| \geq \epsilon_n \big) \to 0. \quad \text{(S7.52)}$$

To that end, we have to prove

$$
\begin{aligned}
&\mathbb{P}\big(|\widehat{t}_1 - t_1| \geq \epsilon_n\big)\\
&\leq \mathbb{P}\big(\widehat{t}_1 \geq t_1 + \epsilon_n\big) + \mathbb{P}\big(\widehat{t}_1 \leq t_1 - \epsilon_n\big)\\
&\leq \mathbb{P}\big(\max_{t \geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_1(t)\|_{(s_0,2)} \geq \|\widetilde{\boldsymbol{C}}_1(t_1)\|_{(s_0,2)}\big) + \mathbb{P}\big(\max_{t \leq t_1-\epsilon_n} \|\widetilde{\boldsymbol{C}}_1(t)\|_{(s_0,2)} \geq \|\widetilde{\boldsymbol{C}}_1(t_1)\|_{(s_0,2)}\big).
\end{aligned} \tag{S7.53}
$$

Hence, to prove $\mathbb{P}\big(|\widehat{t}_1 - t_1| \geq \epsilon_n\big) \to 0$, it is equivalent to prove

$$
\begin{aligned}
&\mathbb{P}\big(\underbrace{\max_{t \geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_1(t)\|_{(s_0,2)} - \|\widetilde{\boldsymbol{C}}_1(t_1)\|_{(s_0,2)} \leq 0}_{A_1}\big)\\
&+\mathbb{P}\big(\underbrace{\max_{t \leq t_1-\epsilon_n} \|\widetilde{\boldsymbol{C}}_1(t)\|_{(s_0,2)} - \|\widetilde{\boldsymbol{C}}_1(t_1)\|_{(s_0,2)} \leq 0}_{A_2}\big) \to 1.
\end{aligned} \tag{S7.54}
$$

Next, we prove $\mathbb{P}(A_1) \to 1$ and $\mathbb{P}(A_2) \to 1$. By the symmetry, we only consider $\mathbb{P}(A_1) \to 1$. Define the two events $\mathcal{H}_1$ and $\mathcal{H}_2$:

$$
\begin{aligned}
\mathcal{H}_1 &= \big\{\max_{t \geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_1(t)\|_{(s_0,2)} := \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \{1,\ldots,p\}\\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t)\|_2 = \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M}\\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t)\|_2\big\},\\
\mathcal{H}_2 &= \big\{\|\widetilde{\boldsymbol{C}}_1(t_1)\|_{(s_0,2)} := \max_{\substack{J \subset \{1,\ldots,p\}\\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t_1)\|_2 = \max_{\substack{J \subset \mathcal{M}\\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t_1)\|_2\big\}.
\end{aligned} \tag{S7.55}
$$

The following Lemma 16 shows that $\mathcal{H}_1$ and $\mathcal{H}_2$ occur with high probability. The proof of Lemma 16 is provided in Section S8.5.

Lemma 16. *Under* **Assumptions A, B, C.2, E.2 - E.4**, *we have*

$$
\mathbb{P}(\mathcal{H}_1) \to 1 \quad \textit{and} \quad \mathbb{P}(\mathcal{H}_2) \to 1. \tag{S7.56}
$$

Now, under $\mathcal{H}_1 \cap \mathcal{H}_2$, we have:

$$\begin{aligned}
\mathbb{P}(A_1) &= \mathbb{P}\big(\max_{t \geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_1(t)\|_{(s_0,2)} - \|\widetilde{\boldsymbol{C}}_1(t_1)\|_{(s_0,2)} \leq 0\big) \\
&= \mathbb{P}\big(\max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t)\|_2 - \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t_1)\|_2 \leq 0\big) \\
&= \mathbb{P}\big(\max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t)\|^2 - \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t_1)\|^2 \leq 0\big).
\end{aligned}$$

Note that under $\mathbf{H}_1$, we have the following decomposition

$$\widetilde{\boldsymbol{C}}_1(t) = \boldsymbol{C}_1^I(t) + \boldsymbol{\delta}(t) + \boldsymbol{R}(t), \tag{S7.57}$$

where $\boldsymbol{C}_1^I(t)$, $\boldsymbol{\delta}(t)$ and $\boldsymbol{R}(t)$ are defined in (S7.99) and (S7.100), respectively.

Similarly, we have

$$\widetilde{\boldsymbol{C}}_1(t_1) = \boldsymbol{C}_1^I(t_1) + \boldsymbol{\delta}(t_1) + \boldsymbol{R}(t_1), \tag{S7.58}$$

by replacing t by t_1. To prove $\mathbb{P}(A_1) \to 1$, we consider $\max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t)\|^2 - \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t_1)\|^2 \leq 0$. By the fact that $\max a_i - \max b_i \leq \max(a_i - b_i)$ for any $\{a_i\}$ and $\{b_i\}$, we have:

$$\begin{aligned}
&\max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t)\|^2 - \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t_1)\|^2 \\
&\leq \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \big(\|\Pi_J\big(\boldsymbol{C}_1^I(t) + \boldsymbol{\delta}(t) + \boldsymbol{R}(t)\big)\|^2 - \|\Pi_J\big(\boldsymbol{C}_1^I(t_1) + \boldsymbol{\delta}(t_1) + \boldsymbol{R}(t_1)\big)\|^2\big) \\
&\leq A_{1.1} + A_{1.2} + A_{1.3} + A_{1.4} + A_{1.5} + A_{1.6},
\end{aligned}$$

where

$$
\begin{aligned}
&A_{1.1} := \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\|\Pi_J \boldsymbol{C}_1^I(t)\|^2 + \|\Pi_J \boldsymbol{C}_1^I(t_1)\|^2\}, \\
&A_{1.2} := \frac{1}{3} \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\|\Pi_J \boldsymbol{\delta}(t)\|^2 - \|\Pi_J \boldsymbol{\delta}(t_1)\|^2\}, \\
&A_{1.3} := \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\|\Pi_J \boldsymbol{R}(t)\|^2 + \|\Pi_J \boldsymbol{R}(t_1)\|^2\}, \\
&A_{1.4} := 2 \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\Pi_J \boldsymbol{C}_1^I(t)^\top \Pi_J \boldsymbol{R}(t) - \Pi_J \boldsymbol{C}_1^I(t_1)^\top \Pi_J \boldsymbol{R}(t_1)\}, \\
&A_{1.5} \\
&:= \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{2(\Pi_J \boldsymbol{C}_1^I(t)^\top \Pi_J \boldsymbol{\delta}(t) - \Pi_J \boldsymbol{C}_1^I(t_1)^\top \Pi_J \boldsymbol{\delta}(t_1)) + \frac{1}{3}(\|\Pi_J \boldsymbol{\delta}(t)\|^2 - \|\Pi_J \boldsymbol{\delta}(t_1)\|^2)\}, \\
&A_{1.6} \\
&:= \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{2(\Pi_J \boldsymbol{\delta}(t)^\top \Pi_J \boldsymbol{R}(t) - \Pi_J \boldsymbol{\delta}(t_1)^\top \Pi_J \boldsymbol{R}(t_1)) + \frac{1}{3}(\|\Pi_J \boldsymbol{\delta}(t)\|^2 - \|\Pi_J \boldsymbol{\delta}(t_1)\|^2)\}.
\end{aligned}
\tag{S7.59}
$$

Our goal is to prove that $\mathbb{P}(A_{1.1} + A_{1.2} + A_{1.3} + A_{1.4} + A_{1.5} + A_{1.6} \leq 0) \to 1$.

Next, we consider $A_{1.1}, \ldots, A_{1.6}$, respectively. For $A_{1.1}$, we have:

$$
\begin{aligned}
A_{1.1} &\leq 2 \max_{q_0 \leq t \leq 1-q_0} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \boldsymbol{C}_1^I(t)\|^2 \\
&\leq 2 \max_{q_0 \leq t \leq 1-q_0} \max_{\substack{J \subset \{1,\ldots,p\} \\ |J|=s_0}} \|\Pi_J \boldsymbol{C}_1^I(t)\|^2 \\
&\leq 2 \max_{q_0 \leq t \leq 1-q_0} (\|\boldsymbol{C}_1^I(t)\|_{(s_0,2)}^2) \\
&\leq 2 \max_{q_0 \leq t \leq 1-q_0} (s_0^{1/2} \|\boldsymbol{C}_1^I(t)\|_\infty)^2 \\
&\leq C s_0 M^2 \log(pn) := C_1(s_0, M) \log(pn),
\end{aligned}
\tag{S7.60}
$$

where the last inequality comes from Lemma 7. Next, we consider $A_{1.2}$. By the definition of $\boldsymbol{\delta}(t)$ and $\boldsymbol{\delta}(t_1)$ as defined in (S7.99), for $t \geq t_1 + \epsilon_n$ and

$J \subset \mathcal{M}$, we have:

$$\begin{aligned} A_{1.2} &=_{(1)} \frac{1}{3} \max_{t \geq t_1 + \epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} \{ \|\Pi_J \boldsymbol{\delta}(t)\|^2 - \|\Pi_J \boldsymbol{\delta}(t_1)\|^2 \} \\ &=_{(2)} \frac{1}{3} \max_{t \geq t_1 + \epsilon_n} \left(n t_1^2 (t_1 - t)(2 - t - t_1) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2 \right) \\ &=_{(3)} -\frac{1}{3} n \epsilon_n t_1^2 (2 - 2t_1 - \epsilon_n) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2 \\ &\leq_{(4)} -\frac{1}{6} q_0 n \epsilon_n \|\boldsymbol{\Delta}\|_{(s_0,2)}^2, \end{aligned} \tag{S7.61}$$

where the last inequality comes from $t_1 \in [q_0, 1 - q_0]$, and $\epsilon_n = o(1)$. For $A_{1.3}$, by the definition of $\boldsymbol{R}(t)$ and $\boldsymbol{R}(t_1)$, and using Lemmas 7 and 9, we have:

$$A_{1.3} \leq C s_0 s^2 M^2 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2 := C_3(s_0, M) s^2 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2. \tag{S7.62}$$

Next, we consider $A.14$. By the Cauchy-Swartz inequality, we have:

$$\begin{aligned} A_{1.4} &= 2 \max_{t \geq t_1 + \epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} \{ \Pi_J \widetilde{\boldsymbol{C}}_1^I(t)^\top \Pi_J \boldsymbol{R}_1(t) - \Pi_J \widetilde{\boldsymbol{C}}_1^I(t_1)^\top \Pi_J \boldsymbol{R}_1(t_1) \} \\ &\leq_{(1)} 4 \max_{t \in [q_0, 1 - q_0]} \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} |\{ \Pi_J \widetilde{\boldsymbol{C}}_1^I(t)^\top \Pi_J \boldsymbol{R}_1(t)| \\ &\leq_{(2)} 4 \max_{t \in [q_0, 1 - q_0]} \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} \|\{ \Pi_J \widetilde{\boldsymbol{C}}_1^I(t)\|_2 \|\Pi_J \boldsymbol{R}_1(t)\|_2 \\ &\leq_{(3)} 4 \max_{t \in [q_0, 1 - q_0]} \|\widetilde{\boldsymbol{C}}_1^I(t)\|_{(s_0,2)} \times \max_{t \in [q_0, 1 - q_0]} \|\boldsymbol{R}_1(t)\|_{(s_0,2)} \\ &\leq_{(4)} C s_0^{1/2} M \sqrt{\log(pn)} \times s_0^{1/2} \sqrt{n} M^2 \sqrt{\frac{\log(pn)}{n}} s \|\boldsymbol{\Delta}\|_{(s_0,2)} \\ &\leq_{(5)} C s_0 s M^3 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)} := C_4(s_0, M) s \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}, \end{aligned} \tag{S7.63}$$

where (4) comes from Lemma 7 and Lemma 8. Hence, combining (S7.61) -

(S7.62), if ϵ_n satisfies

$$\epsilon_n = C \max\Big\{ \underbrace{C_1(s_0, M)\frac{\log(pn)}{n\|\boldsymbol{\Delta}\|^2_{(s_0,2)}}}_{\text{by } A_{1.1}}, \underbrace{C_3(s_0, M)\frac{s^2\log(pn)}{n}}_{\text{by } A_{1.3}}, \underbrace{\frac{C_4(s_0, M)s\log(pn)}{n\|\boldsymbol{\Delta}\|_{(s_0,2)}}}_{\text{by } A_{1.4}} \Big\} \tag{S7.64}$$

for some big enough constant $C > 0$, with probability tending to one, we have $A_{1.1} + A_{1.2} + A_{1.3} + A_{1.4} \le 0$.

Next, we prove $A_{1.5} + A_{1.6} \le 0$. For $A_{1.5}$, using the triangle inequality, we have:

$$\begin{aligned} A_{1.5} &= \max_{t \ge t_1+\epsilon_n} \max_{\substack{J\subset\mathcal{M} \\ |J|=s_0}} \Big\{ 2(\Pi_J \boldsymbol{C}_1^I(t)^\top \Pi_J \boldsymbol{\delta}(t) - \Pi_J \boldsymbol{C}_1^I(t_1)^\top \Pi_J \boldsymbol{\delta}(t_1)) \\ &\quad - \frac{1}{3}(\|\Pi_J \boldsymbol{\delta}_(t_1)\|^2 - \|\Pi_J \boldsymbol{\delta}(t)\|^2) \Big\} \\ &= A_{1.5.1} + A_{1.5.2}, \end{aligned} \tag{S7.65}$$

where

$$\begin{aligned} &A_{1.5.1} \\ &= \max_{t \ge t_1+\epsilon_n} \max_{\substack{J\subset\mathcal{M} \\ |J|=s_0}} \Big\{ 2\Pi_J \boldsymbol{C}_1^I(t)^\top (\Pi_J \boldsymbol{\delta}(t) - \Pi_J \boldsymbol{\delta}(t_1)) \} - \frac{1}{6}(\|\Pi_J \boldsymbol{\delta}(t_1)\|^2 - \|\Pi_J \boldsymbol{\delta}(t)\|^2) \Big\}, \\ &A_{1.5.2} \\ &= \max_{t \ge t_1+\epsilon_n} \max_{\substack{J\subset\mathcal{M} \\ |J|=s_0}} \Big\{ 2\Pi_J \boldsymbol{\delta}(t_1)^\top (\Pi_J \boldsymbol{C}_1^I(t) - \Pi_J \boldsymbol{C}_1^I(t_1)) - \frac{1}{6}(\|\Pi_J \boldsymbol{\delta}(t_1)\|^2 - \|\Pi_J \boldsymbol{\delta}(t)\|^2) \Big\}. \end{aligned} \tag{S7.66}$$

To bound $A_{1.5}$, we prove $\mathbb{P}(A_{1.5.1} \le 0) \to 1$ and $\mathbb{P}(A_{1.5.2} \le 0) \to 1$, respectively. To bound $A_{1.5.1}$, note that for any fixed $t \ge t_1 + \epsilon_n$ and $J \subset \mathcal{M}$ with

$|J| = s_0$, we have:

$$
\begin{aligned}
&2\Pi_J \boldsymbol{C}_1^I(t)^\top \Pi_J(\boldsymbol{\delta}(t) - \boldsymbol{\delta}_1(t_1)) - \frac{1}{6}(\|\Pi_J\boldsymbol{\delta}(t_1)\|^2 - \|\Pi_J\boldsymbol{\delta}(t)\|^2) \\
&\leq_{(1)} 2\|\Pi_J \boldsymbol{C}_1^I(t)\|_2 \|\Pi_J(\boldsymbol{\delta}(t) - \boldsymbol{\delta}(t_1))\|_2 - \frac{1}{6}\big(nt_1^2(t-t_1)(2-t-t_1)\|\Pi_J\boldsymbol{\Delta}\|^2\big) \\
&\leq_{(2)} 2s_0^{1/2}\|\Pi_J \boldsymbol{C}_1^I(t)\|_\infty \sqrt{n}t_1(t-t_1))\|\boldsymbol{\Delta}\|_{(s_0,2)} - \frac{1}{6}\big(nt_1^2(t-t_1)(1+q_0-t_1)\|\boldsymbol{\Delta}\|^2_{(s_0,2)}\big) \\
&\leq_{(3)} 2s_0^{1/2}\|\boldsymbol{C}_1^I(t)\|_\infty \sqrt{n}t_1(t-t_1))\|\boldsymbol{\Delta}\|_{(s_0,2)} - \frac{1}{6}\big(nt_1^2(t-t_1)(1+q_0-t_1)\|\boldsymbol{\Delta}\|^2_{(s_0,2)}\big).
\end{aligned}
\tag{S7.67}
$$

Hence, by (S7.67), to prove $\mathbb{P}(A_{1.5.1} \leq 0) \to 1$, it is sufficient to prove that

$$
\mathbb{P}\Big(\max_{t\geq t_1+\epsilon_n} \max_{\substack{J\subset\mathcal{M}\\|J|=s_0}} \{2s_0^{1/2}\|\boldsymbol{C}_1^I(t)\|_\infty\sqrt{n}t_1(t-t_1))\|\boldsymbol{\Delta}\|_{(s_0,2)} - \frac{1}{6}\big(nt_1^2(t-t_1)(1+q_0-t_1)\|\boldsymbol{\Delta}\|^2_{(s_0,2)}\big)\} \leq 0\Big) \to 1.
$$

Equivalently, it is sufficient to prove that

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{t\geq t_1+\epsilon_n} \max_{\substack{J\subset\mathcal{M}\\|J|=s_0}} \{2s_0^{1/2}\|\boldsymbol{C}_1^I(t)\|_\infty\sqrt{n}t_1(t-t_1))\|\boldsymbol{\Delta}\|_{(s_0,2)} \\
&\qquad -\frac{1}{6}\big(nt_1^2(t-t_1)(1+q_0-t_1)\|\boldsymbol{\Delta}\|^2_{(s_0,2)}\big)\} \leq 0\Big) \\
&\geq \mathbb{P}\Big(\max_{q_0\leq t\leq 1-q_0} 2s_0^{1/2}\|\boldsymbol{C}_1^I(t)\|_\infty t_1 - \frac{1}{6}\big(\sqrt{n}t_1^2(1+q_0-t_1)\|\boldsymbol{\Delta}\|_{(s_0,2)}\big) \leq 0\Big) \to 1.
\end{aligned}
$$

Note that by Lemma 7, we have $\max_{q_0\leq t\leq 1-q_0}\{2s_0^{1/2}\|\widetilde{\boldsymbol{C}}_1^I(t)\|_\infty = O_p(s_0^{1/2}M\sqrt{\log(pn)})$.

Moreover, if we choose a big enough constant C^* in (S7.49), we have

$$
\mathbb{P}\Big(\max_{q_0\leq t\leq 1-q_0} 2s_0^{1/2}\|\widetilde{\boldsymbol{C}}_1^I(t)\|_\infty t_1 - \frac{1}{6}\big(\sqrt{n}t_1^2(1+q_0-t_1)\|\boldsymbol{\Delta}\|_{(s_0,2)}\big) \leq 0\Big) \to 1,
$$

which yields $\mathbb{P}(A_{1.5.1}) \to 1$. After bounding $A_{1.5.1}$, we next consider $A_{1.5.2}$.

Note that for any fixed $t \geq t_1 + \epsilon_n$ and $J \subset \mathcal{M}$ with $|J| = s_0$, we have:

$$
\begin{aligned}
&2\Pi_J\boldsymbol{\delta}(t_1)^\top \Pi_J(\boldsymbol{C}_1^I(t) - \boldsymbol{C}_1^I(t_1)) - \frac{1}{6}(\|\Pi_J\boldsymbol{\delta}(t_1)\|^2 - \|\Pi_J\boldsymbol{\delta}(t)\|^2) \\
&\leq_{(1)} 2\|\Pi_J\boldsymbol{\delta}(t_1)\|_2\|\Pi_J(\boldsymbol{C}_1^I(t) - \boldsymbol{C}_1^I(t_1))\|_2 - \frac{1}{6}\big(nt_1^2(t-t_1)(2-t-t_1)\|\Pi_J\boldsymbol{\Delta}\|^2 \\
&\leq 2\sqrt{n}t_1(1-t_1)\|\boldsymbol{\Delta}\|_{(s_0,2)}s_0^{1/2}\|\widetilde{\boldsymbol{C}}_1^I(t) - \widetilde{\boldsymbol{C}}_1^I(t_1)\|_\infty - \frac{1}{6}\big(nt_1^2(t-t_1)(1+q_0-t_1)\|\boldsymbol{\Delta}\|_{(s_0,2)}^2.
\end{aligned} \tag{S7.68}
$$

Note that by the definition of $\boldsymbol{C}_1^I(t)$ and $\boldsymbol{C}_1^I(t_1)$, we have:

$$
\boldsymbol{C}_1^I(t) - \boldsymbol{C}_1^I(t_1) = \frac{1}{\sqrt{n}}\Big(\sum_{i=\lfloor nt_1\rfloor+1}^{\lfloor nt\rfloor}\boldsymbol{X}_i\epsilon_i - \frac{\lfloor nt\rfloor - \lfloor nt_1\rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\epsilon_i\Big). \tag{S7.69}
$$

Hence, combining (S7.68) and (S7.69), we have:

$$
2\Pi_J\boldsymbol{\delta}(t_1)^\top \Pi_J(\boldsymbol{C}_1^I(t) - \boldsymbol{C}_1^I(t_1)) - \frac{1}{6}(\|\Pi_J\boldsymbol{\delta}(t_1)\|^2 - \|\Pi_J\boldsymbol{\delta}(t)\|^2) \leq A_{1.5.2}^I + A_{1.5.2}^{II},
$$

where

$$
\begin{aligned}
A_{1.5.2}^I &= 2t_1(1-t_1)\|\boldsymbol{\Delta}\|_{(s_0,2)}s_0^{1/2}\|\sum_{i=\lfloor nt_1\rfloor+1}^{\lfloor nt\rfloor}\boldsymbol{X}_i\epsilon_i\|_\infty \\
&\qquad -\frac{1}{12}\big(nt_1^2(t-t_1)(1+q_0-t_1)\|\boldsymbol{\Delta}\|_{(s_0,2)}^2, \\
A_{1.5.2}^{II} &= 2t_1(1-t_1)\|\boldsymbol{\Delta}\|_{(s_0,2)}s_0^{1/2}\|\frac{\lfloor nt\rfloor - \lfloor nt_1\rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\epsilon_i\|_\infty \\
&\qquad -\frac{1}{12}\big(nt_1^2(t-t_1)(1+q_0-t_1)\|\boldsymbol{\Delta}\|_{(s_0,2)}^2.
\end{aligned} \tag{S7.70}
$$

Considering (S7.66), (S7.68), (S7.69), and (S7.70), to prove $\mathbb{P}(A_{1.5.2}) \to 1$, it is sufficient to prove $\mathbb{P}(\max_t \max_J A_{1.5.2}^I \leq 0) \to 1$ and $\mathbb{P}(\max_t \max_J A_{1.5.2}^{II} \leq$

$0) \to 1$. For $A^{I}_{1.5.2}$, we have to prove

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{t \geq t_1 + \epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} \Big\{ 2t_1(1-t_1)\|\boldsymbol{\Delta}\|_{(s_0,2)} s_0^{1/2} \| \sum_{i=\lfloor nt_1 \rfloor+1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \epsilon_i \|_\infty \\
&\qquad - \frac{1}{6}\big(n t_1^2 (t - t_1)(1 + q_0 - t_1) \|\boldsymbol{\Delta}\|^2_{(s_0,2)} \Big\} \leq 0 \Big) \\
&= \mathbb{P}\Big(\max_{t \geq t_1 + \epsilon_n} \| \frac{1}{\lfloor nt \rfloor - \lfloor nt_1 \rfloor} \sum_{i=\lfloor nt_1 \rfloor+1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \epsilon_i \|_\infty \leq C(t_1, q_0) s_0^{-1/2} \|\boldsymbol{\Delta}\|_{(s_0,2)} \Big) \to 1.
\end{aligned}
\tag{S7.71}
$$

Note that by Lemma 7, we have

$$
\max_{t \geq t_1 + \epsilon_n} \| \frac{1}{\lfloor nt \rfloor - \lfloor nt_1 \rfloor} \sum_{i=\lfloor nt_1 \rfloor+1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \epsilon_i \|_\infty = O_p(M n^{1/4} \frac{\log(pn)}{n \epsilon_n}).
$$

Hence, if we choose

$$
\epsilon_n = C_5(s_0, M) \frac{\log(pn)}{n^{3/4} \|\boldsymbol{\Delta}\|_{(s_0,2)}}
$$

for some big enough constant $C_5(s_0, M) > 0$, we have (S7.71) holds, which yields $\mathbb{P}(\max_t \max_J A^{I}_{1.5.2} \leq 0) \to 1$. Similarly, we can prove $\mathbb{P}(\max_t \max_J A^{II}_{1.5.2} \leq 0) \to 1$, which yields $\mathbb{P}(A_{1.5.2} \leq 0) \to 1$.

With a very similar proof technique, if we choose $\epsilon_n = C_6(s_0, M) \dfrac{s^2 \log(pn)}{n}$ for some big enough constant $C_6 > 0$, we can prove

$$
\mathbb{P}(A_{1.6} \leq 0) \to 1.
$$

Combining the previous results, if ϵ_n satisfies

$$\begin{aligned}\epsilon_n = C\max\Big\{ & \underbrace{C_1(s_0,M)\frac{\log(pn)}{n\|\mathbf{\Delta}\|_{(s_0,2)}^2}}_{\text{by A}_{1.1}}, \underbrace{\frac{C_3(s_0,M)s^2\log(pn)}{n}}_{\text{by A}_{1.3}}, \underbrace{\frac{C_4(s_0,M)s\log(pn)}{n\|\mathbf{\Delta}\|_{(s_0,2)}}}_{\text{by A}_{1.4}}, \\ & \underbrace{\frac{C_5(s_0,M)\log(pn)}{n^{3/4}\|\mathbf{\Delta}\|_{(s_0,2)}}}_{\text{by A}_{1.5}}, \underbrace{C_6(s_0,M)\frac{s^2\log(pn)}{n}}_{\text{by A}_{1.6}}\Big\} \\ = C(s_0,M) & \frac{\log(pn)}{n\|\mathbf{\Delta}\|_{(s_0,2)}^2} \times \max\Big\{1, s^2\|\mathbf{\Delta}\|_{(s_0,2)}^2, s\|\mathbf{\Delta}\|_{(s_0,2)}, n^{1/4}\|\mathbf{\Delta}\|_{(s_0,2)}\Big\}.\end{aligned} \tag{S7.72}$$

we can prove $\mathbb{P}(A_1) \to 1$. By symmetry, we can prove $\mathbb{P}(A_2) \to 1$, which finishes the proof.

Lastly, we need to discuss the five terms in (S7.72). Note that by Assumption F and the assumption that $\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1 \leq C_{\mathbf{\Delta}}$, we have $s^2\|\mathbf{\Delta}\|_{(s_0,2)}^2 = O(1)$ and $s\|\mathbf{\Delta}\|_{(s_0,2)} = O(1)$. Moreover, by the assumption that $n^{1/4} = o(s)$, we have $n^{1/4}\|\mathbf{\Delta}\|_{(s_0,2)} = o(1)$, which finishes the proof.

□

S7.3.2 Change point estimation for $\alpha = 0$

Proof. For $\alpha = 0$, recall $\boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) := \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{X}_i\big(\mathbf{1}\{Y_i - \widehat{b}_k - \boldsymbol{X}_i^\top\widehat{\boldsymbol{\beta}} \leq 0\} - \tau_k\big)$ as the score function. For each $t \in [q_0, 1-q_0]$, define $\widetilde{\boldsymbol{C}}_0(t) = (\widetilde{C}_{01}(t), \ldots, \widetilde{C}_{0p}(t))^\top$ with

$$\widetilde{\boldsymbol{C}}_0(t) = \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{\lfloor nt\rfloor} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt\rfloor}{n}\sum_{i=1}^{n} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}})\Big). \tag{S7.73}$$

Note that there is no variance estimator in $\widetilde{\boldsymbol{C}}_0(t)$. Recall $\widehat{t}_0 := \arg\max_{t\in[q_0,1-q_0]} \|\widetilde{\boldsymbol{C}}_0(t)\|_{(s_0,2)}$. To prove Theorem 3 with $\alpha = 0$, we need to prove that as $n, p \to \infty$, by choosing a large enough constant $C(s_0, M, \widetilde{\boldsymbol{\tau}})$ in ϵ_n (which will be given in (S7.84)), we have

$$\mathbb{P}\big(|\widehat{t}_0 - t_1| \geq \epsilon_n\big) \to 0. \tag{S7.74}$$

Similar to Section S7.3.1, we have to prove $\mathbb{P}(A_1) \to 1$ and $\mathbb{P}(A_2) \to 1$, where

$$\begin{aligned} A_1 &= \max_{t\geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_0(t)\|_{(s_0,2)} - \|\widetilde{\boldsymbol{C}}_0(t_1)\|_{(s_0,2)} \leq 0, \\ A_2 &= \max_{t\leq t_1-\epsilon_n} \|\widetilde{\boldsymbol{C}}_0(t)\|_{(s_0,2)} - \|\widetilde{\boldsymbol{C}}_0(t_1)\|_{(s_0,2)} \leq 0. \end{aligned} \tag{S7.75}$$

By the symmetry, we only consider $\mathbb{P}(A_1) \to 1$. Define the two events $\mathcal{H}_1$ and $\mathcal{H}_2$:

$$\begin{aligned} \mathcal{H}_1 &= \Big\{ \max_{t\geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_0(t)\|_{(s_0,2)} := \max_{t\geq t_1+\epsilon_n} \max_{\substack{J\subset\{1,\dots,p\}\\|J|=s_0}} \|\Pi_J\widetilde{\boldsymbol{C}}_0(t)\|_2 = \max_{t\geq t_1+\epsilon_n} \max_{\substack{J\subset\mathcal{M}\\|J|=s_0}} \|\Pi_J\widetilde{\boldsymbol{C}}_0(t)\|_2 \Big\}, \\ \mathcal{H}_2 &= \Big\{ \|\widetilde{\boldsymbol{C}}_0(t_1)\|_{(s_0,2)} := \max_{\substack{J\subset\{1,\dots,p\}\\|J|=s_0}} \|\Pi_J\widetilde{\boldsymbol{C}}_0(t_1)\|_2 = \max_{\substack{J\subset\mathcal{M}\\|J|=s_0}} \|\Pi_J\widetilde{\boldsymbol{C}}_0(t_1)\|_2 \Big\}. \end{aligned} \tag{S7.76}$$

Similar to the proof of Lemma 16, we can prove $\mathbb{P}(\mathcal{H}_1 \cap \mathcal{H}_2) \to 1$. Now, under $\mathcal{H}_1 \cap \mathcal{H}_2$, we have:

$$\begin{aligned} \mathbb{P}(A_1) &= \mathbb{P}\big(\max_{t\geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_0(t)\|_{(s_0,2)} - \|\widetilde{\boldsymbol{C}}_0(t_1)\|_{(s_0,2)} \leq 0 \big) \\ &= \mathbb{P}\big(\max_{t\geq t_1+\epsilon_n} \max_{\substack{J\subset\mathcal{M}\\|J|=s_0}} \|\Pi_J\widetilde{\boldsymbol{C}}_0(t)\|_2 - \max_{\substack{J\subset\mathcal{M}\\|J|=s_0}} \|\Pi_J\widetilde{\boldsymbol{C}}_0(t_1)\|_2 \leq 0 \big) \\ &= \mathbb{P}\big(\max_{t\geq t_1+\epsilon_n} \max_{\substack{J\subset\mathcal{M}\\|J|=s_0}} \|\Pi_J\widetilde{\boldsymbol{C}}_0(t)\|^2 - \max_{\substack{J\subset\mathcal{M}\\|J|=s_0}} \|\Pi_J\widetilde{\boldsymbol{C}}_0(t_1)\|^2 \leq 0 \big). \end{aligned}$$

Recall $\sigma^2(0,\widetilde{\boldsymbol{\tau}}) = \sqrt{K^{-2}\sum_{k_1,k_2}\gamma_{k_1k_2}}$, with $\gamma_{k_1k_2} := \min(\tau_{k_1},\tau_{k_2}) - \tau_{k_1}\tau_{k_2}$ and $\boldsymbol{C}_0(t)$ defined in (S7.106). Then, under $\mathbf{H}_1$, we have the following decomposition

$$\begin{aligned}\widetilde{\boldsymbol{C}}_0(t) &= \sigma(0,\widetilde{\boldsymbol{\tau}}) \times \boldsymbol{C}_0(t)\\ &= \sigma(0,\widetilde{\boldsymbol{\tau}}) \times \big(- SNR(0,\widetilde{\boldsymbol{\tau}}) \times \boldsymbol{\delta}(t) + \boldsymbol{C}_0^{(1)}(t) - SNR(0,\widetilde{\boldsymbol{\tau}}) \times \boldsymbol{R}(t) + \boldsymbol{C}_0^{(3)}(t) + \boldsymbol{C}_0^{(4)}(t)\big),\end{aligned} \tag{S7.77}$$

where the second equation comes from the decomposition in (S7.116). By the fact that $\max a_i - \max b_i \le \max(a_i - b_i)$ and $\max(a_i + b_i) \le \max a_i + \max b_i$ for any $\{a_i\}$ and $\{b_i\}$, we have:

$$\begin{aligned}&\max_{t\ge t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_0(t)\|^2 - \max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_0(t_1)\|^2\\ &\le \sigma^2(0,\widetilde{\boldsymbol{\tau}}) \times \max_{t\ge t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\Big(\|\Pi_J\big(- SNR(0,\widetilde{\boldsymbol{\tau}}) \times \boldsymbol{\delta}(t) + \boldsymbol{C}_0^{(1)}(t)\\ &\qquad\qquad -SNR(0,\widetilde{\boldsymbol{\tau}}) \times \boldsymbol{R}(t) + \boldsymbol{C}_0^{(3)}(t) + \boldsymbol{C}_0^{(4)}(t)\big)\|^2\\ &\qquad -\|\Pi_J\big(- SNR(0,\widetilde{\boldsymbol{\tau}}) \times \boldsymbol{\delta}(t_1) + \boldsymbol{C}_0^{(1)}(t_1) - SNR(0,\widetilde{\boldsymbol{\tau}}) \times \boldsymbol{R}(t_1) + \boldsymbol{C}_0^{(3)}(t_1) + \boldsymbol{C}_0^{(4)}(t_1)\big)\|^2\Big)\\ &\le \sigma^2(0,\widetilde{\boldsymbol{\tau}}) \times (A_{1.1} + \cdots + A_{1.15}),\end{aligned}$$

where the fifteen parts $A_{1.1} \cdots A_{1.15}$ are defined as:

$$A_{1.1} := \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\|\Pi_J \boldsymbol{C}_0^{(1)}(t)\|^2 + \|\Pi_J \boldsymbol{C}_0^{(1)}(t_1)\|^2\},$$

$$A_{1.2} := \frac{SNR^2(0,\widetilde{\boldsymbol{\tau}})}{5} \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\|\Pi_J \boldsymbol{\delta}(t)\|^2 - \|\Pi_J \boldsymbol{\delta}(t_1)\|^2\},$$

$$A_{1.3} := SNR^2(0,\widetilde{\boldsymbol{\tau}}) \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\|\Pi_J \boldsymbol{R}(t)\|^2 + \|\Pi_J \boldsymbol{R}(t_1)\|^2\},$$

$$A_{1.4} := 2SNR(0,\widetilde{\boldsymbol{\tau}}) \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{-\Pi_J \boldsymbol{C}_0^{(1)}(t)^\top \Pi_J \boldsymbol{R}(t) + \Pi_J \boldsymbol{C}_0^{(1)}(t_1)^\top \Pi_J \boldsymbol{R}(t_1)\},$$

$$A_{1.5} := 2SNR(0,\widetilde{\boldsymbol{\tau}}) \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \Big\{ - \Pi_J \boldsymbol{C}_0^{(1)}(t)^\top \Pi_J \boldsymbol{\delta}(t) + \Pi_J \boldsymbol{C}_0^{(1)}(t_1)^\top \Pi_J \boldsymbol{\delta}(t_1) + \frac{SNR^2(0,\widetilde{\boldsymbol{\tau}})}{5} (\|\Pi_J \boldsymbol{\delta}(t)\|^2 - \|\Pi_J \boldsymbol{\delta}(t_1)\|^2) \Big\},$$

$$A_{1.6} := 2SNR^2(0,\widetilde{\boldsymbol{\tau}}) \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \Big\{ \Pi_J \boldsymbol{R}(t)^\top \Pi_J \boldsymbol{\delta}(t) - \Pi_J \boldsymbol{R}(t_1)^\top \Pi_J \boldsymbol{\delta}(t_1) + \frac{SNR^2(0,\widetilde{\boldsymbol{\tau}})}{5} (\|\Pi_J \boldsymbol{\delta}(t)\|^2 - \|\Pi_J \boldsymbol{\delta}(t_1)\|^2) \Big\},$$

$$A_{1.7} := \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\|\Pi_J \boldsymbol{C}_0^{(3)}(t)\|^2 + \|\Pi_J \boldsymbol{C}_0^{(3)}(t_1)\|^2\},$$

$$A_{1.8} := \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\|\Pi_J \boldsymbol{C}_0^{(4)}(t)\|^2 + \|\Pi_J \boldsymbol{C}_0^{(4)}(t_1)\|^2\},$$

$$A_{1.9} := 2 \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\Pi_J \boldsymbol{C}_0^{(1)}(t)^\top \Pi_J \boldsymbol{C}_0^{(3)}(t) - \Pi_J \boldsymbol{C}_0^{(1)}(t_1)^\top \Pi_J \boldsymbol{C}_0^{(3)}(t_1)\},$$

$$A_{1.10} := 2 \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\Pi_J \boldsymbol{C}_0^{(1)}(t)^\top \Pi_J \boldsymbol{C}_0^{(4)}(t) - \Pi_J \boldsymbol{C}_0^{(1)}(t_1)^\top \Pi_J \boldsymbol{C}_0^{(4)}(t_1)\},$$

$$A_{1.11} := 2SNR(0,\widetilde{\boldsymbol{\tau}}) \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{-\Pi_J \boldsymbol{R}(t)^\top \Pi_J \boldsymbol{C}_0^{(3)}(t) + \Pi_J \boldsymbol{R}(t_1)^\top \Pi_J \boldsymbol{C}_0^{(3)}(t_1)\},$$

$$A_{1.12} := 2SNR(0,\widetilde{\boldsymbol{\tau}}) \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{-\Pi_J \boldsymbol{R}(t)^\top \Pi_J \boldsymbol{C}_0^{(4)}(t) + \Pi_J \boldsymbol{R}(t_1)^\top \Pi_J \boldsymbol{C}_0^{(4)}(t_1)\},$$

$$A_{1.13} := 2 \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\Pi_J \boldsymbol{C}_0^{(3)}(t)^\top \Pi_J \boldsymbol{C}_0^{(4)}(t) - \Pi_J \boldsymbol{C}_0^{(3)}(t_1)^\top \Pi_J \boldsymbol{C}_0^{(4)}(t_1)\},$$

$$A_{1.14} := 2SNR(0,\widetilde{\boldsymbol{\tau}}) \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \Big\{ - \Pi_J \boldsymbol{C}_0^{(3)}(t)^\top \Pi_J \boldsymbol{\delta}(t) + \Pi_J \boldsymbol{C}_0^{(3)}(t_1)^\top \Pi_J \boldsymbol{\delta}(t_1) + \frac{SNR^2(0,\widetilde{\boldsymbol{\tau}})}{5} (\|\Pi_J \boldsymbol{\delta}(t)\|^2 - \|\Pi_J \boldsymbol{\delta}(t_1)\|^2) \Big\}$$

$$A_{1.15} := 2SNR(0,\widetilde{\boldsymbol{\tau}}) \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \Big\{ - \Pi_J \boldsymbol{C}_0^{(4)}(t)^\top \Pi_J \boldsymbol{\delta}(t) + \Pi_J \boldsymbol{C}_0^{(4)}(t_1)^\top \Pi_J \boldsymbol{\delta}(t_1) + \frac{SNR^2(0,\widetilde{\boldsymbol{\tau}})}{5} (\|\Pi_J \boldsymbol{\delta}(t)\|^2 - \|\Pi_J \boldsymbol{\delta}(t_1)\|^2) \Big\}.$$

Next, we aim to prove that $\mathbb{P}(A_{1.1} + \cdots + A_{1.15} \leq 0) \to 1$. The proof proceeds into five steps: **Step 1:** We aim to prove that, with probability tending to 1,

$$A_{1.1}+A_{1.2}+A_{1.3}+A_{1.4}+A_{1.7}+A_{1.8}+A_{1.9}+A_{1.10}+A_{1.11}+A_{1.12}+A_{1.13} \leq 0.$$

The main idea of step 1 is to obtain the upper bound for each item. Note that similar to the proofs in Section S7.3.1, we can directly prove that:

$$\begin{aligned}
A_{1.1} &\leq C_1(s_0, M, \widetilde{\boldsymbol{\tau}}) \log(pn), \\
A_{1.2} &\leq -\frac{SNR^2(0, \widetilde{\boldsymbol{\tau}})}{10} q_0 n \epsilon_n \|\boldsymbol{\Delta}\|_{(s_0,2)}^2, \\
A_{1.3} &\leq C_3(s_0, M, \widetilde{\boldsymbol{\tau}}) SNR^2(0, \widetilde{\boldsymbol{\tau}}) s^2 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2, \\
A_{1.4} &\leq C_4(s_0, M, \widetilde{\boldsymbol{\tau}}) SNR(0, \widetilde{\boldsymbol{\tau}}) s \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}.
\end{aligned}$$

For $A_{1.7}$, by (S7.121), we have:

$$\begin{aligned}
A_{1.7} &:= \max_{t \geq t_1 + \epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} \{\|\Pi_J \boldsymbol{C}_0^{(3)}(t)\|^2 + \|\Pi_J \boldsymbol{C}_0^{(3)}(t_1)\|^2\} \\
&\leq 2 \max_{q_0 \leq t \leq 1 - q_0} \|\boldsymbol{C}_0^{(3)}(t)\|_{(s_0,2)}^2 \\
&\leq C_7(s_0, M, \widetilde{\boldsymbol{\tau}}) s^4 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}^4.
\end{aligned}$$

For $A_{1.8}$, by (S7.122), we have:

$$\begin{aligned}
A_{1.8} &:= \max_{t \geq t_1 + \epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} \{\|\Pi_J \boldsymbol{C}_0^{(4)}(t)\|^2 + \|\Pi_J \boldsymbol{C}_0^{(4)}(t_1)\|^2\} \\
&\leq 2 \max_{q_0 \leq t \leq 1 - q_0} \|\boldsymbol{C}_0^{(4)}(t)\|_{(s_0,2)}^2 \\
&\leq C_8(s_0, M, \widetilde{\boldsymbol{\tau}}) s^2 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2.
\end{aligned}$$

For $A_{1.9}$, by the Cauchy-Swartz inequality, Lemma 7, and (S7.121), we have:

$$
\begin{aligned}
A_{1.9} &= 2 \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \{\Pi_J \boldsymbol{C}_0^{(1)}(t)^\top \Pi_J \boldsymbol{C}_0^{(3)}(t) - \Pi_J \widetilde{\boldsymbol{C}}_0^{(1)}(t_1)^\top \Pi_J \boldsymbol{C}_0^{(3)}(t_1)\} \\
&\leq_{(1)} 4 \max_{t \in [q_0, 1-q_0]} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} |\{\Pi_J \boldsymbol{C}_0^{(1)}(t)^\top \Pi_J \boldsymbol{C}_0^{(3)}(t)| \\
&\leq_{(2)} 4 \max_{t \in [q_0, 1-q_0]} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\{\Pi_J \boldsymbol{C}_0^{(1)}(t)\|_2 \|\Pi_J \boldsymbol{C}_0^{(3)}(t)\|_2 \\
&\leq_{(3)} 4 \max_{t \in [q_0, 1-q_0]} \|\boldsymbol{C}_0^{(1)}(t)\|_{(s_0,2)} \times \max_{t \in [q_0, 1-q_0]} \|\boldsymbol{C}_0^{(3)}(t)\|_{(s_0,2)} \\
&\leq_{(4)} C s_0^{1/2} M \sqrt{\log(pn)} \times s_0^{1/2} \sqrt{\log(pn)} s^2 \|\boldsymbol{\Delta}\|_{(s_0,2)}^2 \\
&\leq_{(5)} C_9(s_0, M, \widetilde{\boldsymbol{\tau}}) s^2 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2.
\end{aligned}
\tag{S7.78}
$$

Similarly, for $A_{1.10} - A_{1.13}$, by (S7.117), (S7.121), (S7.122), and the Cauchy-Swartz inequality, we can prove that:

$$
\begin{aligned}
A_{1.10} &\leq C_{10}(s_0, M, \widetilde{\boldsymbol{\tau}}) s \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}, \\
A_{1.11} &\leq C_{11}(s_0, M, \widetilde{\boldsymbol{\tau}}) SNR(0, \widetilde{\boldsymbol{\tau}}) s^3 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}^3, \\
A_{1.12} &\leq C_{12}(s_0, M, \widetilde{\boldsymbol{\tau}}) SNR(0, \widetilde{\boldsymbol{\tau}}) s^2 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2, \\
A_{1.13} &\leq C_{13}(s_0, M, \widetilde{\boldsymbol{\tau}}) s^3 \log(pn) \|\boldsymbol{\Delta}\|_{(s_0,2)}^3.
\end{aligned}
$$

Note that by Assumption F and the assumption that $\|\boldsymbol{\Delta}\|_1 \leq C_{\boldsymbol{\Delta}}$, we have $s^2 \|\boldsymbol{\Delta}\|_{(s_0,2)}^2 = O(1)$, $s^3 \|\boldsymbol{\Delta}\|_{(s_0,2)}^3 = O(1)$, and $s \|\boldsymbol{\Delta}\|_{(s_0,2)} = O(1)$. Hence, for the above results, by Assumption E.2, up to some constants, the upper bounds of $A_{1.1}$ dominates the others. Hence, if ϵ_n satisfies:

$$
\epsilon_n \geq C(s_0, M, \widetilde{\boldsymbol{\tau}}) \frac{\log(pn)}{n SNR^2(0, \widetilde{\boldsymbol{\tau}}) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2}, \tag{S7.79}
$$

for some big enough $C > 0$, w.p.a.1, we have

$$A_{1.1}+A_{1.2}+A_{1.3}+A_{1.4}+A_{1.7}+A_{1.8}+A_{1.9}+A_{1.10}+A_{1.11}+A_{1.12}+A_{1.13} \leq 0.$$

Step 2: We aim to prove that $\mathbb{P}(A_{1.5} \leq 0) \to 1$. With a very similar proof procedure as (S7.65) - (S7.71) in Section S7.3.1, if ϵ_n satisfies:

$$\epsilon_n \geq C_5(s_0, M, \widetilde{\boldsymbol{\tau}}) \frac{\log(pn)}{nSNR^2(0, \widetilde{\boldsymbol{\tau}})\|\boldsymbol{\Delta}\|^2_{(s_0,2)}}, \tag{S7.80}$$

for some big enough $C_5 > 0$, then, w.p.a.1, we have $A_{1.5} \leq 0$.

Step 3: We aim to prove that $\mathbb{P}(A_{1.6} \leq 0) \to 1$. Note that with a very similar proof procedure as (S7.65) - (S7.71) in Section S7.3.1, if ϵ_n satisfies:

$$\epsilon_n = C_6(s_0, M, \widetilde{\boldsymbol{\tau}}) \frac{s^2 \log(pn)}{n} := C_6(s_0, M, \widetilde{\boldsymbol{\tau}}) \frac{\log(pn)}{n\|\boldsymbol{\Delta}\|^2_{(s_0,2)}} s^2 \|\boldsymbol{\Delta}\|^2_{(s_0,2)}, \tag{S7.81}$$

for some big enough $C_6 > 0$, then, w.p.a.1, we have $A_{1.6} \leq 0$.

Step 4: We aim to prove that $\mathbb{P}(A_{1.14} \leq 0) \to 1$. Using similar analysis as in (S7.119) - (S7.120) and with a very similar proof procedure but some tedious modifications of (S7.65) - (S7.71) in Section S7.3.1, if ϵ_n satisfies:

$$\epsilon_n \geq C_{14}(s_0, M, \widetilde{\boldsymbol{\tau}}) \frac{\log(pn)}{nSNR^2(0, \widetilde{\boldsymbol{\tau}})\|\boldsymbol{\Delta}\|^2_{(s_0,2)}} s^4 \|\boldsymbol{\Delta}\|^4_{(s_0,2)}, \tag{S7.82}$$

for some big enough $C_{14} > 0$, then, w.p.a.1, we have $A_{1.14} \leq 0$.

Step 5: We aim to prove $\mathbb{P}(A_{1.15} \leq 0) \to 1$. Using some tedious modifications of Lemma 17, with a very similar proof procedure as (S7.65) - (S7.71) in Section S7.3.1, if ϵ_n satisfies:

$$\epsilon_n \geq C_{15}(s_0, M, \widetilde{\boldsymbol{\tau}}) \frac{\log(pn)}{nSNR^2(0, \widetilde{\boldsymbol{\tau}})\|\boldsymbol{\Delta}\|^2_{(s_0,2)}} s^2 \|\boldsymbol{\Delta}\|^2_{(s_0,2)}, \tag{S7.83}$$

for some big enough $C_{15} > 0$, then w.p.a.1, we have $A_{1.15} \leq 0$.

Lastly, considering (S7.79) - (S7.83), if ϵ_n satisfies:

$$\epsilon_n \geq C^*(s_0, M, \widetilde{\boldsymbol{\tau}}) \frac{\log(pn)}{nSNR^2(0, \widetilde{\boldsymbol{\tau}})\|\boldsymbol{\Delta}\|^2_{(s_0,2)}}, \tag{S7.84}$$

for some big enough $C^* > 0$, we have $\mathbb{P}(A_{1.1} + \cdots + A_{1.15} \leq 0) \to 1$, which yields $\mathbb{P}(A_1) \to 1$. Similarly, we can prove $\mathbb{P}(A_2) \to 1$, which finishes the proof of Theorem 3 with $\alpha = 0$. □

S7.3.3 Change point estimation for $\alpha \in (0, 1)$

Proof. For $\alpha \in (0, 1)$, recall $\boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) := (1-\alpha)\frac{1}{K}\sum_{k=1}^{K} \boldsymbol{X}_i\big(\mathbf{1}\{Y_i - \widehat{b}_k - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}} \leq 0\} - \tau_k\big) - \alpha \boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}})$ as the weighted score function. For $\alpha \in (0, 1)$ and each $t \in [q_0, 1-q_0]$, define $\widetilde{\boldsymbol{C}}_\alpha(t) = (\widetilde{C}_{\alpha 1}(t), \ldots, \widetilde{C}_{\alpha p}(t))^\top$ with

$$\widetilde{\boldsymbol{C}}_\alpha(t) = \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}})\Big). \tag{S7.85}$$

Note that there is no variance estimator in $\widetilde{\boldsymbol{C}}_\alpha(t)$. Recall $\widehat{t}_\alpha := \underset{t \in [q_0, 1-q_0]}{\arg\max} \|\widetilde{\boldsymbol{C}}_\alpha(t)\|_{(s_0,2)}$. To prove Theorem 3 with $\alpha \in (0, 1)$, we need to prove that as $n, p \to \infty$, by choosing a large enough constant $C(s_0, M, q_0, \alpha)$ in ϵ_n (which will be given in (S7.90)), we have

$$\mathbb{P}\big(|\widehat{t}_\alpha - t_1| \geq \epsilon_n\big) \to 0. \tag{S7.86}$$

Similar to Sections S7.3.1 and S7.3.2, we have to prove $\mathbb{P}(A_1) \to 1$ and $\mathbb{P}(A_2) \to 1$, where

$$\begin{aligned} A_1 &= \max_{t \geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_\alpha(t)\|_{(s_0,2)} - \|\widetilde{\boldsymbol{C}}_\alpha(t_1)\|_{(s_0,2)} \leq 0, \\ A_2 &= \max_{t \leq t_1-\epsilon_n} \|\widetilde{\boldsymbol{C}}_\alpha(t)\|_{(s_0,2)} - \|\widetilde{\boldsymbol{C}}_\alpha(t_1)\|_{(s_0,2)} \leq 0. \end{aligned} \tag{S7.87}$$

By the symmetry, we only consider $\mathbb{P}(A_1) \to 1$. Similar to the previous two sections, define the two events $\mathcal{H}_1$ and $\mathcal{H}_2$:

$$\begin{aligned} \mathcal{H}_1 &= \Big\{ \max_{t \geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_\alpha(t)\|_{(s_0,2)} := \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \{1,\ldots,p\} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t)\|_2 = \max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t)\|_2 \Big\}, \\ \mathcal{H}_2 &= \Big\{ \|\widetilde{\boldsymbol{C}}_\alpha(t_1)\|_{(s_0,2)} := \max_{\substack{J \subset \{1,\ldots,p\} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t_1)\|_2 = \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t_1)\|_2 \Big\}. \end{aligned} \tag{S7.88}$$

Similar to the proof of Lemma 16, we can prove $\mathbb{P}(\mathcal{H}_1 \cap \mathcal{H}_2) \to 1$. Now, under $\mathcal{H}_1 \cap \mathcal{H}_2$, we have:

$$\begin{aligned} \mathbb{P}(A_1) &= \mathbb{P}\Big(\max_{t \geq t_1+\epsilon_n} \|\widetilde{\boldsymbol{C}}_\alpha(t)\|_{(s_0,2)} - \|\widetilde{\boldsymbol{C}}_\alpha(t_1)\|_{(s_0,2)} \leq 0 \Big) \\ &= \mathbb{P}\Big(\max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t)\|_2 - \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t_1)\|_2 \leq 0 \Big) \\ &= \mathbb{P}\Big(\max_{t \geq t_1+\epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t)\|^2 - \max_{\substack{J \subset \mathcal{M} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t_1)\|^2 \leq 0 \Big). \end{aligned}$$

Recall $\sigma^2(\alpha, \widetilde{\boldsymbol{\tau}}) := \mathrm{Var}[(1-\alpha)e_i(\widetilde{\boldsymbol{\tau}}) - \alpha\epsilon_i]$ and $\boldsymbol{C}_\alpha(t)$ defined in (S7.128). Then, under $\mathbf{H}_1$, we have the following decomposition

$$\begin{aligned}
\widetilde{\boldsymbol{C}}_\alpha(t) &= \sigma(\alpha, \widetilde{\boldsymbol{\tau}}) \times \boldsymbol{C}_\alpha(t) \\
&= \sigma(\alpha, \widetilde{\boldsymbol{\tau}}) \times \big(\widetilde{\boldsymbol{C}}_\alpha^I(t) - SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \boldsymbol{\delta}(t) - SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \boldsymbol{R}(t) \\
&+ (1-\alpha)\boldsymbol{C}_0^{(3)}(t) + (1-\alpha)\boldsymbol{C}_0^{(4)}(t)\big),
\end{aligned} \tag{S7.89}$$

where the second equation comes from the decomposition in (S7.129). By the fact that $\max a_i - \max b_i \le \max(a_i - b_i)$ and $\max(a_i + b_i) \le \max a_i + \max b_i$ for any $\{a_i\}$ and $\{b_i\}$, we have:

$$\begin{aligned}
&\Big\{ \max_{t \ge t_1 + \epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t)\|^2 - \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_\alpha(t_1)\|^2 \le 0 \Big\} \\
&\subset \Big\{ \sigma^2(\alpha, \widetilde{\boldsymbol{\tau}}) \times \max_{t \ge t_1 + \epsilon_n} \max_{\substack{J \subset \mathcal{M} \\ |J| = s_0}} \Big(\|\Pi_J \big(\widetilde{\boldsymbol{C}}_\alpha^I(t) - SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \boldsymbol{\delta}(t) - SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \boldsymbol{R}(t) \\
&+ (1-\alpha)\boldsymbol{C}_0^{(3)}(t) + (1-\alpha)\boldsymbol{C}_0^{(4)}(t)\big)\|^2 \\
&\qquad - \|\Pi_J \big(\widetilde{\boldsymbol{C}}_\alpha^I(t_1) - SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \boldsymbol{\delta}(t_1) - SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \boldsymbol{R}(t_1) \\
&+ (1-\alpha)\boldsymbol{C}_0^{(3)}(t_1) + (1-\alpha)\boldsymbol{C}_0^{(4)}(t_1)\big)\|^2 \Big) \le 0 \Big\}.
\end{aligned}$$

Note that similar to Section S7.3.2, for the above inequality, we can decompose it into fifteen parts. Moreover, using the obtained bounds in Sections S7.3.1 and S7.3.2, if ϵ_n satisfies:

$$\epsilon_n \ge C^*(s_0, M, \widetilde{\boldsymbol{\tau}}, \alpha) \frac{\log(pn)}{nSNR^2(\alpha, \widetilde{\boldsymbol{\tau}}) \|\boldsymbol{\Delta}\|_{(s_0,2)}^2}, \tag{S7.90}$$

for some big enough $C^* > 0$, it is not hard to prove that $\mathbb{P}(A_1) \to 1$ and $\mathbb{P}(A_2) \to 1$, which finishes the proof of Theorem 3 with $\alpha \in (0, 1)$.

□

S7.4 Proof of Theorem 4

Let $r_\alpha(n) = \sqrt{s\log(pn)/n}$ if $\alpha = 1$ and $r_\alpha(n) = s\sqrt{\frac{\log(pn)}{n}} \vee s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}$ if $\alpha \in [0, 1)$. In this section, we aim to prove the consistency of $\widehat{\sigma}^2(\alpha, \widetilde{\boldsymbol{\tau}})$ in the sense that

$$|\widehat{\sigma}^2(\alpha, \widetilde{\boldsymbol{\tau}}) - \sigma^2(\alpha, \widetilde{\boldsymbol{\tau}})| = O_p(r_\alpha(n)). \tag{S7.91}$$

We consider the proof in two cases:

Case 1: the signal jump satisfies $SNR(\alpha, \widetilde{\boldsymbol{\tau}})\|\boldsymbol{\Delta}\|_{(s_0,2)} \gg \sqrt{\log(pn)/n}$.

In this case, by Theorem 3, w.p.a.1, we have:

$$\left|n\widehat{t}_\alpha - nt_1\right| = o(n).$$

Recall $n_- := \{i : i \le nh\widehat{t}_\alpha\}$ and $n_+ := \{i : \widehat{t}_\alpha n + (1-h)(1-\widehat{t}_\alpha)n \le i \le n\}$ for some $0 < h < 1$. Hence, by Theorem 3, w.p.a.1, the samples in n_- are before the true change point t_1 and those in n_+ are after t_1. Hence, we can use a very similar proof technique as in Section S7.1 to yield (S7.91).

Case 2: the signal jump satisfies $SNR(\alpha, \widetilde{\boldsymbol{\tau}})\|\boldsymbol{\Delta}\|_{(s_0,2)} = O(\sqrt{\log(pn)/n})$.

In this case, the change point estimator $\widehat{t}_\alpha$ can be an arbitrary number which satisfies $\widehat{t}_\alpha \in [q_0, 1-q_0]$. Note that in this case, the signal jump $\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}$

is very small in the sense that:

$$\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1 = O(s\sqrt{\frac{\log(pn)}{n}}), \quad \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_2 = O(\sqrt{s\frac{\log(pn)}{n}}).$$

In this case, using some modifications of Theorem 1 in Section S7.1, we can still prove

$$|\widehat{\sigma}^2(\alpha, \widetilde{\boldsymbol{\tau}}) - \sigma^2(\alpha, \widetilde{\boldsymbol{\tau}})| = O_p(r_\alpha(n)). \tag{S7.92}$$

Since the modifications are lengthy, to save space, we omit the details.

S7.5 Proof of Theorem 5

Throughout the following proofs, we assume $\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_\infty \geq \sqrt{\log(p)/n}$ and $\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2 \geq \sqrt{s\log(p)/n}$, as well as $\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_1 \geq s\sqrt{\log(p)/n}$. Next, we give the power results for $\alpha = 1$, $\alpha = 0$ and $\alpha \in (0,1)$, respectively. For simplicity, we will omit the subscript α whenever needed.

S7.5.1 Power analysis for $\alpha = 1$

Firstly, we consider the oracle case that assumes the variance is known by letting $\widehat{\sigma}^2(\alpha, \widetilde{\boldsymbol{\tau}}) = \sigma^2$, where $\sigma^2 := \mathrm{Var}[\epsilon]$. In addition, for the case of $\alpha = 1$, we have $SNR(\alpha, \widetilde{\boldsymbol{\tau}}) := SNR(1, \widetilde{\boldsymbol{\tau}}) = 1/\sigma$, where $\sigma^2 = \mathrm{Var}(\epsilon)$. Without loss of generality, we assume $\sigma^2 = 1$. The proof of Theorem 5 proceeds in two steps. In Step 1, we obtain the upper bound of $c_{T_1^b}(1-\gamma)$, where $c_{T_1^b}(1-\gamma)$ is the $1-\gamma$ th quantile of T_1^b, which is defined as

$$c_{T_1^b}(1-\gamma) := \inf\big\{t : \mathbb{P}(T_1^b \leq t) \geq 1-\gamma\big\}. \tag{S7.93}$$

In Step 2, using the obtained upper bound, we get the lower bound of $\mathbb{P}\big(T_1 \geq c_{T_1^b}(1-\gamma)\big)$ and prove

$$\mathbb{P}\big(T_1 \geq c_{T_1^b}(1-\gamma)\big) \to 1, \text{ as } n, p \to \infty. \tag{S7.94}$$

Note that $\{\Psi_{\gamma,1} = 1\} \Leftrightarrow \{T_1 \geq \widehat{c}_{T_\alpha^b}(1-\gamma)\}$, where

$$\widehat{c}_{T_1^b}(1-\gamma) := \inf\Big\{t : (B+1)^{-1}\sum_{b=1}^{B} \mathbf{1}\{T_1^b \leq t\} \geq 1-\gamma\Big\}. \tag{S7.95}$$

Finally, using the fact that $\widehat{c}_{T_1^b}(1-\gamma)$ is the estimation for $c_{T_1^b}(1-\gamma)$ based on the bootstrap samples, we complete the proof. Now, we consider the two steps in detail.

Step 1: By the definition of T_1^b, we have: $T_1^b = \max_{q_0 \leq t \leq 1-q_0} \|\boldsymbol{C}_1^b(t)\|_{(s_0,2)}$, where

$$\boldsymbol{C}_1^b(t) := \frac{1}{\sqrt{n}v(1,\widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \epsilon_i^b - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \epsilon_i^b\Big),$$

with ϵ_i^b being i.i.d $N(0,1)$, $v(1,\widetilde{\boldsymbol{\tau}}) := \mathrm{Var}[e_i^b] = 1$. Our next goal is to obtain an upper bound of $c_{T_1^b}(1-\gamma)$. To this end, for any $1 \leq i \leq n$, $1 \leq j \leq p$, and $\lfloor nq_0 \rfloor \leq k \leq n - \lfloor nq_0 \rfloor$, we define $W_{ijk}^b = X_{ij}\epsilon_i^b a_{ik}$, where

$a_{ik} := \mathbf{1}\{i \le k\} - k/n$. Using the above notations, for T_1, we have:

$$T_1 = \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_1^b(t)\|_{(s_0,2)} \le s_0^{1/2} \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_1^b(t)\|_\infty$$
$$= s_0^{1/2} \frac{1}{\sqrt{n}} \underbrace{\max_{\substack{1 \le j \le p, \\ \lfloor nq_0 \rfloor \le k \le n - \lfloor nq_0 \rfloor}} |\sum_{i=1}^n W_{ijk}^b|}_{Z}.$$

Hence, according to the above inequality, let $c_Z(1-\gamma)$ be the $1-\gamma$-th quantile of Z, then we have:

$$c_{T_1^b}(1-\gamma) \le s_0^{1/2} \frac{1}{\sqrt{n}} c_Z(1-\gamma). \tag{S7.96}$$

Next, we obtain an upper bound of $c_Z(1-\gamma)$. The main technique is to use Lemma 1 and Lemma 2. Let $M = \max_{i,j,k} |W_{ijk}|$ and $\sigma_*^2 = \max_{jk} \sum_i \mathbb{E}[W_{ijk}^2]$. Then, we have

$$\sigma_*^2 = \max_{jk} \sum_i \mathbb{E}[W_{ijk}^2] = \max_{jk} \mathbb{E}[X_{ij}\epsilon_i^b a_{ik}]^2 \le nM^2(1-q_0)^2 \le C_1(M, q_0)n,$$

where the last inequality uses the fact that $|X_{ij}| \le M$ and $|a_{ik}| \le 1 - q_0$. For $\mathbb{E}[M^2]$, we have:

$$\mathbb{E}[M^2] = \mathbb{E}[\max_{ijk} |X_{ij}\epsilon_i^b a_{ik}|^2] \le M^2(1-q_0)^2 \mathbb{E}\big[\max_i |\epsilon_i^b|^2\big] \le C_2(M, q_0)\log(n),$$

where the last inequality comes from Example 3.5.6 in Embrechts et al. (2013). Let $n' = n - 2\lfloor nq_0 \rfloor$. Using the above results, by Lemma 1, we have:

$$\mathbb{E}[Z] \le C(\sigma_* \sqrt{\log(pn')} + \sqrt{\mathbb{E}[M^2]} \log pn') \le C_3(M, q_0)\sqrt{n \log(pn)}.$$

Note that X_{ij} and e_i^b are all sub-Gaussian random variables, which implies $\|X_{ij}\epsilon_i^b\|_{\psi_1}$ exists. Hence, we have:

$$\|M\|_{\psi_1} := \| \max_{i,j,k} |X_{ij}\epsilon_i^b a_{ik}|\|_{\psi_1} \leq C\log(pnn'+1) \max_{i,j,k} \|X_{ij}\epsilon_i^b a_{ik}\|_{\psi_1} \leq C_4(M, q_0)\log(pn).$$

By Lemma 2, taking $\eta = 1$ and $\beta = 1$, we have:

$$\mathbb{P}(Z \geq 2\mathbb{E}[Z] + t) \leq \exp(-\frac{t^2}{3C_1(M, q_0)n}) + 3\exp\Big(-\frac{t}{C_4(M, q_0)\log(pn)}\Big).$$

Taking $t = 2(t_1 \vee t_2)$, where t_1 and t_2 satisfy

$$-\frac{t_1^2}{3C_1(M, q_0)n} = \log(\gamma/2) \quad \text{and} \quad -\frac{t_2}{C_4(M, q_0)\log(pn)} = \log(\gamma/6),$$

we have

$$\mathbb{P}(Z \geq 2\mathbb{E}[Z] + t) \leq \gamma.$$

By noting that $t := 2(t_1 \vee t_2) \leq C_5(M, q_0)\sqrt{n\log(1/\gamma)}$ and $\mathbb{E}[Z] \leq C_3(M, q_0)\sqrt{n\log(pn)}$, we have:

$$c_Z(1-\gamma) = 2\mathbb{E}[Z] + t$$
$$\leq 2C_3(M, q_0)\sqrt{n\log(pn)} + C_5(M, q_0)\sqrt{n\log(1/\gamma)} \leq C_6(\sqrt{n\log(pn)} + \sqrt{n\log(1/\gamma)}).$$

Lastly, considering (S7.96), we have:

$$c_{T_1^b}(1-\gamma) \leq C_6(M, q_0)s_0^{1/2}(\sqrt{\log(pn)} + \sqrt{\log(1/\gamma)}), \qquad \text{(S7.97)}$$

where $C_6(M, q_0)$ is some universal constant not depending on n or p.

Step 2: In this step, we aim to prove that $\mathbb{P}\big(T_1 \geq c_{T_1^b}(1-\gamma)\big) \to 1$ as

$n,p \to \infty$. Note that in Step 1, we have obtained the upper bound of $c_{T_1^b}(1-\gamma)$. Hence, it is sufficient to prove that $H_1 \to 1$ as $n,p \to \infty$, where

$$H_1 := \mathbb{P}\big(T_1 \geq C_6(M,q_0)s_0^{1/2}(\sqrt{\log(pn)} + \sqrt{\log(1/\gamma)})\big). \tag{S7.98}$$

To prove $H_1 \to 1$, we need the decomposition of T_1 under $\mathbf{H}_1$. Recall the decomposition of $\boldsymbol{C}_1(t)$ defined in (S7.22). Let the signal jump be

$$\boldsymbol{\delta}(t) := \begin{cases} \sqrt{n}\dfrac{\lfloor nt \rfloor}{n}\dfrac{n-\lfloor nt_1 \rfloor}{n}\boldsymbol{\Sigma}\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big), & \text{if } \ t \leq t_1, \\ \\ \sqrt{n}\dfrac{\lfloor nt_1 \rfloor}{n}\dfrac{n-\lfloor nt \rfloor}{n}\boldsymbol{\Sigma}\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big), & \text{if } \ t > t_1. \end{cases} \tag{S7.99}$$

Then, under $\mathbf{H}_1$, for $\alpha = 1$, we have the following decomposition:

$$\boldsymbol{C}_1(t) = \boldsymbol{C}_1^{\boldsymbol{I}}(t) + SNR(1,\widetilde{\boldsymbol{\tau}}) \times \boldsymbol{\delta}(t) + SNR(1,\widetilde{\boldsymbol{\tau}}) \times \boldsymbol{R}(t),$$

where $\boldsymbol{C}_1(t)$ and $\boldsymbol{R}(t)$ are defined as

$$\boldsymbol{C}_1^{I}(t) := \frac{1}{\sqrt{n}\widehat{\sigma}(\alpha,\widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor}\boldsymbol{X}_i\epsilon_i - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\epsilon_i\Big),\ \boldsymbol{R}(t) := \boldsymbol{R}^{I}(t)\mathbf{1}\{i \leq \lfloor nt_1 \rfloor\} + \boldsymbol{R}^{II}(t)\mathbf{1}\{i > \lfloor nt_1 \rfloor\}, \tag{S7.100}$$

with $\boldsymbol{R}^{I}(t)$ and $\boldsymbol{R}^{II}(t)$ being defined as

$$\begin{aligned} \boldsymbol{R}^{I}(t) := {} & \frac{\lfloor nt \rfloor(n - \lfloor nt \rfloor)}{n^{3/2}}\big(\widehat{\boldsymbol{\Sigma}}(0:t) - \boldsymbol{\Sigma}\big)\big(\boldsymbol{\beta}^{(1)} - \widehat{\boldsymbol{\beta}}\big) \\ & - \frac{\lfloor nt \rfloor(\lfloor nt_1 \rfloor - \lfloor nt \rfloor)}{n^{3/2}}\big(\widehat{\boldsymbol{\Sigma}}(t:t_1) - \boldsymbol{\Sigma}\big)\big(\boldsymbol{\beta}^{(1)} - \widehat{\boldsymbol{\beta}}\big) \\ & - \frac{\lfloor nt \rfloor(n - \lfloor nt_1 \rfloor)}{n^{3/2}}\big(\widehat{\boldsymbol{\Sigma}}(t_1:1) - \boldsymbol{\Sigma}\big)\big(\boldsymbol{\beta}^{(2)} - \widehat{\boldsymbol{\beta}}\big), \end{aligned}$$

and

$$\boldsymbol{R}^{II}(t) := \frac{\lfloor nt_1\rfloor(n-\lfloor nt\rfloor)}{n^{3/2}}\big(\widehat{\boldsymbol{\Sigma}}(0:t_1)-\boldsymbol{\Sigma}\big)\big(\boldsymbol{\beta}^{(1)}-\widehat{\boldsymbol{\beta}}\big)$$
$$-\frac{(n-\lfloor nt\rfloor)(\lfloor nt\rfloor-\lfloor nt_1\rfloor)}{n^{3/2}}\big(\widehat{\boldsymbol{\Sigma}}(t_1:t)-\boldsymbol{\Sigma}\big)\big(\boldsymbol{\beta}^{(2)}-\widehat{\boldsymbol{\beta}}\big)$$
$$-\frac{\lfloor nt\rfloor(n-\lfloor nt\rfloor)}{n^{3/2}}\big(\widehat{\boldsymbol{\Sigma}}(t:1)-\boldsymbol{\Sigma}\big)\big(\boldsymbol{\beta}^{(2)}-\widehat{\boldsymbol{\beta}}\big).$$

To prove $H_1 \to 1$, we need the analysis of $\boldsymbol{C}_1^I(t)$, $\boldsymbol{\delta}(t)$, and $\boldsymbol{R}(t)$, respectively. By definition, for $\boldsymbol{\delta}(t)$, we have: $t_1 = \arg\max_{q_0\le t\le 1-q_0}\|\boldsymbol{\delta}(t)\|_{(s_0,2)}$. In other words, $\|\boldsymbol{\delta}(t)\|_{(s_0,2)}$ obtains its maximum value at the true change point location. For $\boldsymbol{C}_1^I(t)$, by Lemma 7 and the fact that $\|\boldsymbol{v}\|_{(s_0,2)} \le s_0^{1/2}\|\boldsymbol{v}\|_\infty$ for any $\boldsymbol{v}\in\mathbb{R}^p$, we have $\max_{q_0\le t\le 1-q_0}\|\boldsymbol{C}_1^I(t)\|_{(s_0,2)} \le Cs_0^{1/2}M\sqrt{\log(pn)}$ for some constant $C>0$. As for $\boldsymbol{R}(t)$, using the triangle inequality, we have:

$$\max_{q_0\le t\le 1-q_0}\|\boldsymbol{R}(t)\|_{(s_0,2)} \le \max_{q_0\le t\le t_1}\|\boldsymbol{R}^I(t)\|_{(s_0,2)} + \max_{t_1\le t\le 1-q_0}\|\boldsymbol{R}^{II}(t)\|_{(s_0,2)}.$$

For $\max_{q_0\le t\le t_1}\|\boldsymbol{R}^I(t)\|_{(s_0,2)}$, using Lemma 8, with probability at least $1-(pn)^{-C}$, we have

$$\max_{q_0\le t\le t_1}\|\boldsymbol{R}^I(t)\|_{(s_0,2)} \le C_1 s_0^{1/2}\sqrt{\log p}\big(\big\|\boldsymbol{\beta}^{(1)}-\widehat{\boldsymbol{\beta}}\big\|_1+\big\|\boldsymbol{\beta}^{(2)}-\widehat{\boldsymbol{\beta}}\big\|_1\big).$$

Note that by Lemma 10 and the fact that $\boldsymbol{\beta}^* = t_1\boldsymbol{\beta}^{(1)} + (1-t_1)\boldsymbol{\beta}^{(2)}$, we have:

$$
\begin{aligned}
&\left\|\boldsymbol{\beta}^{(1)} - \widehat{\boldsymbol{\beta}}\right\|_1 \\
&\quad \leq_{(1)} \left\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^*\right\|_1 + \left\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\right\|_1 \\
&\quad \leq_{(2)} (1-t_1)\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1 + C_1 s M \sqrt{\frac{\log(pn)}{n}}\big(1 + M\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_1\big) \\
&\quad \leq_{(3)} C_1\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1 \\
&\quad \leq_{(4)} C_1 s\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_\infty \\
&\quad \leq_{(5)} C_1 s\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)} \\
&\quad \leq_{(6)} C_2 s\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)},
\end{aligned}
$$

where (2) comes from Lemma 10, (3) comes from the assumption that $sM^2\sqrt{\log(pn)/n} = o(1)$, (6) comes from **Assumption A**. Similarly, we can prove $\left\|\boldsymbol{\beta}^{(2)} - \widehat{\boldsymbol{\beta}}\right\|_1 = O_p(s\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)})$. Combining this result, we can prove that

$$
\begin{aligned}
\max_{q_0 \leq t \leq 1-q_0} \|\boldsymbol{R}(t)\|_{(s_0,2)} &\leq 2\max\Big(\max_{q_0\leq t\leq t_1}\|\boldsymbol{R}^I(t)\|_{(s_0,2)}, \max_{t_1\leq t\leq 1-q_0}\|\boldsymbol{R}^{II}(t)\|_{(s_0,2)}\Big) \\
&\leq C s_0^{1/2} s\sqrt{\log p}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)}.
\end{aligned}
$$

Using the above bounds of $\boldsymbol{C}_1^I(t)$, $\boldsymbol{\delta}(t)$, and $\boldsymbol{R}(t)$, and by the triangle inequality, we have:

$$\begin{aligned}
T_1 &= \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_1(t)\|_{(s_0,2)} \\
&\ge SNR(1,\widetilde{\boldsymbol{\tau}}) \times \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{\delta}(t)\|_{(s_0,2)} - \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_1^I(t)\|_{(s_0,2)} - \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{R}(t)\|_{(s_0,2)} \\
&\ge \sqrt{n} \times SNR(1,\widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1) \times \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)} - C_1 s_0^{1/2} M \sqrt{\log(pn)} \\
&\qquad - C_2 s_0^{1/2} s \sqrt{\log p} \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)}, \\
&\ge \sqrt{n} \times SNR(1,\widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1) \times \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)}(1-\epsilon_n) - C_1 s_0^{1/2} M \sqrt{\log(pn)},
\end{aligned} \tag{S7.101}$$

where $\epsilon_n := (SNR(1,\widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1))^{-1} s_0^{1/2} s \sqrt{\log(p)/n} = O(s_0^{1/2} s \sqrt{\log(p)/n})$.

Recall H_1 as defined in (S7.98). Hence, to prove $H_1 \to 1$, it is sufficient to prove $H_1' \to 1$, where

$$H_1' = \mathbb{P}\Big(\sqrt{n} \times SNR(1,\widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1) \times \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)} \ge \frac{C s_0^{1/2} M \big(\sqrt{\log(pn)} + \sqrt{\log(1/\gamma)}\big)}{1-\epsilon_n}\Big).$$

By (3.37), one can see that $H_1' \to 1$ as $n, p \to \infty$, which finishes the proof.

Remark 2. Note that for $\alpha = 1$, if we replace $\sigma^2(1,\widetilde{\boldsymbol{\tau}})$ by an estimator $\widehat{\sigma}^2(\alpha,\widetilde{\boldsymbol{\tau}})$ which satisfies: $|\widehat{\sigma}^2(\alpha,\widetilde{\boldsymbol{\tau}}) - \sigma^2(1,\widetilde{\boldsymbol{\tau}})| = o_p(1)$, then under condition (3.37), the power still converges to 1.

S7.5.2 Power analysis for $\alpha = 0$

Proof. Firstly, we assume $\widehat{\sigma}^2(0, \widetilde{\boldsymbol{\tau}}) = \sigma^2(0, \widetilde{\boldsymbol{\tau}})$ by considering the variance as unknown, where $\sigma^2(\alpha, \widetilde{\boldsymbol{\tau}}) = \sqrt{K^{-2}\sum_{k_1,k_2}\gamma_{k_1k_2}}$, with $\gamma_{k_1k_2} := \min(\tau_{k_1}, \tau_{k_2}) - \tau_{k_1}\tau_{k_2}$. In addition, for the case of $\alpha = 0$, we have

$$SNR(0, \widetilde{\boldsymbol{\tau}}) = \frac{\sum_{k=1}^{K} f_\epsilon(b_k^{(0)})}{\sqrt{\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\gamma_{k_1k_2}}}.$$

Similar to Section S7.5.1, the proof of Theorem5 proceeds in four steps. In Step 1, we obtain the upper bound of $c_{T_0^b}(1-\gamma)$, where $c_{T_0^b}(1-\gamma)$ is the $(1-\gamma)$-th quantile of T_0^b, which is defined as

$$c_{T_0^b}(1-\gamma) := \inf\big\{t : \mathbb{P}(T_0^b \le t) \ge 1-\gamma\big\}. \qquad \text{(S7.102)}$$

In Steps 2-4, using the upper bound, we get the lower bound of $\mathbb{P}\big(T_0 \ge c_{T_0^b}(1-\gamma)\big)$ and prove

$$\mathbb{P}\big(T_0 \ge c_{T_0^b}(1-\gamma)\big) \to 1, \text{ as } n, p \to \infty. \qquad \text{(S7.103)}$$

Step 1: By the definition of T_0^b, we have: $T_0^b = \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_0^b(t)\|_{(s_0,2)}$, with

$$\boldsymbol{C}_0^b(t) := \frac{1}{\sqrt{n}v(0, \widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt\rfloor}\boldsymbol{X}_i e_i^b(\widetilde{\boldsymbol{\tau}}) - \frac{\lfloor nt\rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i e_i^b(\widetilde{\boldsymbol{\tau}})\Big),$$

where $e_i^b(\widetilde{\boldsymbol{\tau}}) = K^{-1}\sum_{k=1}^{K} e_i^b(\tau_k)$ with $e_i^b(\tau_k) := \mathbf{1}\{\epsilon_i^b \le \Phi^{-1}(\tau_k)\} - \tau_k$, ϵ_i^b is i.i.d $N(0,1)$, $v(0, \widetilde{\boldsymbol{\tau}}) := \mathrm{Var}[e^b(\widetilde{\boldsymbol{\tau}})]$, and $\Phi(x)$ is the CDF for the standard normal distribution. Note that $|e_i^b(\widetilde{\boldsymbol{\tau}})| \le 1$ by definition. Hence, we can use a very similar proof procedure as in Step 1 in Section S7.5.1 to obtain

$$c_{T_0^b}(1-\gamma) \le C(M, q_0, \widetilde{\boldsymbol{\tau}}) s_0^{1/2}(\sqrt{\log(pn)} + \sqrt{\log(1/\gamma)}), \tag{S7.104}$$

where $C(M, q_0, \widetilde{\boldsymbol{\tau}})$ is some universal constant only depending on M, q_0 and $\widetilde{\boldsymbol{\tau}}$.

Step 2 Decomposition of $\boldsymbol{C}_0(t)$. In this step, we aim to prove that $\mathbb{P}\big(T_0 \ge c_{T_0^b}(1-\gamma)\big) \to 1$ as $n, p \to \infty$. Note that in Step 1, we have obtained the upper bound of $c_{T_\alpha^b}(1-\gamma)$. Hence, it is sufficient to prove that $H_1 \to 1$ as $n, p \to \infty$, where

$$H_1 := \mathbb{P}\big(T_0 \ge C(M, q_0, \widetilde{\boldsymbol{\tau}}) s_0^{1/2}(\sqrt{\log(pn)} + \sqrt{\log(1/\gamma)})\big). \tag{S7.105}$$

To prove $H_1 \to 1$, we need the decomposition of T_0 under $\mathbf{H}_1$. Note that for $\alpha = 0$, with known variance, the score based CUSUM process reduces to:

$$\boldsymbol{C}_0(t) = \frac{1}{\sqrt{n}\sigma(0, \widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}})\Big), \tag{S7.106}$$

where $\boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) := \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{X}_i\big(\mathbf{1}\{Y_i - \widehat{b}_k - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}} \le 0\} - \tau_k\big)$. Define

$$\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) := \frac{1}{K}\sum_{k=1}^{K}\Big(\mathbf{1}\{Y_i - \widehat{b}_k - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}} \le 0\} - \tau_k\Big) := \frac{1}{K}\sum_{k=1}^{K} \widehat{e}_i(\tau_k), \tag{S7.107}$$

where $\widehat{e}_i(\tau_k) := \mathbf{1}\{Y_i - \widehat{b}_k - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}} \le 0\} - \tau_k$. For $\widehat{e}_i(\widetilde{\boldsymbol{\tau}})$, we have the following decomposition:

$$\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) = e_i(\widetilde{\boldsymbol{\tau}}) + \mathbb{E}[\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}})] + \underbrace{\{\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}}) - \mathbb{E}[\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}})]\}}_{V_i(\widetilde{\boldsymbol{\tau}})}, \tag{S7.108}$$

where

$$\begin{aligned} e_i(\widetilde{\boldsymbol{\tau}}) &:= \frac{1}{K}\sum_{k=1}^K e_i(\tau_k), \text{ with } e_i(\tau_k) = \mathbf{1}\{\epsilon_i \le b_k^{(0)}\} - \tau_k, \\ \mathbb{E}[\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}})] &:= \frac{1}{K}\sum_{k=1}^K \mathbb{E}\big[\widehat{e}_i(\tau_k) - e_i(\tau_k)\big], \\ V_i(\widetilde{\boldsymbol{\tau}}) &:= \frac{1}{K}\sum_{k=1}^K V_i(\tau_k), \end{aligned} \tag{S7.109}$$

and

$$V_i(\tau_k) = [\mathbf{1}\{Y_i - \widehat{b}_k - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}} \le 0\} - \mathbf{1}\{\epsilon_i \le b_k^{(0)}\}] - \mathbb{E}[\mathbf{1}\{Y_i - \widehat{b}_k - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}} \le 0\} - \mathbf{1}\{\epsilon_i \le b_k^{(0)}\}].$$

Next, we analyze the three parts in (S7.108). Note that the first term $e_i(\widetilde{\boldsymbol{\tau}})$ is a sum for simple Bernoulli random variables. For the second term, by the Taylor's expansion, we have

$$\mathbb{E}[\widehat{e}_i(\widetilde{\boldsymbol{\tau}}) - e_i(\widetilde{\boldsymbol{\tau}})] = \frac{1}{K}\sum_{k=1}^K \mathbb{E}\big[\widehat{e}_i(\tau_k) - e_i(\tau_k)\big] = \underbrace{\frac{1}{K}\sum_{k=1}^K M_i^{(1)}(\tau_k)}_{M_i^{(1)}(\widetilde{\boldsymbol{\tau}})} + \underbrace{\frac{1}{K}\sum_{k=1}^K M_i^{(2)}(\tau_k)}_{M_i^{(2)}(\widetilde{\boldsymbol{\tau}})}, \tag{S7.110}$$

where

$$\begin{aligned} M_i^{(1)}(\tau_k) := {} & f_\epsilon(b_k^{(0)})\big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(1)})\big)\mathbf{1}\{i \le \lfloor nt_1 \rfloor\} \\ & + f_\epsilon(b_k^{(0)})\big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(2)})\big)\mathbf{1}\{i > \lfloor nt_1 \rfloor\}, \end{aligned} \tag{S7.111}$$

and

$$M_i^{(2)}(\tau_k) := \frac{1}{2}f'_\epsilon(\xi_{ik})\big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(1)})\big)^2\mathbf{1}\{i \le \lfloor nt_1 \rfloor\}$$
$$+\frac{1}{2}f'_\epsilon(\xi_{ik})\big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(2)})\big)^2\mathbf{1}\{i > \lfloor nt_1 \rfloor\}, \tag{S7.112}$$

with ξ_{ik} being some constant that between $b_k^{(0)}$ and $\widehat{b}_k + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(1)})$ (or $b_k + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(2)})$).

Hence, based on the above decomposition, for the composite quantile based score function, its CUSUM process can be decomposed into four parts:

$$\boldsymbol{C}_0(t) = \boldsymbol{C}_0^{(1)}(t) + \boldsymbol{C}_0^{(2)}(t) + \boldsymbol{C}_0^{(3)}(t) + \boldsymbol{C}_0^{(4)}(t), \tag{S7.113}$$

where $\boldsymbol{C}_0^{(1)}(t), \ldots, \boldsymbol{C}_0^{(4)}(t)$ are defined as:

$$\begin{aligned}
\boldsymbol{C}_0^{(1)}(t) &= \frac{1}{\sqrt{n}\sigma(\alpha, \widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor}\boldsymbol{X}_i e_i(\widetilde{\boldsymbol{\tau}}) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i e_i(\widetilde{\boldsymbol{\tau}})\Big),\\
\boldsymbol{C}_0^{(2)}(t) &= \frac{1}{\sqrt{n}\sigma(\alpha, \widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor}\boldsymbol{X}_i M_i^{(1)}(\widetilde{\boldsymbol{\tau}}) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i M_i^{(1)}(\widetilde{\boldsymbol{\tau}})\Big),\\
\boldsymbol{C}_0^{(3)}(t) &= \frac{1}{\sqrt{n}\sigma(\alpha, \widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor}\boldsymbol{X}_i M_i^{(2)}(\widetilde{\boldsymbol{\tau}}) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i M_i^{(2)}(\widetilde{\boldsymbol{\tau}})\Big),\\
\boldsymbol{C}_0^{(4)}(t) &= \frac{1}{\sqrt{n}\sigma(\alpha, \widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor}\boldsymbol{X}_i V_i(\widetilde{\boldsymbol{\tau}}) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i V_i(\widetilde{\boldsymbol{\tau}})\Big).
\end{aligned} \tag{S7.114}$$

Note that $\boldsymbol{C}_0^{(2)}(t)$ consists of the signal jump and is very important for detecting a change point. To see this, recall $M_i^{(1)}(\widetilde{\boldsymbol{\tau}}) = \frac{1}{K}\sum_{k=1}^{K} M_i^{(1)}(\tau_k)$

defined in (S7.110). Then, we have

$$\begin{aligned}\boldsymbol{C}_0^{(2)}(t) &=_{(1)} \frac{1}{\sqrt{n}\sigma(\alpha,\widetilde{\boldsymbol{\tau}})}\big(\sum_{i=1}^{\lfloor nt\rfloor}\boldsymbol{X}_i\frac{1}{K}\sum_{k=1}^{K}M_i^{(1)}(\tau_k)-\frac{\lfloor nt\rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\frac{1}{K}\sum_{k=1}^{K}M_i^{(1)}(\tau_k)\big)\\ &=_{(2)} \frac{1}{K}\sum_{k=1}^{K}\big[\frac{1}{\sqrt{n}\sigma(\alpha,\widetilde{\boldsymbol{\tau}})}\big(\sum_{i=1}^{\lfloor nt\rfloor}\boldsymbol{X}_iM_i^{(1)}(\tau_k)-\frac{\lfloor nt\rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_iM_i^{(1)}(\tau_k)\big)\big]\\ &=_{(3)} \frac{1}{K}\sum_{k=1}^{K}\big[\frac{-f_\epsilon(b_k^{(0)})}{\sigma(0,\widetilde{\boldsymbol{\tau}})}\big(\boldsymbol{\delta}(t)+\boldsymbol{R}(t)\big)\big],\end{aligned} \tag{S7.115}$$

where $\boldsymbol{\delta}(t)$ is defined in (S7.99), and $\boldsymbol{R}(t)$ is defined in (S7.100). Hence, combining (S7.113) - (S7.115), under $\mathbf{H}_1$, the score based CUSUM for the quantile loss can be decomposed into four terms:

$$\boldsymbol{C}_0(t) = -SNR(0,\widetilde{\boldsymbol{\tau}})\times\boldsymbol{\delta}(t)+\boldsymbol{C}_0^{(1)}(t)-SNR(0,\widetilde{\boldsymbol{\tau}})\times\boldsymbol{R}(t)+\boldsymbol{C}_0^{(3)}(t)+\boldsymbol{C}_0^{(4)}(t). \tag{S7.116}$$

Step 3: Obtain the upper bounds for the residuals and random noises in $\boldsymbol{C}_0(t)$.

We first consider $\max_t \|\boldsymbol{C}_0^{(1)}\|_{(s_0,2)}$. By definition, $\boldsymbol{C}_0^{(1)}$ is a partial sum process based on $\boldsymbol{X}_ie_i(\widetilde{\boldsymbol{\tau}})$. Hence, by Lemma 7, we can prove that $\max_{q_0\le t\le 1-q_0}\|\boldsymbol{C}_0^{(1)}(t)\|_{(s_0,2)} = O_p(s_0^{1/2}M\sqrt{\log(pn)})$. For $\max_t \|\boldsymbol{R}(t)\|_{(s_0,2)}$, using Lemma 10, Remark 5, and using a similar proof procedure as in Step 2 of Section S7.5.1, we can prove that

$$\max_{q_0\le t\le 1-q_0}\|\boldsymbol{R}(t)\|_{(s_0,2)} = O_p\big(s_0^{1/2}s\sqrt{\log p}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)})\|_{(s_0,2)}\big). \tag{S7.117}$$

Next, we consider $\max_{q_0\le t\le 1-q_0}\|\boldsymbol{C}_0^{(3)}(t)\|_{(s_0,2)}$. To that end, we need some notations. Let $\underset{\sim}{\boldsymbol{\beta}}^{(1)} := ((\boldsymbol{\beta}^{(1)})^\top, (\boldsymbol{b}^{(0)})^\top)^\top \in \mathbb{R}^{p+K}$, $\underset{\sim}{\boldsymbol{\beta}}^{(2)} := ((\boldsymbol{\beta}^{(2)})^\top, (\boldsymbol{b}^{(0)})^\top)^\top \in \mathbb{R}^{p+K}$, $\underset{\approx}{\boldsymbol{X}} := (\boldsymbol{X}^\top, \mathbf{1}_K) \in \mathbb{R}^{p+K}$, and $\boldsymbol{S}_k := \text{diag}(\mathbf{1}_p, \boldsymbol{e}_k)$, where $\boldsymbol{e}_k \in \mathbb{R}^K$ is a vector with the k-th element being 1 and the others being zeros, and $\mathbf{1}_K$ is a K-dimensional vector with all elements being 1s. Moreover, recall $\boldsymbol{S}$ as defined in (S9.191). Then, by the definition of $\boldsymbol{C}_0^{(3)}(t)$, we have:

$$
\begin{aligned}
&\max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_0^{(3)}(t)\|_{(s_0,2)} \\
&\le_{(1)} s_0^{1/2} \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_0^{(3)}(t)\|_\infty \\
&\le_{(2)} s_0^{1/2} \max_t \frac{\lfloor nt \rfloor}{\sqrt{n}\sigma(\alpha, \widetilde{\boldsymbol{\tau}})} \Big\| \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i M_i^{(2)}(\widetilde{\boldsymbol{\tau}}) \Big\|_\infty \\
&\qquad\qquad + s_0^{1/2} \max_t \frac{\lfloor nt \rfloor}{\sqrt{n}\sigma(\alpha, \widetilde{\boldsymbol{\tau}})} \Big\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i M_i^{(2)}(\widetilde{\boldsymbol{\tau}}) \Big\|_\infty \\
&\le_{(3)} C_1 \sqrt{n} s_0^{1/2} \max_j \max_t \Big| \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} X_{ij} M_i^{(2)}(\widetilde{\boldsymbol{\tau}}) \Big| + C_1 \sqrt{n} s_0^{1/2} \max_j \Big| \frac{1}{n} \sum_{i=1}^{n} X_{ij} M_i^{(2)}(\widetilde{\boldsymbol{\tau}}) \Big| \\
&\le_{(4)} C_1 \sqrt{n} s_0^{1/2} (I \vee II) + C_1 M \sqrt{n} s_0^{1/2} (III \vee IV),
\end{aligned}
$$

where $I, \ldots, IV$ are defined as

$$
\begin{aligned}
I &:= \max_{1 \le j \le p} \max_{q_0 \le t \le 1-q_0} \Big| \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} X_{ij} \frac{1}{K} \sum_{k=1}^{K} \frac{1}{2} f_\epsilon'(\xi_{ik}) \big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(1)})\big)^2 \Big|, \\
II &:= \max_{1 \le j \le p} \max_{q_0 \le t \le 1-q_0} \Big| \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} X_{ij} \frac{1}{K} \sum_{k=1}^{K} \frac{1}{2} f_\epsilon'(\xi_{ik}) \big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(2)})\big)^2 \Big|, \\
III &:= \max_j \Big| \frac{1}{n} \sum_{i=1}^{n} X_{ij} \frac{1}{K} \sum_{k=1}^{K} \frac{1}{2} f_\epsilon'(\xi_{ik}) \big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(1)})\big)^2 \Big|, \\
IV &:= \max_j \Big| \frac{1}{n} \sum_{i=1}^{n} X_{ij} \frac{1}{K} \sum_{k=1}^{K} \frac{1}{2} f_\epsilon'(\xi_{ik}) \big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(1)})\big)^2 \Big|.
\end{aligned}
\tag{S7.118}
$$

Next, we consider $I - IV$, respectively. To that end, define

$$M_i^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)}) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}f_\epsilon'(\xi_{ik})\Big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(1)})\Big)^2, \text{ for } i = 1,\ldots,n,$$
$$M_i^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(2)}) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}f_\epsilon'(\xi_{ik})\Big(\widehat{b}_k - b_k^{(0)} + \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(2)})\Big)^2, \text{ for } i = 1,\ldots,n,$$
$$\boldsymbol{M}^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)}) := (M_1^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)}),\ldots,M_n^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)}))^\top,$$
$$\boldsymbol{M}^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(2)}) := (M_1^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(2)}),\ldots,M_n^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(2)}))^\top.$$

For I, we then have:

$$\begin{aligned} I \;&=_{(1)} \max_{1\le j\le p}\max_{q_0\le t\le 1-q_0}\Big|\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}X_{ij}\frac{M_i^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)})}{\|\boldsymbol{M}^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)})\|}\Big|\big\|\boldsymbol{M}^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)})\big\| \\ &\le_{(2)} \max_{\substack{\boldsymbol{w}=(w_1,\ldots,w_n)^\top\\ \|\boldsymbol{w}\|=1}}\max_{1\le j\le p}\max_{q_0\le t\le 1-q_0}\Big|\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}X_{ij}w_i\Big|\big\|\boldsymbol{M}^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)})\big\| \\ &\le_{(3)} O_p\Big(\frac{\sqrt{\log(pn)}}{n}\Big)\big\|\boldsymbol{M}^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)})\big\|, \end{aligned} \tag{S7.119}$$

where (3) comes from Assumption (A.2) and the Hoeffding's inequality.

Hence, to bound I, we need to consider $\|\boldsymbol{M}^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)})\|$. In fact, we have:

$$\begin{aligned}
&\|\boldsymbol{M}^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)})\|^2 \\
&=_{(1)} \sum_{i=1}^{n}[M_i^{(2)}(\widetilde{\boldsymbol{\tau}};\boldsymbol{\beta}^{(1)})]^2 \\
&=_{(2)} \sum_{i=1}^{n}\Big[\frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}f_\epsilon'(\xi_{ik})\Big(\widehat{b}_k-b_k^{(0)}+\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(1)})\Big)^2\Big]^2 \\
&\leq_{(3)} \frac{C_+'^2 n}{4K^2}\max_{1\leq i\leq n}\Big[\sum_{k=1}^{K}\Big(\widehat{b}_k-b_k^{(0)}+\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(1)})\Big)^2\Big]^2 \\
&\leq_{(4)} \frac{C_+'^2 n}{4K^2}\Big[\sum_{k=1}^{K}(|\widehat{b}_k-b_k^{(0)}|+M\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(1)}\|_1)^2\Big]^2 \qquad\qquad (S7.120)\\
&\leq_{(5)} \frac{C_+'^2 n}{4K^2}\Big[KM^2\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}-\underset{\sim}{\boldsymbol{\beta}}^{(1)}\|_1^2\Big]^2 \\
&\leq_{(6)} \frac{M^4C_+'^2 n}{4}\Big[\|\widehat{\underset{\sim}{\boldsymbol{\beta}}}-\underset{\sim}{\boldsymbol{\beta}}^*\|_1+\|\underset{\sim}{\boldsymbol{\beta}}^*-\underset{\sim}{\boldsymbol{\beta}}^{(1)}\|_1\Big]^4 \\
&\leq_{(7)} \frac{M^4C_+'^2 n}{4}\Big[O_p(s\sqrt{\frac{\log(pn)}{n}})+C_1\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\|_1\Big]^4 \\
&\leq_{(8)} O_p(ns^4\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\|_{(s_0,2)}^4),
\end{aligned}$$

where (3) comes from **Assumption D**, (4) comes from **Assumption A**, (7) comes from Lemma 10, (8) comes from $\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\|_1\leq s\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\|_{(s_0,2)}$. Combining (S7.119) and (S7.120), we have

$$I=O_p(s^2\sqrt{\frac{\log(pn)}{n}}\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\|_{(s_0,2)}^2).$$

With a similar proof procedure, we can prove $II, III, IV = O_p(s^2\sqrt{\log(pn)/n}\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\|_{(s_0,2)}^2)$, which yields:

$$\max_{q_0\leq t\leq 1-q_0}\|\boldsymbol{C}_0^{(3)}(t)\|_{(s_0,2)}=O_p(s_0^{1/2}s^2\sqrt{\log(pn)}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)})\|_{(s_0,2)}^2). \qquad (S7.121)$$

Lastly, we consider the control of $\max_{q_0\leq t\leq 1-q_0}\|\boldsymbol{C}_0^{(4)}(t)\|_{(s_0,2)}$. Similar to the proof of Lemma 17, using some tedious modifications, we can prove that:

$$\max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_0^{(4)}(t)\|_{(s_0,2)} = O_p\big(M s_0^{1/2} s\sqrt{\log(pn)}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)})\|_{(s_0,2)}\big). \tag{S7.122}$$

Step 4: Combining the previous results. Recall (S7.104), (S7.105), (S7.116). Using the above bounds of $\boldsymbol{C}_{(0)}^{(1)}(t)$, $\boldsymbol{R}(t)$, $\boldsymbol{C}_{(0)}^{(3)}(t)$, $\boldsymbol{C}_{(0)}^{(4)}(t)$, and by the triangle inequality, w.p.a.1, we have:

$$\begin{aligned} T_0 &= \max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_0(t)\|_{(s_0,2)} \\ &\ge SNR(0,\widetilde{\boldsymbol{\tau}}) \times \max_{q_0\le t\le 1-q_0} \|\boldsymbol{\delta}(t)\|_{(s_0,2)} - \max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_0^{(1)}(t)\|_{(s_0,2)} \\ &\quad -SNR(0,\widetilde{\boldsymbol{\tau}}) \times \max_{q_0\le t\le 1-q_0} \|\boldsymbol{R}(t)\|_{(s_0,2)} - \max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_0^{(3)}(t)\|_{(s_0,2)} - \max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_0^{(4)}(t)\|_{(s_0,2)} \\ &\ge \sqrt{n} \times SNR(0,\widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1) \times \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)})\|_{(s_0,2)} - C_1 s_0^{1/2} M\sqrt{\log(pn)} \\ &\quad -C_2 s_0^{1/2} s\sqrt{\log p}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)})\|_{(s_0,2)} - C_3 s_0^{1/2} s^2\sqrt{\log(pn)}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)})\|_{(s_0,2)}^2 \\ &\qquad -C_4 M s_0^{1/2} s\sqrt{\log(p)}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)})\|_{(s_0,2)} \\ &\ge \sqrt{n} \times SNR(0,\widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1) \times \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)})\|_{(s_0,2)}(1-\epsilon_n) - C_1 s_0^{1/2} M\sqrt{\log(pn)}, \end{aligned} \tag{S7.123}$$

where

$$\epsilon_n := O\big(s_0^{1/2} s\sqrt{\frac{\log p}{n}}\big) \vee O\big(s_0^{1/2} s^2\sqrt{\frac{\log p}{n}}\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\|_{(s_0,2)}\big).$$

Hence, considering (S7.123), to prove (S7.105), it is sufficient to prove $H_1' \to 1$, where

$$\begin{aligned} H_1' = \mathbb{P}\Big(&\sqrt{n} \times SNR(0,\widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1) \times \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)})\|_{(s_0,2)} \\ &\ge \frac{C s_0^{1/2} M\big(\sqrt{\log(pn)}+\sqrt{\log(1/\gamma)}\big)}{1-\epsilon_n}\Big). \end{aligned}$$

By (3.37), it is straightforward to see that $H_1' \to 1$ as $n, p \to \infty$, which finishes the proof.

$\square$

Remark 3. Note that for $\alpha = 0$, if we replace $\sigma^2(0, \widetilde{\boldsymbol{\tau}})$ by an estimator $\widehat{\sigma}^2(0, \widetilde{\boldsymbol{\tau}})$ which satisfies: $|\widehat{\sigma}^2(\alpha, \widetilde{\boldsymbol{\tau}}) - \sigma^2(0, \widetilde{\boldsymbol{\tau}})| = o_p(1)$, then under condition (3.37), the change point test is still consistent.

S7.5.3 Power analysis for $\alpha \in (0, 1)$

Proof. In what follows, we assume $\widehat{\sigma}^2(\alpha, \widetilde{\boldsymbol{\tau}}) = \sigma^2(\alpha, \widetilde{\boldsymbol{\tau}})$ by considering the variance as unknown, where $\sigma^2(\alpha, \widetilde{\boldsymbol{\tau}}) := \mathrm{Var}[(1-\alpha)e_i(\widetilde{\boldsymbol{\tau}}) - \alpha\epsilon_i]$. In addition, for the case of $\alpha \in (0, 1)$, we define

$$SNR(\alpha, \widetilde{\boldsymbol{\tau}}) := \frac{(1-\alpha)\big(\frac{1}{K}\sum_{k=1}^{K} f_\epsilon(b_k^{(0)})\big) + \alpha}{\sigma(\alpha, \widetilde{\boldsymbol{\tau}})}.$$

Similar to Sections S7.5.1 and S7.5.2 , the proof of Theorem 5 proceeds in four steps. In Step 1, we obtain the upper bound of $c_{T_\alpha^b}(1-\gamma)$, where $c_{T_\alpha^b}(1-\gamma)$ is the $(1-\gamma)$-th quantile of T_α^b, which is defined as

$$c_{T_\alpha^b}(1-\gamma) := \inf\big\{t : \mathbb{P}(T_\alpha^b \le t) \ge 1-\gamma\big\}. \tag{S7.124}$$

In Steps 2-4, using the upper bound, we get the lower bound of $\mathbb{P}\big(T_\alpha \ge c_{T_\alpha^b}(1-\gamma)\big)$ and prove

$$\mathbb{P}\big(T_\alpha \ge c_{T_\alpha^b}(1-\gamma)\big) \to 1, \text{ as } n, p \to \infty. \tag{S7.125}$$

Step 1: By the definition of T_α^b, we have: $T_\alpha^b = \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_\alpha^b(t)\|_{(s_0,2)}$, with

$$\boldsymbol{C}_\alpha^b(t) := \frac{1}{\sqrt{n}v(\alpha,\widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor}\boldsymbol{X}_i((1-\alpha)e_i^b(\widetilde{\boldsymbol{\tau}})-\alpha e_i^b)-\frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n}\boldsymbol{X}_i((1-\alpha)e_i^b(\widetilde{\boldsymbol{\tau}})-\alpha e_i^b)\Big),$$

where $e_i^b(\widetilde{\boldsymbol{\tau}}) = K^{-1}\sum_{k=1}^{K} e_i^b(\tau_k)$ with $e_i^b(\tau_k) := \mathbf{1}\{\epsilon_i^b \le \Phi^{-1}(\tau_k)\} - \tau_k$, e_i^b is i.i.d $N(0,1)$, and $\Phi(x)$ is the CDF for the standard normal distribution, and $v^2(\alpha,\widetilde{\boldsymbol{\tau}}) := \mathrm{Var}[(1-\alpha)e_i^b(\widetilde{\boldsymbol{\tau}}) - \alpha e_i^b]$.

Note that $(1-\alpha)e_i^b(\widetilde{\boldsymbol{\tau}}) - \alpha e_i^b$ is just a linear combination of a bounded random variable $e_i^b(\widetilde{\boldsymbol{\tau}})$ and a standard normal distribution. Hence, using a very similar proof procedure as in Step 1 in Section S7.5.1, one can prove

$$c_{T_\alpha^b}(1-\gamma) \le C(M, q_0, \widetilde{\boldsymbol{\tau}}, \alpha)s_0^{1/2}(\sqrt{\log(pn)} + \sqrt{\log(1/\gamma)}), \qquad \text{(S7.126)}$$

where $C(M, q_0, \widetilde{\boldsymbol{\tau}}, \alpha)$ is some universal constant only depending on M, q_0, $\widetilde{\boldsymbol{\tau}}$ and α.

Step 2 Decomposition of $\boldsymbol{C}_\alpha(t)$. In this step, we aim to prove that $\mathbb{P}\big(T_\alpha \ge c_{T_\alpha^b}(1-\gamma)\big) \to 1$ as $n, p \to \infty$. Note that in Step 1, we have obtained the upper bound of $c_{T_\alpha^b}(1-\gamma)$. Hence, it is sufficient to prove that $H_1 \to 1$ as $n, p \to \infty$, where

$$H_1 := \mathbb{P}\big(T_\alpha \ge C(M, q_0, \widetilde{\boldsymbol{\tau}}, \alpha)s_0^{1/2}(\sqrt{\log(pn)} + \sqrt{\log(1/\gamma)})\big). \qquad \text{(S7.127)}$$

To prove $H_1 \to 1$, we need the decomposition of T_α under $\mathbf{H}_1$. Note that for $\alpha \in (0,1)$, with known variance, the score based CUSUM process reduces to:

$$\boldsymbol{C}_\alpha(t) = \frac{1}{\sqrt{n}\sigma(\alpha, \widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n} \boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}})\Big), \tag{S7.128}$$

where $\boldsymbol{Z}(\boldsymbol{X}_i, Y_i; \widetilde{\boldsymbol{\tau}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\beta}}) := (1-\alpha)\frac{1}{K}\sum_{k=1}^{K} \boldsymbol{X}_i\big(\mathbf{1}\{Y_i - \widehat{b}_k - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}} \le 0\} - \tau_k\big) - \alpha \boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}})$.

Using the results obtained in Sections S7.5.1 and S7.5.2, we have the following decomposition:

$$\boldsymbol{C}_\alpha(t) = \widetilde{\boldsymbol{C}}_\alpha^I(t) - SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \boldsymbol{\delta}(t) - SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \boldsymbol{R}(t) + (1-\alpha)\boldsymbol{C}_0^{(3)}(t) + (1-\alpha)\boldsymbol{C}_0^{(4)}(t), \tag{S7.129}$$

where $\widetilde{\boldsymbol{C}}_\alpha^I(t)$ is the random noise based partial sum process defined in (S7.47), $\boldsymbol{\delta}(t)$ is the signal jump defined in (S7.99), $\boldsymbol{R}(t)$ is defined in (S7.100), and $\boldsymbol{C}_0^{(3)}(t)$ and $\boldsymbol{C}_0^{(4)}(t)$ are defined in (S7.114).

Step 3: Obtain the upper bounds for the residuals and random noises in $\boldsymbol{C}_\alpha(t)$. We first bound $\max_t \|\widetilde{\boldsymbol{C}}_\alpha^I(t)\|_{(s_0,2)}$. Note that by its definition in (S7.47), we have:

$$\begin{aligned}
&\max_{t\in[q_0,1-q_0]} \|\widetilde{\boldsymbol{C}}^I_\alpha(t)\|_{(s_0,2)}\\
&= \max_{t\in[q_0,1-q_0]} \|(1-\alpha)\widetilde{\boldsymbol{C}}^I_0(t) - \alpha\widetilde{\boldsymbol{C}}^I_1(t)\|_{(s_0,2)}\\
&\le (1-\alpha)\max_{t\in[q_0,1-q_0]} \|\widetilde{\boldsymbol{C}}^I_0(t)\|_{(s_0,2)} + \alpha \max_{t\in[q_0,1-q_0]} \|\widetilde{\boldsymbol{C}}^I_1(t)\|_{(s_0,2)}\\
&\le (1-\alpha)s_0^{1/2}\max_{t\in[q_0,1-q_0]} \|\widetilde{\boldsymbol{C}}^I_0(t)\|_\infty + \alpha s_0^{1/2} \max_{t\in[q_0,1-q_0]} \|\widetilde{\boldsymbol{C}}^I_1(t)\|_\infty\\
&= O_p(s_0^{1/2} M\sqrt{\log(pn)}),
\end{aligned}$$

where the last equation comes from Lemma 7. Next, we consider $\max_t \|\boldsymbol{R}(t)\|_{(s_0,2)}$. Using Lemma 11, Remark 6, and using a similar proof procedure as Step 2 in Section S7.5.1, we have

$$\max_{q_0\le t\le 1-q_0} \|\boldsymbol{R}(t)\|_{(s_0,2)} = O_p\big(s_0^{1/2} s\sqrt{\log p}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)}\big).$$

For $\boldsymbol{C}^{(3)}_0(t)$ and $\boldsymbol{C}^{(4)}_0(t)$, using the obtained upper bounds in (S7.121) and (S7.122), we have:

$$\begin{aligned}
&\max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}^{(3)}_0(t)\|_{(s_0,2)} = O_p(s_0^{1/2}s^2\sqrt{\log(pn)}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})\|^2_{(s_0,2)}), \quad \text{and}\\
&\max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}^{(4)}_0(t)\|_{(s_0,2)} = O_p\big(Ms_0^{1/2}s\sqrt{\log(pn)}\|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})\|_{(s_0,2)}\big).
\end{aligned}$$

Step 4: Combining the previous results. Recall (S7.126), (S7.127), (S7.116). Using the above bounds of $\widetilde{\boldsymbol{C}}^I_\alpha(t)$, $\boldsymbol{R}(t)$, $\boldsymbol{C}^{(3)}_{(0)}(t)$, $\boldsymbol{C}^{(4)}_{(0)}(t)$, and by the triangle inequality, w.p.a.1, we have:

$$
\begin{aligned}
T_\alpha &= \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_\alpha(t)\|_{(s_0,2)} \\
&\ge SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{\delta}(t)\|_{(s_0,2)} - \max_{q_0 \le t \le 1-q_0} \|\widetilde{\boldsymbol{C}}^I_\alpha(t)\|_{(s_0,2)} \\
&\quad - SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{R}(t)\|_{(s_0,2)} - (1-\alpha) \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}^{(3)}_0(t)\|_{(s_0,2)} \\
&\quad - (1-\alpha) \max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}^{(4)}_0(t)\|_{(s_0,2)} \\
&\ge \sqrt{n} \times SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1) \times \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)} - C_1 s_0^{1/2} M \sqrt{\log(pn)} \\
&\quad\ - C_2 s_0^{1/2} s \sqrt{\log p} \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)} - C_3(1-\alpha) s_0^{1/2} s^2 \sqrt{\log(pn)} \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})\|^2_{(s_0,2)} \\
&\qquad\quad - C_4(1-\alpha) M s_0^{1/2} s \sqrt{\log(p)} \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})\|_{(s_0,2)} \\
&\ge \sqrt{n} \times SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1) \times \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)}(1-\epsilon_n) - C_1 s_0^{1/2} M \sqrt{\log(pn)},
\end{aligned}
\tag{S7.130}
$$

where

$$
\epsilon_n := O\big(s_0^{1/2} s \sqrt{\frac{\log p}{n}}\big) \vee O\big(s_0^{1/2} s^2 \sqrt{\frac{\log p}{n}} \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_{(s_0,2)}\big).
$$

Hence, considering (S7.123), to prove (S7.105), it is sufficient to prove $H_1' \to 1$, where

$$
\begin{aligned}
H_1' = \mathbb{P}\Big(&\sqrt{n} \times SNR(\alpha, \widetilde{\boldsymbol{\tau}}) \times t_1(1-t_1) \times \|\boldsymbol{\Sigma}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})\|_{(s_0,2)} \\
&\ge \frac{C s_0^{1/2} M \big(\sqrt{\log(pn)} + \sqrt{\log(1/\gamma)}\big)}{1-\epsilon_n}\Big).
\end{aligned}
$$

By (3.37), it is straightforward to see that $H_1' \to 1$ as $n, p \to \infty$, which finishes the proof.

Remark 4. Note that for $\alpha \in (0,1)$, if we replace $\sigma^2(\alpha, \widetilde{\boldsymbol{\tau}})$ by an estimator

$\widehat{\sigma}^2(\alpha, \widetilde{\boldsymbol{\tau}})$ which satisfies: $|\widehat{\sigma}^2(\alpha, \widetilde{\boldsymbol{\tau}}) - \sigma^2(\alpha, \widetilde{\boldsymbol{\tau}})| = o_p(1)$, then under condition (3.37), the power still converges to 1.

$\square$

S8 Proofs of lemmas in Section S7

S8.1 Proof of Lemma 12

Proof. In this section, we prove Lemma 12. In other words, we will prove

$$\mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_1(t) - \widetilde{\boldsymbol{C}}_1^I(t)\big\|_{(s_0,2)} \ge \epsilon\big) = o(1). \tag{S8.131}$$

Using the triangle inequality, we have

$$\mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_1(t) - \widetilde{\boldsymbol{C}}_1^I(t)\big\|_{(s_0,2)} \ge \epsilon\big) \le D_1 + D_2, \tag{S8.132}$$

where D_1 and D_2 are defined as

$$\begin{aligned} D_1 &:= \mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_1(t) - \boldsymbol{C}_1^I(t)\big\|_{(s_0,2)} \ge \epsilon/2\big), \\ D_2 &:= \mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_1^I(t) - \widetilde{\boldsymbol{C}}_1^I(t)\big\|_{(s_0,2)} \ge \epsilon/2\big). \end{aligned} \tag{S8.133}$$

By (S8.132), to prove (S8.131), we need to bound D_1 and D_2, respectively.

Step 1: Obtain the upper bound for D_1. We first consider D_1. To this end, we define

$$\mathcal{E} = \big\{c_\epsilon^2/4 \le \widehat{\sigma}^2 \le 4C_\epsilon^2\big\}, \tag{S8.134}$$

where c_ϵ and C_ϵ are in **Assumption B.** By introducing $\mathcal{E}$, we have

$$D_1 \le \mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_1(t) - \boldsymbol{C}_1^I(t)\big\|_{(s_0,2)} \ge \epsilon/2 \cap \mathcal{E}\big) + \mathbb{P}(\mathcal{E}^c). \tag{S8.135}$$

By Theorem 1, we have $\mathbb{P}(\mathcal{E}^c) = o(1)$ as $n, p \to \infty$. Under the event $\mathcal{E}$, we have

$$\begin{aligned} &\mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_1(t) - \boldsymbol{C}_1^I(t)\big\|_{(s_0,2)} \ge \epsilon/2 \cap \mathcal{E}\big) \\ &= \mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_1^{II}(t)\big\|_{(s_0,2)} \ge \epsilon/2 \cap \mathcal{E}\big), \end{aligned} \tag{S8.136}$$

where $\boldsymbol{C}_1^{II}(t)$ is defined in (S7.23). Hence, under the event $\mathcal{E}$, we have

$$\begin{aligned} &\mathbb{P}\big(\max_{q_0 \le t \le 1-q_0} \|\boldsymbol{C}_1^{II}(t)\|_{(s_0,2)} \ge \frac{\epsilon}{2} \cap \mathcal{E}\big) \\ &=_{(1)} \mathbb{P}\Big(\max_t \Big\|\frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i\boldsymbol{X}_i^\top(\boldsymbol{\beta}^{(0)} - \widehat{\boldsymbol{\beta}}) - \frac{\lfloor nt \rfloor}{n}\sum_{i=1}^{n} \boldsymbol{X}_i\boldsymbol{X}_i^\top(\boldsymbol{\beta}^{(0)} - \widehat{\boldsymbol{\beta}})\Big\|_{(s_0,2)} \ge \frac{c_\epsilon \epsilon}{4} \cap \mathcal{E}\Big) \\ &\le_{(2)} \mathbb{P}\Big(\max_t \Big\|\frac{\lfloor nt \rfloor}{\sqrt{n}}\big(\widehat{\boldsymbol{\Sigma}}(1:t) - \widehat{\boldsymbol{\Sigma}}(1:n)\big)(\boldsymbol{\beta}^{(0)} - \widehat{\boldsymbol{\beta}})\Big\|_{(s_0,2)} \ge \frac{c_\epsilon \epsilon}{4} \cap \mathcal{E}\Big) \\ &\le_{(3)} \mathbb{P}\Big(\max_t \Big\|\frac{\lfloor nt \rfloor}{\sqrt{n}}\big(\widehat{\boldsymbol{\Sigma}}(1:t) - \widehat{\boldsymbol{\Sigma}}(1:n)\big)(\boldsymbol{\beta}^{(0)} - \widehat{\boldsymbol{\beta}})\Big\|_{\infty} \ge s_0^{-1/2}\frac{c_\epsilon \epsilon}{4} \cap \mathcal{E}\Big) \\ &\le_{(4)} \mathbb{P}\Big(\max_t \Big\|\frac{\lfloor nt \rfloor}{\sqrt{n}}\big(\widehat{\boldsymbol{\Sigma}}(1:t) - \widehat{\boldsymbol{\Sigma}}(1:n)\big)\Big\|_{\infty}\big\|(\boldsymbol{\beta}^{(0)} - \widehat{\boldsymbol{\beta}})\big\|_1 \ge s_0^{-1/2}\frac{c_\epsilon \epsilon}{4}\Big) \\ &\le_{(5)} \mathbb{P}\Big(\max_t \big\|\big(\widehat{\boldsymbol{\Sigma}}(1:t) - \widehat{\boldsymbol{\Sigma}}(1:n)\big)\big\|_{\infty}\big\|(\boldsymbol{\beta}^{(0)} - \widehat{\boldsymbol{\beta}})\big\|_1 \ge n^{-1/2}s_0^{-1/2}\frac{c_\epsilon \epsilon}{4}\Big), \end{aligned} \tag{S8.137}$$

where (3) comes from the fact that $\|\boldsymbol{v}\|_{(s_0,2)} \le s_0^{1/2}\|\boldsymbol{v}\|_\infty$ for any $\boldsymbol{v} \in \mathbb{R}^p$, (4) comes from the fact that $\|\mathbf{A}\boldsymbol{v}\|_\infty \le \|\mathbf{A}\|_\infty\|\boldsymbol{v}\|_1$ for any matrix $\mathbf{A}$ and vector $\boldsymbol{v}$. By Lemma 8, we have $\max_t \big\|\big(\widehat{\boldsymbol{\Sigma}}(1:t) - \widehat{\boldsymbol{\Sigma}}(1:n)\big)\big\|_\infty = O_p(M^2\sqrt{\log(p)/n})$. Moreover, under $\mathbf{H}_0$, for the lasso estimator $\widehat{\boldsymbol{\beta}}$, using Lemma 10, we have $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_1 = O_p(s\sqrt{\log(p/n)})$. Hence, combining (S8.137) and letting $\epsilon := Cs_0^{1/2}sM^2\log(p)/\sqrt{n}$ for some big enough constant

$C > 0$, we have:

$$\mathbb{P}\Big(\max_{q_0 \le t \le 1-q_0} \big\| \boldsymbol{C}_1^{II}(t) \big\|_{(s_0,2)} \ge \frac{\epsilon}{2} \cap \mathcal{E}\Big) = o(1).$$

Step 2: Obtain the upper bound for D_2. By definition, we have

$$\begin{aligned}
D_2 &:= \mathbb{P}\Big(\max_{q_0 \le t \le 1-q_0} \big\| \boldsymbol{C}_1^{I}(t) - \widetilde{\boldsymbol{C}}_1^{I}(t) \big\|_{(s_0,2)} \ge \epsilon/2\Big) \\
&= \mathbb{P}\Big(\big|\frac{1}{\widehat{\sigma}} - \frac{1}{\sigma}\big| \max_{q_0 \le t \le 1-q_0} \big\| \frac{1}{\sqrt{n}} \big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \epsilon_i \big) \big\|_{(s_0,2)} \ge \epsilon/2\Big) \\
&= \mathbb{P}\Big(\underbrace{\big|\frac{\sigma}{\widehat{\sigma}} - 1\big|}_{I_1} \underbrace{\max_{q_0 \le t \le 1-q_0} \big\| \frac{1}{\sqrt{n}\sigma} \big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \epsilon_i \big) \big\|_{(s_0,2)}}_{I_2} \ge \epsilon/2\Big).
\end{aligned}$$

Hence, to bound D_2, we need to bound I_1 and I_2, respectively. To bound I_1, define $\widetilde{I}_1 = |1 - \frac{\widehat{\sigma}}{\sigma}|$. Using the fact that $a^2 - b^2 = (a-b)(a+b)$, we have:

$$\widetilde{I}_1 = \Big| \frac{\widehat{\sigma}^2 - \sigma^2}{\sigma(\sigma + \widehat{\sigma})} \Big| \le_{(1)} \Big| \frac{\widehat{\sigma}^2 - \sigma^2}{\sigma^2} \Big| \le_{(2)} C_1 \Big| \widehat{\sigma}^2 - \sigma^2 \Big| \le_{(3)} \le C_2 \sqrt{s \frac{\log(pn)}{n}},$$

where (2) comes from **Assumption B**, (3) comes from Theorem 1. By Lemma C.1 in Zhou et al. (2018), we have: $I_1 \le C\widetilde{I}_1$. Next, we consider I_2. Using Lemma 7, and the fact that $\|\boldsymbol{v}\|_{(s_0,2)} \le s_0^{1/2} \|\boldsymbol{v}\|_\infty$ for any $\boldsymbol{v} \in \mathbb{R}^p$, we have:

$$I_2 = O_p\big(M s_0^{1/2} \sqrt{\log(pn)}\big).$$

Hence, we have $I_1 I_2 = O_p(s_0^{1/2} s^{1/2} M \frac{\log(pn)}{\sqrt{n}})$.

Lastly, combining Steps 1 and 2, if we choose $\epsilon := C s_0^{1/2} s M^2 \log(p)/\sqrt{n}$ for some big constant $C > 0$, we have $D_1 + D_2 = o(1)$, which finishes the proof. □

S8.2 Proof of Lemma 13

Proof. In this section, we aim to prove $\sup_{z>0} I_z = o_p(1)$, where

$$I_z := \big|\mathbb{P}\big(\max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\|_{(s_0,2)} > z\big) - \mathbb{P}\big(\max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_1^{b}(t)\|_{(s_0,2)} > z|\mathcal{X}\big)\big|.$$

To that end, let $R = Cs_0 n$ and $L = \sup_{z\in(0,+\infty)} I_z$. Then, we can write L as

$$L = \max(L_1, L_2),$$

where $L_1 = \sup_{z\in(0,R]} I_z$ and $L_2 = \sup_{z\in(R,\infty)} I_z$. Therefore, to prove $L = o_p(1)$, we need to bound L_1 and L_2, respectively. We first bound $L_2 = \sup_{z\in(R,\infty)} I_z$. Considering that for any $\boldsymbol{v} \in \mathbb{R}^p$, $\|\boldsymbol{v}\|_{(s_0,2)} \le s_0^{1/2}\|\boldsymbol{v}\|_\infty \le s_0\|\boldsymbol{v}\|_\infty$ holds, we have

$$L_2 = \sup_{z\in(R,\infty)} I_z \le \mathbb{P}\big(\max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\|_\infty > Cn\big)+\mathbb{P}\big(\max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_1^{b}(t)\|_\infty > Cn|\mathcal{X}\big).$$

By the exponential inequality and similar to the proof of Lemma 7, we can prove that

$$\max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_1^{\boldsymbol{G}}(t)\|_\infty = O_p(M\sqrt{\log(pn)}), \text{and } \max_{q_0\le t\le 1-q_0} \|\boldsymbol{C}_1^{b}(t)\|_\infty|\mathcal{X} = O_p(M\sqrt{\log(pn)}),$$

which yields

$$L_2 = \sup_{z\in(R,\infty)} I_z = o_p(1). \tag{S8.138}$$

After bounding L_2 in (S8.138), we now bound $L_1 := \sup_{z\in(0,R]} I_z$. Let $\mathcal{E}^{R,p} = \{\boldsymbol{x} \in \mathbb{R}^p : \|\boldsymbol{x}\| \le R\}$ and $V^{z,p}_{(s_0,2)} = \{\boldsymbol{x} \in \mathbb{R}^p : \|\boldsymbol{x}\|_{(s_0,2)} \le z\}$.

Considering $\|\boldsymbol{x}\| \leq p^{1/2}\|\boldsymbol{x}\|_\infty \leq p^{1/2}\|\boldsymbol{x}\|_{(s_0,2)}$ for any $\boldsymbol{x} \in \mathbb{R}^p$, we have $V_{(s_0,2)}^{z,p} \subset \mathcal{E}^{Rp^{1/2},p}$ for $z \leq R$. Therefore, considering Lemma 5, there is a m-generated convex set A^m and a $\epsilon > 0$ such that

$$A^m \subset V_{(s_0,2)}^{z,p} \subset A^{m,Rp^{1/2}\epsilon} \text{ and } m \leq p^{s_0}(\frac{\gamma}{\sqrt{\epsilon}}\ln(\frac{1}{\epsilon}))^{s_0^2} \ \big(\text{by } V_{(s_0,2)}^{z,p} \subset \mathcal{E}^{Rp^{1/2},p} \text{ for } z \leq R\big). \tag{S8.139}$$

By setting $\epsilon = (pn)^{-3/2}$, we have $\epsilon' = Rp^{1/2}\epsilon = Cs_0p^{-1}n^{-1/2}$. By (S8.139), for $z \in (0, R]$, we have

$$I_z \leq I_{z,1} + I_{z,2},$$

where

$$I_{z,1} = \max\Big(\mathbb{P}\big(\bigcap_{q_0 \leq t \leq 1-q_0} \boldsymbol{C}_1^{\boldsymbol{G}}(t) \in A^{m,\epsilon'} \setminus A^m\big), \mathbb{P}\big(\bigcap_{q_0 \leq t \leq 1-q_0} \boldsymbol{C}_1^{b}(t) \in A^{m,\epsilon'} \setminus A^m|\mathcal{X}\big)\Big),$$

$$\begin{aligned} I_{z,2} = \ & \max\Big(\big|\mathbb{P}\big(\bigcap_{q_0 \leq t \leq 1-q_0} \boldsymbol{C}_1^{\boldsymbol{G}}(t) \in A^{m,\epsilon'}\big) - \mathbb{P}\big(\bigcap_{q_0 \leq t \leq 1-q_0} \boldsymbol{C}_1^{b}(t) \in A^{m,\epsilon'}|\mathcal{X}\big)\big|, \\ & \big|\mathbb{P}\big(\bigcap_{q_0 \leq t \leq 1-q_0} \boldsymbol{C}_1^{\boldsymbol{G}}(t) \in A^{m}\big) - \mathbb{P}\big(\bigcap_{q_0 \leq t \leq 1-q_0} \boldsymbol{C}_1^{b}(t) \in A^{m}|\mathcal{X}\big)\big|\Big). \end{aligned}$$

Next, we consider $I_{z,1}$ and $I_{z,2}$, respectively. Recall $\epsilon' = Cs_0p^{-1}n^{-1/2}$. For $I_{z,1}$, by Lemma 4 and the definitions of A^m and $A^{m,\epsilon'}$ in (S6.5) and (S8.139), for all $z \in (0, R]$, we have

$$I_{z,1} \leq Cs_0p^{-1}n^{-1/2}\sqrt{\log(m(n-2\lfloor nq_0\rfloor))} \leq Cs_0^2p^{-1}n^{-1/2}\sqrt{\log(pn)} = o_p(1). \tag{S8.140}$$

Recall $\mathcal{V}_{s_0} := \{\boldsymbol{v} \in \mathbb{S}^{q-1} : \|\boldsymbol{v}\| = 1, \|\boldsymbol{v}\|_0 \leq s_0\}$ and $\widehat{\boldsymbol{\Sigma}}(0:t)$ defined in

(S4.2). We then have

$$\begin{aligned}
&\sup_{q_0\le t_1,t_2\le 1-q_0}\ \sup_{\boldsymbol{v}_1,\boldsymbol{v}_2\in\mathcal{V}_{s_0}} \Big|\boldsymbol{v}_1^\top\Big(\mathbb{E}\big[\boldsymbol{C}_1^{\boldsymbol{G}}(t_1)\boldsymbol{C}_1^{\boldsymbol{G}}(t_2)^\top\big]-\mathbb{E}\big[\boldsymbol{C}_1^{b}(t_1)\boldsymbol{C}^{b}(t_2)^\top\big|\mathcal{X}\big]\Big)\boldsymbol{v}_2\Big|\\
&\le_{(1)} \sup_{q_0\le t_1,t_2\le 1-q_0}\Big|\mathbb{E}\big[\boldsymbol{C}_1^{\boldsymbol{G}}(t_1)\boldsymbol{C}_1^{\boldsymbol{G}}(t_2)^\top\big]-\mathbb{E}\big[\boldsymbol{C}_1^{b}(t_1)\boldsymbol{C}^{b}(t_2)^\top\big|\mathcal{X}\big]\Big\|_\infty\big\|\boldsymbol{v}_1\big\|_1\big\|\boldsymbol{v}_2\big\|_1\\
&\le_{(2)} s_0\sup_{q_0\le t_1,t_2\le 1-q_0}\big\|\min(t_1,t_2)(\widehat{\boldsymbol{\Sigma}}(0:\min(t_1,t_2))-\boldsymbol{\Sigma})\\
&\qquad -t_1t_2(\widehat{\boldsymbol{\Sigma}}(0:t_1)-\boldsymbol{\Sigma})-t_1t_2(\widehat{\boldsymbol{\Sigma}}(0:t_2)-\boldsymbol{\Sigma})+t_1t_2(\widehat{\boldsymbol{\Sigma}}(0:1)-\boldsymbol{\Sigma})\big\|_\infty,
\end{aligned} \tag{S8.141}$$

the last inequality in (S8.141) comes from the Cauchy-Schwartz inequality, and the fact $\boldsymbol{v}_1,\boldsymbol{v}_2\in\mathcal{V}_{s_0}$. Therefore, based on (S8.141), using Theorem 4.1 and Remark 4.1 in Chernozhukov et al. (2017) and Lemma 8, with probability tending to one, we have

$$I_{z,2}\le C\Big(s_0M^2\sqrt{\frac{\log(pn)}{n}}\Big)^{1/3}\log^{2/3}\big(m(n-2\lfloor nq_0\rfloor)\big)\le C\Big(\frac{s_0^{10}\log^7(pn)}{n}\Big)^{1/6}. \tag{S8.142}$$

Considering (S8.142), by **Assumptions A, E.1**, we have $I_{z,2}=o_p(1)$ for all $z\in(0,R]$.

Finally, combining (S8.138), (S8.140), and (S8.142), we have $I_z=o_p(1)$ uniformly holds for $z\ge 0$, which finishes the proof for Lemma 13.

□

S8.3 Proof of Lemma 14

Proof. In this section, we prove Lemma 14. In other words, we aim to prove

$$\mathbb{P}\Big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_0(t)-\widetilde{\boldsymbol{C}}_0^I(t)\big\|_{(s_0,2)}\ge\epsilon\Big)=o(1), \tag{S8.143}$$

where $\boldsymbol{C}_0(t)$ is defined in (S7.37), and $\widetilde{\boldsymbol{C}}_0^I(t)$ is defined in (S7.40). Using the triangle inequality, we have

$$\mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_0(t)-\widetilde{\boldsymbol{C}}_0^I(t)\big\|_{(s_0,2)}\ge\epsilon\big)\le D_1+D_2, \tag{S8.144}$$

where D_1 and D_2 are defined as

$$\begin{aligned} D_1 &:= \mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_0(t)-\boldsymbol{C}_0^I(t)\big\|_{(s_0,2)}\ge\epsilon/2\big),\\ D_2 &:= \mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_0^I(t)-\widetilde{\boldsymbol{C}}_0^I(t)\big\|_{(s_0,2)}\ge\epsilon/2\big). \end{aligned} \tag{S8.145}$$

By (S8.144), to prove (S8.143), we need to bound D_1 and D_2, respectively.

Step 1: Obtain the upper bound for D_1. We first consider D_1. To this end, we define

$$\mathcal{E}=\big\{\sigma^2/2\le\widehat{\sigma}^2\le 2\sigma^2\big\}, \tag{S8.146}$$

where $\sigma^2 := \mathrm{Var}[e_i(\widetilde{\boldsymbol{\tau}})]$ is the true variance. By introducing $\mathcal{E}$, we have

$$D_1\le\mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_0(t)-\boldsymbol{C}_0^I(t)\big\|_{(s_0,2)}\ge\epsilon/2\cap\mathcal{E}\big)+\mathbb{P}(\mathcal{E}^c). \tag{S8.147}$$

By Theorem 1, we have $\mathbb{P}(\mathcal{E}^c)=o(1)$ as $n,p\to\infty$. Under the event $\mathcal{E}$, we have

$$\begin{aligned} &\mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_0(t)-\boldsymbol{C}_0^I(t)\big\|_{(s_0,2)}\ge\epsilon/2\cap\mathcal{E}\big)\\ &=\mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_0^{II}(t)\big\|_{(s_0,2)}\ge\epsilon/2\cap\mathcal{E}\big), \end{aligned} \tag{S8.148}$$

where $\boldsymbol{C}_0^{II}(t)$ is defined in (S7.38). Before controlling $\boldsymbol{C}_0^{II}(t)$, given $\mathcal{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, we need to decompose $\boldsymbol{C}_0^{II}(t)|\mathcal{X}$ into two terms:

$$\boldsymbol{C}_0^{II}(t) = \boldsymbol{C}_0^{II,1}(t) + \boldsymbol{C}_0^{II,2}(t), \tag{S8.149}$$

where $\boldsymbol{C}_0^{II,1}(t)$ and $\boldsymbol{C}_0^{II,2}(t)$ are defined as

$$\begin{aligned}
\boldsymbol{C}_0^{II,1}(t) = \frac{1}{\sqrt{n}\widehat{\sigma}(0,\widetilde{\boldsymbol{\tau}})}\Big(&\sum_{i=1}^{\lfloor nt\rfloor}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(\mathbb{E}[\mathbf{1}\{\epsilon_i \le \underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k} + b_k^{(0)}\}] - \mathbb{E}[\mathbf{1}\{\epsilon_i \le b_k^{(0)}\}]\big) \\
&-\frac{\lfloor nt\rfloor}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(\mathbb{E}[\mathbf{1}\{\epsilon_i \le \underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k} + b_k^{(0)}\}] - \mathbb{E}[\mathbf{1}\{\epsilon_i \le b_k^{(0)}\}]\big)\Big), \\
\boldsymbol{C}_0^{II,2}(t) = \frac{1}{\sqrt{n}\widehat{\sigma}(0,\widetilde{\boldsymbol{\tau}})}\Big(&\sum_{i=1}^{\lfloor nt\rfloor}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(g_{ik}(\underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}) - g_{ik}(0)\big) \\
&-\frac{\lfloor nt\rfloor}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(g_{ik}(\underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}) - g_{ik}(0)\big)\Big),
\end{aligned} \tag{S8.150}$$

where $g_{ik}(t) := \mathbf{1}\{\epsilon_i \le b_k^{(0)} + t\} - \mathbb{P}\{\epsilon_i \le b_k^{(0)} + t\}$. Next, we control $\boldsymbol{C}_0^{II,1}(t)$ and $\boldsymbol{C}_0^{II,2}(t)$, respectively.

Let $F_\epsilon(t) := \mathbb{P}(\epsilon \le t)$ be the CDF for ϵ and f_ϵ be its density function. For $\boldsymbol{C}_0^{II,1}(t)$, by its definition, we have:

$$\begin{aligned}
\boldsymbol{C}_0^{II,1}(t) = \frac{1}{\sqrt{n}\widehat{\sigma}(0,\widetilde{\boldsymbol{\tau}})}\Big(&\sum_{i=1}^{\lfloor nt\rfloor}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(F_\epsilon(\underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k} + b_k^{(0)}) - F_\epsilon(b_k^{(0)})\big) \\
&-\frac{\lfloor nt\rfloor}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(F_\epsilon(\underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k} + b_k^{(0)}) - F_\epsilon(b_k^{(0)})\big)\Big).
\end{aligned} \tag{S8.151}$$

Using the Taylor's expansion, we have:

$$
\begin{aligned}
&F_\epsilon(\underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k} + b_k^{(0)}) - F_\epsilon(b_k^{(0)}) \\
&= f_\epsilon(b_k^{(0)}) \underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k} + \frac{1}{2} f'_\epsilon(\xi_{ik}) (\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2 \\
&= f_\epsilon(b_k^{(0)}) \boldsymbol{X}_i^\top \widehat{\boldsymbol{\Delta}} + f_\epsilon(b_k^{(0)})(\widehat{b_k} - b_k^{(0)}) + \frac{1}{2} f'_\epsilon(\xi_{ik}) (\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2 \\
&= f_\epsilon(b_k^{(0)}) \boldsymbol{X}_i^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}) + f_\epsilon(b_k^{(0)})(\widehat{b_k} - b_k^{(0)}) + \frac{1}{2} f'_\epsilon(\xi_{ik}) (\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2,
\end{aligned}
$$

where ξ_{ik} is some random variable between $b_k^{(0)}$ and $\underset{\sim}{\boldsymbol{X}_i}^\top \underset{\sim}{\widehat{\boldsymbol{\Delta}}_k} + b_k^{(0)}$. Hence, by the above expansion, $\boldsymbol{C}_0^{II,1}(t)$ can be decomposed into three terms:

$$
\boldsymbol{C}_0^{II,1}(t) = \boldsymbol{C}_0^{II,1,1}(t) + \boldsymbol{C}_0^{II,1,2}(t) + \boldsymbol{C}_0^{II,1,3}(t), \tag{S8.152}
$$

where $\boldsymbol{C}_0^{II,1,1}(t) - \boldsymbol{C}_0^{II,1,3}(t)$ are defined as

$$
\begin{aligned}
\boldsymbol{C}_0^{II,1,1}(t) &= \frac{1}{\sqrt{n}\widehat{\sigma}(0, \widetilde{\boldsymbol{\tau}})} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \boldsymbol{X}_i^\top \frac{1}{K} \sum_{k=1}^{K} f_\epsilon(b_k^{(0)})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}) \Big) \\
&\quad - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top \frac{1}{K} \sum_{k=1}^{K} f_\epsilon(b_k^{(0)})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}) \Big), \\
\boldsymbol{C}_0^{II,1,2}(t) &= \frac{1}{\sqrt{n}\widehat{\sigma}(0, \widetilde{\boldsymbol{\tau}})} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \frac{1}{K} \sum_{k=1}^{K} f_\epsilon(b_k^{(0)})(\widehat{b_k} - b_k^{(0)}) \Big) \\
&\quad - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \frac{1}{K} \sum_{k=1}^{K} f_\epsilon(b_k^{(0)})(\widehat{b_k} - b_k^{(0)})) \Big), \\
\boldsymbol{C}_0^{II,1,3}(t) &= \frac{1}{\sqrt{n}\widehat{\sigma}(0, \widetilde{\boldsymbol{\tau}})} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \frac{1}{K} \sum_{k=1}^{K} \frac{1}{2} f'_\epsilon(\xi_{ik}) (\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2 \Big) \\
&\quad - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \frac{1}{K} \sum_{k=1}^{K} \frac{1}{2} f'_\epsilon(\xi_{ik}) (\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2) \Big).
\end{aligned} \tag{S8.153}
$$

Hence, to bound $\boldsymbol{C}_0^{II,1}(t)$, we need to bound $\boldsymbol{C}_0^{II,1,1}(t) - \boldsymbol{C}_0^{II,1,3}(t)$ respectively. For $\boldsymbol{C}_0^{II,1,1}(t)$, under the event $\mathcal{E}$, with probability tending to 1, we have:

$$
\begin{aligned}
&\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,1,1}(t)\|_{(s_0,2)} \\
&\leq_{(1)} C s_0^{1/2} \max_t \Big\| \frac{\lfloor nt \rfloor}{\sqrt{n}} \big(\widehat{\boldsymbol{\Sigma}}(1:t) - \widehat{\boldsymbol{\Sigma}}(1:n)\big) \frac{1}{K}\sum_{k=1}^{K} f_\epsilon(b_k^{(0)})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}) \Big\|_\infty \\
&\leq_{(2)} C s_0^{1/2} \max_t \Big\| \frac{\lfloor nt \rfloor}{\sqrt{n}} \big(\widehat{\boldsymbol{\Sigma}}(1:t) - \widehat{\boldsymbol{\Sigma}}(1:n)\big) \|_\infty \times \| \frac{1}{K}\sum_{k=1}^{K} f_\epsilon(b_k^{(0)})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}) \Big\|_1 \\
&\leq_{(3)} C s_0^{1/2} M^2 \sqrt{\log(p)} \Big\| \frac{1}{K}\sum_{k=1}^{K} f_\epsilon(b_k^{(0)})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}) \Big\|_1 \\
&\leq_{(4)} C s_0^{1/2} M^2 \sqrt{\log(p)} \|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}) \|_1 \\
&\leq_{(5)} C s_0^{1/2} M^2 \sqrt{\log(p)}\, s \sqrt{\log(p)/n},
\end{aligned}
\tag{S8.154}
$$

where (1) comes from (S8.146), (3) comes from Lemma 8, (4) comes from Assumption D.2, and (5) comes from Lemma 10. With a similar procedure, we can prove that

$$
\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,1,2}(t)\|_{(s_0,2)} = O_p(s_0^{1/2} s M^2 \log(p)/\sqrt{n}). \tag{S8.155}
$$

Next, we consider $\boldsymbol{C}_0^{II,1,3}(t)$. Using $(a+b)^2 \leq 2(a^2+b^2)$, we have

$$
\Big| \frac{1}{K}\sum_{k=1}^{K} \frac{1}{2} f'_\epsilon(\xi_{ik})(\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2 \Big| \leq \frac{C'_+}{2K}\sum_{k=1}^{K} (\widehat{\delta}_k + \boldsymbol{X}_i^\top \widehat{\boldsymbol{\Delta}})^2 \leq C'_+ \big(\frac{1}{K}\sum_{k=1}^{K} \widehat{\delta}_k^2 + \widehat{\boldsymbol{\Delta}}^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \widehat{\boldsymbol{\Delta}} \big), \tag{S8.156}
$$

where $\widehat{\delta}_k = \widehat{b}_k - b^{(0)}$, $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)}$. Hence, using the above result, under

$\mathcal{E}$, we have

$$
\begin{aligned}
&\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,1,3}(t)\|_{(s_0,2)} \\
&\leq_{(1)} Cs_0^{1/2}\Big\| \max_t \frac{\lfloor nt\rfloor}{\sqrt{n}}\big(\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\boldsymbol{X}_i\frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}f_\epsilon'(\xi_{ik})(\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2\big)\Big\|_\infty \\
&\qquad\qquad +Cs_0^{1/2}\Big\| \max_t \frac{\lfloor nt\rfloor}{\sqrt{n}}\big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}f_\epsilon'(\xi_{ik})(\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2\big)\Big\|_\infty \\
&\leq_{(2)} C\sqrt{n}s_0^{1/2}\max_t\Big(\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\|\boldsymbol{X}_i\|_\infty\Big|\frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}f_\epsilon'(\xi_{ik})(\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2\Big|\Big) \\
&\qquad\qquad +C\sqrt{n}s_0^{1/2}\Big(\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{X}_i\|_\infty\Big|\frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}f_\epsilon'(\xi_{ik})(\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}^\top \underset{\sim}{\boldsymbol{X}_i})^2\Big|\Big) \\
&\leq_{(3)} CC_+'\sqrt{n}Ms_0^{1/2}(\frac{1}{K}\|\widehat{\boldsymbol{\delta}}\|^2+\max_t\widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Sigma}}(0:t)\widehat{\boldsymbol{\Delta}}) \\
&\qquad\qquad +CC_+'\sqrt{n}Ms_0^{1/2}(\frac{1}{K}\|\widehat{\boldsymbol{\delta}}\|^2+\widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Sigma}}(0:1)\widehat{\boldsymbol{\Delta}}) \\
&\leq_{(4)} CC_+'\sqrt{n}Ms_0^{1/2}(\frac{1}{K}\|\widehat{\boldsymbol{\delta}}\|^2+\max_t|\widehat{\boldsymbol{\Delta}}^\top(\widehat{\boldsymbol{\Sigma}}(0:t)-\boldsymbol{\Sigma})\widehat{\boldsymbol{\Delta}}|+\widehat{\boldsymbol{\Delta}}^\top\boldsymbol{\Sigma}\widehat{\boldsymbol{\Delta}} \\
&\qquad\qquad +CC_+'\sqrt{n}Ms_0^{1/2}(\frac{1}{K}\|\widehat{\boldsymbol{\delta}}\|^2+|\widehat{\boldsymbol{\Delta}}^\top(\widehat{\boldsymbol{\Sigma}}(0:1)-\boldsymbol{\Sigma})\widehat{\boldsymbol{\Delta}})|+\widehat{\boldsymbol{\Delta}}^\top\boldsymbol{\Sigma}\widehat{\boldsymbol{\Delta}} \\
&\leq_{(5)} CC_+'\sqrt{n}Ms_0^{1/2}\Big(\frac{1}{K}\|\widehat{\boldsymbol{\delta}}\|^2+\lambda_{\max}(\boldsymbol{\Sigma})\|\widehat{\boldsymbol{\Delta}}\|^2+M^2s\sqrt{\frac{\log(pn)}{n}}\|\widehat{\boldsymbol{\Delta}}\|^2\Big) \\
&\leq_{(6)} Cs_0^{1/2}sM^2\log(p)/\sqrt{n},
\end{aligned}
\tag{S8.157}
$$

where (1) comes from (S8.146) and the triangle inequality, (2) comes from $\max_t\lfloor nt\rfloor/\sqrt{n}\leq\sqrt{n}$, (3) comes from **Assumption D** and (S8.156), and (5) comes from the fact that $|\widehat{\boldsymbol{\Delta}}^\top(\widehat{\boldsymbol{\Sigma}}(0:1)-\boldsymbol{\Sigma})\widehat{\boldsymbol{\Delta}})|\leq\|\widehat{\boldsymbol{\Delta}}\|_1^2\|\widehat{\boldsymbol{\Sigma}}(0:1)-\boldsymbol{\Sigma}\|_\infty\leq s\|\widehat{\boldsymbol{\Delta}}\|_2^2\|\widehat{\boldsymbol{\Sigma}}(0:1)-\boldsymbol{\Sigma}\|_\infty$ and Lemmas 8, (6) comes from Lemma 10 and the fact that $sM^2\log(p)/\sqrt{n}=o(1)$.

Hence, combining (S8.153), (S8.154), (S8.155), (S8.157), we obtain that

$$\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,1}(t)\|_{(s_0,2)} \leq Cs_0^{1/2}sM^2\log(p)/\sqrt{n}. \tag{S8.158}$$

After bounding $\boldsymbol{C}_0^{II,1,3}(t)$, we next consider $\boldsymbol{C}_0^{II,2}(t)$. The following lemma provides the desired bound. The proof of Lemma 17 is given in Section S10.1.

Lemma 17. Suppose Assumptions A, D, E.2 - E.4 hold. *Then, with probability tending to 1, we have:*

$$\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,2}(t)\|_{(s_0,2)} \leq Cs_0^{1/2}(s\log(pn))^{3/4}/n^{1/4},$$

for some big enough constant $C > 0$.

Hence, combining (S8.158) and Lemma 17, we have:

$$\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,2}(t)\|_{(s_0,2)} \leq Cs_0^{1/2}M^2\frac{(s\log(pn))^{3/4}}{n^{1/4}}.$$

Step 2: Obtain the upper bound for D_2. By Theorem 1, and similar to the proof of Step 2 in Section S8.1, we can prove that

$$\max_{q_0\leq t\leq 1-q_0} \left\|\boldsymbol{C}_0^I(t) - \widetilde{\boldsymbol{C}}_0^I(t)\right\|_{(s_0,2)} = r_0(n)\times O_p\big(Ms_0^{1/2}\sqrt{\log(pn)}\big)$$

where $r_0(n) = s\sqrt{\dfrac{\log(pn)}{n}} \vee s^{\frac{1}{2}}(\dfrac{\log(pn)}{n})^{\frac{3}{8}}$.

Lastly, combining Steps 1 and 2, if we choose $\epsilon := Cs_0^{1/2}(s\log(pn))^{3/4}/n^{1/4}$ for some big constant $C > 0$, we have $D_1 + D_2 = o(1)$, which finishes the proof. □

S8.4 Proof of Lemma 15

Proof. In this section, we prove Lemma 15. In other words, we aim to prove

$$\mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_\alpha(t)-\widetilde{\boldsymbol{C}}_\alpha^I(t)\big\|_{(s_0,2)}\ge\epsilon\big)=o(1), \tag{S8.159}$$

where $\boldsymbol{C}_\alpha(t)$ and $\widetilde{\boldsymbol{C}}_\alpha^I(t)$ are defined in (S7.44) and (S7.47). By the triangle inequality, we have

$$\mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_\alpha(t)-\widetilde{\boldsymbol{C}}_\alpha^I(t)\big\|_{(s_0,2)}\ge\epsilon\big)\le D_1+D_2, \tag{S8.160}$$

where D_1 and D_2 are defined as

$$\begin{aligned}D_1&:=\mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_\alpha(t)-\boldsymbol{C}_\alpha^I(t)\big\|_{(s_0,2)}\ge\epsilon/2\big),\\ D_2&:=\mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_\alpha^I(t)-\widetilde{\boldsymbol{C}}_\alpha^I(t)\big\|_{(s_0,2)}\ge\epsilon/2\big).\end{aligned} \tag{S8.161}$$

By (S8.160), to prove (S8.159), we need to bound D_1 and D_2, respectively.

Step 1: Obtain the upper bound for D_1. We first consider D_1. To this end, we define

$$\mathcal{E}=\big\{\sigma^2/2\le\widehat{\sigma}^2\le 2\sigma^2\big\}, \tag{S8.162}$$

where $\sigma^2:=\mathrm{Var}[(1-\alpha)e_i(\widetilde{\boldsymbol{\tau}})+\alpha\epsilon_i]$ is the true variance. By introducing $\mathcal{E}$, we have

$$D_1\le\mathbb{P}\big(\max_{q_0\le t\le 1-q_0}\big\|\boldsymbol{C}_\alpha(t)-\boldsymbol{C}_\alpha^I(t)\big\|_{(s_0,2)}\ge\epsilon/2\cap\mathcal{E}\big)+\mathbb{P}(\mathcal{E}^c). \tag{S8.163}$$

By Theorem 1, we have $\mathbb{P}(\mathcal{E}^c) = o(1)$ as $n, p \to \infty$. Under the event $\mathcal{E}$, we have

$$\mathbb{P}\Big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_\alpha(t) - \boldsymbol{C}_\alpha^I(t)\big\|_{(s_0,2)} \ge \epsilon/2 \cap \mathcal{E}\Big) \;\; = \mathbb{P}\Big(\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_\alpha^{II}(t)\big\|_{(s_0,2)} \ge \epsilon/2 \cap \mathcal{E}\Big), \tag{S8.164}$$

where $\boldsymbol{C}_\alpha^{II}(t)$ is defined in (S7.46), which is decomposed into two parts:

$$\boldsymbol{C}_\alpha^{II}(t) = (1-\alpha)\boldsymbol{C}_0^{II}(t) + \alpha \boldsymbol{C}_1^{II}(t), \tag{S8.165}$$

where $\boldsymbol{C}_1^{II}(t)$ is defined in (S7.23), and $\boldsymbol{C}_0^{II}(t)$ is defined in (S7.38). Note that by the proofs of Lemmas S8.1 and S8.3, we have proved that:

$$\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_1^{II}(t)\big\|_{(s_0,2)} = O_p(M^2 s_0^{1/2} s \frac{\log(pn)}{\sqrt{n}}),$$
$$\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_0^{II}(t)\big\|_{(s_0,2)} = O_p(M^2 s_0^{1/2} \frac{(s\log(pn))^{3/4}}{n^{1/4}}).$$

Moreover, by **Assumption E.2**, we have $s\frac{\log(pn)}{\sqrt{n}} \ll \frac{(s\log(pn))^{3/4}}{n^{1/4}}$, which implies that:

$$\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_\alpha^{II}(t)\big\|_{(s_0,2)} = O_p\Big(M^2 s_0^{1/2} \frac{(s\log(pn))^{3/4}}{n^{1/4}}\Big).$$

Step 2: Obtain the upper bound for D_2. By Theorem 1, and similar to the proof of Step 2 in Sections S8.1 and S8.3, we can prove that

$$\max_{q_0 \le t \le 1-q_0} \big\|\boldsymbol{C}_\alpha^I(t) - \widetilde{\boldsymbol{C}}_\alpha^I(t)\big\|_{(s_0,2)} = r_\alpha(n) \times O_p\big(M s_0^{1/2}\sqrt{\log(pn)}\big),$$

where $r_\alpha(n) = s\sqrt{\frac{\log(pn)}{n}} \vee s^{\frac{1}{2}}(\frac{\log(pn)}{n})^{\frac{3}{8}}$. Note that by **Assumption E.2**, we have

$$r_\alpha(n) \times M s_0^{1/2}\sqrt{\log(pn)} \ll M^2 s_0^{1/2}\frac{(s\log(pn))^{3/4}}{n^{1/4}}.$$

Hence, combining Steps 1 and 2, if we choose $\epsilon := C s_0^{1/2}(s\log(pn))^{3/4}/n^{1/4}$ for some big constant $C > 0$, we have $D_1 + D_2 = o(1)$, which finishes the proof. □

S8.5 Proof of Lemma 16

Proof. Note that the proof for $\mathcal{H}_1$ and $\mathcal{H}_2$ is similar. We only give the proof of $\mathcal{H}_1$. The proof proceeds in two steps: In Step 1, we obtain the upper bounds of $\max_{t \geq t_1+\epsilon_n} \max_{\substack{J\subset\{1,\ldots,p\} \\ |J|=s_0}} \|\Pi_J \widetilde{\boldsymbol{C}}_1(t) - \Pi_J \boldsymbol{\delta}_1(t)\|_2$. In Step 2, using the upper bound and some regular inequalities, we finish the proof.

Step 1: By the decomposition of $\widetilde{\boldsymbol{C}}_1(t)$ as in (S7.57), with probability tending to one, we have:

$$\begin{aligned}
&\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\{1,\ldots,p\}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)-\Pi_J\boldsymbol{\delta}(t)\|_2\\
&\leq_{(1)}\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\{1,\ldots,p\}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1^I(t)+\Pi_J\boldsymbol{R}(t)\|_2\\
&\leq_{(2)}\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\{1,\ldots,p\}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1^I(t)\|_2+\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\{1,\ldots,p\}\\|J|=s_0}}\|\Pi_J\boldsymbol{R}(t)\|_2\\
&=_{(3)}\max_{t\geq t_1+\epsilon_n}\|\widetilde{\boldsymbol{C}}_1^I(t)\|_{(s_0,2)}+\max_{t\geq t_1+\epsilon_n}\|\boldsymbol{R}(t)\|_{(s_0,2)}\\
&\leq_{(4)}s_0^{1/2}\max_{q_0\leq t\leq 1-q_0}\|\widetilde{\boldsymbol{C}}_1^I(t)\|_\infty+s_0^{1/2}\max_{t\geq t_1+\epsilon_n}\|\boldsymbol{R}(t)\|_\infty\\
&\leq_{(5)}\underbrace{C^*(s_0^{1/2}M\sqrt{\log(pn)}+s_0^{1/2}s\sqrt{\log(pn)}\|\boldsymbol{\Delta}\|_{(s_0,2)})}_{t^*}.
\end{aligned}$$

Recall $\mathcal{M}=\{j:\beta_j^{(1)}\neq\beta_j^{(2)}\}$. Note that

$$\begin{aligned}
&\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\{1,\ldots,p\}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)-\Pi_J\boldsymbol{\delta}(t)\|_2\\
&=\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)-\Pi_J\boldsymbol{\delta}(t)\|_2+\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}^c\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)-\Pi_J\boldsymbol{\delta}(t)\|_2.
\end{aligned}$$

Using the fact that $|\max_i\|\boldsymbol{a}_i\|_2-\max_i\|\boldsymbol{b}_i\|_2|\leq\max_i|\|\boldsymbol{a}_i\|_2-\|\boldsymbol{b}_i\|_2|\leq\max_i|\|\boldsymbol{a}_i-\boldsymbol{b}_i\|_2|$ for any vectors $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$, we have:

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)\|_2-\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\boldsymbol{\delta}(t)\|_2\Big|\leq t^*\Big)\to 1\\
&\text{and }\ \mathbb{P}\Big(\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}^c\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)\|_2\leq t^*\Big)\to 1.
\end{aligned}\tag{S8.166}$$

Step 2: Note that $\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\boldsymbol{\delta}(t)\|_2\Big|=\sqrt{n}t_1(1-t_1-\epsilon_n)\|\boldsymbol{\Delta}\|_{(s_0,2)}$. By chooing a big enough constant in (S7.49), we have $\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\boldsymbol{\delta}(t)\|_2\Big|\geq$

$2t^*$. Moreover, by (S8.166), we see that:

$$\begin{aligned}
&\mathbb{P}\Big(\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)\|_2-\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}^c\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)\|_2\leq 0\Big)\\
&\leq_{(1)}\mathbb{P}\Big(\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)\|_2-\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}^c\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)\|_2\\
&\qquad\qquad\leq\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\boldsymbol{\delta}(t)\|_2-t^*-t^*\Big)\\
&\leq_{(2)}\mathbb{P}\Big(\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)\|_2\leq\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}\\|J|=s_0}}\|\Pi_J\boldsymbol{\delta}(t)\|_2-t^*\Big)\\
&\qquad+\mathbb{P}\Big(-\max_{t\geq t_1+\epsilon_n}\max_{\substack{J\subset\mathcal{M}^c\\|J|=s_0}}\|\Pi_J\widetilde{\boldsymbol{C}}_1(t)\|_2\leq -t^*\Big)\to 0,\quad\text{as}\quad (n,p)\to\infty,
\end{aligned}$$

which finishes the proof. □

S9 Proofs of useful lemmas in Section S6

S9.1 Proof of Lemma 6

Proof. In this section, we aim to prove (S6.7). Firstly, we define $\mathcal{E}^R=\{\boldsymbol{x}\in\mathbb{R}^p:\|\boldsymbol{x}\|\leq R\}$ and $V^z_{(s_0,2)}=\{\boldsymbol{x}\in\mathbb{R}^p:\|\boldsymbol{x}\|_{(s_0,2)}\leq z\}$. Then, by the definition of $V^z_{(s_0,2)}$, we have

$$\begin{aligned}
&\sup_{z\in(0,\infty)}\Big|\mathbb{P}\big(\max_{k_0\leq k\leq n-k_0}\|\boldsymbol{S}^{\boldsymbol{Z}}(k)\|_{(s_0,2)}\leq z\big)-\mathbb{P}\big(\max_{k_0\leq k\leq n-k_0}\|\boldsymbol{S}^{\boldsymbol{G}}(k)\|_{(s_0,2)}\leq z\big)\Big|\\
=\;&\sup_{z\in(0,\infty)}\underbrace{\Big|\mathbb{P}\big(\bigcap_{k_0\leq k\leq n-k_0}\big\{\boldsymbol{S}^{\boldsymbol{Z}}(k)\in V^z_{(s_0,2)}\big)\big\}-\mathbb{P}\big(\bigcap_{k_0\leq k\leq n-k_0}\big\{\boldsymbol{S}^{\boldsymbol{G}}(k)\in V^z_{(s_0,2)}\big)\big\}\Big|}_{A_z}.
\end{aligned}\tag{S9.167}$$

By interting $\mathcal{E}^R$ and $(\mathcal{E}^R)^c$ in A_z, we have $A_z \le A_z^{(1)} + A_z^{(2)}$, where

$$A_z^{(1)} := \Big|\mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\{\boldsymbol{S}^{\boldsymbol{Z}}(k)\in(\mathcal{E}^R)^c\cap V^z_{(s_0,2)}\}\Big)-\mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\{\boldsymbol{S}^{\boldsymbol{G}}(k)\in(\mathcal{E}^R)^c\cap V^z_{(s_0,2)}\}\Big)\Big|,$$

$$A_z^{(2)} := \Big|\mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\{\boldsymbol{S}^{\boldsymbol{Z}}(k)\in V^z_{(s_0,2)}\cap\mathcal{E}^R\}\Big)-\mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\{\boldsymbol{S}^{\boldsymbol{G}}(k)\in V^z_{(s_0,2)}\cap\mathcal{E}^R\}\Big)\Big|. \tag{S9.168}$$

Next, we bound $A_z^{(1)}$ and $A_z^{(2)}$ respectively. For $A_z^{(1)}$, using the triangle inequality, we have

$$A_z^{(1)} \le \mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\big\{\boldsymbol{S}^{\boldsymbol{Z}}(k)\in(\mathcal{E}^R)^c\cap V^z_{(s_0,2)}\big\}\Big)+\mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\big\{\boldsymbol{S}^{\boldsymbol{G}}(k)\in(\mathcal{E}^R)^c\cap V^z_{(s_0,2)}\big\}\Big).$$

Recall $\boldsymbol{S}^{\boldsymbol{Z}}$ and $\boldsymbol{S}^{\boldsymbol{G}}$ in (S6.6). Let $a_{ik} = \mathbf{1}\{i \le k\} - k/n$ for $i = 1\ldots, n$ and $k_0 \le k \le n - k_0$. We then have $\boldsymbol{S}^{\boldsymbol{Z}}(k) = n^{-1/2}\sum_{i=1}^n \boldsymbol{Z}_i a_{ik}$ and $\boldsymbol{S}^{\boldsymbol{G}}(k) = n^{-1/2}\sum_{i=1}^n \boldsymbol{G}_i a_{ik}$. Moreover, by the definition of $k_0 = \lfloor nq_0\rfloor$, we have $q_0 \le |a_{ik}| \le 1 - q_0$ for $i = 1, \ldots, n$ and $k_0 \le k \le n - k_0$. Hence, we have

$$\begin{aligned}
&\mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\big\{\boldsymbol{S}^{\boldsymbol{Z}}(k)\in(\mathcal{E}^R)^c\cap V^z_{(s_0,2)}\big\}\Big)\\
&\le \mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\big\{\boldsymbol{S}^{\boldsymbol{Z}}(k)\in(\mathcal{E}^R)^c\big\}\Big)\\
&\le \mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\Big\|n^{-1/2}\sum_{i=1}^n \boldsymbol{Z}_i a_{ik}\Big\|_2 \ge R\Big)\\
&\le \mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0} n^{-1/2}\sum_{i=1}^n \|\boldsymbol{Z}_i\|_2 \ge R\Big)\\
&\le \sum_{i=1}^n \mathbb{P}\Big(\sum_{j=1}^p Z_{ij}^2 \ge \frac{R^2}{n}\Big).
\end{aligned} \tag{S9.169}$$

By Assumption **(M2)** and Markov's inequality, we further have:

$$\sum_{i=1}^{n} \mathbb{P}\Big(\sum_{j=1}^{p} Z_{ij}^2 \geq \frac{R^2}{n}\Big) \leq \frac{n\sum_{i=1}^{n}\sum_{j=1}^{p} \mathbb{E}Z_{ij}^2}{R^2} \leq \frac{np\max_{1\leq j\leq p}\sum_{i=1}^{n} \mathbb{E}Z_{ij}^2}{R^2} \leq \frac{n^2pK^2}{R^2}. \tag{S9.170}$$

Hence, taking $R^2 = n^{5/2}p$ and combining (S9.169) and (S9.170), we have

$$\mathbb{P}\Big(\bigcap_{k_0\leq k\leq n-k_0} \big\{\boldsymbol{S}^{\boldsymbol{Z}}(k) \in (\mathcal{E}^R)^c \cap V^z_{(s_0,2)}\big\}\Big) \leq C\frac{1}{\sqrt{n}}.$$

Similarly, for $\boldsymbol{S}^{\boldsymbol{G}}(k)$, we have $\mathbb{P}\big(\bigcap_{k_0\leq k\leq n-k_0} \big\{\boldsymbol{S}^{\boldsymbol{G}}(k) \in (\mathcal{E}^R)^c \cap V^z_{(s_0,2)}\big\}\big) \leq C\frac{1}{\sqrt{n}}$. The above results yield that $A_z^{(1)} \leq C\frac{1}{\sqrt{n}}$.

After bounding $A_z^{(1)}$, we next consider $A_z^{(2)}$. By Lemma 5, there exists an m-generated convex set A^m such that

$$A^m \subset \mathcal{E}^{R,p} \cap V^{z,p}_{(s_0,2)} \subset A^{m,R\epsilon} \ \text{ and } \ m \leq p^{s_0}\Big(\frac{\gamma}{\sqrt{\epsilon}}\ln(\frac{1}{\epsilon})\Big)^{s_0^2}.$$

By letting

$$\bar{\rho}_1 := \big|\mathbb{P}\big(\bigcap_{k_0\leq k\leq n-k_0} \big(\boldsymbol{S}^{\boldsymbol{Z}}(k) \in A^m\big)\big) - \mathbb{P}\big(\bigcap_{k_0\leq k\leq n-k_0} \big(\boldsymbol{S}^{\boldsymbol{G}}(k) \in A^m\big)\big)\big|,$$

$$\bar{\rho}_2 := \big|\mathbb{P}\big(\bigcap_{k_0\leq k\leq n-k_0} \big(\boldsymbol{S}^{\boldsymbol{Z}}(k) \in A^{m,R\epsilon}\big)\big) - \mathbb{P}\big(\bigcap_{k_0\leq k\leq n-k_0} \big(\boldsymbol{S}^{\boldsymbol{G}}(k) \in A^{m,R\epsilon}\big)\big)\big|,$$

we have

$$\begin{aligned}
&\mathbb{P}\big(\bigcap_{k_0\leq k\leq n-k_0} \big\{\boldsymbol{S}^{\boldsymbol{Z}}(k) \in \mathcal{E}^R \cap V^z_{(s_0,2)}\big\}\big) \\
&\leq \mathbb{P}\big(\bigcap_{k_0\leq k\leq n-k_0} \big(\boldsymbol{S}^{\boldsymbol{Z}}(k) \in A^{m,R\epsilon}\big)\big) \ \big(\text{by } \ \mathcal{E}^R \cap V^z_{(s_0,2)} \subset A^{m,R\epsilon}\big) \\
&\leq \underbrace{\mathbb{P}\Big(\bigcap_{k_0\leq k\leq n-k_0} \big(\boldsymbol{S}^{\boldsymbol{G}}(k) \in A^{m,R\epsilon}\big)\Big)}_{P_z} + \max(\bar{\rho}_1, \bar{\rho}_2).
\end{aligned} \tag{S9.171}$$

Using Assumption (**M1**), by the definition of $A^{m,R\epsilon}$ in (S6.5) and Lemma 4, we have

$$\begin{aligned} P_z &= \mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\bigcap_{\boldsymbol{v}\in\mathcal{V}(A^m)}\big(\boldsymbol{S}^{\boldsymbol{G}}(k)^\top\boldsymbol{v}\le S_{A^m}(\boldsymbol{v})+R\epsilon\big)\Big) \\ &\le \mathbb{P}\Big(\bigcap_{\substack{k_0\le k\le n-k_0\\ \boldsymbol{v}\in\mathcal{V}(A^m)}}\big(\boldsymbol{S}^{\boldsymbol{G}}(k)^\top\boldsymbol{v}\le S_{A^m}(\boldsymbol{v})\big)\Big) \\ &+\mathbb{P}\Big(\bigcap_{\substack{k_0\le k\le n-k_0\\ \boldsymbol{v}\in\mathcal{V}(A^m)}}\big(S_{A^m}(\boldsymbol{v})\le\boldsymbol{S}^{\boldsymbol{G}}(k)^\top\boldsymbol{v}\le S_{A^m}(\boldsymbol{v})+R\epsilon\big)\Big) \\ &\le \mathbb{P}\Big(\bigcap_{k_0\le k\le n-k_0}\big(\boldsymbol{S}^{\boldsymbol{G}}(k)\in\mathcal{E}^R\cap V^z_{(s_0,2)}\big)\Big)+CR\epsilon\sqrt{\log nm}\Big(\text{by } A^m\subset\mathcal{E}^{R,d}\cap V^z_{(s_0,2)}\Big). \end{aligned} \tag{S9.172}$$

Therefore, by (S9.171) and (S9.172), we have

$$\begin{aligned} &\mathbb{P}\big(\bigcap_{k_0\le k\le n-k_0}\big(\boldsymbol{S}^{\boldsymbol{Z}}(k)\in\mathcal{E}^R\cap V^z_{(s_0,2)}\big)\big) \\ &\quad\le \mathbb{P}\big(\bigcap_{k_0\le k\le n-k_0}\big(\boldsymbol{S}^{\boldsymbol{G}}(k)\in\mathcal{E}^R\cap V^z_{(s_0,2)}\big)\big)+CR\epsilon\sqrt{\log nm}+\max(\bar{\rho}_1,\bar{\rho}_2). \end{aligned} \tag{S9.173}$$

Similar to the procedures in (S9.171), (S9.172), and (S9.173), we also have

$$\begin{aligned} &\mathbb{P}\big(\bigcap_{k_0\le k\le n-k_0}\big(\boldsymbol{S}^{\boldsymbol{Z}}(k)\in\mathcal{E}^R\cap V^z_{(s_0,2)}\big)\big) \\ &\quad\ge \mathbb{P}\big(\bigcap_{k_0\le k\le n-k_0}\big(\boldsymbol{S}^{\boldsymbol{G}}(k)\in\mathcal{E}^{R,d}\cap V^{z,d}_{(s_0,p)}\big)\big)-CR\epsilon\sqrt{\log nm}-\max(\bar{\rho}_1,\bar{\rho}_2). \end{aligned} \tag{S9.174}$$

Therefore, by (S9.168), (S9.173), and (S9.174), we obtain

$$A_z^{(2)}\le\max(\bar{\rho}_1,\bar{\rho}_2)+CR\epsilon\sqrt{\log nm}. \tag{S9.175}$$

Next, we consider $\bar{\rho}_1$ and $\bar{\rho}_2$. For $\bar{\rho}_1$, we have

$$\bar{\rho}_1 := \Big|\mathbb{P}\big(\bigcap_{k_0\le k\le n-k_0}\bigcap_{\boldsymbol{v}\in\mathcal{V}(A^m)} \boldsymbol{S}^{\boldsymbol{Z}}(k)^\top \boldsymbol{v} \le S_{A^m}(\boldsymbol{v})\big)$$
$$-\mathbb{P}\big(\bigcap_{k_0\le k\le n-k_0}\bigcap_{\boldsymbol{v}\in\mathcal{V}(A^m)} \boldsymbol{S}^{\boldsymbol{G}}(k)^\top \boldsymbol{v} \le S_{A^m}(\boldsymbol{v})\big)\Big|.$$

Define $\widetilde{Z}_i(k,\boldsymbol{v}) = \boldsymbol{v}^\top \boldsymbol{Z}_i a_{ik}$ and $\widetilde{G}_i(k,\boldsymbol{v}) = \boldsymbol{v}^\top \boldsymbol{G}_i a_{ik}$ for $i = 1,\dots,n$, $k = k_0,\dots,n-k_0$ and $\boldsymbol{v}\in\mathcal{V}(A^m)$. By letting

$$S^{\widetilde{Z}_i(k,v)} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{Z}_i(k,v), \text{ and } S^{\widetilde{G}_i(k,v)} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{G}_i(k,v),$$

we have

$$\bar{\rho}_1 := \Big|\mathbb{P}\big(S^{\widetilde{Z}_i(k,v)} \le S_{A^m}(\boldsymbol{v}), k_0\le k\le n-k_0, \boldsymbol{v}\in\mathcal{V}(A^m)\big)$$
$$-\mathbb{P}\big(S^{\widetilde{G}_i(k,v)} \le S_{A^m}(\boldsymbol{v}), k_0\le k\le n-k_0, \boldsymbol{v}\in\mathcal{V}(A^m)\big)\Big|,$$

which is a high dimensional Gaussian approximation for hyperrectangle in terms of $\{\widetilde{Z}_i(k,v)\}$. To use Proposition 2.1 in Chernozhukov et al. (2017), we need to verify that under Assumptions (**M1**)-(**M3**), $\widetilde{Z}_i(k,v) = \boldsymbol{v}^\top \boldsymbol{Z}_i a_{ik}$ satisfies Conditions (M.1), (M.2) and (E.2) in Chernozhukov et al. (2017). In fact, by Assumption (**M1**), we have $\inf_{k,\boldsymbol{v}} \mathbb{E}\widetilde{Z}_i(k,\boldsymbol{v})^2 \ge b$ holds for $i = 1,\dots,n$, which implies Condition (M.1). Moreover, for $\boldsymbol{v}\in\mathcal{V}(A^m)$, let $J(\boldsymbol{v})$ be the set of non-zero coordinates of $\boldsymbol{v}$ with $|J(\boldsymbol{v})| \le s_0$. Using Hölder's inequality, for any vector $\boldsymbol{a} = (a_1,\dots,a_p)^\top$, we have $(\sum_{j\in J(\boldsymbol{v})}|a_j|)^{2+\ell} \le$

$s_0^{1+\ell}\sum_{j\in J(\boldsymbol{v})}|a_j|^{2+\ell}$. This implies that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|\widetilde{Z}_i(k,\boldsymbol{v})|^{2+\ell} \\
&\le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|\boldsymbol{v}^\top \boldsymbol{Z}_i|^{2+\ell} \quad (\text{by } q_0 \le |a_{ik}| \le 1-q_0) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\sum_{j\in J(\boldsymbol{v})}|Z_{ij}|^{2+\ell} \quad (|J(\boldsymbol{v})| \le s_0 \text{ and } \|\boldsymbol{v}\|=1) \\
&\le s_0^{1+\ell}\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in J(\boldsymbol{v})}\mathbb{E}|Z_{ij}|^{2+\ell} \\
&\le s_0^{2+\ell}\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|Z_{ij}|^{2+\ell} \\
&\le s_0^{2+\ell}K^\ell := (B_n)^\ell, \quad (\text{by Assumption } (\mathbf{M2})),
\end{aligned}$$

where $B_n := Ks_0^{(2+\ell)/\ell}$. Hence, Condition (M.2) holds by taking $B_n := Ks_0^{(2+\ell)/\ell}$. Lastly, we verify Condition (E.2). In fact, we have

$$\begin{aligned}
&\mathbb{E}\Big(\big(\max_{\substack{k_0\le k\le n-k_0\\ \boldsymbol{v}\in\mathcal{V}(A^m)}}|\widetilde{Z}_i(k,\boldsymbol{v})|\big)^q\Big) \\
&\le \mathbb{E}\Big(\big(\max_{\substack{k_0\le k\le n-k_0\\ \boldsymbol{v}\in\mathcal{V}(A^m)}}|\widetilde{Z}_i(k,\boldsymbol{v})|\big)^q\Big) \\
&\le \mathbb{E}\Big(\big(\max_{\boldsymbol{v}\in\mathcal{V}(A^m)}|\boldsymbol{v}^\top\boldsymbol{Z}_i|\big)^q\Big) \quad (\text{by } q_0 \le |a_{ik}| \le 1-q_0) \\
&\le \mathbb{E}\Big(\big(\max_{\boldsymbol{v}\in\mathcal{V}(A^m)}|\textstyle\sum_{j\in J(\boldsymbol{v})}Z_{ij}|\big)^q\Big) \\
&\le s_0^q\mathbb{E}\Big(\big(\max_{1\le j\le p}|Z_{ij}|\big)^q\Big) := (B_n')^q,
\end{aligned}$$

where $B_n' := s_0K$. Hence, Condition (E.2) in Chernozhukov et al. (2017) holds by taking $B_n' := s_0K$. Lastly, taking $\widetilde{B}_n = s_0^3K$, we have

$$\max_{\substack{k_0\le k\le n-k_0\\ \boldsymbol{v}\in\mathcal{V}(A^m)}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|\widetilde{Z}_i(k,v)|^{2+\ell} \le (\widetilde{B}_n)^\ell \quad \text{for } \ell=1,2;$$

$$\text{and} \quad \max\nolimits_{1\le i\le n}\mathbb{E}\Big(\big(\max_{\substack{k_0\le k\le n-k_0\\ \boldsymbol{v}\in\mathcal{V}(A^m)}}|\widetilde{Z}_i(k,v)|\big)^q\Big) \le (\widetilde{B}_n)^q.$$

Let $D_n^{(1)} = \Big(\dfrac{s_0^6 K^2 \log^7(mn^2)}{n}\Big)^{1/6}$ and $D_n^{(2)} = \Big(\dfrac{s_0^6 K^2 \log^3(mn^2)}{n^{1-2/q}}\Big)^{1/3}$. Using Proposition 2.1 in Chernozhukov et al. (2017), for $\bar{\rho}_1$ and $\bar{\rho}_2$, we have

$$\max(\bar{\rho}_1, \bar{\rho}_2) \le C\Big(D_n^{(1)} + D_n^{(2)}\Big), \tag{S9.176}$$

where C is some universal constant not depending on n or p. Combining (S9.167), (S9.168), (S9.175), and (S9.176), we have

$$\sup_{z\in(0,\infty)} A_z \le C_1 \frac{1}{\sqrt{n}} + C_2 R\epsilon\sqrt{\log nm} + C_3\Big(D_n^{(1)} + D_n^{(2)}\Big). \tag{S9.177}$$

Recall $R := n^{5/4}p^{1/2}$ and $m \le p^{s_0}\Big(\frac{\gamma}{\sqrt{\epsilon}}\ln(\dfrac{1}{\epsilon})\Big)^{s_0^2}$. By letting $\epsilon = (pn^2)^{-1}$, we have

$$R\epsilon\sqrt{\log mn} \preceq \Big(\frac{s_0^6 K^2 \log^7(mn^2)}{n}\Big)^{1/6}, \text{ and } R\epsilon\sqrt{\log mn} \preceq \Big(\frac{s_0^6 K^2 \log^3(mn^2)}{n^{1-2/q}}\Big)^{1/3}. \tag{S9.178}$$

Moreover, using the Assumption that $s_0^3 K^{2/7}\log(pn) = O(n^{\xi_1})$ for some $0 < \xi_1 < 1/7$ and $s_0^4 K^{2/3}\log(pn) = O(n^{\xi_2})$ for some $0 < \xi_2 < \frac{1}{3}(1-2/q)$, we have

$$D_n^{(1)} + D_n^{(2)} \le n^{-\xi_0}, \quad \text{for some } \xi_0 > 0. \tag{S9.179}$$

Lastly, combining (S9.177), (S9.178) and (S9.179), we finish the proof of Lemma 6. □

S9.2 Proof of Lemma 7

Proof. Let $a_{ik} = \mathbf{1}\{i \le k\} - k/n$ for $i = 1 \ldots, n$ and $\underline{k}_n \le k \le n - \overline{k}_n$ with $\underline{k}_n := \lfloor na_n \rfloor$ and $\overline{k}_n := \lfloor nb_n \rfloor$. Define $Z_{ij}(k) = X_{ij}\epsilon_i a_{ik}$ for $i = 1, \ldots, n$, $j = 1, \ldots, p$ and $k = \underline{k}_n, \ldots, n - \overline{k}_n$. By definition, we have:

$$\begin{aligned}&\max_{t\in[a_n,1-b_n]}\max_{1\le j\le p}\Big|\frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{\lfloor nt\rfloor} X_{ij}\epsilon_i - \frac{\lfloor nt\rfloor}{n}\sum_{i=1}^{n} X_{ij}\epsilon_i\Big)\Big| \\ &= \max_{\underline{k}_n\le k\le n-\overline{k}_n}\max_{1\le j\le p}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{ij}(k)\Big|.\end{aligned} \tag{S9.180}$$

Note that by **Assumption A, C**, we have $\mathbb{E}|Z_{ij}(k)|^{2+\ell} \le a_{ik}^{2+\ell} M^{2+\ell} K^{\ell}$ for $\ell = 1, 2$. Let $M = \max_{i,j,k} |Z_{ij}(k)|$ and $\sigma^2 = \max_{j,k} \sum_i \mathbb{E}[Z_{ij}^2]$. Then, by Lemma 2, we have:

$$\begin{aligned}&\mathbb{E}\Big[\max_{\underline{k}_n\le k\le n-\overline{k}_n}\max_{1\le j\le p}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{ij}(k)\Big|\Big] \\ &\le \frac{C}{\sqrt{n}}\Big(\sigma\sqrt{\log p(n-\underline{k}_n-\overline{k}_n)} + \sqrt{\mathbb{E}[M^2]}\log p(n-\underline{k}_n-\overline{k}_n)\Big).\end{aligned}$$

For σ^2, using Hölder's inequality, we have $\sigma^2 \le_{(1)} C\sum_{i=1}^{n} a_{ik}^2 M_n^2 \le_{(2)} CnM_n^2$, where (2) comes from $a_n \le |a_{ik}| \le 1 - b_n$. For $\mathbb{E}[M^2]$, by definition, we have:

$$\mathbb{E}[M^2] = \mathbb{E}[\max_{i,j,k} |X_{ij}\epsilon_i a_{ik}|^2] \le_{(1)} M^2\mathbb{E}[\max_i |\epsilon_i|^2] \le_{(2)} CM^2 n^{1/2},$$

where (1) comes from **Assumption A**, (2) comes from **Assumption C** and Theorem 3 in Downey (1990). Hence, we have

$$\mathbb{E}\Big[\max_{\underline{k}_n\le k\le n-\overline{k}_n}\max_{1\le j\le p}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{ij}(k)\Big|\Big] \le CM\sqrt{\log p(n-\underline{k}_n-\overline{k}_n)}. \tag{S9.181}$$

Hence, using Lemma 2, taking $\eta = 1$, $s = 2$ and $t = C^* M \sqrt{\log(p(n - \underline{k}_n - \overline{k}_n))}$ for some big enough constant $C^* > 0$, we have:

$$\mathbb{P}\Big(\max_{\underline{k}_n \le k \le n - \overline{k}_n} \max_{1 \le j \le p} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ij}(k) \Big| \ge C_1 M \sqrt{\log(p(n - \underline{k}_n - \overline{k}_n))} \Big) \le C_2 n^{-1/2},$$

which completes the proof. □

S9.3 Proof of Lemma 9

Proof. In this section, we prove Lemma 9. Note that Lemma 9 applies to both $\mathbf{H}_0$ and $\mathbf{H}_1$. To cover the above two cases in a unified way, we prove the results by assuming there is a change point t_1 such that $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}$ if $i \le \lfloor nt_1 \rfloor$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^{(2)}$ if $i > \lfloor nt_1 \rfloor$. Note that under $\mathbf{H}_0$, we can always set $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$ even though t_1 is not identifiable. Now, we are ready to prove Lemma 9.

Recall $\boldsymbol{\beta}^*$ is the minimizer under the population level which is defined as:

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\Big[(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta})^2 \Big].$$

By the first-order condition, we can see that $\boldsymbol{\beta}^*$ satisfies:

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\Big[\boldsymbol{X}_i (Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^*) \Big] = \mathbf{0}_p. \tag{S9.182}$$

Moreover, since the model is linear, $\boldsymbol{\beta}^* \in \mathbb{R}^p$ has the following explicit form, which is a linear combination of $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$:

$$\boldsymbol{\beta}^* = t_1\boldsymbol{\beta}^{(1)} + (1-t_1)\boldsymbol{\beta}^{(2)}.$$

Note that under $\mathbf{H}_0$ with $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$, we have $\boldsymbol{\beta}^* = \boldsymbol{\beta}^{(1)}$. In this case, $\boldsymbol{\beta}^*$ is the true parameter for the linear model. Recall $\widehat{\boldsymbol{\beta}}$ is the minimizer of the empirical loss defined in (2.10). Hence, we have:

$$\begin{aligned}
&\frac{1}{2n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^\top\widehat{\boldsymbol{\beta}})^2 - \frac{1}{2n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*)^2\\
&\quad =_{(1)} \frac{1}{2n}\sum_{i=1}^{n}\big(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^* - (\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\big)^2 - \frac{1}{2n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*)^2\\
&\quad =_{(2)} \frac{1}{2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top\widehat{\boldsymbol{\Sigma}}(0:1)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*)\\
&\quad =_{(3)} \frac{1}{2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top\widehat{\boldsymbol{\Sigma}}(0:1)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\\
&\qquad -(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top\frac{1}{n}\sum_{i=1}^{n}\Big(\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*) - \mathbb{E}\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*)\Big),
\end{aligned} \tag{S9.183}$$

where $\widehat{\boldsymbol{\Sigma}}(0:1) := \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^\top$, and (3) comes from the first order condition in (S9.182). Hence, by the fact that $\widehat{\boldsymbol{\beta}}$ is the minimizer of (2.10), we have:

$$\frac{1}{2n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^\top\widehat{\boldsymbol{\beta}})^2 - \frac{1}{2n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*)^2 + \lambda\big(\|\widehat{\boldsymbol{\beta}}\|_1 - \|\boldsymbol{\beta}^*\|_1\big) \leq 0,$$

where $\text{MSE}(\widehat{\boldsymbol\beta}) := \big\|\mathbf{X}(\widehat{\boldsymbol\beta}-\boldsymbol\beta^*)\big\|^2$. Moreover, by (S9.183), we have:

$$\begin{aligned}
&\frac{1}{2}(\widehat{\boldsymbol\beta}-\boldsymbol\beta^*)^\top\widehat{\boldsymbol\Sigma}(0:1)(\widehat{\boldsymbol\beta}-\boldsymbol\beta^*)\\
&-(\widehat{\boldsymbol\beta}-\boldsymbol\beta^*)^\top\frac{1}{n}\sum_{i=1}^n\Big(\boldsymbol X_i(Y_i-\boldsymbol X_i^\top\boldsymbol\beta^*)-\mathbb{E}\boldsymbol X_i(Y_i-\boldsymbol X_i^\top\boldsymbol\beta^*)\Big)+\lambda\big(\|\widehat{\boldsymbol\beta}\|_1-\|\boldsymbol\beta^*\|_1\big)\\
&\quad\le\frac{1}{2}\text{MSE}(\widehat{\boldsymbol\beta})-\|\frac{1}{n}\sum_{i=1}^n\Big(\boldsymbol X_i(Y_i-\boldsymbol X_i^\top\boldsymbol\beta^*)-\mathbb{E}\boldsymbol X_i(Y_i-\boldsymbol X_i^\top\boldsymbol\beta^*)\Big)\|_\infty\|\widehat{\boldsymbol\beta}-\boldsymbol\beta^*\|_1\\
&\qquad\qquad+\lambda\big(\|\widehat{\boldsymbol\beta}\|_1-\|\boldsymbol\beta^*\|_1\big)\le 0.
\end{aligned}\tag{S9.184}$$

Moreover, by the fact that $Y_i=\epsilon_i+\boldsymbol\beta^{(1)}\mathbf{1}\{i\le\lfloor nt_1\rfloor\}+\boldsymbol\beta^{(2)}\mathbf{1}\{i>\lfloor nt_1\rfloor\}$,

we have:

$$\begin{aligned}
&\|\frac{1}{n}\sum_{i=1}^n\Big(\boldsymbol X_i(Y_i-\boldsymbol X_i^\top\boldsymbol\beta^*)-\mathbb{E}\boldsymbol X_i(Y_i-\boldsymbol X_i^\top\boldsymbol\beta^*)\Big)\|_\infty\\
&=_{(1)}\Big\|\frac{1}{n}\sum_{i=1}^n\boldsymbol X_i\epsilon_i+t_1(1-t_1)(\widehat{\boldsymbol\Sigma}(0:t_1)-\boldsymbol\Sigma)(\boldsymbol\beta^{(1)}-\boldsymbol\beta^{(2)})\\
&\qquad\qquad+t_1(1-t_1)(\widehat{\boldsymbol\Sigma}(t_1:1)-\boldsymbol\Sigma)(\boldsymbol\beta^{(2)}-\boldsymbol\beta^{(1)})\Big\|_\infty\\
&\le_{(2)}\|\frac{1}{n}\sum_{i=1}^n\boldsymbol X_i\epsilon_i\|_\infty+t_1(1-t_1)\|\widehat{\boldsymbol\Sigma}(0:t_1)-\boldsymbol\Sigma\|_\infty\|\boldsymbol\beta^{(1)}-\boldsymbol\beta^{(2)}\|_1\\
&\qquad\qquad+t_1(1-t_1)\|\widehat{\boldsymbol\Sigma}(t_1:1)-\boldsymbol\Sigma\|_\infty\|\boldsymbol\beta^{(1)}-\boldsymbol\beta^{(2)}\|_1\\
&\le_{(3)}C_1M\sqrt{\frac{\log(pn)}{n}}+C_2M^2\sqrt{\frac{\log(pn)}{n}}\|\boldsymbol\beta^{(1)}-\boldsymbol\beta^{(2)}\|_1+C_3M^2\sqrt{\frac{\log(pn)}{n}}\|\boldsymbol\beta^{(1)}-\boldsymbol\beta^{(2)}\|_1\\
&\le_{(4)}C_4M^2\sqrt{\frac{\log(pn)}{n}}\big(1+\|\boldsymbol\beta^{(1)}-\boldsymbol\beta^{(2)}\|_1\big)\le_{(5)}C_5M^2\sqrt{\frac{\log(pn)}{n}},
\end{aligned}\tag{S9.185}$$

where (3) comes from Lemmas 7 and 8, and (5) comes from the assumption that $\|\boldsymbol\beta^{(1)}-\boldsymbol\beta^{(2)}\|=O(1)$. Hence, by letting $\lambda\ge 2C_5M^2\sqrt{\frac{\log(pn)}{n}}$, and combining (S9.184) and (S9.185), we have:

$$\frac{1}{2}\text{MSE}(\widehat{\boldsymbol\beta})-\frac{\lambda}{2}\|\widehat{\boldsymbol\beta}-\boldsymbol\beta^*\|_1+\lambda\big(\|\widehat{\boldsymbol\beta}\|_1-\|\boldsymbol\beta^*\|_1\big)\le 0.$$

Adding $\lambda\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_1$ on both sides of the above inequality, we have:

$$\frac{1}{2}\mathrm{MSE}(\widehat{\boldsymbol{\beta}})+\frac{\lambda}{2}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_1\le\lambda(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_1-\|\widehat{\boldsymbol{\beta}}\|_1+\|\boldsymbol{\beta}^*\|_1). \tag{S9.186}$$

Hence, by (S9.186) and the fact that $\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{J^c(\boldsymbol{\beta}^*)}\|_1-\|(\widehat{\boldsymbol{\beta}})_{J^c(\boldsymbol{\beta}^*)}\|_1+\|(\boldsymbol{\beta}^*)_{J^c(\boldsymbol{\beta}^*)}\|_1=0$, we have

$$\begin{aligned}\frac{1}{2}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_1 &\le\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)}\|_1-\|(\widehat{\boldsymbol{\beta}})_{J(\boldsymbol{\beta}^*)}\|_1+\|(\boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)}\|_1\\ &\le 2\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)}\|_1,\end{aligned}$$

which implies $\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{J^c(\boldsymbol{\beta}^*)}\|_1\le 3\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)}\|_1$. Combining this result and (S9.186), we have:

$$\frac{1}{2}\mathrm{MSE}(\widehat{\boldsymbol{\beta}})+\lambda\big\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)^c}\big\|_1\le 3\lambda\big\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)}\big\|_1.$$

Note that by **Assumptions A, E.2**, the restricted eigenvalue condition holds for $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*$. Hence, using similar proof techniques as in Bickel et al. (2009), we can derive (S6.11). To save space, we omit the details here. □

S9.4 Proof of Lemma 10

Proof. In this section, we prove Lemma 10. Similar to Lemma 9, we prove the results by assuming there is a change point t_1 such that $\boldsymbol{\beta}=\boldsymbol{\beta}^{(1)}$ if $i\le\lfloor nt_1\rfloor$ and $\boldsymbol{\beta}=\boldsymbol{\beta}^{(2)}$ if $i>\lfloor nt_1\rfloor$. Recall $\underset{\sim}{\boldsymbol{\beta}}^*=((\boldsymbol{\beta}^*)^\top,(\boldsymbol{b}^*)^\top)^\top\in\mathbb{R}^{p+K}$ defined in (S6.10). By the first order condition, for $\alpha=0$, $\underset{\sim}{\boldsymbol{\beta}}^*=$

$((\boldsymbol{\beta}^*)^\top, (\boldsymbol{b}^*)^\top)^\top \in \mathbb{R}^{p+K}$ satisfies the following equation:

$$\begin{aligned}&\mathbb{E}\Big[\textstyle\sum_{i=1}^n \sum_{k=1}^K \boldsymbol{X}_i(\mathbf{1}\{Y_i \le \boldsymbol{X}_i^\top \boldsymbol{\beta}^* + b_k^*\} - \tau_k)\Big] = \mathbf{0}_p,\\ &\mathbb{E}\Big[\textstyle\sum_{i=1}^n (\mathbf{1}\{Y_i \le \boldsymbol{X}_i^\top \boldsymbol{\beta}^* + b_k^*\} - \tau_k)\Big] = 0, \text{ for } k = 1, \ldots, K.\end{aligned} \tag{S9.187}$$

By the fact that $Y_i = \epsilon_i + \boldsymbol{\beta}^{(1)}\mathbf{1}\{i \le \lfloor nt_1 \rfloor\} + \boldsymbol{\beta}^{(2)}\mathbf{1}\{i > \lfloor nt_1 \rfloor\}$, for the above equation, we have:

$$\begin{aligned}&t_1\mathbb{E}\Big[\textstyle\sum_{k=1}^K \boldsymbol{X}\big(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}) + b_k^*) - F_\epsilon(b_k^{(0)})\big)\Big]\\ &+t_2\mathbb{E}\Big[\textstyle\sum_{k=1}^K \boldsymbol{X}\big(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(2)}) + b_k^*) - F_\epsilon(b_k^{(0)})\big)\Big] = \mathbf{0}_p,\end{aligned}$$

and for $k = 1, \ldots, K$,

$$t_1\mathbb{E}\Big[\big(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}) + b_k^*) - F_\epsilon(b_k^{(0)})\big)\Big] + t_2\mathbb{E}\Big[\big(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(2)}) + b_k^*) - F_\epsilon(b_k^{(0)})\big)\Big] = 0,$$

where $t_2 := 1 - t_1$. Moreover, let $\underset{\sim}{\boldsymbol{\beta}}^{(1)} := ((\boldsymbol{\beta}^{(1)})^\top, (\boldsymbol{b}^{(0)})^\top)^\top \in \mathbb{R}^{p+K}$, $\underset{\sim}{\boldsymbol{\beta}}^{(2)} := ((\boldsymbol{\beta}^{(2)})^\top, (\boldsymbol{b}^{(0)})^\top)^\top \in \mathbb{R}^{p+K}$, $\underset{\approx}{\boldsymbol{X}} := (\boldsymbol{X}^\top, \mathbf{1}_K) \in \mathbb{R}^{p+K}$, and $\boldsymbol{S}_k := \mathrm{diag}(\mathbf{1}_p, \boldsymbol{e}_k)$, where $\boldsymbol{e}_k \in \mathbb{R}^K$ is a vector with the k-th element being 1 and the others being zeros, and $\mathbf{1}_K$ is a K-dimensional vector with all elements being 1s. With the above notations, for the above equations, we have:

$$\begin{aligned}&t_1\mathbb{E}\Big[\sum_{k=1}^K \boldsymbol{S}_k \underset{\approx}{\boldsymbol{X}}\big(F_\epsilon((\boldsymbol{S}_k\underset{\approx}{\boldsymbol{X}})^\top(\underset{\approx}{\boldsymbol{\beta}}^* - \underset{\sim}{\boldsymbol{\beta}}^{(1)}) + b_k^{(0)}) - F_\epsilon(b_k^{(0)})\big)\Big]\\ &+t_2\mathbb{E}\Big[\sum_{k=1}^K \boldsymbol{S}_k \underset{\approx}{\boldsymbol{X}}\big(F_\epsilon((\boldsymbol{S}_k\underset{\approx}{\boldsymbol{X}})^\top(\underset{\approx}{\boldsymbol{\beta}}^* - \underset{\sim}{\boldsymbol{\beta}}^{(2)}) + b_k^{(0)}) - F_\epsilon(b_k^{(0)})\big)\Big] = \mathbf{0}_{p+K}.\end{aligned}$$

Furthermore, by the Taylor's expansion, we have:

$$t_1 \underbrace{\Big\{ \sum_{k=1}^{K} \mathbb{E}\Big[\int_0^1 (\boldsymbol{S}_k \underset{\sim\sim}{\boldsymbol{X}})(\boldsymbol{S}_k \underset{\sim\sim}{\boldsymbol{X}})^\top f_\epsilon\big(b_k^{(0)} + t((\boldsymbol{S}_k \underset{\sim\sim}{\boldsymbol{X}})^\top (\underset{\sim\sim}{\boldsymbol{\beta}^*} - \underset{\sim}{\boldsymbol{\beta}}^{(1)}))\big) dt \Big] \Big\}}_{\widetilde{\boldsymbol{\Sigma}}^{(1)} \in \mathbb{R}^{(p+K)\times(p+K)}} (\underset{\sim\sim}{\boldsymbol{\beta}^*} - \underset{\sim}{\boldsymbol{\beta}}^{(1)})$$
$$+ t_2 \underbrace{\Big\{ \sum_{k=1}^{K} \mathbb{E}\Big[\int_0^1 (\boldsymbol{S}_k \underset{\sim\sim}{\boldsymbol{X}})(\boldsymbol{S}_k \underset{\sim\sim}{\boldsymbol{X}})^\top f_\epsilon\big(b_k^{(0)} + t((\boldsymbol{S}_k \underset{\sim\sim}{\boldsymbol{X}})^\top (\underset{\sim\sim}{\boldsymbol{\beta}^*} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}))\big) dt \Big] \Big\}}_{\widetilde{\boldsymbol{\Sigma}}^{(2)} \in \mathbb{R}^{(p+K)\times(p+K)}} (\underset{\sim\sim}{\boldsymbol{\beta}^*} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}) = \mathbf{0}_{p+K}. \tag{S9.188}$$

Hence, for $\underset{\sim}{\boldsymbol{\beta}}^*$, by defining $\widetilde{\boldsymbol{\Sigma}}^{(1)}$ and $\widetilde{\boldsymbol{\Sigma}}^{(2)}$, it has the following explicit form:

$$\underset{\sim}{\boldsymbol{\beta}}^* = (t_1 \widetilde{\boldsymbol{\Sigma}}^{(1)} + t_2 \widetilde{\boldsymbol{\Sigma}}^{(2)})^{-1} (t_1 \widetilde{\boldsymbol{\Sigma}}^{(1)} \boldsymbol{\beta}^{(1)} + t_2 \widetilde{\boldsymbol{\Sigma}}^{(2)} \boldsymbol{\beta}^{(2)}).$$

Moreover, using some calculations, we have:

$$\begin{aligned} \underset{\sim}{\boldsymbol{\beta}}^* - \underset{\sim}{\boldsymbol{\beta}}^{(1)} &= (t_1 \widetilde{\boldsymbol{\Sigma}}^{(1)} + t_2 \widetilde{\boldsymbol{\Sigma}}^{(2)})^{-1} t_2 \widetilde{\boldsymbol{\Sigma}}^{(2)} (\underset{\sim}{\boldsymbol{\beta}}^{(2)} - \underset{\sim}{\boldsymbol{\beta}}^{(1)}) \\ \underset{\sim}{\boldsymbol{\beta}}^* - \underset{\sim}{\boldsymbol{\beta}}^{(2)} &= (t_1 \widetilde{\boldsymbol{\Sigma}}^{(1)} + t_2 \widetilde{\boldsymbol{\Sigma}}^{(2)})^{-1} t_1 \widetilde{\boldsymbol{\Sigma}}^{(1)} (\underset{\sim}{\boldsymbol{\beta}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}). \end{aligned} \tag{S9.189}$$

Remark 5. Note that for any matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ and $\boldsymbol{x} \in \mathbb{R}^p$, we have $\|\mathbf{A}\boldsymbol{x}\|_1 \le \|\mathbf{A}\|_{1,1} \|\boldsymbol{x}\|_1$. Hence, if we assume that

$$\Big\|(t_1 \widetilde{\boldsymbol{\Sigma}}^{(1)} + t_2 \widetilde{\boldsymbol{\Sigma}}^{(2)})^{-1} t_2 \widetilde{\boldsymbol{\Sigma}}^{(2)}\Big\|_{1,1} \le C_1 \text{ and } \Big\|(t_1 \widetilde{\boldsymbol{\Sigma}}^{(1)} + t_2 \widetilde{\boldsymbol{\Sigma}}^{(2)})^{-1} t_1 \widetilde{\boldsymbol{\Sigma}}^{(1)}\Big\|_{1,1} \le C_2,$$

we can prove that $\|\underset{\sim}{\boldsymbol{\beta}}^* - \underset{\sim}{\boldsymbol{\beta}}^{(1)}\|_1 \le C_1 \|\underset{\sim}{\boldsymbol{\beta}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}\|_1 = C_1 \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1$ and $\|\underset{\sim}{\boldsymbol{\beta}}^* - \underset{\sim}{\boldsymbol{\beta}}^{(2)}\|_1 \le C_2 \|\underset{\sim}{\boldsymbol{\beta}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}\|_1 = C_2 \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1$ for the above positive constants $C_1, C_2 > 0$.

So far, we have derived the explicit form for $\underset{\sim}{\boldsymbol{\beta}}^*$ and the difference between $\underset{\sim}{\boldsymbol{\beta}}^* - \underset{\sim}{\boldsymbol{\beta}}^{(1)}$ or $\underset{\sim}{\boldsymbol{\beta}}^* - \underset{\sim}{\boldsymbol{\beta}}^{(2)}$, which is very important for proving Lemma

10. Now, we are ready to give the detailed proof. To that end, we define the following parameter space. Let $\underset{\sim}{\boldsymbol{\Delta}} = (\boldsymbol{\Delta}^\top, \boldsymbol{\delta}^\top)^\top \in \mathbb{R}^{p+K}$ with $\boldsymbol{\Delta} \in \mathbb{R}^p$ and $\boldsymbol{\delta} \in \mathbb{R}^K$, we define

$$\mathcal{A} = \left\{ \ (\boldsymbol{\Delta}^\top, \boldsymbol{\delta}^\top)^\top : \|\boldsymbol{\Delta}_{J^c(\boldsymbol{\beta}^*)}\|_1 \leq 3\|\boldsymbol{\Delta}_{J(\boldsymbol{\beta}^*)}\|_1 + \|\boldsymbol{\delta}\|_1 \right\}. \tag{S9.190}$$

For any $\underset{\sim}{\boldsymbol{\beta}} = ((\boldsymbol{\beta})^\top, (\boldsymbol{b})^\top)^\top \in \mathbb{R}^{p+K}$, let $\underset{\sim}{\boldsymbol{\Delta}} = \underset{\sim}{\boldsymbol{\beta}} - \underset{\sim}{\boldsymbol{\beta}}^*$ and $\underset{\sim}{\widehat{\boldsymbol{\Delta}}} = \underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^*$, where $\widehat{\boldsymbol{\beta}}$ is the minimizer of the empirical loss defined in (2.10) with $\alpha = 0$. Define the empirical loss and its expectation:

$$L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}) := \frac{1}{n}\sum_{i=1}^n \frac{1}{K}\sum_{k=1}^K \rho_{\tau_k}(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta} - b_k),$$
$$\text{and } L_K(\underset{\sim}{\boldsymbol{\beta}}) := \mathbb{E}[L_{n,K}(\underset{\sim}{\boldsymbol{\beta}})] = \frac{1}{n}\sum_{i=1}^n \frac{1}{K}\sum_{k=1}^K \mathbb{E}[\rho_{\tau_k}(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta} - b_k)].$$

Then, we can further define the excess risk as:

$$H(\underset{\sim}{\boldsymbol{\Delta}}) = L_K(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\sim}{\boldsymbol{\Delta}}) - L_K(\underset{\sim}{\boldsymbol{\beta}}^*).$$

The proof of Lemma 10 relies on the following three lemmas. Lemma 18 shows that $\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}$ belongs to $\mathcal{A}$ with a large probability. The proof of Lemma 18 is given in Section S10.2.

Lemma 18. *Assume* **Asssumptions A, D, E.2 - E.4** *hold. Then, with probability tending to one, we have*

$$\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^* \in \mathcal{A}.$$

Next, Lemma 19 shows that the excess risk $H(\underset{\sim}{\boldsymbol{\Delta}})$ can be bounded by the quadratic form of $\underset{\sim}{\boldsymbol{\Delta}}$. To show this, define

$$\boldsymbol{S} = \sum_{k=1}^{K} \begin{pmatrix} \boldsymbol{\Sigma}, \boldsymbol{0} \\ \\ \boldsymbol{0}, \operatorname{diag}(\boldsymbol{e}_k) \end{pmatrix} \in \mathbb{R}^{(p+K)\times(p+K)}, \quad \|\underset{\sim}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^2 = \underset{\sim}{\boldsymbol{\Delta}}^\top \boldsymbol{S} \underset{\sim}{\boldsymbol{\Delta}} = \sum_{k=1}^{K} (\boldsymbol{\Delta}^\top \boldsymbol{\Sigma} \boldsymbol{\Delta} + \delta_k^2). \tag{S9.191}$$

Lemma 19. *Assume* **Assumptions A, D, E.2 - E.4** *hold. For any $\underset{\sim}{\boldsymbol{\Delta}} \in \mathcal{A}$, with probability tending to one, we have*

$$H(\underset{\sim}{\boldsymbol{\Delta}}) \geq c_* \min\Big(\frac{\|\underset{\sim}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^2}{4}, \frac{\|\underset{\sim}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}}{4}\Big),$$

where $c_* > 0$ is some universal constant not depending on n or p.

Lastly, Lemma 20 shows that we can uniformly control the difference between the excess risk and its empirical version.

Lemma 20. *Assume* **Asssumptions A, D, E.2 - E.4** *hold. With probability tending to one, we have:*

$$\sup_{\substack{\underset{\sim}{\boldsymbol{\Delta}} \in \mathcal{A} \\ \|\underset{\sim}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}} \leq \xi}} \big|(L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\sim}{\boldsymbol{\Delta}}) - L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)) - (L_K(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\sim}{\boldsymbol{\Delta}}) - L_K(\underset{\sim}{\boldsymbol{\beta}}^*))\big| \leq C^* M \xi \sqrt{s \frac{\log(pn)}{n}},$$

where $C^* > 0$ is some universal constant not depending on n or p and $s := |J(\boldsymbol{\beta}^*)|$.

With the above lemmas, we are ready to prove Lemma 10. Define two *events* $\mathcal{E}_1$ *and* $\mathcal{E}_2$ *as:*

$$\mathcal{E}_1 = \{\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^* \in \mathcal{A}\},$$

$$\mathcal{E}_2 = \Bigg\{ \sup_{\substack{\underset{\sim}{\boldsymbol{\Delta}} \in \mathcal{A} \\ \|\underset{\sim}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}} \le \xi}} \left|(L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\sim}{\boldsymbol{\Delta}}) - L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)) - (L_K(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\sim}{\boldsymbol{\Delta}}) - L_K(\underset{\sim}{\boldsymbol{\beta}}^*))\right| \le C^* M \xi \sqrt{s \frac{\log(pn)}{n}} \Bigg\}.$$

By Lemmas 18 and 20, we have $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \to 1$. *Hence, in what follows, we give the proof under the event* $\mathcal{E}_1 \cap \mathcal{E}_2$. *Let* $\|\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^*\|_{\boldsymbol{S}} = \xi$. *By the optimality of* $\underset{\sim}{\widehat{\boldsymbol{\beta}}}$, *we have:*

$$L_{n,K}(\underset{\sim}{\widehat{\boldsymbol{\beta}}}) - L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*) + \lambda(\|\underset{\sim}{\widehat{\boldsymbol{\beta}}}\|_1 - \|\underset{\sim}{\boldsymbol{\beta}}^*\|_1) \le 0.$$

Moreover, using the above inequality, under $\mathcal{E}_2$, *we have:*

$$\begin{aligned}
&\lambda(\|\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^*\|_1) \\
&\quad \ge_{(1)} L_{n,K}(\underset{\sim}{\widehat{\boldsymbol{\beta}}}) - L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*) \\
&\quad \ge_{(2)} L_K(\underset{\sim}{\widehat{\boldsymbol{\beta}}}) - L_K(\underset{\sim}{\boldsymbol{\beta}}^*) - C^* M \xi \sqrt{s \frac{\log(pn)}{n}} \\
&\quad \ge_{(3)} c_* \min\left(\frac{\xi^2}{4}, \frac{\xi}{4}\right) - C^* M \xi \sqrt{s \frac{\log(pn)}{n}}.
\end{aligned} \tag{S9.192}$$

Note that under $\mathcal{E}_1$, *we have:*

$$\begin{aligned}
&\|\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^*\|_1 \\
&\quad \le 4\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)}\|_1 + \|\widehat{\boldsymbol{b}} - \boldsymbol{b}^*\|_1 \\
&\quad \le 4\sqrt{s}\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)}\|_2 + \sqrt{K}\|\widehat{\boldsymbol{b}} - \boldsymbol{b}^*\|_2 \\
&\quad \le 4\sqrt{s}\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2 + \sqrt{K}\|\widehat{\boldsymbol{b}} - \boldsymbol{b}^*\|_2 \\
&\quad \le C_E\sqrt{s}\|\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^*\|_{\boldsymbol{S}},
\end{aligned} \tag{S9.193}$$

where $C_E > 0$ *is some universal constant. Note that we can choose* $\lambda = C_\lambda\sqrt{\log(p)/n}$ *for some big enough constant* $C_\lambda > 0$*. Combining (S9.192) and (S9.193), we have:*

$$c_* \min\left(\frac{\xi^2}{4}, \frac{\xi}{4}\right) - C^* M\xi\sqrt{s\frac{\log(pn)}{n}} - C_\lambda\xi\sqrt{s\frac{\log(pn)}{n}} \leq 0,$$

which implies

$$c_*\frac{\xi}{4} - C^* M\xi\sqrt{s\frac{\log(pn)}{n}} - C_\lambda\xi\sqrt{s\frac{\log(pn)}{n}} \leq 0,$$

or

$$c_*\frac{\xi^2}{4} - C^* M\xi\sqrt{s\frac{\log(pn)}{n}} - C_\lambda\xi\sqrt{s\frac{\log(pn)}{n}} \leq 0.$$

Note that $\sqrt{s\frac{\log(pn)}{n}} = o(1)$*. Hence, only the second case applies. As a result, we have:*

$$\xi = \|\widehat{\underset{\sim}{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^*\|_{\boldsymbol{S}} \leq CM\sqrt{s\frac{\log(pn)}{n}}. \tag{S9.194}$$

Lastly, by (S9.194) and some trivial calculations, we can directly derive (S6.12).

□

S9.5 Proof of Lemma 11

Proof. In this section, we prove Lemma 11. Similar to Lemmas 9 and 10, we give the results by assuming there is a change point t_1 such that $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}$ if

$i \le \lfloor nt_1 \rfloor$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^{(2)}$ if $i > \lfloor nt_1 \rfloor$. Note that the results are still applicable even if there is no change point.

Before the proof, we need some discussion about $\underset{\sim}{\boldsymbol{\beta}}^* = ((\boldsymbol{\beta}^*)^\top, (\boldsymbol{b}^*)^\top)^\top \in \mathbb{R}^{p+K}$, which is defined in (S6.10). By the first order condition, for $\alpha \in (0,1)$, we can see that $\underset{\sim}{\boldsymbol{\beta}}^* = ((\boldsymbol{\beta}^*)^\top, (\boldsymbol{b}^*)^\top)^\top \in \mathbb{R}^{p+K}$ satisfies the following equation:

$$
\begin{aligned}
&(1-\alpha)\mathbb{E}\Big[\sum_{i=1}^{n}\sum_{k=1}^{K}\boldsymbol{X}_i(\mathbf{1}\{Y_i \le \boldsymbol{X}_i^\top\boldsymbol{\beta}^* + b_k^*\} - \tau_k)\Big] - \alpha\sum_{i=1}^{n}\mathbb{E}\Big[\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*)\Big] = \mathbf{0}_p,\\
&(1-\alpha)\mathbb{E}\Big[\sum_{i=1}^{n}(\mathbf{1}\{Y_i \le \boldsymbol{X}_i^\top\boldsymbol{\beta}^* + b_k^*\} - \tau_k)\Big] = 0, \text{ for } k = 1, \ldots, K.
\end{aligned}
\tag{S9.195}
$$

Note that $Y_i = \epsilon_i + \boldsymbol{\beta}^{(1)}\mathbf{1}\{i \le \lfloor nt_1 \rfloor\} + \boldsymbol{\beta}^{(2)}\mathbf{1}\{i > \lfloor nt_1 \rfloor\}$. Similar to the analysis in Section S9.4, for the above equation, we have:

$$
\begin{aligned}
&t_1\Big\{(1-\alpha)\mathbb{E}\Big[\sum_{k=1}^{K}\boldsymbol{X}\big(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}) + b_k^*) - F_\epsilon(b_k^{(0)})\big)\Big] + \alpha\mathbb{E}\Big[\boldsymbol{X}\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)})\Big]\Big\}\\
&+t_2\Big\{(1-\alpha)\mathbb{E}\Big[\sum_{k=1}^{K}\boldsymbol{X}\big(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(2)}) + b_k^*) - F_\epsilon(b_k^{(0)})\big)\Big] + \alpha\mathbb{E}\Big[\boldsymbol{X}\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(2)})\Big]\Big\}\\
&= \mathbf{0}_p,
\end{aligned}
$$

and for $k = 1, \ldots, K$,

$$
\begin{aligned}
&t_1\mathbb{E}\Big[\textstyle\sum_{k=1}^{K}\big(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}) + b_k^*) - F_\epsilon(b_k^{(0)})\big)\Big]\\
&+t_2\mathbb{E}\Big[\textstyle\sum_{k=1}^{K}\big(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(2)}) + b_k^*) - F_\epsilon(b_k^{(0)})\big)\Big] = 0,
\end{aligned}
$$

where $t_2 := 1 - t_1$. Moreover, let

$$
\widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Sigma}, \mathbf{0} \\ \mathbf{0}, \mathbf{0} \end{pmatrix} \in \mathbb{R}^{(p+K)\times(p+K)}. \tag{S9.196}
$$

Then, using similar analysis as in Section S9.4, we have:

$$t_1 \underbrace{\left[(1-\alpha)\widetilde{\mathbf{\Sigma}}^{(1)} + \alpha\widetilde{\mathbf{\Sigma}}\right]}_{\breve{\mathbf{\Sigma}}^{(1)}} (\underset{\approx}{\boldsymbol{\beta}}^{*} - \underset{\sim}{\boldsymbol{\beta}}^{(1)}) + t_2 \underbrace{\left[(1-\alpha)\widetilde{\mathbf{\Sigma}}^{(2)} + \alpha\widetilde{\mathbf{\Sigma}}\right]}_{\breve{\mathbf{\Sigma}}^{(2)}} (\underset{\approx}{\boldsymbol{\beta}}^{*} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}) = \mathbf{0}_{p+K},$$

where $\widetilde{\mathbf{\Sigma}}^{(1)}$ and $\widetilde{\mathbf{\Sigma}}^{(2)}$ are defined in (S9.188). Hence, for $\underset{\sim}{\boldsymbol{\beta}}^{*}$, it has the following explicit form:

$$\underset{\sim}{\boldsymbol{\beta}}^{*} = (t_1\breve{\mathbf{\Sigma}}^{(1)} + t_2\breve{\mathbf{\Sigma}}^{(2)})^{-1}(t_1\breve{\mathbf{\Sigma}}^{(1)}\boldsymbol{\beta}^{(1)} + t_2\breve{\mathbf{\Sigma}}^{(2)}\boldsymbol{\beta}^{(2)}).$$

Moreover, using some calculations, we have:

$$\begin{aligned}\underset{\sim}{\boldsymbol{\beta}}^{*} - \underset{\sim}{\boldsymbol{\beta}}^{(1)} &= (t_1\breve{\mathbf{\Sigma}}^{(1)} + t_2\breve{\mathbf{\Sigma}}^{(2)})^{-1}t_2\breve{\mathbf{\Sigma}}^{(2)}(\underset{\sim}{\boldsymbol{\beta}}^{(2)} - \underset{\sim}{\boldsymbol{\beta}}^{(1)}),\\ \underset{\sim}{\boldsymbol{\beta}}^{*} - \underset{\sim}{\boldsymbol{\beta}}^{(2)} &= (t_1\breve{\mathbf{\Sigma}}^{(1)} + t_2\breve{\mathbf{\Sigma}}^{(2)})^{-1}t_1\breve{\mathbf{\Sigma}}^{(1)}(\underset{\sim}{\boldsymbol{\beta}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}).\end{aligned} \tag{S9.197}$$

Remark 6. If we assume that

$$\left\|(t_1\breve{\mathbf{\Sigma}}^{(1)} + t_2\breve{\mathbf{\Sigma}}^{(2)})^{-1}t_2\breve{\mathbf{\Sigma}}^{(2)}\right\|_{1,1} \leq C_1 \text{ and } \left\|(t_1\breve{\mathbf{\Sigma}}^{(1)} + t_2\breve{\mathbf{\Sigma}}^{(2)})^{-1}t_1\breve{\mathbf{\Sigma}}^{(1)}\right\|_{1,1} \leq C_2,$$

we can prove that $\|\underset{\sim}{\boldsymbol{\beta}}^{*} - \underset{\sim}{\boldsymbol{\beta}}^{(1)}\|_1 \leq C_1\|\underset{\sim}{\boldsymbol{\beta}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}\|_1 = C_1\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1$ and $\|\underset{\sim}{\boldsymbol{\beta}}^{*} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}\|_1 \leq C_2\|\underset{\sim}{\boldsymbol{\beta}}^{(1)} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}\|_1 = C_2\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1$ for the above positive constants $C_1, C_2 > 0$.

For $\alpha \in (0,1)$, we have derived the explicit form for $\underset{\sim}{\boldsymbol{\beta}}^{*}$ and the difference between $\underset{\sim}{\boldsymbol{\beta}}^{*} - \underset{\sim}{\boldsymbol{\beta}}^{(1)}$ or $\underset{\sim}{\boldsymbol{\beta}}^{*} - \underset{\sim}{\boldsymbol{\beta}}^{(2)}$, which is very important for proving Lemma 11. Now, we are ready to give the proof.

Recall the parameter space $\mathcal{A}$ defined in (S9.190). For any $\underset{\sim}{\boldsymbol{\beta}} = ((\boldsymbol{\beta})^\top, (\boldsymbol{b})^\top)^\top \in \mathbb{R}^{p+K}$, let $\underset{\approx}{\mathbf{\Delta}} = \underset{\sim}{\boldsymbol{\beta}} - \underset{\sim}{\boldsymbol{\beta}}^{*}$ and $\underset{\approx}{\widehat{\mathbf{\Delta}}} = \underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^{*}$, where $\widehat{\boldsymbol{\beta}}$ is the minimizer of the empirical loss defined in (2.10) with $\alpha \in (0,1)$. Define the empirical loss and its expectation:

$$L_{n,K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}) := (1-\alpha)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\rho_{\tau_k}(Y_i - b_i - \boldsymbol{X}_i^{\top}\boldsymbol{\beta}) + \frac{\alpha}{2n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^{\top}\boldsymbol{\beta})^2,$$
$$\text{and } L_K^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}) := \mathbb{E}[L_{n,K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}})]. \tag{S9.198}$$

Then, for each $\alpha \in (0,1)$, we can further define the excess risk as:

$$H^{\alpha}(\underset{\approx}{\boldsymbol{\Delta}}) = L_K^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\approx}{\boldsymbol{\Delta}}) - L_K^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^*).$$

Similar to Section S9.4, the proof of Lemma 11 relies on the following three lemmas. Specifically, Lemma 21 shows that $\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}$ belongs to $\mathcal{A}$ with a large probability. Lemma 22 shows that the excess risk $H^{\alpha}(\underset{\approx}{\boldsymbol{\Delta}})$ can be bounded by the quadratic form of $\underset{\approx}{\boldsymbol{\Delta}}$. Lastly, Lemma 23 shows that we can uniformly control the difference between the excess risk and its empirical version. The proofs of those lemmas are given in Sections S10.5 - S10.7.

Lemma 21. *Assume* **Assumptions A, B, C.2, D, E.2 - E.4** *hold.Then, with probability tending to one, we have*

$$\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^* \in \mathcal{A}.$$

Lemma 22. *Assume* **Assumptions A, B, C.2, D, E.2 - E.4** *hold. For any* $\underset{\approx}{\boldsymbol{\Delta}} \in \mathcal{A}$*, with probability tending to one,*

$$H^{\alpha}(\underset{\approx}{\boldsymbol{\Delta}}) \geq c_* \min\Big(\frac{\|\underset{\approx}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^2}{4}, \frac{\|\underset{\approx}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}}{4}\Big),$$

where $c_* > 0$ is some universal constant not depending on n or p.

Lemma 23. *Assume* **Assumptions A, B, C.2, D, E.2 - E.4** *hold. With probability tending to one, we have:*

$$\sup_{\substack{\boldsymbol{\Delta}\in\mathcal{A}\\ \|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\leq\xi}} \left|(L_{n,K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^*+\underset{\sim}{\boldsymbol{\Delta}})-L_{n,K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^*))-(L_{K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^*+\underset{\sim}{\boldsymbol{\Delta}})-L_{K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^*))\right| \leq C^* M\xi\sqrt{s\frac{\log(pn)}{n}},$$

where $C^* > 0$ is some universal constant not depending on n or p and $s := |J(\boldsymbol{\beta}^*)|$.

With the above three lemmas, using similar proof procedures as in Section S9.4, we can directly prove Lemma 11. To save space, we omit the details here.

□

S10 Additional lemmas

S10.1 Proof of Lemma 17

Proof. Recall

$$\boldsymbol{C}_0^{II,2}(t) = \frac{1}{\sqrt{n}\widehat{\sigma}(\alpha,\widetilde{\boldsymbol{\tau}})}\Big(\sum_{i=1}^{\lfloor nt\rfloor}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(g_{ik}(\underset{\sim}{\boldsymbol{X}_i}^\top\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}) - g_{ik}(0)\big)$$
$$-\frac{\lfloor nt\rfloor}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(g_{ik}(\underset{\sim}{\boldsymbol{X}_i}^\top\underset{\sim}{\widehat{\boldsymbol{\Delta}}_k}) - g_{ik}(0)\big),$$

where $g_{ik}(t) := \mathbf{1}\{\epsilon_i \le b_k^{(0)} + t\} - \mathbb{P}\{\epsilon_i \le b_k^{(0)} + t\}$. Note that under $\mathcal{E}$, we have:

$$\begin{aligned}
&\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,2}(t)\|_{(s_0,2)} \\
&\le C\sqrt{n}s_0^{1/2} \max_{t\in[q_0,1-q_0]} \Big\|\Big(\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(g_{ik}(\undertilde{\boldsymbol{X}_i}^\top\undertilde{\widehat{\boldsymbol{\Delta}}_k}) - g_{ik}(0)\big)\Big\|_\infty \\
&\quad + C\sqrt{n}s_0^{1/2}\Big\|\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i\big(g_{ik}(\undertilde{\boldsymbol{X}_i}^\top\undertilde{\widehat{\boldsymbol{\Delta}}_k}) - g_{ik}(0)\big)\Big\|_\infty \\
&\le C\sqrt{n}s_0^{1/2}\max_t\max_j\Big|\Big(\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\frac{1}{K}\sum_{k=1}^{K}X_{ij}\big(g_{ik}(\undertilde{\boldsymbol{X}_i}^\top\undertilde{\widehat{\boldsymbol{\Delta}}_k}) - g_{ik}(0)\big)\Big| \\
&\quad + C\sqrt{n}s_0^{1/2}\max_j\Big|\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}X_{ij}\big(g_{ik}(\undertilde{\boldsymbol{X}_i}^\top\undertilde{\widehat{\boldsymbol{\Delta}}_k}) - g_{ik}(0)\big)\Big|.
\end{aligned}$$

Define

$$\begin{aligned}
\psi_j(\epsilon_i, \boldsymbol{X}_i; \undertilde{\boldsymbol{\Delta}_k}) &:= X_{ij}(\mathbf{1}\{\epsilon_i \le \undertilde{\boldsymbol{X}_i}^\top\undertilde{\boldsymbol{\Delta}_k} + b_k^{(0)}\} - \mathbf{1}\{\epsilon_i \le b_k^{(0)}\}) \\
&= X_{ij}(\mathbf{1}\{\epsilon_i \le \boldsymbol{X}_i^\top\boldsymbol{\Delta} + \delta_k + b_k^{(0)}\} - \mathbf{1}\{\epsilon_i \le b_k^{(0)}\}),
\end{aligned}$$

where $\boldsymbol{\Delta} := \boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}$ and $\delta_k := b_k - b_k^{(0)}$ for $1 \le k \le K$. Hence, by definition, conditional on $\mathcal{X} := (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, we have:

$$\begin{aligned}
&\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,2}(t)\|_{(s_0,2)}|\mathcal{X} \\
&\le C\sqrt{n}s_0^{1/2}\underbrace{\max_t\max_j\Big|\Big(\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\frac{1}{K}\sum_{k=1}^{K}(\psi_j(\epsilon_i, \boldsymbol{X}_i; \undertilde{\widehat{\boldsymbol{\Delta}}_k}) - \mathbb{E}[\psi_j(\epsilon_i, \boldsymbol{X}_i; \undertilde{\widehat{\boldsymbol{\Delta}}_k})])\Big)\Big|}_{I} \\
&\quad + C\sqrt{n}s_0^{1/2}\underbrace{\max_j\Big|\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}(\psi_j(\epsilon_i, \boldsymbol{X}_i; \undertilde{\widehat{\boldsymbol{\Delta}}_k}) - \mathbb{E}[\psi_j(\epsilon_i, \boldsymbol{X}_i; \undertilde{\widehat{\boldsymbol{\Delta}}_k})])\Big)\Big|}_{II}.
\end{aligned} \tag{S10.199}$$

Hence, to bound $\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,2}(t)\|_{(s_0,2)}|\mathcal{X}$, we need to consider I and II, respectively. We first consider I. To that end, conditional on $\mathcal{X} :=$

$(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)$, define the function:

$$\begin{aligned} G_{t,j}(\underset{\sim}{\boldsymbol{\Delta}}) &= \frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}\frac{1}{K}\sum_{k=1}^{K}(\psi_j(\epsilon_i,\boldsymbol{X}_i;\underset{\sim}{\boldsymbol{\Delta}_k}) - \mathbb{E}[\psi_j(\epsilon_i,\boldsymbol{X}_i;\underset{\sim}{\boldsymbol{\Delta}_k})]) \\ &:= \frac{1}{n'}\sum_{i=1}^{n'}\frac{1}{K}\sum_{k=1}^{K}(\psi_j(\epsilon_i,\boldsymbol{X}_i;\underset{\sim}{\boldsymbol{\Delta}_k}) - \mathbb{E}[\psi_j(\epsilon_i,\boldsymbol{X}_i;\underset{\sim}{\boldsymbol{\Delta}_k})]) \\ &:= G_{n',j}(\underset{\sim}{\boldsymbol{\Delta}}), \end{aligned}$$

where $n' := \lfloor nt \rfloor$. Moreover, for the sparsity parameter s of $\boldsymbol{\beta}^{(0)}$, and some big enough real numbers $\xi_1,\xi_2,\xi_3 > 0$, define the parameter space:

$$\begin{aligned} &\mathcal{R}(\xi_1,\xi_2,\xi_3) \\ &:= \left\{\underset{\sim}{\boldsymbol{\Delta}} = (\boldsymbol{\Delta}^\top,\boldsymbol{\delta}^\top)^\top : \|\boldsymbol{\Delta}\|_0 \le \xi_1 s, \|\boldsymbol{\Delta}\|_2 \le \xi_2\sqrt{s\frac{\log(pn)}{n}}, \|\boldsymbol{\delta}\|_2 \le \xi_3\sqrt{s\frac{\log(pn)}{n}}\right\}. \end{aligned}$$

By Lemma 10, with probability tending to 1, we have $\widehat{\underset{\sim}{\boldsymbol{\Delta}}} \in \mathcal{R}(\xi_1,\xi_2,\xi_3)$ for some large enough constants $\xi_1,\xi_2,\xi_3 > 0$. Hence, to bound I, it is sufficient to bound:

$$\max_{1\le j\le p}\max_{t\in[q_0,1-q_0]}\sup_{\underset{\sim}{\boldsymbol{\Delta}}\in\mathcal{R}(\xi_1,\xi_2,\xi_3)}|G_{t,j}(\underset{\sim}{\boldsymbol{\Delta}})||\mathcal{X} = \max_{1\le j\le p}\max_{n'\in[\lfloor nq_0\rfloor,n-\lfloor nq_0\rfloor]}\sup_{\underset{\sim}{\boldsymbol{\Delta}}\in\mathcal{R}(\xi_1,\xi_2,\xi_3)}|G_{n',j}(\underset{\sim}{\boldsymbol{\Delta}})||\mathcal{X}.$$

Throughout the following proofs, we assume K is fixed which does not grow with n. To obtain the desired bound, we define the functional class:

$$\mathcal{F} = \left\{f_{\underset{\sim}{\boldsymbol{\Delta}}}(\epsilon,\boldsymbol{X}) = \frac{1}{K}\sum_{k=1}^{K}(\psi_j(\epsilon,\boldsymbol{X};\underset{\sim}{\boldsymbol{\Delta}_k})|\underset{\sim}{\boldsymbol{\Delta}} \in \mathcal{R}(\xi_1,\xi_2,\xi_3)\right\}. \tag{S10.200}$$

Firstly, we obtain the upper bound for each fixed $n' \in [\lfloor nq_0\rfloor, n-\lfloor nq_0\rfloor]$ and $1 \le j \le p$. The main idea is to use Theorem 3.11 in Koltchinskii (2011) (Lemma A.1 in Zhao et al. (2014)) and the Bousquet inequality (Corollary 14.2 in Bühlmann and Van de Geer (2011)) to obtain the tail probability of $\sup_{\boldsymbol{\Delta} \in \mathcal{R}(\xi_1, \xi_2, \xi_3)} |G_{n',j}(\underset{\sim}{\boldsymbol{\Delta}})||\mathcal{X}$. The proofs proceed into five steps.

Step 1: Obtain the envelope for $f_{\boldsymbol{\Delta}}(\epsilon, \boldsymbol{X})$. In fact, by **Assumption A**, we have:

$$
\begin{aligned}
&\sup_{\boldsymbol{\Delta} \in \mathcal{R}(\xi_1, \xi_2, \xi_3)} |f_{\boldsymbol{\Delta}}(\epsilon, \boldsymbol{X})| \\
&\quad = \sup_{\boldsymbol{\Delta} \in \mathcal{R}(\xi_1, \xi_2, \xi_3)} \frac{1}{K} |\sum_{k=1}^{K} (\psi_j(\epsilon, \boldsymbol{X}; \underset{\sim}{\boldsymbol{\Delta}_k})| \\
&\quad = \sup_{\boldsymbol{\Delta} \in \mathcal{R}(\xi_1, \xi_2, \xi_3)} \frac{1}{K} |\sum_{k=1}^{K} (X_j \mathbf{1}\{\epsilon \leq \boldsymbol{X}^\top \boldsymbol{\Delta} + \delta_k + b_k^{(0)}\} - \mathbf{1}\{\epsilon \leq b_k^{(0)}\})| \\
&\quad \leq M,
\end{aligned}
$$

where the last inequality comes from the assumption that $|X_j| \leq M$ for $1 \leq j \leq p$.

Step 2: Obtain the upper bound for $\sigma_{n'}^2 := \sup_{\boldsymbol{\Delta}} \frac{1}{n'} \sum_{i=1}^{n'} \mathbf{Var}[f_{\boldsymbol{\Delta}}(\epsilon_i, \boldsymbol{X}_i)|\mathcal{X}]$, where $\mathcal{X} := \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$. In fact, similar to the proof of Lemma 6.1 in

Zhao et al. (2014), we have:

$$
\begin{aligned}
&\frac{1}{n'}\sum_{i=1}^{n'}\mathrm{Var}[f_{\boldsymbol{\Delta}}(\epsilon_i,\boldsymbol{X}_i)|\mathcal{X}]\\
&\quad\le_{(1)} \frac{1}{n'}\sum_{i=1}^{n'}\frac{6C_+M^2}{K}\sum_{k=1}^{K}|\widetilde{\boldsymbol{X}}_i^\top \underset{\sim}{\boldsymbol{\Delta}}_k|\\
&\quad\le_{(2)} \frac{6C_+M^2}{K}\Big(\frac{1}{n'}\sum_{i=1}^{n'}K|\boldsymbol{X}_i^\top\boldsymbol{\Delta}|+\|\boldsymbol{\delta}\|_1\Big)\\
&\quad\le_{(3)} \frac{6C_+M^2}{K}\Big(K\sqrt{\boldsymbol{\Delta}^\top(\frac{1}{n'}\sum_{i=1}^{n'}\boldsymbol{X}_i\boldsymbol{X}_i^\top)\boldsymbol{\Delta}}+\sqrt{K}\|\boldsymbol{\delta}\|_2\Big)\\
&\quad\le_{(4)} \frac{6C_+M^2}{K}\Big(K\sqrt{\boldsymbol{\Delta}^\top(\frac{1}{n'}\sum_{i=1}^{n'}\boldsymbol{X}_i\boldsymbol{X}_i^\top-\boldsymbol{\Sigma})\boldsymbol{\Delta}+\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}\boldsymbol{\Delta}}+\sqrt{K}\|\boldsymbol{\delta}\|_2\Big)\\
&\quad\le_{(5)} CM^2(\|\boldsymbol{\Delta}\|_2+\|\boldsymbol{\delta}\|_2)\\
&\quad\le_{(6)} CM^2\sqrt{s\frac{\log(pn)}{n}}:=\sigma_{n'}^2,
\end{aligned}
\tag{S10.201}
$$

where (3) comes from the Cauchy-Swartz inequality, (5) comes from Lemma 8.

Step 3: Obtain the covering number of the functional class $\mathcal{F}$ as defined in (S10.200). Let $\mathcal{T}\subset\{1,\dots,p\}$ with $|\mathcal{T}|=\xi_1 s$. Moreover, define the following functional classes:

$$
\begin{aligned}
&\mathcal{F}_k:=\Big\{f_k(\epsilon,\boldsymbol{X})=\mathbf{1}\{\epsilon\le\boldsymbol{X}^\top\boldsymbol{\Delta}+\delta_k+b_k^{(0)}\}-\mathbf{1}\{\epsilon\le b_k^{(0)}\}|\underset{\sim}{\boldsymbol{\Delta}}\in\mathcal{R}(\xi_1,\xi_2,\xi_3),\mathrm{supp}(\boldsymbol{\Delta})\subset\mathcal{T}\Big\},\\
&\mathcal{F}_K:=\Big\{f_K(\epsilon,\boldsymbol{X})=\sum_{k=1}^{K}\big(\mathbf{1}\{\epsilon\le\boldsymbol{X}^\top\boldsymbol{\Delta}+\delta_k+b_k^{(0)}\}-\mathbf{1}\{\epsilon\le b_k^{(0)}\}\big)|\underset{\sim}{\boldsymbol{\Delta}}\in\mathcal{R}(\xi_1,\xi_2,\xi_3),\mathrm{supp}(\boldsymbol{\Delta})\subset\mathcal{T}\Big\}\\
&\mathcal{F}_0:=\Big\{f_0(\epsilon,\boldsymbol{X})=\frac{1}{K}X_j\}\\
&\mathcal{F}_{\mathcal{T}}=\mathcal{F}_K\mathcal{F}_0=\Big\{f_{\mathcal{T}}(\epsilon,\boldsymbol{X})=f_K(\epsilon,\boldsymbol{X})f_0(\epsilon,\boldsymbol{X})|\underset{\sim}{\boldsymbol{\Delta}}\in\mathcal{R}(\xi_1,\xi_2,\xi_3),\mathrm{supp}(\boldsymbol{\Delta})\subset\mathcal{T}\Big\},
\end{aligned}
$$

where $\mathrm{supp}(\boldsymbol{\Delta})$ denotes the support set for $\boldsymbol{\Delta}$. Note that $\mathcal{F}_k$ is a VC-class with VC index smaller than $\xi_1 s+2$, and $|f_k(\epsilon,\boldsymbol{X})|\le 1$, $|f_K(\epsilon,\boldsymbol{X})|\le K$, and $|f_0(\epsilon, \boldsymbol{X})| \leq M/K$. Let $N(\epsilon, \mathcal{F}, L_2(Q))$ be the covering number for some functional class $\mathcal{F}$ under the $L_2(Q)$ distance. Then, by Lemma 24 (ii) in Belloni et al. (2016) and the definition of $\mathcal{F}_{\mathcal{T}}$, we have:

$$
\begin{aligned}
&N(K\epsilon, \mathcal{F}_K, L_2(Q)) \leq [N(\frac{\epsilon}{K}, \mathcal{F}_k, L_2(Q))]^K, \\
&N(\epsilon K\frac{M}{K}, \mathcal{F}_{\mathcal{T}}, L_2(Q)) = N(\epsilon M, \mathcal{F}_{\mathcal{T}}, L_2(Q)) \leq N(\frac{\epsilon K}{2}, \mathcal{F}_K, L_2(Q)), \\
&N(\epsilon M, \mathcal{F}, L_2(Q)) \leq C_p^{\xi_1 s} N(\epsilon M, \mathcal{F}_{\mathcal{T}}, L_2(Q)).
\end{aligned} \tag{S10.202}
$$

Hence, by (S10.202), we have:

$$
N(\epsilon M, \mathcal{F}, L_2(Q)) \leq C_p^{\xi_1 s}[N(\frac{\epsilon}{2K}, \mathcal{F}_k, L_2(Q))]^K. \tag{S10.203}
$$

Furthermore, by Lemma 2.6.7 in Van and Wellner (1996), we have

$$
N(\frac{\epsilon}{2K}, \mathcal{F}_k, L_2(Q))] \leq C(\xi_1 s + 2)(16e)^{\xi_1 s+2}(\frac{2K}{\epsilon})^{2(\xi_1 s+1)}, \tag{S10.204}
$$

where C is some universal constant. Combining (S10.203) and (S10.204), for any probability measure Q, we have:

$$
\begin{aligned}
N(\epsilon M, \mathcal{F}, L_2(Q)) &\leq C_p^{\xi_1 s}[N(\frac{\epsilon}{2K}, \mathcal{F}_k, L_2(Q))]^K \\
&\leq C\Big(\frac{pe}{\xi_1 s}\Big)^{\xi_1 s}(\xi_1 s+2)^K(16e)^{K(\xi_1 s+2)}(\frac{2K}{\epsilon})^{2K(\xi_1 s+1)} \\
&\leq C\Big(\frac{pe}{\xi_1 s}\Big)^{\xi_1 s}\Big(\frac{32eK}{\epsilon}\Big)^{c\xi_1 s},
\end{aligned}
$$

where c and C are some big enough positive constants.

Step 4: Obtain the upper bound of $\mathbb{E}\Big[\sup_{\boldsymbol{\Delta}\in\mathcal{R}(\xi_1,\xi_2,\xi_3)} |G_{n',j}(\underset{\sim}{\boldsymbol{\Delta}})|\big|\mathcal{X}\Big]$.

Recall $\sigma_{n'}$ defined in (S10.201). By Lemma A.1 in Zhao et al. (2014), and using some basic calculations, we have:

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{\boldsymbol{\Delta}\in\mathcal{R}(\xi_1,\xi_2,\xi_3)}|G_{n',j}(\underset{\sim}{\boldsymbol{\Delta}})|\big|\mathcal{X}\Big] \\
&\le \frac{C}{\sqrt{n'}}\mathbb{E}\Big[\int_0^{2\sigma_{n'}}\sqrt{\log N(\epsilon,\mathcal{F},L_2(\mathbb{P}_n|\mathcal{X}))}d\epsilon\Big] \\
&\le \frac{C}{\sqrt{n'}}\mathbb{E}\Big[\int_0^{2\sigma_{n'}}\sqrt{\sup_Q\log N(\epsilon,\mathcal{F},L_2(Q))}d\epsilon\Big] \\
&\le \frac{C}{\sqrt{n'}}\int_0^{2\sigma_{n'}}\sqrt{s\log(\frac{p}{\epsilon})}d\epsilon \\
&\le C\sigma_{n'}\sqrt{\frac{s\log(p\vee n')}{n'}} := r_{n'}.
\end{aligned}$$

Step 5: Obtain the tail bound of $\sup_{\boldsymbol{\Delta}\in\mathcal{R}(\xi_1,\xi_2,\xi_3)}|G_{n',j}(\underset{\sim}{\boldsymbol{\Delta}})|\mathcal{X}$. In fact, by the Bousquet inequality (Corollary 14.2 in Bühlmann and Van de Geer (2011)), we have:

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\boldsymbol{\Delta}\in\mathcal{R}(\xi_1,\xi_2,\xi_3)}|G_{n',j}(\underset{\sim}{\boldsymbol{\Delta}})| \ge r_{n'} + t\sqrt{2(\sigma_{n'}^2+2Mr_{n'})} + \frac{2t^2M}{3}|\mathcal{X}\Big) \\
&\le_{(1)} \exp(-n't^2) \le_{(2)} \exp(-q_0nt^2),
\end{aligned}$$

where (2) comes from $n' = \lfloor nt \rfloor$ with $t \in [q_0, 1-q_0]$. It is straightforward to see that if we take $t = C^*\sqrt{\log(pn)/n}$ for some big enough constant $C^* > 0$, we have:

$$\mathbb{P}\Big(\sup_{\boldsymbol{\Delta}\in\mathcal{R}(\xi_1,\xi_2,\xi_3)}|G_{n',j}(\underset{\sim}{\boldsymbol{\Delta}})| \ge C_1(\frac{s\log(pn)}{n})^{\frac{3}{4}}|\mathcal{X}\Big) \le (pn)^{-C_2}.$$

The above result yieds that:

$$\mathbb{P}\Big(\max_{n',j}\sup_{\boldsymbol{\Delta}\in\mathcal{R}(\xi_1,\xi_2,\xi_3)}|G_{n',j}(\underset{\sim}{\boldsymbol{\Delta}})| \ge C_1(\frac{s\log(pn)}{n})^{\frac{3}{4}}|\mathcal{X}\Big) \le (pn)^{-C_3},$$

which proves that $I = O_p((\frac{s\log(pn)}{n})^{\frac{3}{4}})$, where I is defined in (S10.199). With a similar proof technique, we can also prove $II = O_p((\frac{s\log(pn)}{n})^{\frac{3}{4}})$.

Combining with (S10.199), we have proved that:

$$\max_{t\in[q_0,1-q_0]} \|\boldsymbol{C}_0^{II,2}(t)\|_{(s_0,2)}|\mathcal{X} = O_p(s_0^{1/2}(s\log(pn))^{3/4}/n^{1/4}),$$

which finishes the proof of Lemma 17. □

S10.2 Proof of Lemma 18

Proof. Recall $\underset{\sim}{\boldsymbol{\beta}} = ((\boldsymbol{\beta})^\top, (\boldsymbol{b})^\top)^\top \in \mathbb{R}^{p+K}$ and $L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}) := \frac{1}{n}\sum_{i=1}^n \frac{1}{K}\sum_{k=1}^K \rho_{\tau_k}(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta} - b_k)$. Define

$$\nabla L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}) = \frac{\partial L_{n,K}(\underset{\sim}{\boldsymbol{\beta}})}{\partial \underset{\sim}{\boldsymbol{\beta}}} \in \mathbb{R}^{p+K},$$
$$\nabla_1 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}) = \frac{\partial L_{n,K}(\underset{\sim}{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \in \mathbb{R}^{p},$$
$$\nabla_2 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}) = \frac{\partial L_{n,K}(\underset{\sim}{\boldsymbol{\beta}})}{\partial \boldsymbol{b}} \in \mathbb{R}^{K}.$$

Hence, if we define $a_{i,k}^* = \mathbf{1}\{Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^* - b_k^* \le 0\} - \tau_k$ for $i = 1, \dots, n$ and $k = 1, \dots, K$, we have:

$$\nabla_1 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*) = \frac{1}{n}\sum_{i=1}^n \frac{1}{K}\sum_{k=1}^K \boldsymbol{X}_i a_{i,k}^*, \quad \text{and} \quad \nabla_2 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{a}_i^*,$$

where $\boldsymbol{a}_i^* = (a_{i,1}, \dots, a_{i,K})^\top \in \mathbb{R}^K$. The proof of Lemma 18 proceeds into two steps.

Step 1: Obtain the upper bounds of $\|\nabla_1 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty$ and $\|\nabla_2 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty$.

We first consider $\|\nabla_2 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty$. In fact, we have:

$$
\begin{aligned}
&\|\nabla_2 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \\
&=_{(1)} \max_{1\le k\le K} \Big|\frac{1}{n}\sum_{i=1}^n a^*_{i,k}\Big| \\
&=_{(2)} \max_{1\le k\le K} \Big|\frac{1}{n}\sum_{i=1}^n (a^*_{i,k} - \mathbb{E}[a^*_{i,k}])\Big| \\
&\le_{(3)} \max_{1\le k\le K} t_1 \underbrace{\Big|\frac{1}{nt_1}\sum_{i=1}^{nt_1}\big(\mathbf{1}\{\epsilon_i \le \boldsymbol{X}_i^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}) + b^*_k\} - \mathbb{E}(F_\epsilon(\boldsymbol{X}_i^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}) + b^*_k)\big)\Big|}_{I} \\
&+ \max_{1\le k\le K} t_2 \underbrace{\Big|\frac{1}{nt_2}\sum_{i=nt_1+1}^{n}\big(\mathbf{1}\{\epsilon_i \le \boldsymbol{X}_i^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(2)}) + b^*_k\} - \mathbb{E}(F_\epsilon(\boldsymbol{X}_i^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(2)}) + b^*_k)\big)\Big|}_{II},
\end{aligned}
$$

where (2) comes from the first order condition in (S9.187). Hence, to control $\|\nabla_2 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty$, we need to consider I and II. Let $Z_{i,k} := \mathbf{1}\{\epsilon_i \le \boldsymbol{X}_i^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}) + b^*_k\} - \mathbb{E}(F_\epsilon(\boldsymbol{X}_i^\top(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{(1)}) + b^*_k)$. Note that $\mathbb{E}[Z_{i,k}] = 0$ and $-1 \le Z_{i,k} \le 1$. Hence, by the Hoeffding's inequality, we can prove that $(I \vee II) \le C_1\sqrt{\log(p)/n}$ w.p.a.1. Hence, we prove $\|\nabla_2 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \le C_1\sqrt{\log(p)/n}$ w.p.a.1. for some $C_1 > 0$. Next, we consider $\|\nabla_1 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty$. In fact, we have:

$$
\begin{aligned}
&\|\nabla_1 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \\
&=_{(1)} \max_{1\le j\le p} \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{X}_i a^*_{i,k}\Big| \\
&=_{(2)} \max_{1\le j\le p} \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}(X_{ij}a^*_{i,k}-\mathbb{E}[X_{ij}a^*_{i,k}])\Big| \\
&\le_{(3)} \max_{1\le j\le p}\max_{1\le k\le K} \Big|\frac{1}{n}\sum_{i=1}^{n}(X_{ij}a^*_{i,k}-\mathbb{E}[X_{ij}a^*_{i,k}])\Big| \\
&\le_{(4)} \underbrace{\max_{1\le j\le p}\max_{1\le k\le K} t_1\Big|\frac{1}{nt_1}\sum_{i=1}^{nt_1}(X_{ij}a^*_{i,k}-\mathbb{E}[X_{ij}a^*_{i,k}])\Big|}_{III} \\
&\quad + \underbrace{\max_{1\le j\le p}\max_{1\le k\le K} t_2\Big|\frac{1}{nt_2}\sum_{i=nt_1+1}^{n}(X_{ij}a^*_{i,k}-\mathbb{E}[X_{ij}a^*_{i,k}])\Big|}_{IV},
\end{aligned}
$$

where (2) comes from the first order condition in (S9.187). Let $W_{ijk} = X_{ij}a^*_{i,k} - \mathbb{E}[X_{ij}a^*_{i,k}]$. Conditional on $\mathbf{X}$, for fixed j,k, we have $-M \le -|X_{ij}| \le W_{ijk} \le |X_{ij}| \le M$ and $\mathbb{E}[W_{ijk}] = 0$. Hence, by the Hoeffding's inequality, we can see that $(III \vee IV) \le C_2 M\sqrt{\log(p)/n}$ w.p.a.1, which yields $\|\nabla_1 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \le C_2 M\sqrt{\log(p)/n}$ w.p.a.1.

Step 2: Let $\lambda \ge 2M(C_1 \vee C_2)\sqrt{\log(p)/n}$, where C_1 and C_2 are defined in Step 1. Hence, we have $\|\nabla_1 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \le \lambda/2$ and $\|\nabla_2 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \le \lambda/2$ w.p.a.1. By the convexity of $L_{n,K}(\underset{\sim}{\boldsymbol{\beta}})$, we have:

$$
L_{n,K}(\underset{\sim}{\widehat{\boldsymbol{\beta}}}) - L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*) \ge \nabla L_{n,K}^\top(\boldsymbol{\beta}^*)(\underset{\sim}{\widehat{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}^*) = \nabla_1 L_{n,K}^\top(\underset{\approx}{\boldsymbol{\beta}}^*)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + \nabla_2 L_{n,K}^\top(\underset{\approx}{\boldsymbol{\beta}}^*)(\widehat{\boldsymbol{b}} - \boldsymbol{b}^*).
$$

Combining the above inequality and by the optimality of $\widehat{\underset{\sim}{\boldsymbol{\beta}}}$, we have:

$$\begin{aligned} 0 &\le L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*) - L_{n,K}(\widehat{\underset{\sim}{\boldsymbol{\beta}}}) + \lambda(\|\boldsymbol{\beta}^*\|_1 - \|\widehat{\boldsymbol{\beta}}\|_1) \\ &\le \|\nabla_1 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 + \|\nabla_2 L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \|\widehat{\boldsymbol{b}} - \boldsymbol{b}^*\|_1 + \lambda(\|\boldsymbol{\beta}^*\|_1 - \|\widehat{\boldsymbol{\beta}}\|_1) \\ &\le \frac{\lambda}{2}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 + \frac{\lambda}{2}\|\widehat{\boldsymbol{b}} - \boldsymbol{b}^*\|_1 + \lambda(\|\boldsymbol{\beta}^*\|_1 - \|\widehat{\boldsymbol{\beta}}\|_1). \end{aligned}$$

Adding $\frac{\lambda}{2}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1$ on both sides of the above inequality, and using the same proof as in Section S9.3, we can derive that:

$$\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{J^c(\boldsymbol{\beta}^*)}\|_1 \le 3\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)}\|_1 + \|\widehat{\boldsymbol{b}} - \boldsymbol{b}^*\|,$$

which finishes the proof.

□

S10.3 Proof of Lemma 19

Proof. By the well-known Knight's equation that: $\rho_\tau(x-y) - \rho_\tau(x) = -y(\tau - \mathbf{1}\{x \le 0\}) + \int_0^y \mathbf{1}\{x \le s\} - \mathbf{1}\{x \le 0\} ds$, and the definition of $H(\underset{\sim}{\boldsymbol{\Delta}})$, we have:

$$\begin{aligned} H(\underset{\sim}{\boldsymbol{\Delta}}) &= L_K(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\sim}{\boldsymbol{\Delta}}) - L_K(\underset{\sim}{\boldsymbol{\beta}}^*) \\ &= \frac{1}{n}\sum_{i=1}^n \frac{1}{K}\sum_{k=1}^K \mathbb{E}\big[\rho_{\tau_k}(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - b_k^* - (\boldsymbol{X}_i^\top \boldsymbol{\Delta} + \delta_k)) - \rho_{\tau_k}(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - b_k^*)\big] \\ &= I + II, \end{aligned}$$

where

$$
\begin{aligned}
I &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[(\boldsymbol{X}_i^\top\boldsymbol{\Delta}+b_k)(\mathbf{1}\{Y_i-\boldsymbol{X}_i^\top\boldsymbol{\beta}^*-b_k^*\le 0\}-\tau_k)]\\
&= \boldsymbol{\Delta}^\top\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\boldsymbol{X}_i(\mathbf{1}\{Y_i-\boldsymbol{X}_i^\top\boldsymbol{\beta}^*-b_k^*\le 0\}-\tau_k)]\\
&\qquad +\frac{1}{K}\sum_{k=1}^{K}b_k\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(\mathbf{1}\{Y_i-\boldsymbol{X}_i^\top\boldsymbol{\beta}^*-b_k^*\le 0\}-\tau_k)]
\end{aligned}
$$

and

$$
\begin{aligned}
II &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\int_0^{(\boldsymbol{X}_i^\top\boldsymbol{\Delta}+\delta_k)}(\mathbf{1}\{Y_i-\boldsymbol{X}_i^\top\boldsymbol{\beta}^*-b_k^*\le s\}-\mathbf{1}\{Y_i-\boldsymbol{X}_i^\top\boldsymbol{\beta}^*-b_k^*\le 0\})\\
&= \underbrace{t_1\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\int_0^{(\boldsymbol{X}^\top\boldsymbol{\Delta}+\delta_k)}(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(1)})+b_k^*+s)-F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(1)})+b_k^*))}_{III}\\
&\quad \underbrace{+t_2\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\int_0^{(\boldsymbol{X}^\top\boldsymbol{\Delta}+\delta_k)}(F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(2)})+b_k^*+s)-F_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(2)})+b_k^*))}_{IV}.
\end{aligned}
$$

Note that by the first order condition of $\underset{\sim}{\boldsymbol{\beta}}^*$ in (S9.187), we have $I=0$. Recall $\boldsymbol{S}_k := \mathrm{diag}(\mathbf{1}_p,\boldsymbol{e}_k)$, $\underset{\approx}{\boldsymbol{X}} := (\boldsymbol{X}^\top,\mathbf{1}_K)\in\mathbb{R}^{p+K}$, and $\boldsymbol{S} := \sum_{k=1}^{K}\mathbb{E}[(\boldsymbol{S}_k\underset{\approx}{\boldsymbol{X}})(\boldsymbol{S}_k\underset{\approx}{\boldsymbol{X}})^\top]$ defined in (S9.191). For III, by the Taylor's expansion, we have:

$$
\begin{aligned}
III &=_{(1)} t_1\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\Big[\int_0^{(\boldsymbol{S}_k\boldsymbol{X})^\top\boldsymbol{\Delta}} f_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(1)})+b_k^*)s+\frac{s^2}{2}f_\epsilon'(W)ds\Big]\\
&\ge_{(2)} t_1\frac{1}{K}\frac{C_-}{2}\sum_{k=1}^{K}\mathbb{E}[|(\boldsymbol{S}_k\underset{\approx}{\boldsymbol{X}})^\top\underset{\approx}{\boldsymbol{\Delta}}|^2]-t_1\frac{1}{K}\frac{C_+'}{6}\sum_{k=1}^{K}\mathbb{E}[|(\boldsymbol{S}_k\underset{\approx}{\boldsymbol{X}})^\top\underset{\approx}{\boldsymbol{\Delta}}|^3]\\
&\ge_{(3)} t_1\frac{1}{K}\frac{C_-}{2}\|\underset{\approx}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^2-t_1\frac{1}{K}\frac{C_+'m_0}{6}\|\underset{\approx}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^3,
\end{aligned}
$$

where W in (1) is some random variable between $\boldsymbol{X}^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(1)})+b_k^*+s$ and $\boldsymbol{X}^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(1)})+b_k^*$, (2) follows from the assumption that $\inf_{1\le k\le K} f_\epsilon(\boldsymbol{X}^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta}^{(1)})+b_k^*)\ge C_-$ and $|f_\epsilon'(t)|\le C_+'$, (3) follows from the assumption that $\sum_{k=1}^{K}\mathbb{E}[|(\boldsymbol{S}_k\underset{\approx}{\boldsymbol{X}})^\top\underset{\approx}{\boldsymbol{\Delta}}|^3]\le m_0\|\underset{\approx}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^3$ for some $m_0>0$. Similarly, for IV, we have

$IV \geq t_2 \frac{1}{K} \frac{C_-}{2} \|\underset{\sim\sim}{\mathbf{\Delta}}\|_{\mathbf{S}}^2 - t_2 \frac{1}{K} \frac{C'_+ m_0}{6} \|\underset{\sim\sim}{\mathbf{\Delta}}\|_{\mathbf{S}}^3$, which implies the final result:

$$\begin{aligned} H(\underset{\sim\sim}{\mathbf{\Delta}}) &=_{(1)} I + II \\ &=_{(2)} III + IV \\ &\geq_{(3)} \frac{1}{K} \frac{C_-}{2} \|\underset{\sim\sim}{\mathbf{\Delta}}\|_{\mathbf{S}}^2 - \frac{1}{K} \frac{C'_+ m_0}{6} \|\underset{\sim\sim}{\mathbf{\Delta}}\|_{\mathbf{S}}^3 \\ &\geq_{(4)} c_* \min\Big(\frac{\|\underset{\sim\sim}{\mathbf{\Delta}}\|_{\mathbf{S}}^2}{4}, \frac{\|\underset{\sim\sim}{\mathbf{\Delta}}\|_{\mathbf{S}}}{4}\Big), \end{aligned}$$

where (4) is very similar to the proof of Lemma C.1 in Zhao et al. (2014), which is omitted. □

S10.4 Proof of Lemma 20

Proof. Let $r_{i,k} = Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - b_k^*$ and

$$\begin{aligned} U_i(\mathbf{\Delta}, \boldsymbol{\delta}) &= \frac{1}{K} \sum_{k=1}^K \big[\rho_{\tau_k}(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - b_k^* - (\boldsymbol{X}_i^\top \mathbf{\Delta} + \delta_k)) - \rho_{\tau_k}(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - b_k^*)\big] \\ &= \frac{1}{K} \sum_{k=1}^K \big[\rho_{\tau_k}(r_{i,k} - (\boldsymbol{X}_i^\top \mathbf{\Delta} + \delta_k)) - \rho_{\tau_k}(r_{i,k})\big]. \end{aligned} \tag{S10.205}$$

Hence, using the above notations, we have:

$$\big(L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\sim\sim}{\mathbf{\Delta}}) - L_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)) - (L_K(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\sim\sim}{\mathbf{\Delta}}) - L_K(\underset{\sim}{\boldsymbol{\beta}}^*)\big) = \frac{1}{n} \sum_{i=1}^n \Big[U_i(\mathbf{\Delta}, \boldsymbol{\delta}) - \mathbb{E}[U_i(\mathbf{\Delta}, \boldsymbol{\delta})]\Big].$$

By the lipschitz continuity of $|\rho_\tau(t) - \rho_\tau(s)| \leq |s - t|$, we have

$$U_i(\mathbf{\Delta}, \boldsymbol{\delta}) - \mathbb{E}[U_i(\mathbf{\Delta}, \boldsymbol{\delta})] \leq \frac{2}{K} \sum_{k=1}^K |\boldsymbol{X}_i^\top \mathbf{\Delta} + \delta_k| := C_i(\mathbf{\Delta}, \boldsymbol{\delta}). \tag{S10.206}$$

Let $Z = \sup_{\mathbf{\Delta} \in \mathcal{A}, \|\mathbf{\Delta}\|_{\mathbf{S}} \leq \xi} \Big|\frac{1}{n} \sum_{i=1}^n \Big[U_i(\mathbf{\Delta}, \boldsymbol{\delta}) - \mathbb{E}[U_i(\mathbf{\Delta}, \boldsymbol{\delta})]\Big|$. In what follows, we will use the Massart's inequality (Theorem 14.2 in Bühlmann and Van de

Geer (2011)) to obtain the tail bound:

$$\mathbb{P}\Big(Z > \mathbb{E}Z + t\Big) \le \exp\Big(-\frac{nt^2}{8\sigma^2}\Big), \tag{S10.207}$$

where $\sup_{\boldsymbol{\Delta}\in\mathcal{A},\|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\le\xi} \frac{1}{n}\sum_{i=1}^{n} C_i^2(\boldsymbol{\Delta},\boldsymbol{\delta}) \le \sigma^2$. Hence, to use Massart's inequality, we need two steps.

Step 1: Obtain the upper bound for σ^2. Recall $\boldsymbol{S}_k := \mathrm{diag}(\mathbf{1}_p, \boldsymbol{e}_k)$, $\underset{\sim}{\boldsymbol{X}}_i := (\boldsymbol{X}_i^\top, \mathbf{1}_K) \in \mathbb{R}^{p+K}$, and $\boldsymbol{S} := \sum_{k=1}^{K} \mathbb{E}[(\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}})(\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}})^\top]$. With probability tending to 1, we have:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} C_i^2(\boldsymbol{\Delta},\boldsymbol{\delta}) &=_{(1)} \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{2}{K}\sum_{k=1}^{K}|\boldsymbol{X}_i^\top\boldsymbol{\Delta}+\delta_k|\Big)^2 \\
&\le_{(2)} \frac{4}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}(\boldsymbol{X}_i^\top\boldsymbol{\Delta}+\delta_k)^2 \\
&\le_{(3)} \frac{8}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}((\boldsymbol{X}_i^\top\boldsymbol{\Delta})^2+\delta_k^2) \\
&=_{(4)} \frac{8}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}((\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)^\top\underset{\sim}{\boldsymbol{\Delta}})^2 \\
&=_{(5)} \frac{8}{K}\underset{\sim}{\boldsymbol{\Delta}}^\top\Big[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}((\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)(\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)^\top\Big]\underset{\sim}{\boldsymbol{\Delta}} \\
&=_{(6)} \frac{8}{K}\underset{\sim}{\boldsymbol{\Delta}}^\top\boldsymbol{S}\underset{\sim}{\boldsymbol{\Delta}} + \frac{8}{K}\underset{\sim}{\boldsymbol{\Delta}}^\top\Big[\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}((\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)(\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)^\top-\boldsymbol{S}\Big]\underset{\sim}{\boldsymbol{\Delta}} \\
&\le_{(7)} \frac{8}{K}\|\underset{\sim}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^2 + \frac{8}{K}\|\underset{\sim}{\boldsymbol{\Delta}}\|_1^2\Big\|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}((\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)(\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)^\top-\boldsymbol{S}\Big\|_\infty \\
&\le_{(8)} \frac{8}{K}\|\underset{\sim}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^2 + \frac{8}{K}M\sqrt{\log(p)/n}\|\underset{\sim}{\boldsymbol{\Delta}}\|_1^2 \\
&\le_{(9)} \frac{8}{K}\|\underset{\sim}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^2 + O(Ms\sqrt{\log(p)/n})\|\underset{\sim}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^2 \\
&\le_{(10)} \frac{9}{K}\xi^2,
\end{aligned} \tag{S10.208}$$

where (2) follows from the Cauchy-Swarchz inequality, (3) follows from $(a+b)^2 \leq 2a^2+2b^2$, (8) follows from the large deviation for $\Big\|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}((\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)(\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)^\top - \boldsymbol{S}\Big\|_\infty$, (9) follows from the fact that $\|\underset{\sim}{\boldsymbol{\Delta}}\|_1 \leq 4\|\boldsymbol{\Delta}_{J(\boldsymbol{\beta}^*)}\|_1 + \|\boldsymbol{\delta}\|_1$ and the Cauchy-Swarchz inequality, and (10) comes from the assumption that $Ms\sqrt{\log(p)/n} = o(1)$.

Step 2: Obtain the upper bound for $\mathbb{E}[Z]$. Let $e_1, \ldots, e_n$ be i.i.d Rademacher random variables with $\mathbb{P}(e_i = 1) = \mathbb{P}(e_i = -1) = 1/2$. In fact, by the symmetrization procedure (Theorem 14.3 in Bühlmann and Van de Geer (2011)) and the contraction principle (Theorem 14.4 in Bühlmann and Van de Geer (2011)), we have:

$$
\begin{aligned}
\mathbb{E}[Z] \quad &=_{(1)} \mathbb{E}\Big[\sup_{\boldsymbol{\Delta}\in\mathcal{A},\|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\leq\xi}\Big|\frac{1}{n}\sum_{i=1}^{n}(U_i(\boldsymbol{\Delta},\boldsymbol{\delta}) - \mathbb{E}U_i(\boldsymbol{\Delta},\boldsymbol{\delta}))\Big|\Big] \\
&\leq_{(2)} 2\mathbb{E}\Big[\sup_{\boldsymbol{\Delta}\in\mathcal{A},\|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\leq\xi}\Big|\frac{1}{n}\sum_{i=1}^{n}(e_iU_i(\boldsymbol{\Delta},\boldsymbol{\delta}))\Big|\Big] \\
&=_{(3)} 2\mathbb{E}\Big[\sup_{\boldsymbol{\Delta}\in\mathcal{A},\|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\leq\xi}\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}e_i[\rho_{\tau_k}(r_{i,k} - (\boldsymbol{X}_i^\top\boldsymbol{\Delta} + \delta_k)) - \rho_{\tau_k}(r_{i,k})]\Big|\Big] \\
&\leq_{(4)} 2\max_{1\leq k\leq K}\mathbb{E}\Big[\sup_{\boldsymbol{\Delta}\in\mathcal{A},\|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\leq\xi}\Big|\frac{1}{n}\sum_{i=1}^{n}e_i[\rho_{\tau_k}(r_{i,k} - (\boldsymbol{X}_i^\top\boldsymbol{\Delta} + \delta_k)) - \rho_{\tau_k}(r_{i,k})]\Big|\Big] \\
&\leq_{(5)} 4\max_{1\leq k\leq K}\mathbb{E}\Big[\sup_{\boldsymbol{\Delta}\in\mathcal{A},\|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\leq\xi}\Big|\frac{1}{n}\sum_{i=1}^{n}e_i[\boldsymbol{X}_i^\top\boldsymbol{\Delta} + \delta_k]\Big|\Big] \\
&=_{(6)} 4\max_{1\leq k\leq K}\mathbb{E}\Big[\sup_{\boldsymbol{\Delta}\in\mathcal{A},\|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\leq\xi}\Big|\frac{1}{n}\sum_{i=1}^{n}e_i(\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)^\top\underset{\sim}{\boldsymbol{\Delta}}\Big|\Big] \\
&\leq_{(7)} 4\underbrace{\Big[\sup_{\boldsymbol{\Delta}\in\mathcal{A},\|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\leq\xi}\|\underset{\sim}{\boldsymbol{\Delta}}\|_1\Big]}_{I}\times\underbrace{\max_{1\leq k\leq K}\mathbb{E}\Big\|\frac{1}{n}\sum_{i=1}^{n}e_i(\boldsymbol{S}_k\underset{\sim}{\boldsymbol{X}}_i)\Big\|_\infty}_{II}.
\end{aligned}
\tag{S10.209}
$$

Hence, to control $\mathbb{E}[Z]$, it is sufficient to consider I and II, respectively. For I, using the fact that $\|\underset{\approx}{\boldsymbol{\Delta}}\|_1 \le 4\|\boldsymbol{\Delta}_{J(\boldsymbol{\beta}^*)}\|_1 + \|\boldsymbol{\delta}\|_1$ and the Cauchy-Swarchz inequality, we have $I \le C_1\sqrt{s}\xi$ for some $C_1 > 0$. For II, we can prove that $II \le C_2 M\sqrt{\log(p)/n}$ for some $C_2 > 0$. Hence, combining (S10.207), (S10.208), and (S10.209), if we take $t = C_3\xi\sqrt{\log(p)/n}$ for some large enough $C_3 > 0$, with w.p.a.1, we have $Z \le C_3 M\xi\sqrt{s\log(p)/n}$, which finishes the proof. □

S10.5 Proof of Lemma 21

Proof. Recall $\underset{\sim}{\boldsymbol{\beta}} = ((\boldsymbol{\beta})^\top, (\boldsymbol{b})^\top)^\top \in \mathbb{R}^{p+K}$ and $L^\alpha_{n,K}(\underset{\sim}{\boldsymbol{\beta}})$ defined in (S9.198). Recall $a^*_{i,k} := \mathbf{1}\{Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^* - b^*_k \le 0\} - \tau_k$ for $i = 1,\dots,n$ and $k = 1,\dots,K$. Define

$$\nabla L^\alpha_{n,K}(\underset{\sim}{\boldsymbol{\beta}}) = \frac{\partial L^\alpha_{n,K}(\underset{\sim}{\boldsymbol{\beta}})}{\partial \underset{\sim}{\boldsymbol{\beta}}} \in \mathbb{R}^{p+K},$$

$$\nabla_1 L^\alpha_{n,K}(\underset{\sim}{\boldsymbol{\beta}}) = \frac{\partial L^\alpha_{n,K}(\underset{\sim}{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \in \mathbb{R}^{p},$$

$$\nabla_2 L^\alpha_{n,K}(\underset{\sim}{\boldsymbol{\beta}}) = \frac{\partial L^\alpha_{n,K}(\underset{\sim}{\boldsymbol{\beta}})}{\partial \boldsymbol{b}} \in \mathbb{R}^{K}.$$

Then, we have:

$$\nabla_1 L^\alpha_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*) = (1-\alpha)\frac{1}{n}\sum_{i=1}^n \frac{1}{K}\sum_{k=1}^K \boldsymbol{X}_i a^*_{i,k} - \alpha\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*),$$

$$\text{and} \quad \nabla_2 L^\alpha_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{a}^*_i,$$

where $\boldsymbol{a}^*_i = (a_{i,1},\dots,a_{i,K})^\top \in \mathbb{R}^K$. The proof of Lemma 18 proceeds into two steps.

Step 1: Obtain the upper bounds of $\|\nabla_1 L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty$ and $\|\nabla_2 L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty$. We first consider $\|\nabla_2 L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty$. In fact, by Step 1 in Section S10.2, we have $\|\nabla_2 L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \le C_1\sqrt{\log(p)/n}$ w.p.a.1. for some $C_1 > 0$. Next, we consider $\|\nabla_1 L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty$. In fact, we have:

$$
\begin{aligned}
&\|\nabla_1 L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \\
&=_{(1)} \|(1-\alpha)\frac{1}{n}\sum_{i=1}^n \frac{1}{K}\sum_{k=1}^K \boldsymbol{X}_i a^*_{i,k} - \alpha\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*)\|_\infty \\
&=_{(2)} \|(1-\alpha)\frac{1}{n}\sum_{i=1}^n \frac{1}{K}\sum_{k=1}^K \big(\boldsymbol{X}_i a^*_{i,k} - \mathbb{E}(\boldsymbol{X}_i a^*_{i,k})\big) \\
&\qquad -\alpha\frac{1}{n}\sum_{i=1}^n \big(\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*) - \mathbb{E}(\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*))\big)\|_\infty \\
&\le_{(3)} (1-\alpha)\underbrace{\|\frac{1}{n}\sum_{i=1}^n \frac{1}{K}\sum_{k=1}^K \big(\boldsymbol{X}_i a^*_{i,k} - \mathbb{E}(\boldsymbol{X}_i a^*_{i,k})\big)\|_\infty}_{I} \\
&\qquad +\alpha\underbrace{\|\frac{1}{n}\sum_{i=1}^n \big(\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*) - \mathbb{E}(\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*))\big)\|_\infty}_{II},
\end{aligned}
$$

where (1) comes from the first order condition in (S9.195). By Step 2 in Section S10.2, we have $I \le C_2 M\sqrt{\log(p)/n}$ w.p.a.1. Next, we consider II. In fact, by noting that $Y_i = \epsilon_i + \boldsymbol{\beta}^{(1)}\mathbf{1}\{i \le \lfloor nt_1\rfloor\} + \boldsymbol{\beta}^{(2)}\mathbf{1}\{i > \lfloor nt_1\rfloor\}$, we have:

$$
\begin{aligned}
II \;&=_{(1)} \Big\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\epsilon_i + t_1(\widehat{\boldsymbol{\Sigma}}(0:t_1) - \boldsymbol{\Sigma})(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^*) + t_2(\widehat{\boldsymbol{\Sigma}}(t_1:1) - \boldsymbol{\Sigma})(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^*)\Big\|_\infty \\
&\le_{(2)} \Big\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\epsilon_i\|_\infty + t_1\|\widehat{\boldsymbol{\Sigma}}(0:t_1) - \boldsymbol{\Sigma}\|_\infty\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^*\|_1 + t_2\|\widehat{\boldsymbol{\Sigma}}(t_1:1) - \boldsymbol{\Sigma}\|_\infty\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^*\|_1 \\
&=_{(3)} O_p\big(M\sqrt{\frac{\log(pn)}{n}}\big) + O_p\big(M^2\sqrt{\frac{\log(pn)}{n}}\big)\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^*\|_1 + O_p\big(M^2\sqrt{\frac{\log(pn)}{n}}\big)\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^*\|_1 \\
&=_{(4)} O_p\big(M\sqrt{\frac{\log(pn)}{n}}\big) + O_p\big(M^2\sqrt{\frac{\log(pn)}{n}}\big)\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1 =_{(5)} O_p\big(M^2\sqrt{\frac{\log(pn)}{n}}\big),
\end{aligned}
$$

where (3) comes from Lemmas 7 and 8, (4) and (5) come from Remark 6 and the assumption that $\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1 \leq C_{\boldsymbol{\Delta}}$. Hence, combining the above bounds, w.p.a.1, we have:

$$\|\nabla_1 L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \leq C_1 M^2 \sqrt{\frac{\log(pn)}{n}}, \|\nabla_2 L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^*)\|_\infty \leq C_2 M^2 \sqrt{\frac{\log(pn)}{n}}.$$

Step 2: Let $\lambda \geq 2M^2(C_1 \vee C_2)\sqrt{\log(p)/n}$, where C_1 and C_2 are defined in Step 1. Using a similar proof procedure as in Step 2 of Section S10.2, we can derive that:

$$\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{J^c(\boldsymbol{\beta}^*)}\|_1 \leq 3\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{J(\boldsymbol{\beta}^*)}\|_1 + \|\widehat{\boldsymbol{b}} - \boldsymbol{b}^*\|,$$

which finishes the proof.

□

S10.6 Proof of Lemma 22

Proof. Recall

$$L^{\alpha}_K(\underset{\sim}{\boldsymbol{\beta}}) := (1-\alpha)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\rho_{\tau_k}(Y_i - b_i - \boldsymbol{X}_i^\top \boldsymbol{\beta})] + \frac{\alpha}{2n}\sum_{i=1}^{n}\mathbb{E}(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta})^2.$$

Note that $L^{\alpha}_K(\underset{\sim}{\boldsymbol{\beta}})$ is a combination of the composite quantile loss and the squared loss. Moreover, the excess risk for the squared loss is lower bounded by a squared form. Hence, combining the results in Section S10.3, we can prove that there exists some $c_* > 0$ such that

$$H^{\alpha}(\underset{\approx}{\boldsymbol{\Delta}}) \geq c_* \min\Big(\frac{\|\underset{\approx}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}^2}{4}, \frac{\|\underset{\approx}{\boldsymbol{\Delta}}\|_{\boldsymbol{S}}}{4}\Big).$$

To save space, we omit the details. □

S10.7 Proof of Lemma 23

Define

$$\begin{aligned} U_i^{(1)}(\boldsymbol{\Delta}, \boldsymbol{\delta}) &= \frac{1}{K}\sum_{k=1}^{K}\left[\rho_{\tau_k}(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - b_k^* - (\boldsymbol{X}_i^\top \boldsymbol{\Delta} + \delta_k)) - \rho_{\tau_k}(Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - b_k^*)\right] \\ U_i^{(2)}(\boldsymbol{\Delta}, \boldsymbol{\delta}) &= (Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - b_k^* - (\boldsymbol{X}_i^\top \boldsymbol{\Delta} + \delta_k))^2 - (Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}^* - b_k^*)^2. \end{aligned} \tag{S10.210}$$

Hence, using the above notations, we have:

$$\begin{aligned} &\left(L_{n,K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\approx}{\boldsymbol{\Delta}}) - L_{n,K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^*)) - (L_{K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^* + \underset{\approx}{\boldsymbol{\Delta}}) - L_{K}^{\alpha}(\underset{\sim}{\boldsymbol{\beta}}^*)\right) \\ &\quad = (1-\alpha)\frac{1}{n}\sum_{i=1}^{n}\left[U_i^{(1)}(\boldsymbol{\Delta}, \boldsymbol{\delta}) - \mathbb{E}[U_i^{(1)}(\boldsymbol{\Delta}, \boldsymbol{\delta})]\right] + \frac{\alpha}{2}\frac{1}{n}\sum_{i=1}^{n}\left[U_i^{(2)}(\boldsymbol{\Delta}, \boldsymbol{\delta}) - \mathbb{E}[U_i^{(2)}(\boldsymbol{\Delta}, \boldsymbol{\delta})]\right]. \end{aligned}$$

To prove Lemma 23, it is sufficient to bound I and II, where:

$$\begin{aligned} I &= \sup_{\boldsymbol{\Delta}\in\mathcal{A}, \|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\le\xi}\left|\frac{1}{n}\sum_{i=1}^{n}U_i^{(1)}(\boldsymbol{\Delta}, \boldsymbol{\delta}) - \mathbb{E}[U_i^{(1)}(\boldsymbol{\Delta}, \boldsymbol{\delta})]\right|, \\ II &= \sup_{\boldsymbol{\Delta}\in\mathcal{A}, \|\boldsymbol{\Delta}\|_{\boldsymbol{S}}\le\xi}\left|\frac{1}{n}\sum_{i=1}^{n}U_i^{(2)}(\boldsymbol{\Delta}, \boldsymbol{\delta}) - \mathbb{E}[U_i^{(2)}(\boldsymbol{\Delta}, \boldsymbol{\delta})]\right|. \end{aligned} \tag{S10.211}$$

Note that in Section S10.4, we have proved that $I = O_p(\xi\sqrt{s\log(p)/n})$. Hence, it only remains to consider II. Let $\underset{\sim}{\boldsymbol{\beta}} = \underset{\sim}{\boldsymbol{\beta}}^* + \underset{\approx}{\boldsymbol{\Delta}}$. Then, it is equivalent to consider :

$$\begin{aligned} II = \sup_{\underset{\sim}{\boldsymbol{\beta}}-\underset{\sim}{\boldsymbol{\beta}}^*\in\mathcal{A}, \|\underset{\sim}{\boldsymbol{\beta}}-\underset{\sim}{\boldsymbol{\beta}}^*\|_{\boldsymbol{S}}\le\xi}&\left|\frac{1}{n}\sum_{i=1}^{n}[(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta})^2 - (Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*)^2]\right. \\ &\left. -\mathbb{E}[(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta})^2 - (Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^*)^2]\right| \end{aligned} \tag{S10.212}$$

$$\le II.1 + II.2 + II.3,$$

where

$$II.1 := \sup_{\substack{\beta-\beta^*\in\mathcal{A}\\ \|\beta-\beta^*\|_{S}\le\xi}} \Big|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i \boldsymbol{X}_i^\top(\boldsymbol{\beta}^*-\boldsymbol{\beta})\Big|,$$

$$II.2 := \sup_{\substack{\beta-\beta^*\in\mathcal{A}\\ \|\beta-\beta^*\|_{S}\le\xi}} t_1\Big|\frac{1}{nt_1}\sum_{i=1}^{nt_1}\big[(\boldsymbol{X}_i^\top\boldsymbol{\beta}-\boldsymbol{X}_i^\top\boldsymbol{\beta}^{(1)})^2-(\boldsymbol{X}_i^\top\boldsymbol{\beta}^*-\boldsymbol{X}_i^\top\boldsymbol{\beta}^{(1)})^2\big]$$
$$-\mathbb{E}\big[(\boldsymbol{X}_i^\top\boldsymbol{\beta}-\boldsymbol{X}_i^\top\boldsymbol{\beta}^{(1)})^2-(\boldsymbol{X}_i^\top\boldsymbol{\beta}^*-\boldsymbol{X}_i^\top\boldsymbol{\beta}^{(1)})^2\big]\Big|,$$

$$II.3 := \sup_{\substack{\beta-\beta^*\in\mathcal{A}\\ \|\beta-\beta^*\|_{S}\le\xi}} t_2\Big|\frac{1}{nt_2}\sum_{i=nt_1+1}^{n}\big[(\boldsymbol{X}_i^\top\boldsymbol{\beta}-\boldsymbol{X}_i^\top\boldsymbol{\beta}^{(2)})^2-(\boldsymbol{X}_i^\top\boldsymbol{\beta}^*-\boldsymbol{X}_i^\top\boldsymbol{\beta}^{(2)})^2\big]$$
$$-\mathbb{E}\big[(\boldsymbol{X}_i^\top\boldsymbol{\beta}-\boldsymbol{X}_i^\top\boldsymbol{\beta}^{(2)})^2-(\boldsymbol{X}_i^\top\boldsymbol{\beta}^*-\boldsymbol{X}_i^\top\boldsymbol{\beta}^{(2)})^2\big]\Big|.$$

Next, we consider $II.1 - II.3$, respectively. For $II.1$, we have

$$\begin{aligned}
II.1 &\le_{(1)} \|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i\boldsymbol{X}_i^\top\|_\infty \sup_{\underset{\sim}{\beta}-\underset{\sim}{\beta}^*\in\mathcal{A},\|\underset{\sim}{\beta}-\underset{\sim}{\beta}^*\|_{S}\le\xi}\|\boldsymbol{\beta}^*-\boldsymbol{\beta}\|_1\\
&\le_{(2)} CM\sqrt{\frac{\log(pn)}{n}} \sup_{\underset{\sim}{\beta}-\underset{\sim}{\beta}^*\in\mathcal{A},\|\underset{\sim}{\beta}-\underset{\sim}{\beta}^*\|_{S}\le\xi}\|\boldsymbol{\beta}^*-\boldsymbol{\beta}\|_1\\
&\le_{(3)} CM\sqrt{\frac{\log(pn)}{n}} \sup_{\underset{\sim}{\beta}-\underset{\sim}{\beta}^*\in\mathcal{A},\|\underset{\sim}{\beta}-\underset{\sim}{\beta}^*\|_{S}\le\xi}(4\|(\boldsymbol{\beta}^*-\boldsymbol{\beta})_{J(\boldsymbol{\beta}^*)}\|_1+\|\boldsymbol{b}-\boldsymbol{b}^*\|_1)\\
&\le_{(4)} CM\sqrt{\frac{\log(pn)}{n}} \sup_{\underset{\sim}{\beta}-\underset{\sim}{\beta}^*\in\mathcal{A},\|\underset{\sim}{\beta}-\underset{\sim}{\beta}^*\|_{S}\le\xi}(4\sqrt{s}\|(\boldsymbol{\beta}^*-\boldsymbol{\beta})_{J(\boldsymbol{\beta}^*)}\|_2+\sqrt{K}\|\boldsymbol{b}-\boldsymbol{b}^*\|_2)\\
&\le_{(5)} CM\xi\sqrt{s\frac{\log(pn)}{n}},
\end{aligned}$$

where (2) comes from Lemma 7, (3) follows from the definition of $\mathcal{A}$, (5) follows from the definiteness of $\boldsymbol{S}$.

Our next goal is to bound $II.2$. To that end, we suppose that there exists some universal constant $\eta > 0$ such that for all $\underset{\sim}{\boldsymbol{\beta}}$ satisfying $\|\underset{\sim}{\boldsymbol{\beta}} -$

$\underset{\sim}{\beta}^*\|_{\boldsymbol{S}} \leq \xi$, we have:

$$|\boldsymbol{X}^\top \boldsymbol{\beta} - \boldsymbol{X}^\top \boldsymbol{\beta}^{(1)}| \leq \eta.$$

Note that this is a very common assumption for proving the concentration inequality for squared error loss (see Bühlmann and Van de Geer (2011)). Define the functional class:

$$\gamma(\boldsymbol{X}^\top \boldsymbol{\beta}) = \frac{(\boldsymbol{X}^\top \boldsymbol{\beta} - \boldsymbol{X}^\top \boldsymbol{\beta}^{(1)})^2 - (\boldsymbol{X}^\top \boldsymbol{\beta}^* - \boldsymbol{X}^\top \boldsymbol{\beta}^{(1)})^2}{2\eta}.$$

By definition, we can see that $|\gamma(\boldsymbol{X}^\top \boldsymbol{\beta}) - \gamma(\boldsymbol{X}^\top \boldsymbol{\beta}')| \leq |\boldsymbol{X}^\top \boldsymbol{\beta} - \boldsymbol{X}^\top \boldsymbol{\beta}'|$, which is 1-Lipschitz continous. Moreover, by defining $\gamma(\boldsymbol{X}^\top \boldsymbol{\beta})$, $II.2$ reduces to:

$$II.2 = 2\eta t_1 \underbrace{\sup_{\substack{\boldsymbol{\beta}-\boldsymbol{\beta}^* \in \mathcal{A} \\ \|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|_{\boldsymbol{S}} \leq \xi}} \left| \frac{1}{nt_1} \sum_{i=1}^{nt_1} [\gamma(\boldsymbol{X}_i^\top \boldsymbol{\beta}) - \gamma(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*)] - \mathbb{E}[\gamma(\boldsymbol{X}_i^\top \boldsymbol{\beta}) - \gamma(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*)] \right.}_{Z}.$$

Note that

$$\left|\gamma(\boldsymbol{X}_i^\top \boldsymbol{\beta}) - \gamma(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*) - \mathbb{E}[\gamma(\boldsymbol{X}_i^\top \boldsymbol{\beta}) - \gamma(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*)]\right| \leq 2|\boldsymbol{X}_i^\top (\boldsymbol{\beta} - \boldsymbol{\beta}^*)| := C_i(\boldsymbol{\beta}).$$

In what follows, we will use the Massart's inequality (Theorem 14.2 in Bühlmann and Van de Geer (2011)) to obtain the tail bound:

$$\mathbb{P}\Big(Z > \mathbb{E}Z + t\Big) \leq \exp\Big(-\frac{nt^2}{8\sigma^2}\Big), \tag{S10.213}$$

where $\sup_{\underset{\sim}{\beta}-\underset{\sim}{\beta}^* \in \mathcal{A}, \|\underset{\sim}{\beta}-\underset{\sim}{\beta}^*\|_{\boldsymbol{S}} \leq \xi} \frac{1}{n} \sum_{i=1}^{n} C_i^2(\boldsymbol{\beta}) \leq \sigma^2$. Hence, to use Massart's inequality, we need two steps.

Step 1: Obtain the upper bound for σ^2. In fact, w.p.a.1, we have:

$$\begin{aligned}
&\sup_{\beta-\beta^*\in\mathcal{A},\|\beta-\beta^*\|_{\mathcal{S}}\le\xi} \frac{1}{n}\sum_{i=1}^{n} C_i^2(\boldsymbol{\beta}) \\
&=_{(1)} 4\sup_{\beta-\beta^*\in\mathcal{A},\|\beta-\beta^*\|_{\mathcal{S}}\le\xi} (\boldsymbol{\beta}-\boldsymbol{\beta}^*)^\top\widehat{\boldsymbol{\Sigma}}(0:1)(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \\
&=_{(2)} 4\sup_{\beta-\beta^*\in\mathcal{A},\|\beta-\beta^*\|_{\mathcal{S}}\le\xi} |(\boldsymbol{\beta}-\boldsymbol{\beta}^*)^\top(\widehat{\boldsymbol{\Sigma}}(0:1)-\boldsymbol{\Sigma})(\boldsymbol{\beta}-\boldsymbol{\beta}^*)| \\
&\qquad +4\sup_{\beta-\beta^*\in\mathcal{A},\|\beta-\beta^*\|_{\mathcal{S}}\le\xi} |(\boldsymbol{\beta}-\boldsymbol{\beta}^*)^\top\boldsymbol{\Sigma}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)| \\
&\le_{(3)} 4\sup_{\beta-\beta^*\in\mathcal{A},\|\beta-\beta^*\|_{\mathcal{S}}\le\xi} \|\widehat{\boldsymbol{\Sigma}}(0:1)-\boldsymbol{\Sigma})\|_\infty\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|_1^2 \\
&\qquad +4\lambda_{\max}\sup_{\beta-\beta^*\in\mathcal{A},\|\beta-\beta^*\|_{\mathcal{S}}\le\xi} \|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|^2 \\
&\le_{(4)} C_1M^2\sqrt{\frac{\log(pn)}{n}}s\xi^2 + C_2\xi^2 \le_{(5)} C_3\xi^2 := \sigma^2.
\end{aligned}$$

Step 2: Obtain the upper bound for $\mathbb{E}[Z]$. Let $e_1,\ldots,e_n$ be i.i.d Rademacher random variables with $\mathbb{P}(e_i=1)=\mathbb{P}(e_i=-1)=1/2$. In fact, by the symmetrization procedure (Theorem 14.3 in Bühlmann and Van de Geer (2011)) and the contraction principle (Theorem 14.4 in Bühlmann and Van de Geer (2011)), we have:

$$\begin{aligned}
\mathbb{E}[Z] &\le 2\mathbb{E}\Big[\sup_{\beta-\beta^*\in\mathcal{A},\|\beta-\beta^*\|_{\mathcal{S}}\le\xi}\Big|\frac{1}{nt_1}\sum_{i=1}^{nt_1} e_i(\gamma(\boldsymbol{X}_i^\top\boldsymbol{\beta})-\gamma(\boldsymbol{X}_i^\top\boldsymbol{\beta}^*))\Big|\Big] \\
&\le 4\mathbb{E}\Big[\sup_{\beta-\beta^*\in\mathcal{A},\|\beta-\beta^*\|_{\mathcal{S}}\le\xi}\Big|\frac{1}{nt_1}\sum_{i=1}^{nt_1} e_i(\boldsymbol{X}_i^\top\boldsymbol{\beta}-\boldsymbol{X}_i^\top\boldsymbol{\beta}^*)\Big|\Big] \\
&\le 4\sup_{\beta-\beta^*\in\mathcal{A},\|\beta-\beta^*\|_{\mathcal{S}}\le\xi}\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\|_1\mathbb{E}\Big\|\frac{1}{nt_1}\sum_{i=1}^{nt_1} e_i\boldsymbol{X}_i\|_\infty \le C\xi\sqrt{s}M\sqrt{\frac{\log(pn)}{n}},
\end{aligned}$$

where the last inequality comes from the Hoeffding's inequality. Hence, combining Steps 1 and 2, taking $t = C\xi\sqrt{\log(p)/n}$ for some big enough constant $C > 0$, we have, w.p.a.1, $Z = O(\xi\sqrt{s\log(p)/n})$, which implies $II.2 = O_p(\xi\sqrt{s\log(p)/n})$.

Similarly, we can prove $II.3 = O_p(\xi\sqrt{s\log(p)/n})$. Considering (S10.211) and (S10.212), we have proved

$$\sup_{\substack{\boldsymbol{\Delta}\in\mathcal{A}\\ \|\boldsymbol{\Delta}\|_{\mathcal{S}}\le\xi}} \left|(L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^{*}+\underset{\approx}{\boldsymbol{\Delta}})-L^{\alpha}_{n,K}(\underset{\sim}{\boldsymbol{\beta}}^{*}))-(L^{\alpha}_{K}(\underset{\sim}{\boldsymbol{\beta}}^{*}+\underset{\approx}{\boldsymbol{\Delta}})-L^{\alpha}_{K}(\underset{\sim}{\boldsymbol{\beta}}^{*}))\right| = O_p\big(M\xi\sqrt{s\frac{\log(pn)}{n}}\big),$$

which finishes the proof.

References

Belloni, A., M. Chen, and V. Chernozhukov (2016). Quantile graphical models : Prediction and conditional independence with applications to financial risk management. arXiv preprint: 1607.00286.

Belloni, A. and V. Chernozhukov (2011). l1-penalized quantile regression in high-dimensional sparse models. The Annals of Statistics 39(1), 82–130.

Bickel, P. J., Y. Ritov, A. B. Tsybakov, et al. (2009). Simultaneous analysis of lasso and dantzig selector. The Annals of Statistics 37(4), 1705–1732.

Bühlmann, P. and S. Van de Geer (2011). Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Heidelberg Dordrecht London New York.

Chernozhukov, V., D. Chetverikov, K. Kato, et al. (2017). Central limit theorems and bootstrap in high dimensions. The Annals of Probability 45(4), 2309–2352.

Downey, P. J. (1990). Distribution-free bounds on the expectation of the maximum with scheduling applications. Operations Research Letters 9(3), 189–201.

Embrechts, P., C. Klüppelberg, and T. Mikosch (2013). Modelling extremal events: for insurance and finance, Volume 33. Springer Science & Business Media.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. The Annals of Statistics 42(6), 2243–2281.

Leonardi, F. and P. Bühlmann (2016). Computationally efficient change point detection for high-dimensional regression. arXiv preprint: 1601.03704.

Nazarov, F. (2003). On the maximal perimeter of a convex set in R^n with respect to a Gaussian measure. Geometric aspects of functional analysis 1807, 169–187.

Van, D. and J. A. Wellner (1996). Weak convergence and empirical processes. Springer,.

Wang, D., Z. Zhao, K. Z. Lin, and R. Willett (2021). Statistically and computationally efficient change point localization in regression settings. Journal of Machine Learning Research 22(248), 1–46.

Xu, H., D. Wang, Z. Zhao, and Y. Yu (2022). Change point inference in high-dimensional regression models under temporal dependence. arXiv preprint arXiv:2207.12453.

Zhang, B., J. Geng, and L. Lai (2015). Change-point estimation in high dimensional linear regression models via sparse group lasso. In 2015 53rd Annual Allerton Conference on

Communication, Control, and Computing (Allerton), pp. 815–821. IEEE.

Zhao, T., M. Kolar, and H. Liu (2014). A general framework for robust testing and confidence regions in high-dimensional quantile regression. arXiv preprint: 1601.03704.

Zhou, C., W.-X. Zhou, X.-S. Zhang, and H. Liu (2018). A unified framework for testing high dimensional parameters: a data-adaptive approach. arXiv preprint: 1601.03704.