Statistica Sinica

Simultaneous Change Point Detection and Identification for High Dimensional Linear Models

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

[1,2]*Fudan University, China;* [3]*University of North Carolina at Chapel Hill, U.S.A*

Abstract: In this article, we consider change point inference for high dimensional linear models. For change point detection, given any subgroup of variables, we propose a new method for testing the homogeneity of corresponding regression coefficients across the observations. Under some regularity conditions, the proposed new testing procedure controls the type I error asymptotically and is powerful against sparse alternatives and enjoys certain optimality. For change point identification, an "argmax" based change point estimator is proposed which is shown to be consistent for the true change point location. Moreover, combining with the binary segmentation technique, we further extend our new method for detecting and identifying multiple change points. Extensive simulation studies justify the validity of our new method and an application to the Alzheimer's disease data analysis further demonstrates its competitive performance.

Key words and phrases: Change point inference; High dimensions; Linear regression; Multiplier bootstrap; Subgroups.

1. Introduction

Driven by the great improvement of data collection and storage capacity, high dimensional linear regression models, have attracted a lot of attentions because of its simplicity for interpreting the effect of different variables in predicting the response. Specifically, we are interested in the following model:

$$Y = \boldsymbol{X}^\top \boldsymbol{\beta} + \epsilon,$$

where $Y \in \mathbb{R}$ is the response variable, $\boldsymbol{X} = (X_1, \ldots, X_p) \in \mathbb{R}^p$ is the covariate vector, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a p-dimensional unknown vector of coefficients, and $\epsilon \in \mathbb{R}$ is the error term.

For high dimensional linear regression, the L_1-penalized technique lasso (Tibshirani (1996)) is a popular method for estimating $\boldsymbol{\beta}$. In the past decades, lots of research attentions both in machine learning and statistics have been focused on studying theoretical properties of lasso and other penalized methods. Most of the existing literature on high dimensional linear regression focuses on the case with a homogeneous linear model, where the regression coefficients are assumed invariant across the observations. With many modern complex datasets for analysis in practice, data heterogeneity is a common challenge in many real applications such as economy and genetics. In some applications, the regression coefficients may have a sudden change at some unknown time point, which is called a change point. Typical examples include racial segregation and crime prediction in sociology; and financial contagion in economy. For these problems, methods and theories designed for independently and identically (*i.i.d.*) distributed settings are no longer applicable. As a result, ignoring these structural breaks in machine learning applications may lead to misleading results and wrong decision making. For the regression change point problem, a fundamental question is whether the underlying regression model remains homogenous across the observations. To address this issue, in this article, we investigate change point inference for high dimensional linear models. Specifically, let $(Y_i, \boldsymbol{X}_i)_{i=1}^n$ be n ordered independent realizations of $(Y, \boldsymbol{X})$. We aim to detect whether the regression coefficients have a change point during the observations. In particular, let $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ be two p-dimensional vectors of coefficients with $\boldsymbol{\beta}^{(1)} = (\beta_1^{(1)}, \ldots, \beta_p^{(1)})^\top$ and $\boldsymbol{\beta}^{(2)} = (\beta_1^{(2)}, \ldots, \beta_p^{(2)})^\top$. We consider the following linear regression model with a possible change point:

$$Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(1)} \mathbf{1}\{1 \le i \le k_*\} + \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(2)} \mathbf{1}\{k_* + 1 \le i \le n\} + \epsilon_i, \tag{1.1}$$

where k_* is the possible but unknown change point location and $(\epsilon_i)_{i=1}^n$ are the error terms. In this paper, we assume $k_* = \lfloor nt_0 \rfloor$ for some $t_0 \in (0, 1)$. For any given subgroup $\mathcal{G} \subset \{1, \ldots, p\}$, the first goal is to test

$$\begin{aligned} &\mathbf{H}_{0,\mathcal{G}} : \beta_s^{(1)} = \beta_s^{(2)} \ \text{ for all } s \in \mathcal{G} \ \text{ v.s.} \\ &\mathbf{H}_{1,\mathcal{G}} : \text{There exist } s \in \mathcal{G} \text{ and } k_* \in \{1, \ldots, n-1\} \text{ s.t. } \beta_s^{(1)} \neq \beta_s^{(2)}. \end{aligned} \tag{1.2}$$

In other words, under $\mathbf{H}_{0,\mathcal{G}}$, the regression coefficients in each subgroup $\mathcal{G}$ are homogeneous across the observations, and under $\mathbf{H}_{1,\mathcal{G}}$ there is a change point at an unknown time point k_* such that the regression coefficients have a sudden change after k_*. Our second goal of the paper is to identify the change point location once we reject $\mathbf{H}_{0,\mathcal{G}}$ in (1.2). In this paper, we assume that the number of coefficients can be much larger than the number of observations, i.e., $p \succeq n$, which is known as a high dimensional problem.

For the low dimensional setting with a fixed p and $p < n$, change point inference for linear regression models has been well-studied. For example, Quandt (1960) considered testing (1.2) for a simple regression model with $p = 2$. Based on that, several techniques were proposed in the literature. Among them are maximum likelihood ratio tests (Horváth, 1995), partial sums of regression residuals (Bai and Perron, 1998), and the union intersection test (Horváth and Shao, 1995). Moreover, as a special case of linear regression models, Chan et al. (2014) considered change point detection for the autoregressive model. As compared to the broad literature in the low dimensional setting, methods and theory for high dimensional change point inference of (1.1) have not been investigated much until recently. For instance, Lee et al. (2016) considered a high dimensional regression model with a possible change point due to a covariate threshold. Based on the

L_1/L_0 regularization, Kaul et al. (2019) proposed a two-step algorithm for the detection and estimation of parameters in a high-dimensional change point regression model. As extensions to multiple structural breaks in high dimensional linear models, Leonardi and Bühlmann (2016) proposed fast algorithms for multiple change point estimation based on dynamic programming and binary search algorithms. In addition, Zhang et al. (2015) developed an approach for estimating multiple change points based on sparse group lasso. Wang et al. (2021) proposed a projection-based algorithm for estimating multiple change points. Recently, Cho and Owens (2022); Bai and Safikhani (2023) constructed estimates for the multiple change points in high-dimensional regression models based on methods of moving window and blocked fused lasso. Kaul et al. (2021); Xu et al. (2022) respectively considered the problem of making confidence interval for the change point in the context of high dimensional mean vector-based models and linear regression models. Chen et al. (2023) proposed a new method for determing the number of change points with false discovery rate controls. Other related papers include He et al. (2023); Wang et al. (2022).

It is worth noting that the majority of above mentioned papers mainly focused on the estimation of regression coefficients as well as the change point locations by assuming a pre-existing change point in the model. To our best knowledge, the testing problem of (1.2) has not been considered yet. How to make effective change point detection remains to be an urgent but challenging task. To fill this gap, in this article, we consider change point inference in the context of high dimensional linear models.

The main contributions of this paper are as follows. For any prespecified subgroup $\mathcal{G} \subset \{1, \ldots, p\}$, we propose a new method for testing the homogeneity of corresponding regression coefficients across the observations. For change point detection, the proposed test statistic $T_{\mathcal{G}}$ is constructed based on a weighted L_∞ aggregation, both temporally and spatially, of the process $\{Z_j(\lfloor nt \rfloor)\}_{j\in\mathcal{G}, t\in[\tau_0, 1-\tau_0]}$, where $Z_j(\lfloor nt \rfloor) = \breve{\beta}_j^{(0,t)} - \breve{\beta}_j^{(t,1)}$ with $\breve{\beta}_j^{(0,t)}$ and $\breve{\beta}_j^{(t,1)}$ denoting the de-biased lasso estimators for coordinate j before and after time point $\lfloor nt \rfloor$, respectively. It is shown that $T_{\mathcal{G}}$ is powerful against sparse alternatives with only a few entries in $\mathcal{G}$ having a change point. To approximate its limiting null distribution, a multiplier bootstrap procedure is introduced. The proposed bootstrap can automatically account for the dependence structures of $\{Z_j(\lfloor nt \rfloor)\}_{j\in\mathcal{G}, t\in[\tau_0, 1-\tau_0]}$ and allow the group size $|\mathcal{G}|$ to grow exponentially with the sample size n. Furthermore, to identify the change point location, for each time point $\lfloor nt \rfloor$, we first aggregate the coordinates with the L_∞-norm, then a change point estimator $\widehat{t}_{0,\mathcal{G}}$ is obtained by taking "argmax" with respect to t of the above aggregated process with some proper weights. In addition to single change point detection, by combining with the binary segmentation technique (Vostrikova, 1981), we extend our new algorithm for detecting multiple change points which enjoys better performance than the existing methods.

In terms of theoretical investigation, with mild moment conditions on the covariates and errors in the regression model, we justify the validity of our proposed method in terms of change point detection and identification. In particular, our bootstrap procedure consistently approximates the limiting null distribution of $T_{\mathcal{G}}$, which implies that the proposed new test preserves the pre-specified significance level asymptotically. Furthermore, under $\mathbf{H}_{1,\mathcal{G}}$, our new method is sensitive to sparse alternatives and can reject the null hypothesis with probability tending to one. It is worth mentioning that Xia et al. (2018) considered two sample tests for high dimensional linear regression models. They derived some conditions for consistently distinguishing two sample regression models, which are shown to be minimax optimal. Our requirement for detecting a change point under $\mathbf{H}_{1,\mathcal{G}}$ has the same order as the condition derived in Xia et al. (2018). As for the change point estimation, we prove that our proposed argmax-based change point estimator is consistent for t_0 with an estimation error rate of $\left|\widehat{t}_{0,\mathcal{G}} - t_0\right| = O_p\big(\frac{\log(|\mathcal{G}|n)}{n\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}^2}\big)$, where $\boldsymbol{\delta} := \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}$

with $\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} =: \max_{j\in\mathcal{G}} |\beta_j^{(1)} - \beta_j^{(2)}|$. Hence, the above estimation result shows that our proposed change point estimator is consistent as long as $\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_{\mathcal{G},\infty} \gg \sqrt{\log(|\mathcal{G}|n)/n}$ and allows the overall sparsity of regression coefficients and the group's magnitude $|\mathcal{G}|$ to grow simultaneously with the sample size n. We demonstrate that our new testing procedure is relatively simple to implement and extensive numerical studies provide strong support to our theory. Moreover, an R package called "RegCpt" is developed to implement our proposed new algorithms.

The rest of this paper is organized as follows. In Section 2, we introduce our new methodology for Problem (1.2). In Section 3, some theoretical results are derived in terms of change point detection and identification. In Section 4, extensive numerical studies are investigated. The detailed proofs of the main theorems, additional numerical studies and an interesting application to the Alzheimer's disease dataset are given in the Appendix.

For $\boldsymbol{v} = (v_1, \ldots, v_p)^\top \in \mathbb{R}^p$, define its L_p norm as $\|\boldsymbol{v}\|_p = (\sum_{j=1}^d |v_j|^p)^{1/p}$ for $1 \le p \le \infty$. For $p = \infty$, define $\|\boldsymbol{v}\|_\infty = \max_{1\le j\le d} |v_j|$. For a subset $\mathcal{G} \subset \{1, \ldots, p\}$, denote $\|\boldsymbol{v}\|_{\mathcal{G},\infty}$ by $\max_{j\in\mathcal{G}} |v_j|$. For any set $\mathcal{S}$, denote its cardinality by $|\mathcal{S}|$. For two real numbered sequences a_n and b_n, we set $a_n = O(b_n)$ if there exits a constant C such that $|a_n| \le C|b_n|$ for a sufficiently large n; $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$; $a_n \asymp b_n$ if there exists constants c and C such that $c|b_n| \leq |a_n| \leq C|b_n|$ for a sufficiently large n. For a sequence of random variables $\{\xi_1, \xi_2, \ldots\}$, we denote $\xi_n = o_p(1)$ if $\xi_n \xrightarrow{\mathbb{P}} 0$. Define $\lfloor x \rfloor$ as the largest integer less than or equal to x for $x \geq 0$.

2. Methodology

2.1 New test statistic

We present our methodology for testing the existence of a change point in Model (1.1). To this end, we first introduce some basic model settings. Recall the regression model

$$Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(1)} \mathbf{1}\{1 \leq i \leq \lfloor nt_0 \rfloor\} + \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(2)} \mathbf{1}\{\lfloor nt_0 \rfloor + 1 \leq i \leq n\} + \epsilon_i. \quad (2.3)$$

Denote $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ as a $n \times 1$ response vector, $\mathbf{X}$ is a $n \times p$ design matrix with $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,p})^\top$ being its i-th row for $1 \leq i \leq n$, and $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^\top$ is the error vector. For the unknown $p \times 1$ regression vectors $\boldsymbol{\beta}^{(1)} = (\beta_1^{(1)}, \ldots, \beta_p^{(1)})^\top$ and $\boldsymbol{\beta}^{(2)} = (\beta_1^{(2)}, \ldots, \beta_p^{(2)})^\top$, 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. Define $\mathbf{\Sigma} = (\Sigma_{i,j}) = \mathrm{Cov}(\boldsymbol{X}_1)$ as the covariance matrix of $\boldsymbol{X}_1$ and $\mathbf{\Theta} = (\theta_{i,j})$ as the inverse of $\mathbf{\Sigma}$. For $\mathbf{\Theta}$, let $s_j = |\{1 \leq k \leq p : \theta_{j,k} \neq 0, k \neq j\}|$. In addition to the above notations, we assume that the change point does not happen at the beginning or end of data observations. In other words, there exists some $\tau_0 \in (0, 0.5)$ such that $t_0 \in [\tau_0, 1-\tau_0]$ holds. Note that the search boundary scales with n by allowing $\tau_0 \to 0$.

To propose our method, we first introduce the de-sparsified (de-biased) lasso estimator, which was proposed in Van de Geer et al. (2014) and Zhang and Zhang (2014). Specifically, for Model (2.3), let $\widehat{\boldsymbol{\beta}}_n$ be a lasso estimator from $\widehat{\boldsymbol{\beta}}^n = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2/n + 2\lambda_n \|\boldsymbol{\beta}\|_1$, where λ_n is the non-negative regularization parameter. Then for a homogeneous model with no change points, the de-biased lasso estimator is defined:

$$\breve{\boldsymbol{\beta}}^n = \widehat{\boldsymbol{\beta}}^n + \widehat{\boldsymbol{\Theta}}\mathbf{X}^\top\big(\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^n\big)/n, \tag{2.4}$$

where $\widehat{\boldsymbol{\Theta}}$ is some appropriate estimator for $\boldsymbol{\Theta}$. Essentially, the de-biased lasso estimator $\breve{\boldsymbol{\beta}}_n$ is a lasso solution by plugging in a Karush-Kuhn-Tucker (KKT) condition. It has been widely used for constructing confidence intervals and statistical tests for high dimensional parameters, and proven to be asymptotically optimal in terms of semiparametric efficiency.

Remark 1. In this paper, we adopt the node-wise estimation for obtaining $\widehat{\boldsymbol{\Theta}}$, as proposed in Meinshausen and Bühlmann (2006). The main idea is to perform regression on each variable using the remaining ones. In particular, denote $\mathbf{X}^j$ as the j-th column of $\mathbf{X}$ and $\mathbf{X}^{-j}$ as the remaining columns. For each $j = 1\ldots,p$, define

$$\widehat{\boldsymbol{\gamma}}_j = \mathop{\arg\min}_{\boldsymbol{\gamma}\in\mathbb{R}^{p-1}} \Big(\|\mathbf{X}^j - \mathbf{X}^{-j}\boldsymbol{\gamma}\|_2^2/n + 2\lambda_{(j)}\|\boldsymbol{\gamma}\|_1 \Big), \tag{2.5}$$

with $\widehat{\boldsymbol{\gamma}}_j = \{\widehat{\gamma}_{j,k} : k = 1\ldots,p, k \neq j\}$. Denote by $\widehat{\mathbf{C}} = (\widehat{c}_{i,j})_{i,j}^p$ with $\widehat{c}_{i,i} = 1$ and $\widehat{c}_{i,j} = -\gamma_{i,j}$ for $i \neq j$. Let $\widehat{\tau}_j^2 = \|\mathbf{X}^j - \mathbf{X}^{-j}\widehat{\boldsymbol{\gamma}}_j\|_2^2/n + \lambda_{(j)}\|\widehat{\boldsymbol{\gamma}}_j\|_1$ and $\widehat{\mathbf{T}}^2 = \mathrm{diag}\{\widehat{\tau}_1^2,\ldots,\widehat{\tau}_p^2\}$. The node-wise lasso estimator for $\boldsymbol{\Theta}$ is defined as

$$\widehat{\boldsymbol{\Theta}} = \widehat{\mathbf{T}}^{-2}\widehat{\mathbf{C}}. \tag{2.6}$$

It is shown that $\widehat{\boldsymbol{\Theta}}$ enjoys good properties in estimation accuracy. More importantly, it is possible to use parallel computation for calculating $\widehat{\boldsymbol{\Theta}}$, which is more appropriate for modern statistical applications with large scale datasets.

Since there is a possible but unknown change point in Model (2.3), we can not use (2.4) directly to make statistical inferences on $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$. The main challenge comes from the unknown change point t_0. To overcome this difficulty, instead of only calculating a single de-biased lasso estimator $\breve{\boldsymbol{\beta}}^n$, we need to construct the de-biased lasso-based process. To that end, we need some notations. For any $0 \leq s < t \leq 1$, define

$$\begin{aligned}
&\boldsymbol{Y}_{(s,t)} = (Y_{\lfloor ns\rfloor+1},\ldots,Y_{\lfloor nt\rfloor})^\top, \quad \boldsymbol{\epsilon}_{(s,t)} = (\epsilon_{\lfloor ns\rfloor+1},\ldots,\epsilon_{\lfloor nt\rfloor})^\top,\\
&\mathbf{X}_{(s,t)} = (\boldsymbol{X}_{\lfloor ns\rfloor+1},\ldots,\boldsymbol{X}_{\lfloor nt\rfloor})^\top, \widehat{\boldsymbol{\Sigma}}_{(s,t)} = \frac{1}{\lfloor nt\rfloor - \lfloor ns\rfloor + 1}\sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor} \boldsymbol{X}_i\boldsymbol{X}_i^\top.
\end{aligned}$$

To motivate our testing statistic, for each fixed $t \in [\tau_0, 1-\tau_0]$, we define

$$\boldsymbol{\beta}^{(0,t)} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\,\mathbb{E}\big\|\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\boldsymbol{\beta}\big\|_2^2, \boldsymbol{\beta}^{(t,1)} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\,\mathbb{E}\big\|\boldsymbol{Y}_{(t,1)} - \mathbf{X}_{(t,1)}\boldsymbol{\beta}\big\|_2^2. \tag{2.7}$$

By definition, $\boldsymbol{\beta}^{(0,t)}$ and $\boldsymbol{\beta}^{(t,1)}$ are the best regression coefficients for predicting $\boldsymbol{Y}_{(0,t)}$ and $\boldsymbol{Y}_{(t,1)}$ under the squared error loss, respectively. More importantly, suppose there is a change point t_0 in the linear model (2.3). According to the search location t and the true change point location t_0, the underlying true parameters can have the following explicit form:

$$\boldsymbol{\beta}^{(0,t)} = \boldsymbol{\beta}^{(1)}\mathbf{1}\{t \in [\tau_0, t_0]\} + \big(\frac{\lfloor nt_0\rfloor}{\lfloor nt\rfloor}\boldsymbol{\beta}^{(1)} + \frac{\lfloor nt\rfloor - \lfloor nt_0\rfloor}{\lfloor nt\rfloor}\boldsymbol{\beta}^{(2)}\big)\mathbf{1}\{t \in [t_0, 1-\tau_0]\},$$

and

$$\boldsymbol{\beta}^{(t,1)} = \big(\frac{\lfloor nt_0\rfloor - \lfloor nt\rfloor}{n - \lfloor nt\rfloor}\boldsymbol{\beta}^{(1)} + \frac{n - \lfloor nt_0\rfloor}{n - \lfloor nt\rfloor}\boldsymbol{\beta}^{(2)}\big)\mathbf{1}\{t \in [\tau_0, t_0]\} + \boldsymbol{\beta}^{(2)}\mathbf{1}\{t \in [t_0, 1-\tau_0]\}.$$

From the population level, we can define the theoretical signal jump process:

$$\begin{aligned}\boldsymbol{\delta}_n(t) &:= \sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}(\boldsymbol{\beta}^{(0,t)} - \boldsymbol{\beta}^{(t,1)})\\ &= \sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt_0\rfloor^*}{n}\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big)\mathbf{1}\{t \in [\tau_0, t_0]\}\\ &+ \sqrt{n}\frac{\lfloor nt_0\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big)\mathbf{1}\{t \in [t_0, 1-\tau_0]\},\end{aligned} \tag{2.8}$$

where $\lfloor nt\rfloor^* := n - \lfloor nt\rfloor$.

The signal function in (2.8) has some interesting properties. First, under $\mathbf{H}_{0,\mathcal{G}}$ of no change points, it reduces to a vector of zeros at each time point $\lfloor nt \rfloor$. Second, under $\mathbf{H}_{1,\mathcal{G}}$, $\boldsymbol{\delta}_n(t)$ is at most $(s^{(1)} + s^{(2)})$-sparse since we require sparse regression coefficients in the model. Third, we can see that $\|\boldsymbol{\delta}_n(t)\|_{\mathcal{G},\infty}$ with $t \in [\tau_0, 1-\tau_0]$ obtains its maximum value at the true change point location t_0. Hence, to make change point inference for high dimensional linear models, the key point is how to propose a test statistic that can estimate $\boldsymbol{\delta}_n(t)$ well under $\mathbf{H}_{1,\mathcal{G}}$, and has some theoretically tractable limiting null distributions under $\mathbf{H}_{0,\mathcal{G}}$. A natural idea is to use the lasso estimators directly. Specifically, for each time point, we obtain the lasso estimators $\widehat{\boldsymbol{\beta}}^{(0,t)} = (\widehat{\beta}_1^{(0,t)}, \ldots, \widehat{\beta}_p^{(0,t)})^\top$ and $\widehat{\boldsymbol{\beta}}^{(t,1)} = (\widehat{\beta}_1^{(t,1)}, \ldots, \widehat{\beta}_p^{(t,1)})^\top$:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(0,t)} &= \operatorname*{arg\,min}_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2\lfloor nt \rfloor}\left\|\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\boldsymbol{\beta}\right\|_2^2 + \lambda_1(t)\|\boldsymbol{\beta}\|_1, \\
\widehat{\boldsymbol{\beta}}^{(t,1)} &= \operatorname*{arg\,min}_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2\lfloor nt \rfloor^*}\left\|\boldsymbol{Y}_{(t,1)} - \mathbf{X}_{(t,1)}\boldsymbol{\beta}\right\|_2^2 + \lambda_2(t)\|\boldsymbol{\beta}\|_1,
\end{aligned} \tag{2.9}$$

where $\lambda_1(t)$ and $\lambda_2(t)$ are some regularity parameters to account for the data heterogeneity. It is well known that due to the ℓ_1 regularized penalization in (2.9), the lasso estimators are typically biased and do not have a tractable limiting null distribution. As a result, some"de-biased" process is needed. The main idea is to plug into the KKT conditions under both $\mathbf{H}_{0,\mathcal{G}}$ and $\mathbf{H}_{1,\mathcal{G}}$ for the change point model. To give an insight into the de-biased process for change point detection, in what follows, we assume $\mathbf{H}_{1,\mathcal{G}}$ holds.

Firstly, we consider the case that the search location satisfies $t \in [\tau_0, t_0]$. Let $\widehat{\boldsymbol{\kappa}}_1(t) \in \mathbb{R}^p$ and $\widehat{\boldsymbol{\kappa}}_2(t) \in \mathbb{R}^p$ be the subdifferentials of $\|\boldsymbol{\beta}\|_1$ for the first and second optimization problems in (2.9), respectively. Then, by the KKT condition, we have:

$$
\begin{aligned}
&-\mathbf{X}_{(0,t)}^{\top}(\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\widehat{\boldsymbol{\beta}}^{(0,t)})/\lfloor nt \rfloor + \lambda_1(t)\widehat{\boldsymbol{\kappa}}_1(t) = \mathbf{0}, \\
&-\mathbf{X}_{(t,1)}^{\top}(\boldsymbol{Y}_{(t,1)} - \mathbf{X}_{(t,1)}\widehat{\boldsymbol{\beta}}^{(t,1)})/\lfloor nt \rfloor^{*} + \lambda_2(t)\widehat{\boldsymbol{\kappa}}_2(t) = \mathbf{0}.
\end{aligned} \tag{2.10}
$$

Note that for $t \in [\tau_0, t_0]$, the samples $\{\boldsymbol{Y}_{(0,t)}, \mathbf{X}_{(0,t)}\}$ are homogeneous with regression coefficients being $\boldsymbol{\beta}^{(0,t)} = \boldsymbol{\beta}^{(1)}$. Hence, similar to the analysis in Van de Geer et al. (2014), for the first term in (2.10), for $t \in [\tau_0, t_0]$, we have the following decomposition:

$$
\widehat{\boldsymbol{\beta}}^{(0,t)} + \widehat{\boldsymbol{\Theta}}\lambda_1(t)\widehat{\boldsymbol{\kappa}}_1(t) - \boldsymbol{\beta}^{(1)} = \mathbf{X}_{(0,t)}^{\top}\boldsymbol{\epsilon}_{(0,t)}/\lfloor nt \rfloor \overbrace{-(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I})(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(1)})}^{\boldsymbol{\Delta}_{(0,t)}^{I}}. \tag{2.11}
$$

For the second term in (2.10), we note that the samples $\{\boldsymbol{Y}_{(t,1)}, \mathbf{X}_{(t,1)}\}$ with $t \in [\tau_0, t_0]$ are heterogeneous due to the change point at t_0. Observe that

$$
\mathbf{X}_{(t,1)}^{\top} = (\mathbf{X}_{(t,t_0)}^{\top}, \mathbf{X}_{(t_0,1)}^{\top}), \text{ and } \mathbf{Y}_{(t,1)} = ((\mathbf{X}_{(t,t_0)}\boldsymbol{\beta}^{(1)})^{\top} + \boldsymbol{\epsilon}_{(t,t_0)}^{\top}, (\mathbf{X}_{(t_0,1)}\boldsymbol{\beta}^{(2)})^{\top} + \boldsymbol{\epsilon}_{(t_0,1)}^{\top})^{\top}.
$$

Then, the KKT condition for the second equation in (2.10) becomes:

$$
\begin{aligned}
&\lambda_2(t)\widehat{\boldsymbol{\kappa}}_2(t) \\
&= \mathbf{X}_{(t,t_0)}^{\top}\mathbf{X}_{(t,t_0)}(\boldsymbol{\beta}^{(1)} - \widehat{\boldsymbol{\beta}}^{(t,1)})/\lfloor nt \rfloor^{*} + \mathbf{X}_{(t_0,1)}^{\top}\mathbf{X}_{(t_0,1)}(\boldsymbol{\beta}^{(2)} - \widehat{\boldsymbol{\beta}}^{(t,1)})/\lfloor nt \rfloor^{*} + \mathbf{X}_{(t,1)}^{\top}\boldsymbol{\epsilon}_{(t,1)}/\lfloor nt \rfloor^{*} \\
&= \widehat{\boldsymbol{\Sigma}}_{(t,1)}(\boldsymbol{\beta}^{(2)} - \widehat{\boldsymbol{\beta}}^{(t,1)}) + \frac{\lfloor nt_0 \rfloor - \lfloor nt \rfloor}{\lfloor nt \rfloor^{*}}\widehat{\boldsymbol{\Sigma}}_{(t,t_0)}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}) + \mathbf{X}_{(t,1)}^{\top}\boldsymbol{\epsilon}_{(t,1)}/\lfloor nt \rfloor^{*}.
\end{aligned} \tag{2.12}
$$

Multiplying $\widehat{\boldsymbol{\Theta}}$ on both sides of (2.12), for the case of $t \in [\tau_0, t_0]$, we have:

$$\begin{aligned}
&\widehat{\boldsymbol{\beta}}^{(t,1)} + \widehat{\boldsymbol{\Theta}}\lambda_2(t)\widehat{\boldsymbol{\kappa}}_2(t) - \underbrace{\Big(\frac{\lfloor nt_0\rfloor - \lfloor nt\rfloor}{\lfloor nt\rfloor^*}\boldsymbol{\beta}^{(1)} + \frac{n-\lfloor nt_0\rfloor}{\lfloor nt\rfloor^*}\boldsymbol{\beta}^{(2)}\Big)}_{\boldsymbol{\beta}^{(t,1)}} \\
&= \underbrace{-(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,1)} - \mathbf{I})(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(2)}) - \frac{\lfloor nt_0\rfloor - \lfloor nt\rfloor}{\lfloor nt\rfloor^*}(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,t_0)} - \mathbf{I})(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})}_{\boldsymbol{\Delta}^{I}_{(t,1)}} + \mathbf{X}^\top_{(t,1)}\boldsymbol{\epsilon}_{(t,1)}/\lfloor nt\rfloor^*.
\end{aligned} \tag{2.13}$$

Secondly, for the case of $t \in [t_0, 1-\tau_0]$, using a very similar analysis, we can prove that:

$$\begin{aligned}
&\widehat{\boldsymbol{\beta}}^{(0,t)} + \widehat{\boldsymbol{\Theta}}\lambda_1(t)\widehat{\boldsymbol{\kappa}}_1(t) - \underbrace{\Big(\frac{\lfloor nt_0\rfloor}{\lfloor nt\rfloor}\boldsymbol{\beta}^{(1)} + \frac{\lfloor nt\rfloor - \lfloor nt_0\rfloor}{\lfloor nt\rfloor}\boldsymbol{\beta}^{(2)}\Big)}_{\boldsymbol{\beta}^{(0,t)}} = \mathbf{X}^\top_{(0,t)}\boldsymbol{\epsilon}_{(0,t)}/\lfloor nt\rfloor + \boldsymbol{\Delta}^{II}_{(0,t)}, \\
&\widehat{\boldsymbol{\beta}}^{(t,1)} + \widehat{\boldsymbol{\Theta}}\lambda_2(t)\widehat{\boldsymbol{\kappa}}_2(t) - \underbrace{\boldsymbol{\beta}^{(2)}}_{\boldsymbol{\beta}^{(t,1)}} = \mathbf{X}^\top_{(t,1)}\boldsymbol{\epsilon}_{(t,1)}/\lfloor nt\rfloor^* + \boldsymbol{\Delta}^{II}_{(t,1)},
\end{aligned} \tag{2.14}$$

where the two terms $\boldsymbol{\Delta}^{II}_{(0,t)}$ are $\boldsymbol{\Delta}^{II}_{(t,1)}$ are defined as

$$\begin{aligned}
\boldsymbol{\Delta}^{II}_{(0,t)} &:= -\frac{\lfloor nt\rfloor - \lfloor nt_0\rfloor}{\lfloor nt\rfloor}\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t_0,t)} - \mathbf{I}\big)\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big) - \big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(1)}\big), \\
\boldsymbol{\Delta}^{II}_{(t,1)} &:= -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,1)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(2)}\big).
\end{aligned}$$

Combining the results in (2.10)-(2.14), for each $t \in [\tau_0, 1-\tau_0]$, we then construct the de-biased lasso estimators $\breve{\boldsymbol{\beta}}^{(0,t)} = (\breve{\beta}_1^{(0,t)}, \ldots, \breve{\beta}_p^{(0,t)})^\top$ and $\breve{\boldsymbol{\beta}}^{(t,1)} = (\breve{\beta}_1^{(t,1)}, \ldots, \breve{\beta}_p^{(t,1)})^\top$ as follows:

$$\begin{aligned}
\breve{\boldsymbol{\beta}}^{(0,t)} &= \widehat{\boldsymbol{\beta}}^{(0,t)} + \widehat{\boldsymbol{\Theta}}\mathbf{X}^\top_{(0,t)}\big(\mathbf{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\widehat{\boldsymbol{\beta}}^{(0,t)}\big)/\lfloor nt\rfloor, \\
\breve{\boldsymbol{\beta}}^{(t,1)} &= \widehat{\boldsymbol{\beta}}^{(t,1)} + \widehat{\boldsymbol{\Theta}}\mathbf{X}^\top_{(t,1)}\big(\mathbf{Y}_{(t,1)} - \mathbf{X}_{(t,1)}\widehat{\boldsymbol{\beta}}^{(t,1)}\big)/\lfloor nt\rfloor^*.
\end{aligned} \tag{2.15}$$

The construction of our new test statistic comes from our important new derivation (2.15). In particular, under some regularity conditions, the difference between $\breve{\boldsymbol{\beta}}^{(0,t)}$ and $\breve{\boldsymbol{\beta}}^{(t,1)}$ has the following decomposition:

$$\sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\big(\breve{\boldsymbol{\beta}}^{(0,t)}-\breve{\boldsymbol{\beta}}^{(t,1)}\big) = \underbrace{\boldsymbol{\delta}_n(t)}_{\textbf{Signal function}} + \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\widehat{\boldsymbol{\Theta}}\boldsymbol{X}_i\epsilon_i}_{\textbf{Random noise}} + \underbrace{\sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}(\boldsymbol{R}^{(0,t)}-\boldsymbol{R}^{(t,1)})}_{\textbf{Random bias}}, \tag{2.16}$$

where $\boldsymbol{\delta}_n(t)$ is defined in (2.8), and $\boldsymbol{R}^{(0,t)}$ and $\boldsymbol{R}^{(t,1)}$ are the residuals:

$$\boldsymbol{R}_{(0,t)} = \boldsymbol{\Delta}^{I}_{(0,t)}\mathbf{1}\{t\in[\tau_0,t_0]\} + \boldsymbol{\Delta}^{II}_{(0,t)}\mathbf{1}\{t\in[t_0,1-\tau_0]\},$$
$$\boldsymbol{R}_{(t,1)} = \boldsymbol{\Delta}^{I}_{(t,1)}\mathbf{1}\{t\in[\tau_0,t_0]\} + \boldsymbol{\Delta}^{II}_{(t,1)}\mathbf{1}\{t\in[t_0,1-\tau_0]\}.$$

The above de-biased lasso-based process enjoys several advantages for making change point inference. Firstly, under $\mathbf{H}_{0,\mathcal{G}}$ of no change points, it is the combination of a partial sum-based process plus a random bias term. The latter one can be shown to be negligible. Moreover, under $\mathbf{H}_{1,\mathcal{G}}$, we can see that the de-biased lasso-based process is an asymptotically unbiased estimator for the signal function defined in (2.8), allowing us to make change point detection and identification. The derivation of (2.16) is different from the original de-biased lasso estimator in (2.4) and requires a fundamental modification of Bickel et al. (2009)) to account for data heterogeneity. More details can be found in the Appendix.

Motived by the above observation, for any given subgroup $\mathcal{G}\subset\{1,\ldots,p\}$, a natural test statistic for the hypothesis (1.2) is defined as

$$\widetilde{T}_{\mathcal{G}} = \max_{t\in[\tau_0,1-\tau_0]} \max_{j\in\mathcal{G}} \sqrt{n}\frac{\lfloor nt \rfloor}{n}\Big(1-\frac{\lfloor nt \rfloor}{n}\Big)\Big|\breve{\beta}_j^{(0,t)} - \breve{\beta}_j^{(t,1)}\Big|.$$

For any given subgroup $\mathcal{G}$, the proposed new statistic $\widetilde{T}_{\mathcal{G}}$ searches all possible locations of time points. It is demonstrated that $\widetilde{T}_{\mathcal{G}}$ is powerful against sparse alternatives with only a few entries in $\mathcal{G}$ having a change point, and a large value of $\widetilde{T}_{\mathcal{G}}$ leads to a rejection of $\mathbf{H}_{0,\mathcal{G}}$.

2.2 Weighted variance estimation

In Section 2.1, we introduced $\widetilde{T}_{\mathcal{G}}$ for the hypothesis (1.2). Considering the variability of the design matrix $\mathbf{X}$ and the error term $\boldsymbol{\epsilon}$, the test statistic $\widetilde{T}_G$ is heterogeneous. Hence, we need to take its variance into account and standardize it. In this paper, we adopt a weighted variance estimator. Specifically, let $\widehat{\boldsymbol{\Omega}} = (\widehat{\omega}_{i,j})_{i,j}^p = \widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_n\widehat{\boldsymbol{\Theta}}^\top$ with $\widehat{\boldsymbol{\Sigma}}_n := \mathbf{X}^\top\mathbf{X}/n$. For each $t \in [\tau_0, 1-\tau_0]$, denote

$$\widehat{\sigma}_\epsilon^2(t) = \frac{1}{n}\big(\big\|\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\widehat{\boldsymbol{\beta}}^{(0,t)}\big\|_2^2 + \big\|\boldsymbol{Y}_{(t,1)} - \mathbf{X}_{(t,1)}\widehat{\boldsymbol{\beta}}^{(t,1)}\big\|_2^2\big). \qquad (2.17)$$

Under $\mathbf{H}_{0,\mathcal{G}}$ of no change points in the model, we can prove that

$$\max_{\tau_0\le t\le 1-\tau_0} \max_{1\le j\le p} |\widehat{\sigma}_\epsilon^2(t)\widehat{\omega}_{j,j} - \sigma_\epsilon^2\omega_{j,j}| = o_p(1).$$

Under $\mathbf{H}_{1,\mathcal{G}}$, however, $\widehat{\sigma}_\epsilon^2(t)$ is not a consistent estimator for σ_ϵ^2 because of the unknown change point t_0. Furthermore, as discussed in Shao and Zhang

(2010), an inappropriate variance estimator may lead to non-monotonic power performance. In order to form a powerful test statistic, it is necessary to construct consistent variance estimation for $\mathbf{H}_{0,\mathcal{G}}$ and $\mathbf{H}_{1,\mathcal{G}}$. To address this issue, we need to deal with the unknown change point first. In particular, for a given subgroup $\mathcal{G}$, define

$$H_{\mathcal{G}}(t) = \max_{j \in \mathcal{G}} \frac{\lfloor nt \rfloor}{n}\Big(1 - \frac{\lfloor nt \rfloor}{n}\Big)\Big|\breve{\beta}_j^{(0,t)} - \breve{\beta}_j^{(t,1)}\Big|, \text{ with } t \in [\tau_0, 1-\tau_0].$$

By maximizing $H_{\mathcal{G}}(t)$, we obtain the argmax-based change point estimator:

$$\widehat{t}_{0,\mathcal{G}} = \mathop{\arg\max}_{t \in [\tau_0, 1-\tau_0]} H_{\mathcal{G}}(t). \tag{2.18}$$

Based on (2.18), let $\widehat{t}_0 = \widehat{t}_{0,\mathcal{G}}$ with $\mathcal{G} = \{1, \ldots, p\}$. We put $\widehat{t}_0$ into $\widehat{\sigma}_\epsilon^2(t)$ and get a weighted variance estimator for σ_ϵ^2 as

$$\widehat{\sigma}_\epsilon^2 = \frac{1}{n}\big(\big\|\boldsymbol{Y}_{(0,\widehat{t}_0)} - \mathbf{X}_{(0,\widehat{t}_0)}\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)}\big\|_2^2 + \big\|\boldsymbol{Y}_{(\widehat{t}_0,1)} - \mathbf{X}_{(\widehat{t}_0,1)}\widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)}\big\|_2^2\big). \tag{2.19}$$

As shown in our theoretical analysis, the new variance estimation in (2.19) is consistent under both $\mathbf{H}_{0,\mathcal{G}}$ and $\mathbf{H}_{1,\mathcal{G}}$. The proof is nontrivial since we need to justify the consistency of $\widehat{t}_{0,\mathcal{G}}$ for t_0, which is known to be an important but difficult task for high dimensional linear models (Lee et al. (2016)).

Using the new variance estimator in (2.19), for any given subgroup $\mathcal{G} \subset \{1, \ldots, p\}$, our new test statistic for the hypothesis (1.2) is finally defined as follows:

$$T_{\mathcal{G}} = \max_{t \in [\tau_0, 1-\tau_0]} \max_{j \in \mathcal{G}} \sqrt{n}\frac{\lfloor nt \rfloor}{n}\Big(1 - \frac{\lfloor nt \rfloor}{n}\Big)\Big|\frac{\breve{\beta}_j^{(0,t)} - \breve{\beta}_j^{(t,1)}}{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}}\Big|. \tag{2.20}$$

2.3 Multiplier bootstrap for approximating the null distribution

In Section 2.2, we have proposed the new test statistic $T_{\mathcal{G}}$ for the hypothesis (1.2). It is challenging to directly obtain its limiting null distribution in high dimensions. Bootstrap has been widely used for making statistical inference on high dimensional linear models since the seminal work of Chernozhukov et al. (2013). For high dimensional linear models with change points, however, existing bootstrap techniques are not applicable and it is desirable to design a new method. To overcome this problem, we investigate two types of multiplier bootstrap.

2.3.1 Bootstrap-I

Recall the decomposition in (2.16). Under $\mathbf{H}_{0,\mathcal{G}}$, we have

$$\sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\big(\breve{\boldsymbol{\beta}}^{(0,t)} - \breve{\boldsymbol{\beta}}^{(t,1)}\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt \rfloor}\widehat{\boldsymbol{\Theta}}\boldsymbol{X}_i\epsilon_i + (\boldsymbol{R}^{(0,t)} - \boldsymbol{R}^{(t,1)}).$$

It is shown that under $\mathbf{H}_{0,\mathcal{G}}$, the residual-based process $\{\boldsymbol{R}^{(0,t)} - \boldsymbol{R}^{(t,1)}, t \in [\tau_0, 1-\tau_0]\}$ is asymptotically negligible and the partial sum-based process $\{n^{-1/2}\sum_{i=1}^{\lfloor nt \rfloor}\widehat{\boldsymbol{\Theta}}\boldsymbol{X}_i\epsilon_i, t \in [\tau_0, 1-\tau_0]\}$ determines the limiting null distribution of $T_{\mathcal{G}}$, which is known as the leading term. This motivates us to first consider the following bootstrap method:

Step 1: For the b-th bootstrap, generate *i.i.d.* random variables $\epsilon_1^b, \ldots, \epsilon_n^b$ with $\epsilon_i^b \sim N(0,1)$.

Step 2: Calculate the testing statistic for the b-th bootstrap by

$$W_{\mathcal{G}}^b = \max_{t\in[\tau_0,1-\tau_0]} \max_{j\in\mathcal{G}} \sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\widehat{\omega}_{j,j}^{-1/2}\Big|\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\widehat{\mathbf{\Theta}}_j^\top\mathbf{X}_i\epsilon_i^b - \frac{1}{\lfloor nt\rfloor^*}\sum_{i=\lfloor nt\rfloor+1}^{n}\widehat{\mathbf{\Theta}}_j^\top\mathbf{X}_i\epsilon_i^b\Big|,$$

where $\widehat{\mathbf{\Theta}}_j^\top$ is the j-th row of $\widehat{\mathbf{\Theta}}$.

Step 3: Repeat the above process for B times.

Step 4: Based on the bootstrap samples $\{W_{\mathcal{G}}^1,\ldots,W_{\mathcal{G}}^B\}$, calculate the bootstrap sample-based critical value

$$\widehat{w}_{\mathcal{G},\alpha} = \inf\Big\{t : (B+1)^{-1}\sum_{b=1}^{B}\mathbf{1}\{W_{\mathcal{G}}^b \le t|\mathbf{X},\boldsymbol{Y}\} \ge 1-\alpha\Big\}.$$

Step 5: Reject $\mathbf{H}_{0,\mathcal{G}}$ if and only if $T_{\mathcal{G}} \ge \widehat{w}_{\mathcal{G},1-\alpha}$.

Note that the above bootstrap method essentially bootstraps the partial sum-based process, which has been recently used for change point detection of high dimensional mean vectors in Jirak (2015); Yu and Chen (2021). As shown in our numerical studies, Bootstrap-I suffers from serious size distortions. This phenomenon is due to large biases arising from the residual-based process $\{\boldsymbol{R}^{(0,t)} - \boldsymbol{R}^{(t,1)}, t \in [\tau_0, 1-\tau_0]\}$, which can not be ignored in finite sample performance although it is asymptotically negligible. Hence, for change point detection in high dimensional linear models, substantial modifications are needed and it is desirable to consider a new candidate bootstrap method. To overcome this problem, different from the existing methods, we choose to bootstrap the entire de-biased lasso-based process as shown in the following Bootstrap-II.

2.3.2 Bootstrap-II

The key idea of this bootstrap is to approximate the null limiting distribution under both $\mathbf{H}_{0,\mathcal{G}}$ and $\mathbf{H}_{1,\mathcal{G}}$. In particular, it proceeds as follows:

Step 1: Given $\widehat{\sigma}_\epsilon^2$ in (2.19), for the b-th bootstrap, let $\epsilon_1^b, \ldots, \epsilon_n^b$ be *i.i.d.* random variables following $N(0, \widehat{\sigma}_\epsilon^2)$. Define the b-th bootstrap of response vectors $\boldsymbol{Y}^b = (Y_1^b, \ldots, Y_n^b)^\top$:

$$Y_i^b = \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)} \mathbf{1}\{1 \leq i \leq \lfloor n\widehat{t}_0 \rfloor\} + \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)} \mathbf{1}\{\lfloor n\widehat{t}_0 \rfloor < i \leq n\} + \epsilon_i^b, \quad (2.21)$$

where $\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)}$ and $\widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)}$ are the lasso estimators before and after $\widehat{t}_0$.

Step 2: Denote $\boldsymbol{Y}_{(0,t)}^b = (Y_1^b, \ldots, Y_{\lfloor nt \rfloor}^b)^\top$, and $\boldsymbol{Y}_{(t,1)}^b = (Y_{\lfloor nt \rfloor+1}^b, \ldots, Y_n^b)^\top$. We then define the b-th bootstrap version of the de-biased lasso estimators before and after $\lfloor nt \rfloor$ as $\breve{\boldsymbol{\beta}}^{b,(0,t)} = (\breve{\beta}_1^{b,(0,t)}, \ldots, \breve{\beta}_p^{b,(0,t)})^\top$ and $\breve{\boldsymbol{\beta}}^{b,(t,1)} = (\breve{\beta}_1^{b,(t,1)}, \ldots, \breve{\beta}_p^{b,(t,1)})^\top$, where

$$\begin{aligned}
\breve{\boldsymbol{\beta}}^{b,(0,t)} &:= \widehat{\boldsymbol{\beta}}^{b,(0,t)} + \widehat{\boldsymbol{\Theta}} \mathbf{X}_{(0,t)}^\top \Big(\mathbf{Y}_{(0,t)}^b - \mathbf{X}_{(0,t)} \widehat{\boldsymbol{\beta}}^{b,(0,t)} \Big) / \lfloor nt \rfloor, \\
\breve{\boldsymbol{\beta}}^{b,(t,1)} &:= \widehat{\boldsymbol{\beta}}^{b,(t,1)} + \widehat{\boldsymbol{\Theta}} \mathbf{X}_{(t,1)}^\top \Big(\mathbf{Y}_{(t,1)}^b - \mathbf{X}_{(t,1)} \widehat{\boldsymbol{\beta}}^{b,(t,1)} \Big) / \lfloor nt \rfloor^*,
\end{aligned} \quad (2.22)$$

and $\widehat{\boldsymbol{\beta}}^{b,(0,t)}$ and $\widehat{\boldsymbol{\beta}}^{b,(t,1)}$ are the lasso estimators before and after t using the bootstrap samples $\{\boldsymbol{Y}_{(0,t)}^b, \mathbf{X}_{(0,t)}\}$ and $\{\boldsymbol{Y}_{(t,1)}^b, \mathbf{X}_{(t,1)}\}$.

Step 3: Define the bootstrap sample-based signal function $\widehat{\boldsymbol{\delta}}(t) = (\widehat{\delta}_1(t), \ldots, \widehat{\delta}_p(t))^\top$:

$$\widehat{\boldsymbol{\delta}}(t) = \frac{n - \lfloor n\widehat{t}_0 \rfloor}{n - \lfloor nt \rfloor}\big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)}\big)\mathbf{1}\{t \in [\tau_0, \widehat{t}_0]\} + \frac{\lfloor n\widehat{t}_0 \rfloor}{\lfloor nt \rfloor}\big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)}\big)\mathbf{1}\{t \in [\widehat{t}_0, 1-\tau_0]\}.$$

Step 4: Calculate the b-th bootstrap version for the test statistic $T_{\mathcal{G}}$ by

$$T_{\mathcal{G}}^b = \max_{t \in [\tau_0, 1-\tau_0]} \max_{j \in \mathcal{G}} \sqrt{n}\frac{\lfloor nt \rfloor}{n}\Big(1 - \frac{\lfloor nt \rfloor}{n}\Big)\Big|\frac{\breve{\beta}_j^{b,(0,t)} - \breve{\beta}_j^{b,(t,1)} - \widehat{\delta}_j(t)}{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}}\Big|. \tag{2.23}$$

Step 5: Repeat the above procedures (2.21)-(2.23) for B times and obtain the bootstrap samples $\{T_{\mathcal{G}}^1, \ldots, T_{\mathcal{G}}^B\}$. Let $c_{\mathcal{G},\alpha} := \inf\{t : \mathbb{P}(T_{\mathcal{G}} \leq t) \geq 1 - \alpha\}$ be the theoretical critical value of $T_{\mathcal{G}}$. Using the bootstrap samples $\{T_{\mathcal{G}}^1, \ldots, T_{\mathcal{G}}^B\}$, we estimate $c_{\mathcal{G},\alpha}$ by

$$\widehat{c}_{\mathcal{G},\alpha} = \inf\Big\{t : (B+1)^{-1}\sum_{b=1}^{B} \mathbf{1}\{T_{\mathcal{G}}^b \leq t | \mathbf{X}, \boldsymbol{Y}\} \geq 1 - \alpha\Big\}. \tag{2.24}$$

Step 6: Define the new test for the hypothesis (1.2) as follows:

$$\Phi_{\mathcal{G},\alpha} = \mathbf{1}\{T_{\mathcal{G}} \geq \widehat{c}_{\mathcal{G},\alpha}\}. \tag{2.25}$$

Given a significance level $\alpha \in (0, 1)$ and a prespecified subgroup $\mathcal{G}$, for the hypothesis (1.2), we reject $\mathbf{H}_{0,\mathcal{G}}$ if $\Phi_{\mathcal{G},\alpha} = 1$.

It is shown in theory that the Bootsrap-II-based test statistic $T_{\mathcal{G}}^b$ approximates the limiting null distribution of $T_{\mathcal{G}}$. More importantly, by bootstrapping the whole de-biased lasso-based process, Bootstrap-II enjoys better size performance than Bootstrap-I under various candidate subgroups. This is supported by our extensive numerical studies in Section 4.

3. Theoretical properties

In this section, we examine some theoretical properties of our proposed method including the size, power and the change point estimation results.

3.1 Basic assumptions

To save space, we provide brief descriptions of our assumptions below. More details on basic assumptions for making change point inference on high dimensional linear models can be found in the Supplementary Materials. Assumptions (**A.1**) – (**A.3**) impose some regular conditions on the design matrix as well as the error terms. Assumption (**A.4**) contains basic requirements on model parameters. Assumption (**A.5**) is a technical condition on the regularity parameters in (2.5) and (2.9).

3.2 Main results

We derive some theoretical results of our proposed new test. In Section 3.2.1, we consider the control of Type I error. In Section 3.2.2, we examine the power performance as well as the accuracy of change point estimation.

3.2.1 The validity of test size

Before giving the size results, we first consider the variance estimation. Theorem 1 shows that the pooled weighted variance estimator is uniformly consistent under the null hypothesis. It is crucial for deriving the Gaussian approximation results as in Theorem 2.

Theorem 1. *Suppose Assumptions* $(\mathbf{A}.\mathbf{1})$ – $(\mathbf{A}.\mathbf{5})$ *hold. Under* $\mathbf{H}_{0,\mathcal{G}}$*, for the variance estimator, with probability at least* $1 - C_1(np)^{-C_2}$*, we have*

$$\max_{1\leq j,k\leq p} |\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,k} - \sigma_\epsilon^2 \omega_{j,k}| \leq C_3\big(\sqrt{\frac{\log(n)}{n}} + \max_j \lambda_{(j)}\sqrt{s_j}\big),$$

where $C_1, \ldots, C_3$ *are universal positive constants not depending on* n *or* p*.*

Based on Theorem 1 as well as other regularity conditions, the following Theorem 2 justifies the validity of our bootstrap procedure.

Theorem 2. *Suppose Assumptions* $(\mathbf{A}.\mathbf{1})$ – $(\mathbf{A}.\mathbf{5})$ *hold. Under* $\mathbf{H}_{0,\mathcal{G}}$*, for any given subgroup* $\mathcal{G} \subset \{1, \ldots, p\}$*, we have*

$$\sup_{z\in(0,\infty)} \left|\mathbb{P}(T_{\mathcal{G}} \leq z) - \mathbb{P}(T_{\mathcal{G}}^b \leq z|\{\mathbf{X}, \boldsymbol{Y}\})\right| = o_p(1), \ \text{as } n, p \to \infty.$$

Theorem 2 shows that we can uniformly approximate the distribution of $T_{\mathcal{G}}$ using that of $T_{\mathcal{G}}^b$. As a corollary, the following Corollary 1 shows that our proposed new test can control the Type I error asymptotically for any given pre-specified significant level α.

Corollary 1. *Assume Assumptions* $(\mathbf{A.1})$–$(\mathbf{A.5})$ *hold. Under* $\mathbf{H}_{0,\mathcal{G}}$*, for any given subgroup* $\mathcal{G} \subset \{1,\ldots,p\}$*, we have* $\mathbb{P}(\Phi_{\mathcal{G},\alpha} = 1) \to \alpha$*, as* $n, p, B \to \infty$.

3.2.2 Analysis under $\mathbf{H}_{1,\mathcal{G}}$

After analyzing the theoretical results under the null hypothesis, we next consider the performance under $\mathbf{H}_{1,\mathcal{G}}$. To this end, some additional assumptions are needed.

Assumption $(\mathbf{A.6})$. Let $\boldsymbol{\delta} = \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}$. For the signal jump, we require there exists a constant $c^* \in [0, \infty)$ such that $\lim_{n,p\to\infty} s\|\boldsymbol{\delta}\|_\infty \to c^*$.

Note that Assumption $(\mathbf{A.6})$ is a signal strength requirement for identifying the change point location t_0 with high accuracy. It allows weak signals that can scale to zero as $(n, p) \to \infty$. With the additional assumption as well as those of $(\mathbf{A.1})$ – $(\mathbf{A.5})$, the following Theorem 3 provides a non-asymptotic estimation error bound of $\widehat{t}_{0,\mathcal{G}}$ for t_0.

Theorem 3. *Suppose Assumptions* $(\mathbf{A.1})$ - $(\mathbf{A.6})$ *hold. Assume additionally* $\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} \gg \sqrt{\log(|\mathcal{G}|n)/n}$ *holds. For any given subgroup* $\mathcal{G} \subset \{1,\ldots,p\}$*, under* $\mathbf{H}_{1,\mathcal{G}}$*, with probability at least* $1 - C_1(np)^{-C_2}$*, we have*

$$\left|\widehat{t}_{0,\mathcal{G}} - t_0\right| \le C^* \frac{\log(|\mathcal{G}|n)}{n\|\boldsymbol{\delta}\|^2_{\mathcal{G},\infty}}, \tag{3.26}$$

where C^* *is a universal positive constant not depending on* n *or* p.

Theorem 3 shows that our subgroup-based change point estimator is asymptotically consistent, which allows the group size $|\mathcal{G}|$ to grow with the sample size n as long as $\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} \gg \sqrt{\log(|\mathcal{G}|n)/n}$ holds.

Remark 2. Note that Jirak (2015); Yu and Chen (2021) considered the change point estimation for high dimensional mean vectors. They obtained the change point estimators by taking "argmax" of the corresponding partial sum processes with an estimation error rate of $O_p\big(\log(p)/(n\|\boldsymbol{\Delta}\|_{\min}^2)\big)$, where $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_p)^\top$ is the signal jump of mean vectors before and after the change point and $\|\boldsymbol{\Delta}\|_{\min}$ is the minimum signal jump for the coordinates with a change point. Different from Jirak (2015); Yu and Chen (2021), we adopt a different proof technique and derive an estimation error bound of $O_p\big(\log(p)/(n\|\boldsymbol{\Delta}\|_{\infty}^2)\big)$. Considering $\|\boldsymbol{\Delta}\|_\infty$ can be much larger than $\|\boldsymbol{\Delta}\|_{\min}$, our result is sharper than Jirak (2015); Yu and Chen (2021). More proof details can be found in the Appendix.

After analyzing the change point identification, we next consider the change point detection. Note that for the change point problem, variance estimation under the alternative is a difficult but important task. As pointed out in Shao and Zhang (2010), due to the unknown change point, any improper estimation may lead to non-monotonic power performance. This distinguishes the change point problem substantially from one-sample or two-sample tests where homogenous data are used to construct consistent variance estimation.

Theorem 4 shows that the pooled weighted variance estimation is uniformly consistent under $\mathbf{H}_{1,\mathcal{G}}$. This guarantees that our new testing method has reasonable power performance.

Theorem 4. *Suppose Assumptions* ($\mathbf{A.1}$) - ($\mathbf{A.6}$) *hold. Then, for the weighted variance estimation, under* $\mathbf{H}_{1,\mathcal{G}}$*, we have*

$$\max_{1\le j,k\le p} |\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,k} - \sigma_\epsilon^2 \omega_{j,k}| = o_p(1), \ \ as \ \ n,p \to \infty. \tag{3.27}$$

From the proof of Theorem 4, some interesting observations can be found. On one hand, if the signal strength is too weak such that $\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} = O(\sqrt{\log(pn)/n})$ holds, then the pooled weighted variance estimator $\widehat{\sigma}_\epsilon^2$ is a consistent estimator for σ_ϵ^2 even though we can not guarantee a consistent change point estimator in this case. On the other hand, if the signal strength is big enough such that $\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} \gg \sqrt{\log(pn)/n}$ holds, then a consistent change point estimator $\widehat{t}_{0,\mathcal{G}}$ is needed to guarantee (3.27) holds. These are insightful findings for variance estimation in change point analysis, which is different from the *i.i.d.* case.

Lastly, we discuss the power properties. To this end, we need some additional notations. Recall $\Pi = \{j : \beta_j^{(1)} \neq \beta_j^{(2)}\}$ as the set of coordinates having a change point. Define the oracle signal to noise ratio vector $\boldsymbol{D} = (D_1, \ldots, D_p)^\top$ with

$$D_j := \begin{cases} 0, & \text{for} \quad j \in \Pi^c \\ \left| \dfrac{t_0(1-t_0)(\beta_j^{(2)} - \beta_j^{(1)})}{(\sigma_\epsilon^2 \omega_{j,j})^{1/2}} \right|, & \text{for} \quad j \in \Pi. \end{cases} \tag{3.28}$$

With the above notations and some regularity conditions, the following Theorem 5 shows that we can reject the null hypothesis of no change points with overwhelming probability.

Theorem 5. *Suppose Assumptions* $(\mathbf{A.1})$ – $(\mathbf{A.6})$ *hold. Let* $\epsilon_n = o(1)$*. For any given subgroup* $\mathcal{G} \subset \{1, \ldots, p\}$*, if* $\boldsymbol{D}$ *satisfies*

$$\sqrt{n} \big\| \boldsymbol{D} \big\|_{\mathcal{G},\infty} \geq \frac{C_0}{1-\epsilon_n} \Big(\sqrt{2\log(|\mathcal{G}|n)} + \sqrt{2\log(\alpha^{-1})} \Big), \tag{3.29}$$

under $\mathbf{H}_{1,\mathcal{G}}$*, we have* $\mathbb{P}(\Phi_{\mathcal{G},\alpha} = 1) \to 1$*, as* $n, p, B \to \infty$*, where* C_0 *is a large enough universal positive constant not depending on* n *or* p*.*

Theorem 5 demonstrates that with probability tending to one, our proposed new test can detect the existence of a change point for any given subgroup as long as the corresponding signal to noise ratio satisfies (3.29). Combining (3.28) and (3.29), we note that with a larger signal jump, a smaller noise level, and a closer change point location to the middle of data observations, it is more likely to trigger a rejection of the null hypothesis.

Lastly, we would like to point out that the requirements for identifying and detecting a change point are different. More specifically, from Theorem 3, to correctly identify the location of a change point with desirable accuracy, the signal strength should at least satisfy $\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} \gg \sqrt{\log(|\mathcal{G}|n)/n}$. In contrast, Theorem 5 shows that it is sufficient to detect a change point if $\|\boldsymbol{D}\|_{\mathcal{G},\infty} \geq C\sqrt{\log(|\mathcal{G}|n)/n}$ holds. Hence, we need more stringent conditions for locating a change point than detecting its existence.

4. Numerical studies

We examine the numerical performance of our proposed method and compare it with several existing state-of-art techniques.

We first consider single change point detection. For the design matrix $\mathbf{X}$, we generate $\boldsymbol{X}_i$ ($i.i.d.$) from $N(\mathbf{0}, \boldsymbol{\Sigma})$, where the following two types of covariance structures are investigated: $\boldsymbol{\Sigma} = \mathbf{I}_{p\times p}$ (**Model 1**) and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^*$ with $\boldsymbol{\Sigma}^* = (\sigma^*)_{i,j=1}^p$, where $\sigma^*_{i,j} = 0.5^{|i-j|}$ for $1 \leq i, j \leq p$ (**Model 2**).

To show the bootstrap performance, for each model, the error terms $(\epsilon_i)_{i=1}^n$ are $i.i.d.$ generated from standard normal distributions, standardized Gamma$(4, 1)$ distributions as well as Student's t_5 distributions.

For the regression coefficient $\boldsymbol{\beta}^{(1)}$, for each replication, we generate s non-zero covariates randomly selected from $\mathcal{S} = \{1, \ldots, 50\}$. The corresponding selected coefficients are *i.i.d.* from $\mathrm{U}(0,2)$, and the remaining $p-s$ covariates are 0's. Note that we generate regression coefficients out of $\mathcal{S}$, which is denoted as the active set. Under $\mathbf{H}_{0,\mathcal{G}}$, we set $\boldsymbol{\beta}^{(2)} = \boldsymbol{\beta}^{(1)}$. Throughout the simulations, we consider various combinations of the sample sizes n, data dimensions p, and overall sparsities s by setting $n \in \{200, 300\}$, $p \in \{100, 200, 300, 400\}$ and $s \in \{5, 10\}$. The number of bootstrap replications is $B = 100$. Without additional specifications, all numerical results are based on 2000 replications.

4.1 Empirical sizes

We investigate the empirical sizes. We set the significance levels $\alpha = 1\%, 5\%$. Furthermore, three different types of subgroups are investigated: $\mathcal{G} = \mathcal{S}$, $\mathcal{G} = \mathcal{S}^c$, and $\mathcal{G} = \mathcal{S} \cup \mathcal{S}^c = \{1, \ldots, p\}$. To evaluate the numerical performance, in addition to our proposed methods, we consider four existing well-known techniques for change point detection of high dimensional linear models: the high dimensional lasso-based method in Lee et al. (2016) (Lee2016), the sparse group lasso-based method in Zhang et al. (2015) (SGL), the binary segmentation-based method in Leonardi and Bühlmann (2016) (L&B), and the Variance Projected Wild Binary Segmentation in Wang et al. (2021) (VPWBS).

It is worth noting that under $\mathbf{H}_{0,\mathcal{G}}$ with $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$, SGL and L&B can potentially select the true homogeneous model by identifying the change points at $\{1, n\}$. Hence, we record their rates of false selections as their "empirical sizes". As for Lee2016, their main purpose is to simultaneously estimate the potential single change point as well as the regression coefficients. Therefore, we do not report their empirical sizes and powers here.

Table 1 summarizes the empirical sizes for Models 1 and 2 with different combinations of (n, p, s) under $N(0, 1)$ distributions. We can see that both SGL and L&B are only applicable for the case of the overall subset with $\mathcal{G} = \{1, \ldots, p\}$. In those cases, SGL suffers from serious size distortions with too many false selections. One reasonable explanation is that SGL builds their algorithms on the sparse group lasso which tends to overestimate the number of change points. Moreover, we observe that L&B seems to be conservative although it can select the homogenous model with no false selections. As for our proposed methods, the empirical sizes of Boot-I are out of control (especially for the active set $\mathcal{S}$). This suggests that for change point detection of high dimensional linear models, the residual term of the de-biased lasso-based process can not be ignored, even though it is asymptotically negligible in theory. As compared to Boot-I, Boot-II benefits from bootstrapping the whole de-biased lasso-based process. In

Table 1: Empirical sizes for Models 1 and 2. The errors are generated from $N(0,1)$. The results are based on 2000 replications.

Empirical sizes (%) with $(n,s)=(200,5)$								
Model	$\mathcal{G}$	p	Boot-I ($\alpha=1\%$)	Boot-II ($\alpha=1\%$)	Boot-I ($\alpha=5\%$)	Boot-II ($\alpha=5\%$)	SGL	L&B
$\mathbf{\Sigma}=\mathbf{I}$	$\mathcal{S}$	200	7.61	1.70	18.52	3.86	NA	NA
		400	10.70	1.80	23.05	5.30	NA	NA
	$\mathcal{S}^c$	200	8.23	1.44	15.43	4.06	NA	NA
		400	11.93	0.93	21.60	3.40	NA	NA
	$\mathcal{S}\cup\mathcal{S}^c$	200	7.41	1.03	14.20	2.93	38.89	0.00
		400	12.55	1.39	27.37	3.86	46.67	0.00
$\mathbf{\Sigma}=\mathbf{\Sigma}^*$	$\mathcal{S}$	200	7.61	1.49	14.40	4.73	NA	NA
		400	8.64	1.65	16.26	4.68	NA	NA
	$\mathcal{S}^c$	200	3.50	0.82	12.14	3.09	NA	NA
		400	5.76	0.67	12.76	3.03	NA	NA
	$\mathcal{S}\cup\mathcal{S}^c$	200	4.73	0.82	13.37	3.29	77.78	0.00
		400	7.82	1.23	17.08	3.19	80.00	0.00
Empirical sizes (%) with $(n,s)=(300,10)$								
Model	$\mathcal{G}$	p	Boot-I ($\alpha=1\%$)	Boot-II ($\alpha=1\%$)	Boot-I ($\alpha=5\%$)	Boot-II ($\alpha=5\%$)	SGL	L&B
$\mathbf{\Sigma}=\mathbf{I}$	$\mathcal{S}$	200	12.76	1.83	23.66	3.25	NA	NA
		400	19.55	1.88	33.74	7.35	NA	NA
	$\mathcal{S}^c$	200	8.33	1.02	16.67	3.25	NA	NA
		400	13.79	1.63	26.95	3.06	NA	NA
	$\mathcal{S}\cup\mathcal{S}^c$	200	11.52	0.82	22.43	3.27	56.67	0.00
		400	17.49	2.45	32.30	5.71	62.30	0.00
$\mathbf{\Sigma}=\mathbf{\Sigma}^*$	$\mathcal{S}$	200	10.91	0.62	22.63	2.67	NA	NA
		400	17.07	2.26	28.86	5.56	NA	NA
	$\mathcal{S}^c$	200	4.32	0.41	11.32	1.65	NA	NA
		400	3.66	0.81	10.77	2.44	NA	NA
	$\mathcal{S}\cup\mathcal{S}^c$	200	6.50	1.85	16.06	4.32	56.67	0.00
		400	7.06	0.61	17.57	3.25	55.30	0.00

most cases, the empirical sizes for Boot-II are close to the nominal level across various dimensions and subgroups. Interestingly, it shows that the empirical performance of Boot-II is affected by the candidate subgroups. More specifically, empirical sizes for the active set $\mathcal{S}$ are sometimes larger than the nominal level and the size performance of the non-active set $\mathcal{S}^c$ performs the best among all candidate subgroups. Note that similar findings are also observed in constructing simultaneous confidence intervals in Zhang and Cheng (2017) for the given subgroup $\mathcal{G}$. In addition, we can see that Boot-II can still have satisfactory size performance as the non-zero elements increase slowly from $s = 5$ to $s = 10$.

In the appendix, we report the size performance under standardized Gamma$(4, 1)$ and Student's t_5 distributions in Tables **??** and **??**. In both cases, our proposed method can control the size under the nominal level. This suggests that the bootstrap null distribution is correctly calibrated even for non-normal underlying errors.

4.2 Empirical powers

We next analyze the empirical powers. Denote the signal jump

$$\boldsymbol{\delta} = C\sqrt{\log(p)/n} \times \left(2^3, 2^2, 2^1, 2^0, 2^{-1}\right)^\top .$$

Table 2: Empirical powers (%) under Model 1. The numerical results are based on 2000 replications.

Empirical powers (%) with $\boldsymbol{\delta} = 0.5\sqrt{\log(p)/n} \times (2^3, 2^2, 2^1, 2^0, 2^{-1})$.

			Change point at $k^* = 0.5n$		Change point at $k^* = 0.3n$	
Model	$\mathcal{G}$	p	Boot-II	L&B	Boot-II	L&B
$\mathbf{\Sigma} = \mathbf{I}$	$\mathcal{S}$	200	58.33	NA	36.46	NA
		400	64.93	NA	42.71	NA
	$\mathcal{S}^c$	200	2.08	NA	4.17	NA
		400	3.47	NA	3.82	NA
	$\mathcal{S} \cup \mathcal{S}^c$	200	43.75	0.00	29.17	0.00
		400	40.97	0.00	27.17	0.00

Empirical powers (%) with $\boldsymbol{\delta} = \sqrt{\log(p)/n} \times (2^3, 2^2, 2^1, 2^0, 2^{-1})$.

			Change point at $k^* = 0.5n$		Change point at $k^* = 0.3n$	
Model	$\mathcal{G}$	p	Boot-II	L&B	Boot-II	L&B
$\mathbf{\Sigma} = \mathbf{I}$	$\mathcal{S}$	200	100.00	NA	99.38	NA
		400	99.59	NA	99.38	NA
	$\mathcal{S}^c$	1 200	3.50	NA	3.91	NA
		400	3.09	NA	2.06	NA
	$\mathcal{S} \cup \mathcal{S}^c$	200	100.00	36.87	99.18	29.29
		400	99.38	38.38	99.38	28.28

Empirical powers (%) with $\boldsymbol{\delta} = 2\sqrt{\log(p)/n} \times (2^3, 2^2, 2^1, 2^0, 2^{-1})$.

			Change point at $k^* = 0.5n$		Change point at $k^* = 0.3n$	
Model	$\mathcal{G}$	p	Boot-II	L&B	Boot-II	L&B
$\mathbf{\Sigma} = \mathbf{I}$	$\mathcal{S}$	200	100.00	NA	100.00	NA
		400	100.00	NA	100.00	NA
	$\mathcal{S}^c$	200	2.47	NA	1.65	NA
		400	3.50	NA	2.88	NA
	$\mathcal{S} \cup \mathcal{S}^c$	200	100.00	99.49	100.00	98.48
		400	100.00	100.00	100.00	97.98

We set $n = 200$. We first generate $\boldsymbol{\beta}^{(1)}$ with $s = 5$ non-zero elements following $\mathrm{U}(0, 2)$ distributions out of $\mathcal{S} = \{1, \ldots, 50\}$. Then, we add $\boldsymbol{\delta}$ with $C \in \{0.5, 1, 2\}$ on the corresponding 5 non-zero covariates of $\boldsymbol{\beta}^{(1)}$ to generate $\boldsymbol{\beta}^{(2)}$. Note that in this setting, $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ have a common support.

Tables 2 shows the power results with $n = 200$, where various data dimensions, change point locations, candidate subgroups, and signal strength are considered. Note that we do not report the results of SGL and Boot-I because of their serious size distortions. According to Table 2, we see that our proposed method can detect a change point with a very high probability across various data dimensions when the candidate subgroup has a change point ($\mathcal{G} = \mathcal{S}$ and $\mathcal{G} = \mathcal{S} \cup \mathcal{S}^c$). Interestingly, it is shown that the powers in $\mathcal{S}^c$ are close to the nominal level since the coefficients in $\mathcal{S}^c$ are zeros before and after the change point. As for L&B, we see that it can successfully detect a change point when the signal jump is relatively strong ($C = 2$). However, L&B is not very sensitive to weak signals with $C = 0.5$ and $C = 1$. The above analysis suggests that our proposed method is very powerful to sparse alternatives and is more efficient and flexible than the existing methods for change point detection of high dimensional linear models. Moreover, Table **??** in the appendix shows the power performance similar to Table 2 for Model 2 with banded covariance structures.

In addition to single change point detection, we also combine the proposed new testing method with the binary segmentation technique for detecting multiple change points. The detailed algorithms and numerical performance are provided in the supplementary materials. The results further demonstrate the superior performance of our method over its competitors.

5. Conclusions

In this paper, we propose a new method for change point inference in the context of high dimensional linear models. For any given subgroup $\mathcal{G} \subset \{1, \ldots, p\}$, a L_∞-norm-based test statistic $T_{\mathcal{G}}$ is constructed for testing the homogeneity of regression coefficients across the observations. To approximate its limiting null distribution, a novel multiplier bootstrap procedure is introduced. Our new method is powerful against sparse alternatives with only a few entries in $\mathcal{G}$ having a change point, and allows the group size $|\mathcal{G}|$ to grow exponentially with the sample size n. As for the change point identification, a new change point estimator is obtained by taking "argmax" of the L_∞-aggregated process $H_{\mathcal{G}}(t)$. Theoretically, the change point estimator is shown to be consistent, allowing the overall sparsity s of regression coefficients and the group size $|\mathcal{G}|$ to grow simultaneously with the sample size n. In addition to single change point detection, we further combine our proposed method with the binary segmentation-based technique for detecting and identifying multiple change points. Our new testing method is relatively easy to implement and is justified via extensive numerical studies.

Acknowledgement

The research was supported in part by the National Natural Science Foundation of China 12101132,12331009 (Bin Liu), 11971116 (Zhang), and US National Science Foundation grant DMS-2100729 and National Institute of Health grant R01GM126550 (Y. Liu).

Supplementary Materials

The online supplementary materials provide detailed basic assumptions and proofs of the main theory, and additional numerical results including size, power, multiple change point detection. In addition, an interesting application to the Alzheimer's disease data analysis is also provided.

References

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Bai, Y. and A. Safikhani (2023). A unified framework for change point detection in high-dimensional linear models. *Statistica Sinica 33*, 1–28.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics 37*(4), 1705–1732.

Chan, N. H., C. Y. Yau, and R. Zhang (2014). Group lasso for structural break time series. *Journal of the American Statistical Association 109*(506), 590–599.

Chen, H., H. Ren, F. Yao, and C. Zou (2023). Data-driven selection of the number of change-points via error rate control. *Journal of the American Statistical Association 118*(542), 1415–1428.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics 41*(6), 2786–2819.

Cho, H. and D. Owens (2022). High-dimensional data segmentation in regression settings permitting heavy tails and temporal dependence. *arXiv preprint arXiv:2209.08892*.

He, Z., D. Cheng, and Y. Zhao (2023). Multiple testing of local extrema for detection of structural breaks in piecewise linear models. *arXiv preprint arXiv:2308.04368*.

Horváth, L. (1995). Detecting changes in linear regressions. *Statistics 26*(3), 189–208.

Horváth, L. and Q.-M. Shao (1995). Limit theorems for the union-intersection test. *Journal of Statistical Planning and Inference 44*(2), 133–148.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics 43*(6), 2451–2483.

Kaul, A., S. B. Fotopoulos, V. K. Jandhyala, and A. Safikhani (2021). Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics 15*, 71–134.

Kaul, A., V. K. Jandhyala, and S. B. Fotopoulos (2019). An efficient two step algorithm for high dimensional change point regression models without grid search. *The Journal of Machine Learning Research 20*, 1–40.

Lee, S., M. H. Seo, and Y. Shin (2016). The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(1), 193–210.

Leonardi, F. and P. Bühlmann (2016). Computationally efficient change point detection for high-dimensional regression. *arXiv preprint: 1601.03704*.

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics 34*(3), 1436–1462.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association 55*(290), 324–330.

Shao, X. and X. Zhang (2010). Testing for change points in time series. *Journal of the American Statistical Association 105*(491), 1228–1240.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 58*(1), 267–288.

Van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics 42*(3), 1166–1202.

Vostrikova, L. Y. (1981). Detecting disorder in multidimensional random process. *Soviet Math. Dokl 24*, 55–59.

Wang, D., Z. Zhao, K. Z. Lin, and R. Willett (2021). Statistically and computationally efficient change point localization in regression settings. *The Journal of Machine Learning Research 22*(248), 1–46.

Wang, F., O. Madrid, Y. Yu, and A. Rinaldo (2022). Denoising and change point localisation in piecewise-constant high-dimensional regression coefficients. In *International Conference on Artificial Intelligence and Statistics*, pp. 4309–4338. PMLR.

Xia, Y., T. Cai, and T. T. Cai (2018). Two-sample tests for high-dimensional linear regression with an application to detecting interactions. *Statistica Sinica 28*(1), 63–92.

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*.

Yu, M. and X. Chen (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 83*(2), 247–270.

Zhang, B., J. Geng, and L. Lai (2015). Change-point estimation in high dimensional linear regression models via sparse group lasso. *In: 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 815–821.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(1), 217–242.

Zhang, X. and G. Cheng (2017). Simultaneous inference for high-dimensional linear models. *Journal of the American Statistical Association 112*(518), 757–768.

Department of Statistics and Data Science, School of Management at Fudan University

E-mail: liubin0145@gmail.com

E-mail: xszhang@fudan.edu.cn

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

E-mail: yfliu@email.unc.edu

Supplementary Material to "Simultaneous Change Point Detection and Identification for High Dimensional Linear Models"

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

[1,2]*Department of Statistics and Data Science, School of Management at Fudan University*

[3]*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*

The Appendix provides detailed proofs and additional results of the main paper. In Section S1, we introduce some additional notations. In Section S2, we provide the procedure for detecting multiple change points using our method. In Section S3, we introduce some basic assumptions for deriving the theoretical results. In Section S5, some additional numerical results, including size, power as well as detecting multiple change points, are provided. In Section S6, we apply our method to the Alzheimer's disease data analysis. In Section S7, some useful lemmas are provided. In Section S8, we give the detailed proofs of theoretical results in the main paper. In Sections S9 and S10, we prove the useful lemmas in Section S7 as well as the lemmas used in Section S8.

S1 Some notations

Under $\mathbf{H}_0$, we set $\boldsymbol{\beta}^{(0)} := \boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$ and $s^{(0)} := s^{(1)} = s^{(2)}$. We set $s := s^{(1)} \vee s^{(2)}$. For a given subgroup $\mathcal{G}$, set $\Pi_{\mathcal{G}} = \{j \in \mathcal{G} : \beta_j^{(1)} - \beta_j^{(2)} \neq 0\}$ as the subset of coordinates with a change point. For a vector $\boldsymbol{v} \in \mathbb{R}^p$, we set $\mathcal{M}(\boldsymbol{v})$ as the number of non-zero elements of $\boldsymbol{v}$, i.e. $\mathcal{M}(\boldsymbol{v}) = \sum_{j=1}^{p} \mathbf{1}\{v_j \neq 0\}$. We denote $J(\boldsymbol{v}) = \{1 \leq j \leq p : v_j \neq 0\}$ as the set 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. Denote $\mathcal{X} = \{\mathbf{X}, \boldsymbol{Y}\}$. We use $C_1, C_2, \ldots$ to denote constants that may vary from line to line.

S2 Extensions to multiple change points

So far, we have proposed new methods for detecting a single change point as well as identifying its location using the argmax based estimator. In this section, we aim to extend our new testing procedure for detecting and identifying multiple change points for high dimensional linear models. In particular, suppose there are m change points $k_1, \ldots, k_m$ that divide the linear structures into $m+1$ segments with different regression coefficients:

$$\begin{cases} Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(1)} + \epsilon_i, & \text{for } i = 1, \dots, k_1, \\ Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(2)} + \epsilon_i, & \text{for } i = k_1 + 1, \dots, k_2, \\ \quad\quad \vdots & \\ Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(m)} + \epsilon_i, & \text{for } i = k_{m-1} + 1, \dots, k_m, \\ Y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta}^{(m+1)} + \epsilon_i, & \text{for } i = k_m + 1, \dots, n. \end{cases} \tag{S2.1}$$

Based on Model (S2.1), for any given subgroup $\mathcal{G} \subset \{1, \dots, p\}$, in the case of multiple change points, we consider the following hypothesis:

$$\begin{aligned} &\mathbf{H}'_{0,\mathcal{G}} : \beta_s^{(1)} = \beta_s^{(2)} = \cdots = \beta_s^{(m)} = \beta_s^{(m+1)} \ \text{ for all } s \in \mathcal{G} \ \text{ v.s.} \\ &\mathbf{H}'_{1,\mathcal{G}} : \text{There exist } s \in \mathcal{G} \text{ and at least one } j^* \in \{1, \dots, m\} \text{ s.t. } \beta_s^{(j^*)} \neq \beta_s^{(j^*+1)}. \end{aligned} \tag{S2.2}$$

To solve Problem (S2.2), we combine our bootstrap-based new testing procedure with the well-known binary segmentation technique (Vostrikova, 1981) to simultaneously detect and identify multiple change points. More specifically, for each candidate search interval (s, e), we detect the existence of a change point. If $\mathbf{H}_{0,\mathcal{G}}$ is rejected, we identify the new change point b by taking the argmax in (2.18). Then the interval (s, e) is split into two subintervals (s, b) and (b, e) and we conduct the above procedure on (s, b) and (b, e) separately. This algorithm is stopped until no subinterval can detect a change point. Algorithm S2.1 describes our bootstrap-based multiple change point testing procedure. It is demonstrated by our numerical studies that Algorithm S2.1 can automatically account for the data generating mechanism and simultaneously detect and identify multiple change points, which enjoys better performance than existing techniques.

Algorithm S2.1 : Bootstrap-based binary segmentation procedure for multiple change point detection in high dimensional linear regression models.

Input: Given the data set $\{\mathbf{X}, \boldsymbol{Y}\}$, set the value for τ_0, the number of bootstrap replications B, and the subset $\mathcal{G}$.

Step 1: Initialize the set of change point pairs $\mathcal{T} = \{0, 1\}$.

Step 2: For each pair $\{s, e\}$ in $\mathcal{T}$, detect the existence of a change point. If $\mathbf{H}_{0,\mathcal{G}}$ is rejected, identify the new change point b by taking the argmax in (2.18). Then add new pairs of nodes $\{s, b\}$ and $\{b, e\}$ to $\mathcal{T}$ and update $\mathcal{T}$ as $\mathcal{T} = \mathcal{T} \cup \{s, b\} \cup \{b, e\}$.

Step 3: Repeat Step 2 until no more new pair of nodes can be added. Denote the terminal set of change point pairs by $\mathcal{T}_{\text{final}} = \cup_{i=1}^{\widehat{m}+1}\{\widehat{t}_{i-1}, \widehat{t}_i\}$.

Output: Algorithm S2.1 provides the change point estimator $\widehat{\boldsymbol{t}} = (\widehat{t}_0, ..., \widehat{t}_{\widehat{m}+1})^\top$, where $\widehat{m} = \#\mathcal{T}_{\text{final}} - 1$ and $0 = \widehat{t}_0 < \widehat{t}_1 < ... < \widehat{t}_{\widehat{m}} < \widehat{t}_{\widehat{m}+1} = 1$, including the number and locations.

S3 Basic assumptions

We introduce some basic assumptions for making change point inference on high dimensional linear models. Assumption $(\mathbf{A.1})$ is a basic requirement for the change point location. Assumptions (**A**.**1**) – (**A**.**3**) impose some regular conditions on the design matrix as well as the error terms. Assumption (**A**.**4**) contains basic requirements on model parameters. Assumption (**A**.**5**) is a technical condition on the regularity parameters in (2.5) and (2.9).

Before giving the assumptions, we introduce the concept of the restricted eigenvalue (RE) and uniform restricted eigenvalue (URE) conditions.

Definition 1. (Restricted eigenvalue RE$(s_j, 3)$). For integers s_j such that $1 \leq s_j \leq p-1$, a set of indices $J_0 \subset \{1, \ldots, p-1\}$ with $|J_0| \leq s_j$, define

$$\mathcal{R}^{(j)}(s_j, 3) = \min_{\substack{J_0 \subset \{1,\ldots,p-1\} \\ |J_0| \leq s_j}} \min_{\substack{\boldsymbol{\delta} \neq 0 \\ \|\boldsymbol{\delta}_{J_0^c}\|_1 \leq 3\|\boldsymbol{\delta}_{J_0}\|_1}} \frac{\|\mathbf{X}^{-j}\boldsymbol{\delta}\|_2}{\sqrt{n}\|\boldsymbol{\delta}_{J_0}\|_2}, \text{ with } 1 \leq j \leq p, \quad \text{(S3.3)}$$

where $\mathbf{X}^{-j} \in \mathbb{R}^{n\times(p-1)}$ is a submatrix of $\mathbf{X}$ with the j-th column being removed, and $\boldsymbol{\delta}_{J_0}$ is the vector that has the same coordinates as $\boldsymbol{\delta}$ on J_0 and zero coordinates on the complement J_0^c of J_0.

Definition 2. (Uniform restricted eigenvalue URE$(s, 3, \mathbb{T})$). For integers s such that $1 \leq s \leq p$, a set of indices $J_0 \subset \{1, \ldots, p\}$ with $|J_0| \leq s$, and $\mathbb{T} = [\tau_0, 1-\tau_0]$, define

$$\mathcal{R}_1(s,3,\mathbb{T}) = \min_{t\in\mathbb{T}} \min_{\substack{J_0\subset\{1,\ldots,p\}\\|J_0|\leq s}} \min_{\substack{\boldsymbol{\delta}\neq 0\\ \|\boldsymbol{\delta}_{J_0^c}\|_1\leq 3\|\boldsymbol{\delta}_{J_0}\|_1}} \frac{\|\mathbf{X}_{(0,t)}\boldsymbol{\delta}\|_2}{\sqrt{\lfloor nt\rfloor}\|\boldsymbol{\delta}_{J_0}\|_2}. \tag{S3.4}$$

and

$$\mathcal{R}_2(s,3,\mathbb{T}) = \min_{t\in\mathbb{T}} \min_{\substack{J_0\subset\{1,\ldots,p\}\\|J_0|\leq s}} \min_{\substack{\boldsymbol{\delta}\neq 0\\ \|\boldsymbol{\delta}_{J_0^c}\|_1\leq 3\|\boldsymbol{\delta}_{J_0}\|_1}} \frac{\|\mathbf{X}_{(t,1)}\boldsymbol{\delta}\|_2}{\sqrt{\lfloor nt\rfloor^*}\|\boldsymbol{\delta}_{J_0}\|_2}. \tag{S3.5}$$

Note that Definition 1 is similar to the RE conditions introduced in Bickel et al. (2009) and is mainly used for the node-wise lasso estimators. It is well-known that the RE conditions are among the weakest assumptions on the design matrix and are important for deriving the estimation error bounds of the lasso solutions. See Raskutti et al. (2010); Van De Geer and Bühlmann (2009). Moreover, our testing procedure needs to calculate $\widehat{\boldsymbol{\beta}}^{(0,t)}$ and $\widehat{\boldsymbol{\beta}}^{(t,1)}$ as in (2.9). For each search location $t \in [\tau_0, 1-\tau_0]$, to guarantee $\widehat{\boldsymbol{\beta}}^{(0,t)}$ and $\widehat{\boldsymbol{\beta}}^{(t,1)}$ enjoy desirable properties toward their population counterpart $\boldsymbol{\beta}^{(0,t)}$ and $\boldsymbol{\beta}^{(t,1)}$, we introduce the uniform restricted eigenvalue condition as in Definition 3, which is an extension of the RE condition. With the above two definitions, we are ready to introduce the assumptions, which are summarized as follows:

Assumption (**A**.**1**) The design matrix $\mathbf{X}$ has *i.i.d.* rows following sub-Gaussian distributions. In other words, there exists a positive constant K such that $\sup_{i,j}\mathbb{E}(\exp(|X_{i,j}|^2/K)) \leq 1$ holds.

Assumption (**A**.**2**) The error terms $\{\epsilon_i\}_{i=1}^n$ are *i.i.d.* sub-Gaussian with finite variance σ_ϵ^2. In other words, there exist positive constants K', c_ϵ and C_ϵ such that $\mathbb{E}(\exp(|\epsilon_i|^2/K')) \leq 1$ and $c_\epsilon \leq \mathrm{Var}(\epsilon_i) \leq C_\epsilon$ hold. Furthermore, ϵ_i is independent with $\boldsymbol{X}_i$ for $i = 1, \ldots, n$.

Assumption (**A**.**3**) Assume that there are positive constants κ_1 and κ_2 such that $\max_j \Sigma_{j,j} < \kappa_1 < \infty$ and $\max_j \|\boldsymbol{\theta}_j\|_2 < \kappa_2 < \infty$ hold, where $\boldsymbol{\theta}_j$ is the j-th row of $\boldsymbol{\Theta} = (\theta_{j,k}) := \boldsymbol{\Sigma}^{-1}$. Moreover, for the RE and URE conditions, we require

$$\min_{1\leq j\leq p} \mathcal{R}^{(j)}(s_j, 3) > \kappa_3, \quad \min\big(\mathcal{R}_1(s, 3, \mathbb{T}), \mathcal{R}_2(s, 3, \mathbb{T})\big) > \kappa_4 \tag{S3.6}$$

for some $\kappa_3, \kappa_4 > 0$, where $s_j := |\{1 \leq k \leq p : \theta_{j,k} \neq 0, k \neq j\}|$.

Assumption (**A**.**4**) For the change point model in (2.3), we assume the following:

(a) Assume that $\log(pn) = O(\lfloor n\tau_0 \rfloor^\zeta)$ holds for some $0 < \zeta < 1/7$;

(b)We assume $\lfloor n\tau_0 \rfloor \to \infty$, $\max_{1\leq j\leq p} s_j \dfrac{\log(pn)}{\sqrt{n}} \to 0$ and $s\sqrt{n}\dfrac{\log(pn)}{\lfloor n\tau_0 \rfloor} \to 0$ as $n, p \to \infty$, where $s := s^{(1)} \vee s^{(2)}$;

(c) There exists some constant $\gamma \in (0, 1]$ such that $|\mathcal{G}| = p^\gamma$.

Assumption (**A**.**5**) For the node-wise regression in (2.5), we require the regularization parameter $\lambda_{(j)} \asymp \sqrt{\log(p)/n}$ uniformly in j. For the lasso-based estimators in (2.9), we require

$$\lambda_1(t) \asymp \sqrt{\frac{\log(p)}{\lfloor nt \rfloor}}, \quad \lambda_2(t) \asymp \sqrt{\frac{\log(p)}{\lfloor nt \rfloor^*}}, \text{ for } t \in [\tau_0, 1-\tau_0]. \tag{S3.7}$$

Assumptions (**A.1**) – (**A.3**) are relatively weak conditions on the covariates and error terms. In particular, they require that $\{\boldsymbol{X}_i\}_{i=1}^n$ and $\{\epsilon_i\}_{i=1}^n$ are sub-Gaussian distributed with "well-behaved" sample covariance matrix and non-degenerate variances σ_ϵ^2, which covers a wide broad of distributional patterns and has been commonly adopted in high dimensional data analysis. Assumption (**A.4**) specifies the scaling relationships among parameters ($\{s, s_j, n, p, |\mathcal{G}|\}$) in Model (2.3). More specifically, (a) allows the number of variables (p) can grow exponentially with the number of data observations (n) as long as $\log(pn) = O(\lfloor n\tau_0 \rfloor^\zeta)$ holds; (b) allows the number of active variables (s and s_j) can go to infinity if $\max_{1\le j\le p} s_j \frac{\log(pn)}{\sqrt{n}} \to 0$ and $s\sqrt{n}\frac{\log(pn)}{\lfloor n\tau_0 \rfloor} \to 0$ holds; (c) demonstrates that we can make change point inference on any large scale subgroup $\mathcal{G}$ with $|\mathcal{G}| = p^\gamma$. Lastly, Assumption (**A.5**) imposes some technical conditions on the regularity parameters of lasso and node-wise lasso, which is important for deriving desirable estimation error bounds of the corresponding estimators. It is worth mentioning that (S3.7) automatically accounts for the heterogeneity of the ℓ_1 regularization problem (2.9) and is consistent with the classical conditions as in Bickel et al. (2009) when the data are homogenous (e.g. $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$).

Lastly, the following Proposition 1 shows that the RE and URE conditions in (S3.6) of Assumption (A.2) hold with high probabilities.

Proposition 1. *(i) For integers s_j such that $1 \leq s_j \leq p-1$, a set of indices $J_0 \subset \{1,\ldots,p-1\}$ with $|J_0| \leq s_j$ and $s_j\sqrt{\log(p)/n} = o(1)$. Under Assumption (A.1), if $\mathbf{\Sigma}$ satisfies*

$$\min_{1\leq j\leq p} \min_{\substack{J_0\subset\{1,\ldots,p-1\}\\ |J_0|\leq s_j}} \min_{\substack{\boldsymbol{\delta}\neq 0\\ \|\boldsymbol{\delta}_{J_0^c}\|_1\leq 3\|\boldsymbol{\delta}_{J_0}\|_1}} \frac{\|\mathbf{\Sigma}^{-j,-j}\boldsymbol{\delta}\|_2}{\|\boldsymbol{\delta}_{J_0}\|} \geq 2\kappa_3, \tag{S3.8}$$

for some $\kappa_3 > 0$, then we have:

$$\mathbb{P}(\min_{1\leq j\leq p} \mathcal{R}^{(j)}(s_j,3) > \kappa_3) \geq 1 - C_1(np)^{-C_2},$$

where $\mathbf{\Sigma}^{-j,-j} := \mathbb{E}[\mathbf{X}^{-j}(\mathbf{X}^{-j})^\top/n]$ and C_1, C_2 are universal positive constants not depending on n or p. (ii) Similarly, for integers s such that $1 \leq s \leq p$, a set of indices $J_0 \subset \{1,\ldots,p\}$ with $|J_0| \leq s$ and $s\sqrt{\log(p)/\lfloor n\tau_0\rfloor} = o(1)$. Under Assumption (A.1), if $\mathbf{\Sigma}$ satisfies

$$\min_{\substack{J_0\subset\{1,\ldots,p\}\\ |J_0|\leq s}} \min_{\substack{\boldsymbol{\delta}\neq 0\\ \|\boldsymbol{\delta}_{J_0^c}\|_1\leq 3\|\boldsymbol{\delta}_{J_0}\|_1}} \frac{\|\mathbf{\Sigma}\boldsymbol{\delta}\|_2}{\|\boldsymbol{\delta}_{J_0}\|_2} \geq 2\kappa_4, \tag{S3.9}$$

for some $\kappa_4 > 0$, then we have

$$\mathbb{P}(\min\big(\mathcal{R}_1(s,3,\mathbb{T}),\mathcal{R}_2(s,3,\mathbb{T})\big) > \kappa_4) \geq 1 - C_3(np)^{-C_4},$$

where $C_3, C_4 > 0$ are some universal constants not depending on n or p.

Remark 1. The proof of Proposition 1 is given in the Appendix. A sufficient condition for both (S3.8) and (S3.9) hold is $\lambda_{\min}(\mathbf{\Sigma}) > c$ for some

$c > 0$, where $\lambda_{\min}(\boldsymbol{\Sigma})$ is the smallest eigenvalue of $\boldsymbol{\Sigma}$. Note that the smallest eigenvalue condition is easy to verify and has been widely used in the literature such as Kaul et al. (2019); Wang et al. (2021) for change point analysis of high dimensional linear models. For example, many commonly used covariance matrices such as Toeplitz matrices, blocked diagonal matrices have positive smallest eigenvalue values.

S4 Connection with the existing methods

In this section, we compare our proposed methodology and theorems with several related papers. He et al. (2023) considered multiple testing of local extrema for detection of change points in piecewise linear models. Note that He et al. (2023) essentially studied the change point for the mean function of the univariate Gaussian process while we considered high dimensional linear models. Wang et al. (2022) studied the theoretical properties of the fused lasso procedure in the context of a linear regression model in which the regression coefficients are totally ordered and assumed to be sparse and piecewise constant. It is worth mentioning that although Wang et al. (2022) also assumed that the regression coefficients are piecewise constants, the 'piecewise' here refers to the values of the regression coefficients being piecewise over the coordinate components, with the regression coefficients remaining constant as a whole throughout the entire sample observation process. This is very different from the model we consider, because we assume that the p-dimensional regression coefficients as a whole are piecewise constants with respect to the observation time t. Kaul et al. (2021); Xu et al. (2022) respectively considered the problem of change point inference for ultra-high dimensional mean vector-based models and linear regression models. However, both of these papers focus on constructing corresponding confidence intervals for the unknown change point locations, rather than the problem of change point testing considered in this paper. Lastly, Cho and Owens (2022); Bai and Safikhani (2023) respectively constructed estimates for the location and number of multiple change points in ultra-high-dimensional regression models based on methods of moving window and blocked fused lasso. Unlike these two methods, this paper is primarily concerned with the change point testing problem in regression models. Therefore, this paper needs to construct the debiased lasso based testing statistics to adopt the Gaussian approximation method, so that the testing procedure controls the type I error while maintaining high detection power. It is worth mentioning that the signal-to-noise ratio condition required for change point detection derived in this paper is weaker than that required for the above-mentioned change point estimation methods. Specifically, according to the results of Cho and Owens (2022); Bai and Safikhani (2023), to correctly identify the location of a change point with desirable accuracy, the signal strength should at least satisfy $\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_\infty \gg \sqrt{\log(pn)/n}$. In contrast, our theorem shows that it is sufficient to detect a change point if $\left\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\right\|_\infty \geq C\sqrt{\log(pn)/n}$ holds. Hence, we need more stringent conditions for locating a change point than detecting its existence. We believe that the aforementioned results should be a noteworthy point in change point detection for high-dimensional linear regression models.

S5 Additional numerical results

S5.1 Implementations of the existing techniques

Before reporting additional numerical results, we first demonstrate how to implement the mentioned techniques in this paper.

Implementation of the existing methods: For Lee2016, we use the **package-glmnet** to implement their proposed algorithm. Note that Lee2016 involves a selection of the tuning parameter λ. For each replication, we generate a sequence from 2^{-5} to 2^5 and select the "best" λ by 10-fold cross-validation. For L&B, we use the binary segmentation-based method with parameters suggested by the authors using the **package-glmnet**. Moreover, we use a three folded cross-validation procedure to select the tuning parameters in L&B. For VPWBS, we implement the algorithm using the codes provided by the authors at GitHub (https://github.com/darenwang/VPBS). For SGL, we use the **package-SGL** with parameters in favor of their method and use three folded cross-validation to select the tuning parameters. Note that SGL solves the following optimization problem:

$$\begin{aligned}
&\{\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_n\} \\
&= \underset{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_n \in \mathbb{R}^p}{\arg\min} \sum_{i=1}^n (Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}_i)^2 + \lambda_n \alpha \sum_{i=1}^n \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i-1}\|_2 + \lambda_n (1-\alpha) \sum_{i=1}^n \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i-1}\|_1.
\end{aligned}$$

Based on the above optimization, SGL finds a change point at i^* if $\widehat{\boldsymbol{\beta}}_{i^*} - \widehat{\boldsymbol{\beta}}_{i^*-1} \neq \mathbf{0}$. It is well-known that lasso tends to over select the variables. In addition, SGL essentially solves a group lasso problem by calculating $n \times p$ parameters using only n observations. As a result, SGL may yield false alarms by identifying some $\{i : \boldsymbol{\beta}_i - \boldsymbol{\beta}_{i-1} = \mathbf{0}_p\}$ as a change point. This can be seen by our following empirical size performance in Section 4.1 as well as the multiple change point detection results in Section S5.4. Moreover, we note that this phenomenon was also observed by Wang et al. (2021).

Implementation of our method: As for our proposed method, we use the **package-hdi** to obtain the node-wise lasso estimator $\widehat{\boldsymbol{\Theta}}$. Note that the calculation of the lasso processes $\widehat{\boldsymbol{\beta}}^{(0,t)}$ and $\widehat{\boldsymbol{\beta}}^{(t,1)}$ with $t \in [\tau_0, 1 -$

$\tau_0]$ involves the selection of tuning parameters $\lambda_1(t)$ and $\lambda_2(t)$ defined in (2.9). We select the tuning parameters via three folded cross-validation. Specifically, for each search location $t \in [\tau_0, 1-\tau_0]$, we set

$$\lambda_1(t) = C\sqrt{\frac{\log(p)}{\lfloor nt \rfloor}}, \quad \text{and} \quad \lambda_2(t) = C\sqrt{\frac{\log(p)}{\lfloor nt \rfloor^*}}, \quad \text{with } C \in \{1, 2, \ldots, 8\}.$$

Then, we use the **package-glmnet** to select the best "C" via three folded cross-validation, which enjoys satisfactory performance in change point detection and identification.

S5.2 Additional size performance

In addition to $N(0,1)$, we also report the size performance under standardized Gamma$(4,1)$ (Table S5.1) and Student's t_5 (Table S5.2) distributions which have very similar performance to Table 1 of the main paper. In this case, our proposed method can control the size under the nominal level. This suggests that the bootstrap null distribution is correctly calibrated even for non-normal underlying errors.

S5.3 Additional power performance

Table S5.3 shows the power performance for Model 2 with banded covariance structures of $\boldsymbol{X}$, which is similar to Table 2 in the main paper.

S5.4 Multiple change point detection

So far, we have considered the numerical performance of single change point detection and identification. Next, we investigate multiple change points detection for Problem (S2.2). In this numerical study, we consider two model settings:

Case 2: Alternatives with three change points. In this case, we set $n = 600$ and $p = 200$ with three change points at $k_1 = 180$, $k_2 = 300$, and $k_3 = 420$, respectively. The above three change points divide the data into four segments with different regression coefficients: $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(2)}$, $\boldsymbol{\beta}^{(3)}$, and $\boldsymbol{\beta}^{(4)}$. We first generate $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$. The generating mechanism for $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ is the same as Case 1 in the single change point setting except that we use a signal jump

$$\boldsymbol{\delta}' = C\sqrt{\frac{\log(p)}{n}}\Big(2^4, 2^3, 2^2, 2^1, 2^0\Big)^\top.$$

Then, we set $\boldsymbol{\beta}^{(3)} = \boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(4)} = \boldsymbol{\beta}^{(2)}$. In this case, we set $C \in \{1.5, 3\}$.

Case 3: Alternatives with four change points. In this case, we set $n = 1000$ and $p = 200$ with four change points at $k_1 = 300$, $k_2 = 450$, $k_3 = 550$, and $k_4 = 700$, respectively. The above four change points divide the data into five segments with different regression coefficients: $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(5)}$. We first generate $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ as introduced in Case 2. Then, we set

$\boldsymbol{\beta}^{(3)} = \boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(4)} = \boldsymbol{\beta}^{(2)}$ and $\boldsymbol{\beta}^{(5)} = \boldsymbol{\beta}^{(1)}$. In this case, we set $C \in \{2, 4\}$.

We use Algorithm S2.1 to detect and identify multiple change points and compare our methods with SGL, L&B, and VPWBS. Note that Lee2016 is not applicable here because they only considered single change point detection. Moreover, to evaluate their performance, we report the mean for the number of identified change points (Mean) and the mean adjusted Rand index between the identified change points and the true change points (Adj.Rand) as well as its standard deviations (Sd.Adj.Rand). Note that the adjusted Rand index with a value belonging to $[-1, 1]$ is well adopted for measuring the similarity between two data clusterings. The adjusted Rand index with a value being one means that the data clusterings are exactly the same. The results are reported in Table S5.4. For detecting the number of multiple change points, SGL tends to overestimate the numbers across all model settings. This is consistent with our numerical studies in the size control in Section 4.1. For L&B, it has satisfactory performance when the signal jump is strong with $C = 3$ or $C = 4$. However, L&B fails to detect all relevant three or four change points when the signal-to-noise ratio is low by setting $C = 1.5$ or $C = 2$. This suggests that L&B is not very sensitive to weak signals and this observation is consistent with our previous power analysis in Section 4.2. As for our proposed method, it can correctly detect the three (or four) change points on average even for a small signal jump. For identifying the change point locations, VPWBS has better performance than L&B when the signal is weak and L&B becomes very competitive as the signal becomes stronger. In most cases, the Arg-max based methods can estimate the locations with high accuracy and have better performance than their competitors. This is supported by the high Adj.Rand. Finally, we would like to point out that for all methods, their performance becomes better when the model has a stronger signal jump.

In summary, as compared to the existing works, our bootstrap-assistant method is more efficient and accurate for detecting and identifying multiple change points. Moreover, it is able to detect the structural changes for any given subgroup and is very flexible to use.

S5.5 Computational cost

In this section, we compare the computational cost of the existing methods. In theory, for detecting a single change point, the computational costs for the existing methods are $O(n\text{Lasso}(n,p))$ (Lee2016), $O(n\text{Lasso}(n,p))$ (L&B), $O(Mn\text{GroupLasso}(n,p))$ (VPWBS), $O(\text{GroupLasso}(n,np))$ (SGL), and $(B+1)O(n\text{Lasso}(n,p))$ (our proposed method), where $\text{Lasso}(n,p)$ and $\text{GroupLasso}(n,p)$ denote the computational cost for solving lasso and group lasso problems with the sample size n and the data dimension p, M is the number of random intervals in Wang et al. (2021), and B is the number of bootstrap replications. Empirically, we implement the corresponding program independently on a CPU (Linux) with 2.50GHz and 256G RAM and report the average computational time (seconds) based on 5 replications. Note that the computational cost for our proposed method mainly relies on the bootstrap procedure which can be time-consuming. Since the B bootstrap replications can be done separately, we can use parallel computation in modern computer techniques to further reduce the computational time via implementing the B bootstrap replications in a parallel fashion on different cores of the Linux server. Specifically, for our method, we report the computational cost by using 8, 16, and 32 logical cores, respectively. Figure S5.1 reports the computational time for the existing methods with various $n \in \{200, 400, 600, 800, 1000\}$ (upper) and $p \in \{100, 200, 300, 400\}$ (bottom). In general, Lee2016 and L&B are the most efficient and have very close performance. The computational time for SGL is the most expensive among all methods. For our proposed algorithm, we can see that it has a tolerable computational cost and can even be comparable to its competitors using more cores. Lastly, Figure S5.1 shows that for all methods, the computational time grows linearly with n and p, and it appears that the computational cost is more sensitive to the growth of the sample size n than the data dimension p.

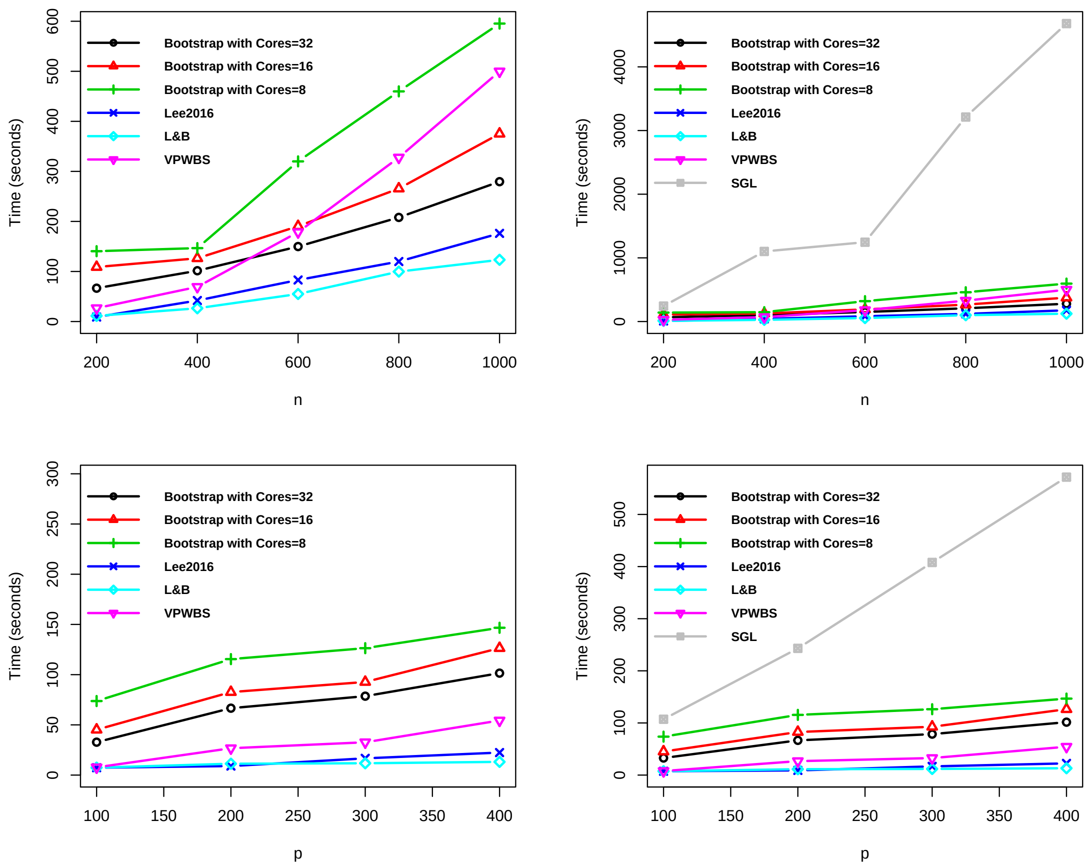

Figure S5.1: Computational time (seconds) for the existing methods based on an average of 5 replications. Upper: Computational time for $p = 200$ and $n \in \{200, 400, 600, 800, 1000\}$ without the plot of SGL (left) and with SGL (right), respectively. Bottom: Computational time for $n = 200$ and $p \in \{100, 200, 300, 400\}$ without the plot of SGL (left) and with SGL (right), respectively.

Table S5.1: Empirical sizes for **Models 1 and 2** under various combinations of (n, p, s). The errors are generated from standardized **Gamma(4,1)** distributions. The results are based on 2000 replications.

			Empirical sizes (%) for Gamma(4,1) with $(n, s) = (200, 5)$					
Model	$\mathcal{G}$	p	Boot-I ($\alpha = 1\%$)	Boot-II ($\alpha = 1\%$)	Boot-I ($\alpha = 5\%$)	Boot-II ($\alpha = 5\%$)	SGL	L&B
$\mathbf{\Sigma} = \mathbf{I}$	$\mathcal{S}$	100	7.00	1.70	14.81	4.63	NA	NA
		200	8.64	1.29	17.70	4.32	NA	NA
		300	9.67	2.11	16.67	5.14	NA	NA
		400	13.99	1.80	23.66	5.14	NA	NA
	$\mathcal{S}^c$	100	4.32	0.98	9.67	3.60	NA	NA
		200	6.38	1.23	15.02	3.81	NA	NA
		300	11.11	1.08	20.99	3.86	NA	NA
		400	13.58	1.80	24.90	4.27	NA	NA
	$\mathcal{S} \cup \mathcal{S}^c$	100	6.17	1.49	15.43	4.73	56.67	0.00
		200	9.05	1.54	17.28	3.96	43.33	0.00
		300	10.91	1.44	23.25	4.22	40.00	0.00
		400	18.31	2.11	30.66	4.94	40.00	0.00
$\mathbf{\Sigma} = \mathbf{\Sigma}^*$	$\mathcal{S}$	100	4.94	1.92	11.73	4.87	NA	NA
		200	6.79	1.58	15.64	4.46	NA	NA
		300	8.23	2.12	17.90	5.81	NA	NA
		400	12.55	2.06	24.07	4.65	NA	NA
	$\mathcal{S}^c$	100	3.91	1.44	10.08	4.03	NA	NA
		200	3.70	1.57	10.29	3.82	NA	NA
		300	7.61	1.30	14.61	3.69	NA	NA
		400	4.73	0.89	15.84	2.73	NA	NA
	$\mathcal{S} \cup \mathcal{S}^c$	100	8.64	1.36	16.87	3.35	51.11	0.00
		200	7.00	1.37	12.96	3.14	40.00	0.00
		300	8.02	1.36	19.55	3.14	50.00	0.00
		400	7.20	1.16	15.02	3.76	37.78	0.00

Table S5.2: Empirical sizes for **Models 1 and 2** under various combinations of (n, p, s). The errors are generated from standardized **Student's** t_5 distributions. The results are based on 2000 replications.

Empirical sizes (%) for Student's t_5 with $(n, s) = (200, 5)$								
Model	$\mathcal{G}$	p	Boot-I ($\alpha = 1\%$)	Boot-II ($\alpha = 1\%$)	Boot-I ($\alpha = 5\%$)	Boot-II ($\alpha = 5\%$)	SGL	L&B
$\mathbf{\Sigma} = \mathbf{I}$	$\mathcal{S}$	100	5.35	1.29	15.23	4.17	NA	NA
		200	9.26	1.95	21.40	5.61	NA	NA
		300	9.05	1.95	20.16	5.30	NA	NA
		400	14.40	2.37	22.84	6.43	NA	NA
	$\mathcal{S}^c$	100	5.97	1.18	10.29	4.22	NA	NA
		200	9.67	1.59	20.99	4.42	NA	NA
		300	10.70	2.16	22.22	4.78	NA	NA
		400	11.93	1.85	21.60	4.48	NA	NA
	$\mathcal{S} \cup \mathcal{S}^c$	100	7.20	1.65	16.05	4.63	61.11	0.00
		200	10.29	1.80	20.78	4.68	45.56	0.00
		300	12.76	1.75	26.13	5.20	50.00	0.00
		400	16.46	2.42	30.45	5.04	54.44	0.00
$\mathbf{\Sigma} = \mathbf{\Sigma}^*$	$\mathcal{S}$	100	6.17	1.33	13.58	3.90	NA	NA
		200	9.05	1.89	18.31	5.38	NA	NA
		300	9.05	2.72	18.31	5.78	NA	NA
		400	10.91	2.04	21.19	5.32	NA	NA
	$\mathcal{S}^c$	100	4.53	1.48	10.29	4.35	NA	NA
		200	3.91	1.64	10.08	4.26	NA	NA
		300	6.79	1.44	14.40	3.65	NA	NA
		400	7.61	1.80	16.46	4.41	NA	NA
	$\mathcal{S} \cup \mathcal{S}^c$	100	6.79	1.64	13.58	5.19	51.11	0.00
		200	5.14	1.59	12.96	4.41	44.44	0.00
		300	9.26	2.10	18.11	4.87	31.11	0.00
		400	9.47	2.05	18.93	4.87	36.67	0.00

Table S5.3: Empirical powers (%) for **Case 1 under Model 2** with various dimensions, candidate subgroups, and change point locations. The sample size is $n = 200$. The significance level is $\alpha = 5\%$. The numerical results are based on 2000 replications.

Empirical powers (%) with $\boldsymbol{\delta} = 0.5\sqrt{\log(p)/n} \times (2^3, 2^2, 2^1, 2^0, 2^{-1})$**.**						
			Change point at $k^* = 0.5n$		Change point at $k^* = 0.3n$	
Model	$\mathcal{G}$	p	Boot-II	L&B	Boot-II	L&B
$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^*$	$\mathcal{S}$	200	49.33	NA	30.27	NA
		400	45.33	NA	33.33	NA
	$\mathcal{S}^c$	200	1.67	NA	3.00	NA
		400	2.67	NA	1.83	NA
	$\mathcal{S} \cup \mathcal{S}^c$	200	34.00	0.00	21.43	0.00
		400	28.00	0.00	18.67	0.00
Empirical powers (%) with $\boldsymbol{\delta} = \sqrt{\log(p)/n} \times (2^3, 2^2, 2^1, 2^0, 2^{-1})$**.**						
			Change point at $k^* = 0.5n$		Change point at $k^* = 0.3n$	
Model	$\mathcal{G}$	p	Boot-II	L&B	Boot-II	L&B
$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^*$	$\mathcal{S}$	200	100.00	NA	99.18	NA
		400	100.00	NA	99.18	NA
	$\mathcal{S}^c$	200	2.06	NA	2.67	NA
		400	2.06	NA	1.65	NA
	$\mathcal{S} \cup \mathcal{S}^c$	200	99.59	60.42	97.53	40.63
		400	99.18	57.29	95.68	47.92
Empirical powers (%) with $\boldsymbol{\delta} = 2\sqrt{\log(p)/n} \times (2^3, 2^2, 2^1, 2^0, 2^{-1})$**.**						
			Change point at $k^* = 0.5n$		Change point at $k^* = 0.3n$	
Model	$\mathcal{G}$	p	Boot-II	L&B	Boot-II	L&B
$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^*$	$\mathcal{S}$	200	100.00	NA	100.00	NA
		400	100.00	NA	100.00	NA
	$\mathcal{S}^c$	200	2.67	NA	1.82	NA
		400	2.26	NA	1.65	NA
	$\mathcal{S} \cup \mathcal{S}^c$	200	100.00	100.00	100.00	99.49
		400	100.00	100.00	100.00	99.49

Table S5.4: Multiple change point detection results for **Models 1 and 2** under **Case 2**. The significance level is $\alpha = 5\%$. The numerical results are based on 100 replications.

Multiple change points with $(n, p) = (600, 200)$ **and three change points at** $(180, 300, 420)$							
			$\mathbf{\Sigma} = \mathbf{I}$			$\mathbf{\Sigma} = \mathbf{\Sigma}^*$	
C	Method	Mean	Adj.Rand	Sd.Adj.Rand	Mean	Adj.Rand	Sd.Adj.Rand
	$\mathcal{G} = \mathcal{S}$						
$C = 1.5$	Arg-max	3.265	0.947	0.056	3.133	0.952	0.043
	L&B	NA	NA	NA	NA	NA	NA
	SGL	NA	NA	NA	NA	NA	NA
	VPWBS	NA	NA	NA	NA	NA	NA
	$\mathcal{G} = \mathcal{S} \cup \mathcal{S}^{\mathbf{c}}$						
	Arg-max	3.177	0.950	0.048	2.983	0.940	0.045
	L&B	1.000	0.398	0.013	1.133	0.439	0.148
	SGL	4.000	0.722	0.111	5.417	0.753	0.083
	VPWBS	2.857	0.899	0.133	2.949	0.918	0.086
	$\mathcal{G} = \mathcal{S}$						
$C = 3$	Arg-max	3.112	0.967	0.034	3.200	0.955	0.049
	L&B	NA	NA	NA	NA	NA	NA
	SGL	NA	NA	NA	NA	NA	NA
	VPWBS	NA	NA	NA	NA	NA	NA
	$\mathcal{G} = \mathcal{S} \cup \mathcal{S}^{\mathbf{c}}$						
	Arg-max	3.104	0.968	0.032	3.250	0.951	0.035
	L&B	3.000	0.991	0.006	3.000	0.992	0.007
	SGL	7.000	0.767	0.093	8.000	0.873	0.118
	VPWBS	2.878	0.945	0.066	2.898	0.944	0.060

Table S5.5: Multiple change point detection results for **Models 1 and 2** under **Case 3**. The significance level is $\alpha = 5\%$. The numerical results are based on 100 replications.

Multiple change points with $(n, p) = (1000, 200)$ **and four change points at** $(300, 450, 550, 700)$							
			$\mathbf{\Sigma} = \mathbf{I}$			$\mathbf{\Sigma} = \mathbf{\Sigma}^*$	
C	Method	Mean	Adj.Rand	Sd.Adj.Rand	Mean	Adj.Rand	Sd.Adj.Rand
	$\mathcal{G} = \mathcal{S}$						
$C = 2$	Arg-max	4.100	0.967	0.047	4.183	0.968	0.036
	L&B	NA	NA	NA	NA	NA	NA
	SGL	NA	NA	NA	NA	NA	NA
	VPWBS	NA	NA	NA	NA	NA	NA
	$\mathcal{G} = \mathcal{S} \cup \mathcal{S}^{\mathbf{c}}$						
	Arg-max	4.067	0.949	0.052	4.200	0.961	0.044
	L&B	1.600	0.589	0.296	1.867	0.688	0.185
	SGL	6.167	0.664	0.054	6.500	0.708	0.104
	VPWBS	3.296	0.882	0.093	3.276	0.882	0.106
	$\mathcal{G} = \mathcal{S}$						
$C = 4$	Arg-max	4.150	0.971	0.031	4.067	0.968	0.029
	L&B	NA	NA	NA	NA	NA	NA
	SGL	NA	NA	NA	NA	NA	NA
	VPWBS	NA	NA	NA	NA	NA	NA
	$\mathcal{G} = \mathcal{S} \cup \mathcal{S}^{\mathbf{c}}$						
	Arg-max	4.050	0.979	0.026	4.183	0.967	0.040
	L&B	3.956	0.988	0.038	4.000	0.994	0.004
	SGL	8.833	0.799	0.111	8.583	0.807	0.112
	VPWBS	3.520	0.932	0.052	3.592	0.939	0.046

S6 Application to Alzheimer's Disease Data Analysis

In this section, we apply our proposed method to analyze data from the Alzheimer's Disease Neuroimaging Initiative (`http://adni.loni.usc.edu/`). It is known that AD accounts for most forms of dementia characterized by progressive cognitive and memory deficits. This makes it a very important health issue which attracts a lot of scientific attentions in recent years. To study AD, Mini-Mental State Examination (MMSE) (Folstein et al., 1975) is a 30-point questionnaire that is commonly used to measure cognitive impairment. According to MMSE, any score of 24 or more (out of 30) indicates a normal cognition. Below this, scores can indicate severe ($\leq$9 points), moderate (10–18 points) or mild (19–23 points) cognitive impairment. Because of the strong relationship between the MMSE score and AD, it can be interesting and useful to predict the MMSE score using some biomarkers for diagnosing the current disease status of AD as well as to identify important predictive biomarkers. According to previous studies (Yu and Liu, 2016; Yu et al., 2020), structural magnetic resonance imaging (MRI) data are very useful for the prediction of the MMSE score. However, these studies typically ignored the effect of other covariates such as ages, education years, or genders on the linear models. Hence, an interesting question is whether there is a change point in the linear structure between the MMSE score and MRI data due to some other covariates. If a change point exists, we would like to identify the location of the change point. To answer these questions, we use our proposed change point detection method to address these issues. We focus on the covariate age which is of particular interest in AD studies. We obtain the dataset for our analysis from the ADNI database. After proper image preprocessing steps such as anterior commissure posterior commissure correction and intensity inhomogeneity correction, we obtain the final dataset with 410 subjects with 225 normal controls and 185 AD patients. For each subject with known age, there is one MMSE score and 93 MRI features corresponding to 93 manually labeled regions of interest (ROI) (Zhang and Shen, 2012). We treat the MMSE score as the response variable and MRI features as predictors in our model. The dataset is first scaled to have mean 0 and variance 1 for the MMSE score and each MRI feature. We are interested in detecting a change point in the linear structure due to the change of ages. Considering potential effect variations of different samples, we randomly select 370 subjects from the whole 410 subjects according to the empirical distribution of ages shown in Figure S6.1 (left) as the training data and use the remaining 40 subjects as the testing data. Then, we sort the training subjects by the value of ages and use our proposed method to detect and identify a change point in the covariate age. We repeat the above process for 50 times. As a comparison, for each random split, we also use lasso to select variables on the training data via 10-fold cross-validation. For this study, we set the significance level at 5%. The number of bootstraps is 200.

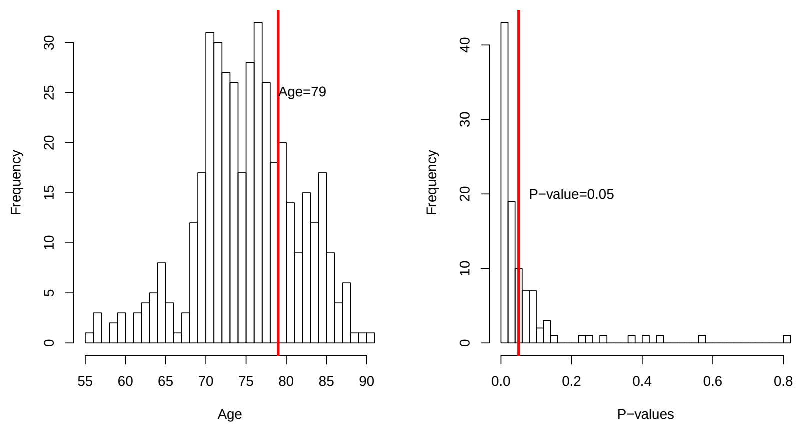

Figure S6.1: Left: Distribution of ages among the 410 subjects. Right: Empirical p-values for change point detection out of 50 random data splits.

Figure S6.1 (right) shows the empirical p-values for the 50 random data splits. Based on our results, 82% of the random splits with an estimated p-value lower than 0.05 have detected a change point. This strongly suggests that there is a change point in the linear structure due to the covariate age. Moreover, for the above 82% random splits, we record the estimated change points in Figure S6.2. We can see that in most cases, the argmax-based estimator identify the change point at the age of 79. The above analysis indicates that the linear structure between the MMSE score and MRI may be different before and after the age of 79. To see this more clearly, among the random splits with a change point, Figure S6.3 reports the features (with estimated coefficients bigger than 0.01) which are selected for more than 80% times before and after the change point, respectively. There are 16 features selected before the change point and 6 features selected after the change point. In other words, those 16 features shown in Figure S6.3 (left) are very predictive for the MMSE score for people with an age smaller or equal to 79. Once the age exceeds 79, it is better to predict the MMSE score using the other 6 features in Figure S6.3 (right). To verify this, for those random splits with a change point, we calculate the mean squared error for the corresponding testing data, based on the selected models using the training data. Figure S6.4 shows the results of our proposed method and lasso. We can see that our proposed method has better prediction performance by segmenting the model by the covariate age, with about 5.34% lower averaged MSE than that of lasso.

Lastly, as for the selected variables, some interesting observations can be made. For example, ROI 83 is predictive for the MMSE score across all ages. ROIs 30 and 69 are only very predictive for the MMSE score under the age of 79 and above 79, respectively. It is known that the 83th

ROI corresponds to the amygdala region, and the 30th and 69th features correspond to the hippocampal regions. According to many previous studies (Zhang and Shen, 2012), those regions are known to be related to AD based on group comparison methods. For these and other selected features, it would be very interesting to investigate their relationship with AD by some group comparison studies according to the segmentation of ages.

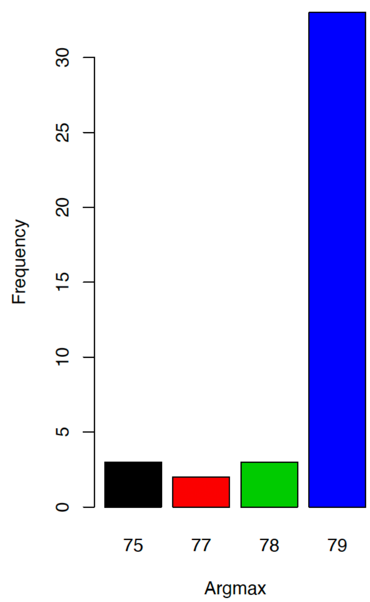

Figure S6.2: Estimated change points for the 82% random splits with change points among the 50 replications.

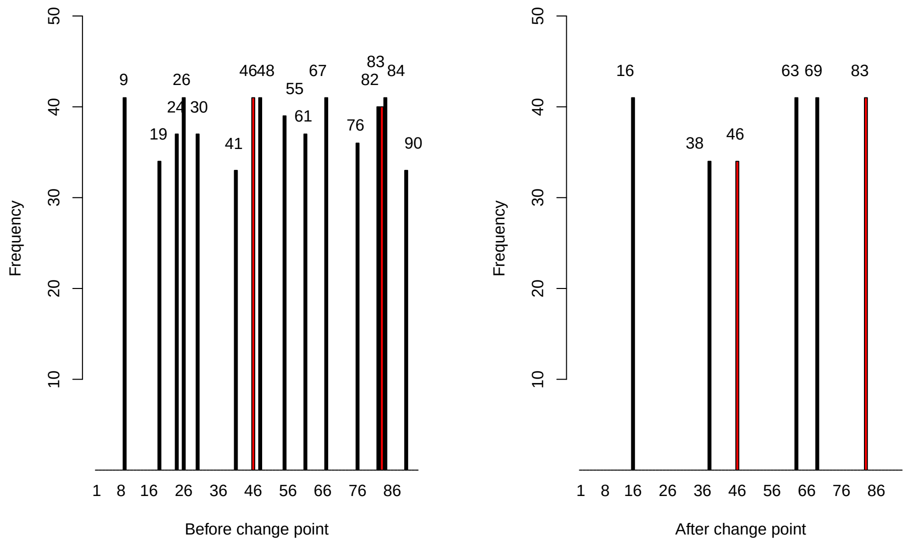

Figure S6.3: Frequency of features selected before the change point (left) and after the change point (right) for the ADNI data out of 50 random splits. Red corresponds to the features that are selected both before and after the change point.

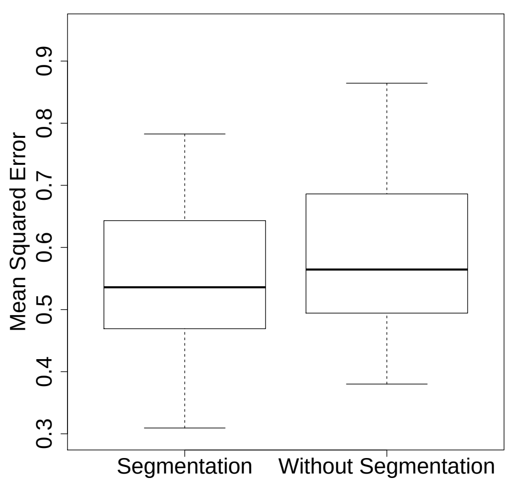

Figure S6.4: Mean squared errors for the prediction of the MMSE score with and without change point models.

S7 Useful lemmas

Let $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ be independent centered random vectors in $\mathbb{R}^p$ with $\boldsymbol{Z}_i = (Z_{i,1}, \ldots, Z_{i,p})^\top$ 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 $\boldsymbol{Z}_i$. We then require the following conditions:

(M1) There is a constant $b > 0$ such that $\inf_{1 \le j \le p} \mathbb{E}(Z_{i,j})^2 \ge b$ for $i = 1, \ldots, n$.

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

(M3) There exists a constant $K' > 0$ such that $\mathbb{E}\big(\exp(|Z_{i,j}|/K')\big) \le 2$ for $j = 1, \ldots, d$ and $i = 1, \ldots, n$.

Lemma 1. *(Liu et al. (2020)) Assume that* $\log(pn) = O(\lfloor n\tau_0 \rfloor^\zeta)$ *holds for some* $0 < \zeta < 1/7$. *Let*

$$\boldsymbol{S}^{\boldsymbol{Z}}(\lfloor nt \rfloor) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{Z}_i \big(\mathbf{1}(i \le \lfloor nt \rfloor) - \lfloor nt \rfloor / n\big), \ \ \boldsymbol{S}^{\boldsymbol{G}}(\lfloor nt \rfloor) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{G}_i \big(\mathbf{1}(i \le \lfloor nt \rfloor) - \lfloor nt \rfloor / n\big)$$

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*

$$\sup_{z \in (0,\infty)} \Big| \mathbb{P}\big(\max_{\tau_0 \le t \le 1-\tau_0} \|\boldsymbol{S}^{\boldsymbol{Z}}(\lfloor nt \rfloor)\|_\infty \le z\big) - \mathbb{P}\big(\sup_{\tau_0 \le t \le 1-\tau_0} \|\boldsymbol{S}^{\boldsymbol{G}}(\lfloor nt \rfloor)\|_\infty \le z\big) \Big| \le C n^{-\zeta_0}, \tag{S7.10}$$

where C is a constant only depending on b, K, and K'.

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

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

where C is a constant only depending on b.

Lemma 3. *(Zhou et al. (2018)) Let $\boldsymbol{W} = (W_1, \ldots, W_p)^\top$ be a random vector with a marginal distribution $N(0, \sigma_i^2)$ $(1 \le i \le p)$. Suppose $\exists A_0 > 0$ such that $\max_i \sigma_i^2 \le A_0^2$. Then, for any $t > 0$, we have*

$$\mathbb{E}\big(\max_{1\le i\le p} |W_i|\big) \le \frac{\log(2p)}{t} + \frac{tA_0^2}{2}.$$

Lemma 4 (Van de Geer et al. (2014))**.** *Suppose Assumptions* $(\mathbf{A.1})$ – $(\mathbf{A.3})$ *hold. Assume additionally $\max_j \sqrt{s_j \log(p)/n} = o(1)$ holds. For the node-wise regression in (2.5), choosing the tuning parameters $\lambda_{(j)} \asymp \sqrt{\log(p)/n}$ uniformly over j, we have*

$$\|\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\|_q = O_p\Big(s_j^{1/q}\sqrt{\frac{\log(p)}{n}}\Big), \text{ for } q = 1, 2. \tag{S7.11}$$

Lemma 5. *Let $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ be independent centered random vectors in $\mathbb{R}^p$ with $\boldsymbol{Z}_i = (Z_{i,1}, \ldots, Z_{i,p})^\top$ for $i = 1, \ldots, n$. Assume that $\mathbf{Z}_i$ follows the sub-exponential distribution. Then, for any given subgroup $\mathcal{G} \subset \{1, \ldots, p\}$,* *with probability at least* $1 - C_1(pn)^{-C_2}$, *we have*

$$\max_{t\in[\tau_0,1-\tau_0]} \max_{j\in\mathcal{G}} \Big| \frac{1}{\sqrt{n}} \Big(\sum_{i=1}^{\lfloor nt \rfloor} Z_{i,j} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} Z_{i,j} \Big) \Big| \le C_3 \sqrt{\log(|\mathcal{G}|n)}, \quad \text{(S7.12)}$$

where C_1, C_2, *and* C_3 *are universal positive constants not depending on* p *or* n.

We next provide some useful results for the lasso estimators from heterogeneous data observations. To this end, for each $t \in [\tau_0, 1-\tau_0]$, define

$$\mathcal{A}(t) = \left\{ \Big\| \frac{1}{\lfloor nt \rfloor} \big(\mathbf{X}_{(0,t)}\big)^\top (\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)} \boldsymbol{\beta}^{(0,t)} \Big\|_\infty \le \lambda^{(1)} \right\},$$

$$\mathcal{B}(t) = \left\{ \Big\| \frac{1}{\lfloor nt \rfloor^*} \big(\mathbf{X}_{(t,1)}\big)^\top (\boldsymbol{Y}_{(t,1)} - \mathbf{X}_{(t,1)} \boldsymbol{\beta}^{(t,1)} \Big\|_\infty \le \lambda^{(2)} \right\}, \quad \text{(S7.13)}$$

where $\lambda^{(1)} := K_1 \sqrt{\frac{\log(p)}{\lfloor nt \rfloor}}$ and $\lambda^{(2)} := K_2 \sqrt{\frac{\log(p)}{\lfloor nt \rfloor^*}}$, and $K_1, \ldots, K_2$ are some universal positive constants not depending on n or p.

The following Lemma 6 provides a basic inequality for the lasso estimators, which is important for deriving the precise estimation error bound as well as prediction error bound (see Lemma 8 below). The proof of Lemma 6 is given in Section S10.2.

Lemma 6. *Suppose Assumptions* (**A.1**) – (**A.3**) *hold. Assume* $\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2 \le C_{\boldsymbol{\Delta}}$ *for some* $C_{\boldsymbol{\Delta}} > 0$. *Recall* $\boldsymbol{\beta}^{(0,t)}$ *and* $\boldsymbol{\beta}^{(t,1)}$ *defined in (2.7). Let* $\widehat{\boldsymbol{\beta}}^{(0,t)}$ *and* $\widehat{\boldsymbol{\beta}}^{(t,1)}$ *be the lasso estimators as defined in (2.9). Then, for each*

$t \in [\tau_0, 1-\tau_0]$, under the event $\mathcal{A}(t) \cap \mathcal{B}(t)$, we have

$$\frac{\left\|\mathbf{X}_{(0,t)}\left(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0,t)}\right)\right\|_2^2}{\lfloor nt \rfloor} + \lambda_1(t)\left\|(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0,t)})_{J^c(\boldsymbol{\beta}^{(0,t)})}\right\|_1 \leq 3\lambda_1(t)\left\|(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0,t)})_{J(\boldsymbol{\beta}^{(0,t)})}\right\|_1, \tag{S7.14}$$

and

$$\frac{\left\|\mathbf{X}_{(t,1)}\left(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(t,1)}\right)\right\|_2^2}{\lfloor nt \rfloor^*} + \lambda_2(t)\left\|(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(t,1)})_{J^c(\boldsymbol{\beta}^{(t,1)})}\right\|_1 \leq 3\lambda_2(t)\left\|(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(t,1)})_{J(\boldsymbol{\beta}^{(t,1)})}\right\|_1, \tag{S7.15}$$

where $\lambda_1(t) := 2\lambda^{(1)}$, $\lambda_2(t) := 2\lambda^{(2)}$.

The following Lemma 7 provides the estimation error bounds for the lasso estimators $\widehat{\boldsymbol{\beta}}^{(0,t)}$ and $\widehat{\boldsymbol{\beta}}^{(t,1)}$ in terms of ℓ_q-norm. The proof of Lemma 7 is given in Section S10.3.

Lemma 7. *Suppose Assumptions* (**A.1**) – (**A.3**) *hold. Assume $\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2 \leq C_{\boldsymbol{\Delta}}$ for some $C_{\boldsymbol{\Delta}} > 0$. Recall $\boldsymbol{\beta}^{(0,t)}$ and $\boldsymbol{\beta}^{(t,1)}$ defined in (2.7). Let $\widehat{\boldsymbol{\beta}}^{(0,t)}$ and $\widehat{\boldsymbol{\beta}}^{(t,1)}$ be the lasso estimators as defined in (2.9). For each $t \in [\tau_0, 1-\tau_0]$, let $s_1(t) := \mathcal{M}(\boldsymbol{\beta}^{(0,t)})$ and $s_2(t) := \mathcal{M}(\boldsymbol{\beta}^{(t,1)})$. Then, under* *the event $\mathcal{A}(t) \cap \mathcal{B}(t)$, we have*

$$\left\|\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0,t)}\right\|_q \le C_1 (s_1(t))^{\frac{1}{q}} \sqrt{\frac{\log p}{\lfloor nt \rfloor}}, \quad \left\|\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(t,1)}\right\|_q \le C_2 (s_2(t))^{\frac{1}{q}} \sqrt{\frac{\log p}{\lfloor nt \rfloor^*}}, \ q = 1, 2,$$

$$\frac{\left\|\mathbf{X}_{(0,t)}\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0,t)}\big)\right\|_2^2}{\lfloor nt \rfloor} \le C_3 s_1(t) \frac{\log p}{\lfloor nt \rfloor}, \quad \frac{\left\|\mathbf{X}_{(t,1)}\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(t,1)}\big)\right\|_2^2}{\lfloor nt \rfloor^*} \le C_4 s_2(t) \frac{\log p}{\lfloor nt \rfloor^*},$$

$$\mathcal{M}(\widehat{\boldsymbol{\beta}}^{(0,t)}) \le C_5 s_1(t), \qquad \mathcal{M}(\widehat{\boldsymbol{\beta}}^{(t,1)}) \le C_6 s_2(t), \tag{S7.16}$$

where $C_1, \ldots, C_6$ are some universal positive constants not depending on n or p.

Lastly, as a by product of Lemma 7, the following Lemma 8 provides the estimation error bounds for $\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(1)}$ and $\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(2)}$ in terms of the ℓ_q-norm, which is frequently used in the proofs.

Lemma 8. *Suppose Assumptions* (**A**.**1**) – (**A**.**3**) *hold. Assume $\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2 \le C_{\boldsymbol{\Delta}}$ for some $C_{\boldsymbol{\Delta}} > 0$. Recall $s := s^{(1)} \vee s^{(2)}$. Let $\widehat{\boldsymbol{\beta}}^{(0,t)}$ and $\widehat{\boldsymbol{\beta}}^{(t,1)}$ be the lasso estimators as defined in (2.9). Then, under the event*

$\mathcal{A}(t) \cap \mathcal{B}(t)$, for each $t \in [\tau_0, 1-\tau_0]$, we have

$$\big\|\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(1)}\big\|_q \leq C_1 \max\Big\{s^{\frac{1}{q}}\sqrt{\frac{\log p}{\lfloor nt\rfloor}}, \frac{\lfloor nt\rfloor - \lfloor nt_0\rfloor}{\lfloor nt\rfloor}\big\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\big\|_q \mathbf{1}\{t \geq t_0\}\Big\}, q = 1, 2,$$

$$\big\|\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(2)}\big\|_q \leq C_2 \max\Big\{s^{\frac{1}{q}}\sqrt{\frac{\log p}{\lfloor nt\rfloor^*}}, \frac{\lfloor nt_0\rfloor - \lfloor nt\rfloor}{\lfloor nt\rfloor^*}\big\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\big\|_q \mathbf{1}\{t \leq t_0\}\Big\}, q = 1, 2,$$

$$\frac{\big\|\mathbf{X}_{(0,t)}\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(1)}\big)\big\|_2^2}{\lfloor nt\rfloor} \leq C_3 \max\Big\{s\frac{\log p}{\lfloor nt\rfloor}, \big(\frac{\lfloor nt\rfloor - \lfloor nt_0\rfloor}{\lfloor nt\rfloor}\big)^2\big\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\big\|_2^2 \mathbf{1}\{t \geq t_0\}\Big\},$$

$$\frac{\big\|\mathbf{X}_{(t,1)}\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(2)}\big)\big\|_2^2}{\lfloor nt\rfloor^*} \leq C_4 \max\Big\{s\frac{\log p}{\lfloor nt\rfloor^*}, \big(\frac{\lfloor nt_0\rfloor - \lfloor nt\rfloor}{\lfloor nt\rfloor^*}\big)^2\big\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\big\|_2^2 \mathbf{1}\{t \leq t_0\}\Big\},$$

$$\mathcal{M}(\widehat{\boldsymbol{\beta}}^{(0,t)}) \leq C_5 s, \quad \mathcal{M}(\widehat{\boldsymbol{\beta}}^{(t,1)}) \leq C_6 s, \tag{S7.17}$$

where $C_1, \ldots, C_6$ are some universal positive constants not depending on n or p.

The following Lemma 9 shows that the results in Lemmas 6 – 8 occur uniformly over $t \in [\tau_0, 1-\tau_0]$ with high probability. The proof of Lemma 9 is given in Section S10.4.

Lemma 9. *Suppose Assumptions* (**A.1**) – (**A.3**) *hold. Assume $\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2 \leq C_{\boldsymbol{\Delta}}$ for some $C_{\boldsymbol{\Delta}} > 0$. Then we have*

$$\mathbb{P}\Big(\bigcap_{t\in[\tau_0, 1-\tau_0]} \big\{\mathcal{A}(t) \cap \mathcal{B}(t)\big\}\Big) \geq 1 - C_1(np)^{-C_2}, \tag{S7.18}$$

where C_1, C_2 are some big enough universal positive constants not depending on n or p.

S8 Proof of main results

S8.1 Proof of Proposition 1

Proof. Note that the proof of Part (i) is easier than Part (ii). To save space, we give the proof of Part (ii). Firstly, we consider $\mathcal{R}_1(s, 3, \mathbb{T})$. The proof proceeds in two steps.

Step 1: we prove $\sup_{t\in[\tau_0,1-\tau_0]} \|\widehat{\mathbf{\Sigma}}_{(0,t)} - \mathbf{\Sigma}\|_\infty = O_p(\sqrt{\log(p)/\lfloor n\tau_0\rfloor})$. For any fixed $t \in [\tau_0, 1-\tau_0]$ and $j, k \in \{1, \ldots, p\}$, by Assumption (A.1), using exponential inequality, we have

$$\mathbb{P}\big(|\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}(X_{ij}X_{ik}-\mathbb{E}[X_{ij}X_{ik}])| \geq x\big) \leq C_1\exp(-C_2\lfloor nt\rfloor x^2) \leq C_1\exp(-C_2\lfloor n\tau_0\rfloor x^2).$$

Hence, taking $x = C_3\sqrt{\log(pn)/\lfloor n\tau_0\rfloor}$ for some big constant $C_3 > 0$, we have:

$$\mathbb{P}\big(|\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}(X_{ij}X_{ik}-\mathbb{E}[X_{ij}X_{ik}])| \geq x\big) \leq C_1(np)^{-C_3}.$$

As a result, we have:

$$
\begin{aligned}
&\mathbb{P}(\sup_{t\in[\tau_0,1-\tau_0]} \|\widehat{\mathbf{\Sigma}}_{(0,t)} - \mathbf{\Sigma}\|_\infty \geq x) \\
&= \mathbb{P}\Big(\bigcup_t \bigcup_{j,k} \big\{ |\frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} (X_{ij}X_{ik} - \mathbb{E}[X_{ij}X_{ik}])| \geq x \big\} \Big) \\
&\leq np^2 \max_{t,j,k} \mathbb{P}\big(|\frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} (X_{ij}X_{ik} - \mathbb{E}[X_{ij}X_{ik}])| \geq x \big) \\
&\leq C_1 (np)^{-C_4},
\end{aligned}
$$

where C_1–C_4 are some big enough universal constants. This yields $\sup_{t\in[\tau_0,1-\tau_0]} \|\widehat{\mathbf{\Sigma}}_{(0,t)} - \mathbf{\Sigma}\|_\infty = O_p(\sqrt{\log(p)/\lfloor n\tau_0 \rfloor})$.

Step 2: For integers s such that $1 \leq s \leq p$, a set of indices $J_0 \subset \{1,\ldots,p\}$ with $|J_0| \leq s$, and any vector $\boldsymbol{\delta}$ satisfying $\|\boldsymbol{\delta}_{J_0^c}\|_1 \leq 3\|\boldsymbol{\delta}_{J_0}\|_1$, we have:

$$
\begin{aligned}
\frac{\boldsymbol{\delta}^\top \widehat{\mathbf{\Sigma}}_{(0,t)} \boldsymbol{\delta}}{|\boldsymbol{\delta}_{J_0}|_2^2} \quad &=_{(1)} \frac{\boldsymbol{\delta}^\top \mathbf{\Sigma} \boldsymbol{\delta}}{|\boldsymbol{\delta}_{J_0}|_2^2} + \frac{\boldsymbol{\delta}^\top (\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}_{(0,t)}) \boldsymbol{\delta}}{|\boldsymbol{\delta}_{J_0}|_2^2}, \\
&\geq_{(2)} \frac{\boldsymbol{\delta}^\top \mathbf{\Sigma} \boldsymbol{\delta}}{|\boldsymbol{\delta}_{J_0}|_2^2} - \frac{\sup_{t\in[\tau_0,1-\tau_0]} \|\widehat{\mathbf{\Sigma}}_{(0,t)} - \mathbf{\Sigma}\|_\infty}{|\boldsymbol{\delta}_{J_0}|_2^2} |\boldsymbol{\delta}|_1^2, \\
&\geq_{(3)} \frac{\boldsymbol{\delta}^\top \mathbf{\Sigma} \boldsymbol{\delta}}{|\boldsymbol{\delta}_{J_0}|_2^2} - \frac{\sup_{t\in[\tau_0,1-\tau_0]} \|\widehat{\mathbf{\Sigma}}_{(0,t)} - \mathbf{\Sigma}\|_\infty}{|\boldsymbol{\delta}_{J_0}|_2^2} (1+c_0)^2 |\boldsymbol{\delta}_{J_0}|_1^2, \\
&\geq_{(4)} \frac{\boldsymbol{\delta}^\top \mathbf{\Sigma} \boldsymbol{\delta}}{|\boldsymbol{\delta}_{J_0}|_2^2} - \sup_{t\in[\tau_0,1-\tau_0]} \|\widehat{\mathbf{\Sigma}}_{(0,t)} - \mathbf{\Sigma}\|_\infty (1+c_0)^2 s. \\
&\geq_{(5)} 4\kappa_4^4 - sO_p(\sqrt{\log(p)/\lfloor n\tau_0 \rfloor}) \geq_{(6)} \kappa_4^2,
\end{aligned}
\tag{S8.19}
$$

where (5) comes from Condition (S3.9) and the result in Step 1, (6) comes from the assumption $s\sqrt{\log(p)/\lfloor n\tau_0 \rfloor} = o(1)$. Lastly, combining Steps 1 and 2, we finish the proof.

$\square$

S8.2 Proof of Theorem 1

Proof. Under $\mathbf{H}_0$, the change point t_0 is not identifiable. Hence, to prove Theorem 1, we need to prove the convergence of $\widehat{\sigma}_\epsilon^2(t)\widehat{\omega}_{j,k}\}$ to $\{\sigma_\epsilon^2\omega_{j,k}\}$ uniformly over $\tau_0 \leq t \leq 1-\tau_0$ and $1 \leq j,k \leq p$, where $\widehat{\sigma}_\epsilon^2(t)$ is defined in (2.17). Note that for each t, j and k,

$$
\begin{aligned}
&|\widehat{\sigma}_\epsilon^2(t)\widehat{\omega}_{j,k} - \sigma_\epsilon^2\omega_{j,k}| \\
&\leq |\widehat{\sigma}_\epsilon^2(t)\widehat{\omega}_{j,k} - \sigma_\epsilon^2\widehat{\omega}_{j,k}| + \sigma_\epsilon^2|\widehat{\omega}_{j,k} - \omega_{j,k}| \\
&\leq |\widehat{\sigma}_\epsilon^2(t) - \sigma_\epsilon^2||\widehat{\omega}_{j,k} - \omega_{j,k}| + |\widehat{\sigma}_\epsilon^2(t) - \sigma_\epsilon^2|\omega_{j,k} + \sigma_\epsilon^2|\widehat{\omega}_{j,k} - \omega_{j,k}| \\
&\leq C(|\widehat{\sigma}_\epsilon^2(t) - \sigma_\epsilon^2| + |\widehat{\omega}_{j,k} - \omega_{j,k}|),
\end{aligned}
\tag{S8.20}
$$

where the last inequality comes from Assumptions ($\mathbf{A.2}$) and ($\mathbf{A.3}$) and C is a universal positive constant not depending on n or p. Hence, by (S8.20), to prove Theorem 1, we need to bound $\max_{t\in[\tau_0,1-\tau_0]}|\widehat{\sigma}_\epsilon^2(t) - \sigma_\epsilon^2|$ and $\max_{1\leq j,k\leq p}|\widehat{\omega}_{j,k} - \omega_{j,k}|$, respectively.

For bounding $\max\limits_{t\in[\tau_0,1-\tau_0]}|\widehat{\sigma}_\epsilon^2(t) - \sigma_\epsilon^2|$, by the definition of $\widehat{\sigma}_\epsilon^2(t)$ in (2.17), under $\mathbf{H}_0$, using some straightforward calculations, we have

$$
\begin{aligned}
&\widehat{\sigma}_{\epsilon}^2(t) - \sigma_{\epsilon}^2 \\
&= n^{-1}\Big(\big\|\boldsymbol{\epsilon}_{(0,t)} + \mathbf{X}_{(0,t)}\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0)}\big)\big\|_2^2\Big) + \big\|\boldsymbol{\epsilon}_{(t,1)} + \mathbf{X}_{(t,1)}\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(0)}\big)\big\|_2^2\Big) - \sigma_{\epsilon}^2, \\
&= n^{-1}\big\|\mathbf{X}_{(0,t)}\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0)}\big)\big\|_2^2 + 2\frac{\lfloor nt\rfloor}{n}\frac{(\boldsymbol{\epsilon}_{(0,t)})^\top \mathbf{X}_{(0,t)}}{\lfloor nt\rfloor}\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0)}\big) \\
&\quad + n^{-1}\big\|\mathbf{X}_{(t,1)}\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(0)}\big)\big\|_2^2 + 2\frac{\lfloor nt\rfloor^*}{n}\frac{(\boldsymbol{\epsilon}_{(t,1)})^\top \mathbf{X}_{(t,1)}}{\lfloor nt\rfloor^*}\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(0)}\big) \\
&\qquad\qquad + n^{-1}\sum_{i=1}^n(\epsilon_i^2 - \sigma_\epsilon^2).
\end{aligned} \tag{S8.21}
$$

By (S8.21), to bound $\max_{t\in[\tau_0,1-\tau_0]}|\widehat{\sigma}_{\epsilon}^2(t) - \sigma_{\epsilon}^2|$, we need to consider the five parts on the RHS of (S8.21), respectively. For the first four parts, by Lemma 8, we have

$$
\begin{aligned}
&\frac{1}{n}\big\|\mathbf{X}_{(0,t)}\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0)}\big)\big\|_2^2 \le \frac{\lfloor nt\rfloor}{n}O_p\Big(s^{(0)}\frac{\log(p)}{\lfloor nt\rfloor}\Big) = O_p\Big(s^{(0)}\frac{\log(p)}{n}\Big), \\
&\frac{1}{n}\big\|\mathbf{X}_{(t,1)}\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(0)}\big)\big\|_2^2 \le \frac{\lfloor nt\rfloor^*}{n}O_p\Big(s^{(0)}\frac{\log(p)}{\lfloor nt\rfloor^*}\Big) = O_p\Big(s^{(0)}\frac{\log(p)}{n}\Big), \\
&\Big|2\frac{\lfloor nt\rfloor}{n}\frac{(\boldsymbol{\epsilon}_{(0,t)})^\top \mathbf{X}_{(0,t)}}{\lfloor nt\rfloor}\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0)}\big)\Big| \le O_p\Big(\lambda^{(1)}\big\|\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0)}\big\|_1\Big) \le O_p\Big(s^{(0)}\frac{\log(p)}{\lfloor nt\rfloor}\Big), \\
&\Big|2\frac{\lfloor nt\rfloor^*}{n}\frac{(\boldsymbol{\epsilon}_{(t,1)})^\top \mathbf{X}_{(t,1)}}{\lfloor nt\rfloor^*}\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(0)}\big)\Big| \le O_p\Big(\lambda^{(3)}\big\|\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(0)}\big\|_1\Big) \le O_p\Big(s^{(0)}\frac{\log(p)}{\lfloor nt\rfloor^*}\Big).
\end{aligned} \tag{S8.22}
$$

Note that $\epsilon_i^2 - \sigma_\epsilon^2$ follows the sub-exponential distribution. For $\sum_{i=1}^n(\epsilon_i^2 - \sigma_\epsilon^2)/n$, under Assumption (**A**.**2**), using Bernstein's inequalities, we can prove

$$\sum_{i=1}^{n}(\epsilon_i^2 - \sigma_\epsilon^2)/n \le O_p\Big(\sqrt{\frac{\log(n)}{n}}\Big). \tag{S8.23}$$

Hence, combining (S8.21), (S8.22), and (S8.23), and using Assumptions (**A**.**1**) – (**A**.**3**), we have

$$\max_{t\in[\tau_0,1-\tau_0]} |\widehat{\sigma}_\epsilon^2(t) - \sigma_\epsilon^2| \le O_p\Big(\sqrt{\frac{\log(n)}{n}}\Big). \tag{S8.24}$$

Next, we bound $\max_{1\le j,k\le p} |\widehat{\omega}_{j,k} - \omega_{j,k}|$. By Lemmas 5.3 and 5.4 in Van de Geer et al. (2014), we have

$$\max_{1\le j,k\le p} |\widehat{\omega}_{j,k} - \omega_{j,k}| = \max_{1\le j,k\le p} |\widehat{\mathbf{\Theta}}_j^\top \widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Theta}}_k - \mathbf{\Theta}_j^\top \mathbf{\Sigma}\mathbf{\Theta}_k| = O_p(\max_j \lambda_{(j)}\sqrt{s_j}). \tag{S8.25}$$

Finally, combining (S8.24) and (S8.25), we have

$$\max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j,k\le p} |\widehat{\sigma}_\epsilon^2(t)\widehat{\omega}_{j,k} - \sigma_\epsilon^2\omega_{j,k}| \le O_p\Big(\sqrt{\frac{\log(n)}{n}} + \max_j \lambda_{(j)}\sqrt{s_j}\Big), \tag{S8.26}$$

which completes the proof of Theorem 1. □

S8.3 Proof of Theorem 2

Proof. In this section, we aim to prove

$$\sup_{z\in(0,\infty)} \Big|\mathbb{P}(T_{\mathcal{G}} \le z) - \mathbb{P}(T_{\mathcal{G}}^b \le z|\mathcal{X})\Big| = o_p(1), \text{ as } n,p\to\infty. \tag{S8.27}$$

The proof proceeds in four steps. In Steps 1 and 2, we decompose $T_{\mathcal{G}}$ and $T_{\mathcal{G}}^b$ into a leading term and a residual term and show that the corresponding residual terms can be asymptotically negligible. In Step 3, we prove that it is possible to approximate the leading term of $T_{\mathcal{G}}$ by that of $T_{\mathcal{G}}^b$. In Step 4, we combine the previous results to complete the proof.

Step 1 (Decomposition of $T_{\mathcal{G}}$). Note that under the null hypothesis of no change point, we have $\beta_s^{(1)} = \beta_s^{(2)} = \beta_s^{(0)}$ for $1 \le s \le p$. By the definition of the de-biased lasso estimators $\breve{\boldsymbol{\beta}}^{(0,t)}$ and $\breve{\boldsymbol{\beta}}^{(t,1)}$ in (2.15), we can write them as follows:

$$\begin{aligned}
\breve{\boldsymbol{\beta}}^{(0,t)} &= \boldsymbol{\beta}^{(0)} + \widehat{\boldsymbol{\Theta}}(\mathbf{X}_{(0,t)})^\top \boldsymbol{\epsilon}_{(0,t)}/\lfloor nt \rfloor + \boldsymbol{\Delta}^{(0,t)}, \\
\breve{\boldsymbol{\beta}}^{(t,1)} &= \boldsymbol{\beta}^{(0)} + \widehat{\boldsymbol{\Theta}}(\mathbf{X}_{(t,1)})^\top \boldsymbol{\epsilon}_{(t,1)}/\lfloor nt \rfloor^* + \boldsymbol{\Delta}^{(t,1)},
\end{aligned} \tag{S8.28}$$

where $\boldsymbol{\Delta}^{(0,t)} = \big(\Delta_1^{(0,t)}, \ldots, \Delta_p^{(0,t)}\big)^\top$ and $\boldsymbol{\Delta}^{(t,1)} = \big(\Delta_1^{(t,1)}, \ldots, \Delta_p^{(t,1)}\big)^\top$ are defined as

$$\begin{aligned}
\boldsymbol{\Delta}^{(0,t)} &:= -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0)}\big), \\
\boldsymbol{\Delta}^{(t,1)} &:= -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,1)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(0)}\big),
\end{aligned} \tag{S8.29}$$

with $\widehat{\boldsymbol{\Sigma}}_{(0,t)} := (\mathbf{X}_{(0,t)})^\top \mathbf{X}_{(0,t)}/\lfloor nt \rfloor$ and $\widehat{\boldsymbol{\Sigma}}_{(t,1)} := (\mathbf{X}_{(t,1)})^\top \mathbf{X}_{(t,1)}/\lfloor nt \rfloor^*$. Denote $\widehat{\boldsymbol{\Theta}}_i$, $\mathbf{X}_{(0,t),i}$, $\mathbf{X}_{(t,1),i}$ as the i-th row of $\widehat{\boldsymbol{\Theta}}$, $\mathbf{X}_{(0,t)}$, and $\mathbf{X}_{(t,1)}$, respectively. Then, for each coordinate j at time point $\lfloor nt \rfloor$, we can write each coordinate of the de-biased lasso estimator in the following form:

$$\breve{\beta}_j^{(0,t)} = \beta_j^{(0)} + \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} \widehat{\boldsymbol{\Theta}}_j^\top \mathbf{X}_i \epsilon_i + \Delta_j^{(0,t)},$$

$$\breve{\beta}_j^{(t,1)} = \beta_j^{(0)} + \frac{1}{\lfloor nt \rfloor^*} \sum_{i=\lfloor nt \rfloor+1}^{n} \widehat{\boldsymbol{\Theta}}_j^\top \mathbf{X}_i \epsilon_i + \Delta_j^{(t,1)}. \tag{S8.30}$$

For each $t \in [\tau_0, 1-\tau_0]$ and $1 \leq j \leq p$, define the coordinate-wise process as

$$C_j(\lfloor nt \rfloor) = \sqrt{n} \frac{\lfloor nt \rfloor}{n} \frac{\lfloor nt \rfloor^*}{n} \frac{\left(\breve{\beta}_j^{(0,t)} - \breve{\beta}_j^{(t,1)}\right)}{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}}. \tag{S8.31}$$

By the definition of $T_{\mathcal{G}}$ in (2.20) , we have $T_{\mathcal{G}} = \max_{t \in [\tau_0, 1-\tau_0]} \max_{j \in \mathcal{G}} |C_j(\lfloor nt \rfloor)|$. Furthermore, by (S8.30), we can decompose $C_j(\lfloor nt \rfloor)$ into two parts:

$$C_j(\lfloor nt \rfloor) = C_j^{\mathrm{I}}(\lfloor nt \rfloor) + C_j^{\mathrm{II}}(\lfloor nt \rfloor), \text{ for } t \in [\tau_0, 1-\tau_0], \quad 1 \leq j \leq p, \tag{S8.32}$$

with

$$C_j^{\mathrm{I}}(\lfloor nt \rfloor) := \sqrt{n} \frac{\lfloor nt \rfloor}{n} \frac{\lfloor nt \rfloor^*}{n} \frac{\left(\frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} \widehat{\boldsymbol{\Theta}}_j^\top \mathbf{X}_i \epsilon_i - \frac{1}{\lfloor nt \rfloor^*} \sum_{i=\lfloor nt \rfloor+1}^{n} \widehat{\boldsymbol{\Theta}}_j^\top \mathbf{X}_i \epsilon_i\right)}{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}},$$

$$C_j^{\mathrm{II}}(\lfloor nt \rfloor) := \sqrt{n} \frac{\lfloor nt \rfloor}{n} \frac{\lfloor nt \rfloor^*}{n} \frac{\left(\Delta_j^{(0,t)} - \Delta_j^{(t,1)}\right)}{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}}, \text{ with } 1 \leq j \leq p \text{ and } t \in [\tau_0, 1-\tau_0]. \tag{S8.33}$$

Note that we can regard $C_j^{\mathrm{I}}(\lfloor nt \rfloor)$ as the leading term and $C_j^{\mathrm{II}}(\lfloor nt \rfloor)$ as the residual term of $C_j(\lfloor nt \rfloor)$. Furthermore, by replacing $\widehat{\sigma}_\epsilon^2$, $\widehat{\omega}_{j,j}$, and $\widehat{\boldsymbol{\Theta}}_j$ by their true values σ_ϵ^2, $\omega_{j,j}$, and $\boldsymbol{\Theta}_j$, we can define the oracle leading term as follows:

$$\widetilde{C}_j^{\mathrm{I}}(\lfloor nt\rfloor) := \sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\frac{\Big(\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\mathbf{\Theta}_j^\top\mathbf{X}_i\epsilon_i - \frac{1}{\lfloor nt\rfloor^*}\sum_{i=\lfloor nt\rfloor+1}^{n}\mathbf{\Theta}_j^\top\mathbf{X}_i\epsilon_i\Big)}{\sqrt{\sigma_\epsilon^2\omega_{j,j}}}. \tag{S8.34}$$

Based on (S8.31), (S8.33), and (S8.34), define the following four vector-valued processes:

$$\boldsymbol{C}(\lfloor nt\rfloor) = \big(C_1(\lfloor nt\rfloor),\ldots,C_p(\lfloor nt\rfloor)\big)^\top, \qquad \boldsymbol{C}^{\mathrm{I}}(\lfloor nt\rfloor) = \big(C_1^{\mathrm{I}}(\lfloor nt\rfloor),\ldots,C_p^{\mathrm{I}}(\lfloor nt\rfloor)\big)^\top, \tag{S8.35}$$
$$\boldsymbol{C}^{\mathrm{II}}(\lfloor nt\rfloor) = \big(C_1^{\mathrm{II}}(\lfloor nt\rfloor),\ldots,C_p^{\mathrm{II}}(\lfloor nt\rfloor)\big)^\top, \quad \widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor) = \big(\widetilde{C}_1^{\mathrm{I}}(\lfloor nt\rfloor),\ldots,\widetilde{C}_p^{\mathrm{I}}(\lfloor nt\rfloor)\big)^\top.$$

The following Lemma 10 shows that the residual term $|C_j^{\mathrm{II}}|$ can be uniformly negligible over $t \in [\tau_0, 1-\tau_0]$ and $1 \le j \le p$. The proof of Lemma 10 is provided in Section S9.1.

Lemma 10. *Assume Assumptions* $(\mathbf{A.1})$ – $(\mathbf{A.5})$ *hold. Under* $\mathbf{H}_0$, *we have*

$$\mathbb{P}\Big(\max_{\tau_0\le t\le 1-\tau_0}\big\|\boldsymbol{C}(\lfloor nt\rfloor) - \widetilde{\boldsymbol{C}}^I(\lfloor nt\rfloor)\big\|_{\mathcal{G},\infty} \ge \epsilon\Big) = o(1), \tag{S8.36}$$

where $\epsilon = C\max(\max_{1\le j\le p} s_j\frac{\log(pn)}{\sqrt{n}}, s\sqrt{n}\frac{\log(pn)}{\lfloor n\tau_0\rfloor})$, *and* C *is a universal constant not depending on* n *or* p.

Step 2 (Decomposition of $T_{\mathcal{G}}^b$). In this step, we analyze the bootstrap version of the test statistic and decompose $T_{\mathcal{G}}^b$ into a leading term and a residual term. To this end, we need some additional notations. For $0 \le t_1 \le t_2 \le 1$, define

$$\boldsymbol{Y}_{(t_1,t_2)} = (Y_{\lfloor nt_1\rfloor+1},\ldots,Y_{\lfloor nt_2\rfloor})^\top, \quad \boldsymbol{\epsilon}_{(t_1,t_2)} = (\epsilon_{\lfloor nt_1\rfloor+1},\ldots,\epsilon_{\lfloor nt_2\rfloor})^\top,$$
$$\mathbf{X}_{(t_1,t_2)} = (\boldsymbol{X}_{\lfloor nt_1\rfloor+1},\ldots,\boldsymbol{X}_{\lfloor nt_2\rfloor})^\top, \quad \widehat{\boldsymbol{\Sigma}}_{(t_1,t_2)} = \frac{(\mathbf{X}_{(t_1,t_2)})^\top \mathbf{X}_{(t_1,t_2)}}{\lfloor nt_2\rfloor - \lfloor nt_1\rfloor + 1}.$$

Note that the decomposition for $T_{\mathcal{G}}^b$ is different from that of $T_{\mathcal{G}}$. The main difficulty is that the bootstrap based samples involve a change point estimator $\widehat{t}_{0,\mathcal{G}}$ and the data are split into two sub-samples (before and after $\widehat{t}_{0,\mathcal{G}}$), which requires a careful discussion about the location. To analyze $\breve{\boldsymbol{\beta}}^{b,(0,t)}$ and $\breve{\boldsymbol{\beta}}^{b,(t,1)}$ in (2.22), we need to consider the following cases:

Case 1 : The search location t at $t \in [\tau_0, \widehat{t}_{0,\mathcal{G}}]$. In this case, since $\breve{\boldsymbol{\beta}}^{b,(0,t)}$ is constructed using homogeneous bootstrap samples, similar to Step 1, we can decompose $\breve{\boldsymbol{\beta}}^{b,(0,t)}$ as:

$$\breve{\boldsymbol{\beta}}^{b,(0,t)} = \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})} + \frac{\widehat{\boldsymbol{\Theta}}(\mathbf{X}_{(0,t)})^\top \boldsymbol{\epsilon}^{b,(0,t)}}{\lfloor nt\rfloor} + \boldsymbol{\Delta}^{b,(0,t),\mathrm{I}}, \tag{S8.37}$$

where $\boldsymbol{\Delta}^{b,(0,t),\mathrm{I}} = (\Delta_1^{b,(0,t),\mathrm{I}},\ldots,\Delta_p^{b,(0,t),\mathrm{I}})^\top$ are defined as

$$\boldsymbol{\Delta}^{b,(0,t),\mathrm{I}} := -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{b,(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big). \tag{S8.38}$$

For $\breve{\boldsymbol{\beta}}^{b,(t,1)}$, since it is constructed using data both before $\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor$ and after $\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor$, using tedious calculations, we can decompose $\breve{\boldsymbol{\beta}}^{b,(t,1)}$ into

$$\breve{\boldsymbol{\beta}}^{b,(t,1)} = \frac{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor - \lfloor nt\rfloor}{\lfloor nt\rfloor^*}\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})} + \frac{n-\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor nt\rfloor^*}\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)} + \frac{\widehat{\boldsymbol{\Theta}}(\mathbf{X}_{(t,1)})^\top \boldsymbol{\epsilon}_{(t,1)}^b}{\lfloor nt\rfloor^*} + \boldsymbol{\Delta}^{b,(t,1),\mathrm{I}}, \tag{S8.39}$$

where $\boldsymbol{\Delta}^{b,(t,1),\mathrm{I}} = (\Delta_1^{b,(t,1),\mathrm{I}}, \ldots, \Delta_p^{b,(t,1),\mathrm{I}})^\top$ are defined as

$$\begin{aligned}
\boldsymbol{\Delta}^{b,(t,1),\mathrm{I}} \quad &:= -\frac{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor - \lfloor nt\rfloor}{\lfloor nt\rfloor^*}\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,\widehat{t}_{0,\mathcal{G}})} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)} - \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big) \\
&\quad -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,1)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{b,(t,1)} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\big).
\end{aligned} \tag{S8.40}$$

Case 2 : The search location t at $t \in [\widehat{t}_{0,\mathcal{G}}, 1-\tau_0]$. Similar to the analysis of Case 1, using some basic calculations, we can decompose $\breve{\boldsymbol{\beta}}^{b,(0,t)}$ and $\breve{\boldsymbol{\beta}}^{b,(t,1)}$ into

$$\begin{aligned}
\breve{\boldsymbol{\beta}}^{b,(0,t)} &= \frac{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor nt\rfloor}\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})} + \frac{\lfloor nt\rfloor - \lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor nt\rfloor}\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)} + \frac{\widehat{\boldsymbol{\Theta}}(\mathbf{X}_{(0,t)})^\top\boldsymbol{\epsilon}^{b,(0,t)}}{\lfloor nt\rfloor} + \boldsymbol{\Delta}^{b,(0,t),\mathrm{II}}, \\
\breve{\boldsymbol{\beta}}^{b,(t,1)} &= \widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)} + \frac{\widehat{\boldsymbol{\Theta}}(\mathbf{X}_{(t,1)})^\top\boldsymbol{\epsilon}^{b}_{(t,1)}}{\lfloor nt\rfloor^*} + \boldsymbol{\Delta}^{b,(t,1),\mathrm{II}},
\end{aligned} \tag{S8.41}$$

where $\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}} = (\Delta_1^{b,(0,t),\mathrm{II}}, \ldots, \Delta_p^{b,(0,t),\mathrm{II}})^\top$ and $\boldsymbol{\Delta}^{b,(t,1),\mathrm{II}} = (\Delta_1^{b,(t,1)\mathrm{II}}, \ldots, \Delta_p^{b,(t,1),\mathrm{II}})^\top$ are defined as

$$\begin{aligned}
\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}} \quad &:= -\frac{\lfloor nt\rfloor - \lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor nt\rfloor}\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(\widehat{t}_{0,\mathcal{G}},t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\big) \\
&\quad -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{b,(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big), \\
\boldsymbol{\Delta}^{b,(t,1),\mathrm{II}} \quad &:= -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,1)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{b,(t,1)} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\big).
\end{aligned} \tag{S8.42}$$

Based on the above decompositions, we next give a unified form of the debiased lasso estimator for the bootstrap-based samples. To this end, define

$\widehat{\boldsymbol{\delta}}(t) = (\widehat{\delta}_1(t), \ldots, \widehat{\delta}_p(t))^\top$:

$$\widehat{\boldsymbol{\delta}}(t) := \begin{cases} \dfrac{n - \lfloor n\widehat{t}_{0,\mathcal{G}} \rfloor}{n - \lfloor nt \rfloor}\Big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\Big), & \text{for } \ t \in [\tau_0, \widehat{t}_{0,\mathcal{G}}], \\ \\ \dfrac{\lfloor n\widehat{t}_{0,\mathcal{G}} \rfloor}{\lfloor nt \rfloor}\Big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\Big), & \text{for } \ t \in [\widehat{t}_{0,\mathcal{G}}, 1 - \tau_0]. \end{cases} \tag{S8.43}$$

Let $\boldsymbol{\Delta}^{b,(0,t)} = (\Delta_1^{b,(0,t)}, \ldots, \Delta_p^{b,(0,t)})^\top$ and $\boldsymbol{\Delta}^{b,(t,1)} = (\Delta_1^{b,(t,1)}, \ldots, \Delta_p^{b,(t,1)})^\top$ with

$$\boldsymbol{\Delta}^{b,(0,t)} := \boldsymbol{\Delta}^{b,(0,t),\mathrm{I}}\mathbf{1}\big\{t \in [\tau_0, \widehat{t}_{0,\mathcal{G}}]\big\} + \boldsymbol{\Delta}^{b,(0,t),\mathrm{II}}\mathbf{1}\big\{t \in [\widehat{t}_{0,\mathcal{G}}, 1 - \tau_0]\big\},$$

$$\boldsymbol{\Delta}^{b,(t,1)} := \boldsymbol{\Delta}^{b,(t,1),\mathrm{I}}\mathbf{1}\big\{t \in [\tau_0, \widehat{t}_{0,\mathcal{G}}]\big\} + \boldsymbol{\Delta}^{b,(t,1),\mathrm{II}}\mathbf{1}\big\{t \in [\widehat{t}_{0,\mathcal{G}}, 1 - \tau_0]\big\}. \tag{S8.44}$$

With above notations, we are ready to analyze $T_{\mathcal{G}}^b$. Similar to the analysis of Step 1, for each coordinate j at time point $\lfloor nt \rfloor$, we define the coordinate-wise process as

$$C_j^b(\lfloor nt \rfloor) = \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\big(\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j}\big)^{-1/2}\big(\breve{\beta}_j^{b,(0,t)} - \breve{\beta}_j^{b,(t,1)} - \widehat{\delta}_j(t)\big). \tag{S8.45}$$

By the definition of $T_{\mathcal{G}}^b$ in (2.23) , we have $T_{\mathcal{G}}^b = \max\limits_{t \in [\tau_0, 1-\tau_0]} \max\limits_{j \in \mathcal{G}} |C_j^b(\lfloor nt \rfloor)|$. Furthermore, by (S8.37), (S8.39), (S8.41), and (S8.44), we can decompose $C_j^b(\lfloor nt \rfloor)$ into

$$C_j^b(\lfloor nt \rfloor) = C_j^{b,\mathrm{I}}(\lfloor nt \rfloor) + C_j^{b,\mathrm{II}}(\lfloor nt \rfloor), \tag{S8.46}$$

with

$$C_j^{b,\mathrm{I}}(\lfloor nt \rfloor) := \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\frac{\left(\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}\widehat{\boldsymbol{\Theta}}_j^\top \mathbf{X}_i\epsilon_i^b - \frac{1}{\lfloor nt \rfloor^*}\sum_{i=\lfloor nt \rfloor+1}^{n}\widehat{\boldsymbol{\Theta}}_j^\top \mathbf{X}_i\epsilon_i^b\right)}{\sqrt{\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j}}},$$

$$C_j^{b,\mathrm{II}}(\lfloor nt \rfloor) := \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\frac{\left(\Delta_j^{b,(0,t)} - \Delta_j^{b,(t,1)}\right)}{\sqrt{\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j}}}, \text{ with } 1 \le j \le p \text{ and } t \in [\tau_0, 1-\tau_0]. \tag{S8.47}$$

By replacing $\widehat{\sigma}_\epsilon^2$, $\widehat{\omega}_{j,j}$, and $\widehat{\boldsymbol{\Theta}}_j$ by their true values σ_ϵ^2, $\omega_{j,j}$, and $\boldsymbol{\Theta}_j$, for the bootstrap based process $C_j^{b,\mathrm{I}}(\lfloor nt \rfloor)$, we can define the oracle leading term as follows:

$$\widetilde{C}_j^{b,\mathrm{I}}(\lfloor nt \rfloor) := \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\frac{\left(\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}\boldsymbol{\Theta}_j^\top \mathbf{X}_i\epsilon_i^b - \frac{1}{\lfloor nt \rfloor^*}\sum_{i=\lfloor nt \rfloor+1}^{n}\boldsymbol{\Theta}_j^\top \mathbf{X}_i\epsilon_i^b\right)}{\sqrt{\sigma_\epsilon^2\omega_{j,j}}}. \tag{S8.48}$$

Based on (S8.46), (S8.47), and (S8.48), define the following four vector-valued processes:

$$\boldsymbol{C}^b(\lfloor nt \rfloor) = \left(C_1^b(\lfloor nt \rfloor), \ldots, C_p^b(\lfloor nt \rfloor)\right)^\top, \qquad \boldsymbol{C}^{b,\mathrm{I}}(\lfloor nt \rfloor) = \left(C_1^{b,\mathrm{I}}(\lfloor nt \rfloor), \ldots, C_p^{b,\mathrm{I}}(\lfloor nt \rfloor)\right)^\top,$$

$$\boldsymbol{C}^{b,\mathrm{II}}(\lfloor nt \rfloor) = \left(C_1^{b,\mathrm{II}}(\lfloor nt \rfloor), \ldots, C_p^{b,\mathrm{II}}(\lfloor nt \rfloor)\right)^\top, \quad \widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt \rfloor) = \left(\widetilde{C}_1^{b,\mathrm{I}}(\lfloor nt \rfloor), \ldots, \widetilde{C}_p^{b,\mathrm{I}}(\lfloor nt \rfloor)\right)^\top. \tag{S8.49}$$

The following Lemma 11 shows that the residual term $C_j^{b,\mathrm{II}}(\lfloor nt \rfloor)$ can be uniformly negligible over $t \in [\tau_0, 1-\tau_0]$ and $1 \le j \le p$. The proof of Lemma 11 is given in Section S9.2.

Lemma 11. *Assume Assumptions* (**A.1**) – (**A.5**) *hold. Under* $\mathbf{H}_0$*, we have*

$$\mathbb{P}\Big(\max_{\tau_0\le t\le 1-\tau_0}\big\|\boldsymbol{C}^b(\lfloor nt\rfloor)-\widetilde{\boldsymbol{C}}^{b,I}(\lfloor nt\rfloor)\big\|_{\mathcal{G},\infty}\ge \epsilon|\mathcal{X}\Big)=o(1), \qquad \text{(S8.50)}$$

where $\epsilon = C\max(\max_{1\le j\le p} s_j\frac{\log(pn)}{\sqrt{n}}, s\sqrt{n}\frac{\log(pn)}{\lfloor n\tau_0\rfloor})$*, and* C *is a universal constant not depending on* n *or* p.

Step 3 (Gaussian approximation). In Step 1, we have defined the oracle leading term $\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)$. Let

$$\mathbf{V}=\mathrm{diag}\big((\omega_{1,1}\sigma_\epsilon^2)^{-\frac{1}{2}},\ldots,(\omega_{p,p}\sigma_\epsilon^2)^{-\frac{1}{2}}\big). \qquad \text{(S8.51)}$$

By the definition of $\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)$ in (S8.35), we can rewrite it in the form of partial sum process:

$$\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{V}\cdot\boldsymbol{\Theta}\mathbf{X}_i\epsilon_i\Big(\mathbf{1}(i\le\lfloor nt\rfloor)-\frac{\lfloor nt\rfloor}{n}\Big),\ \text{with } \tau_0\le t\le 1-\tau_0. \qquad \text{(S8.52)}$$

In Step 2, we have introduced the oracle leading term $\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)$ in (S8.49) for the bootstrap based test statistic. Similar to $\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)$, we can write $\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)$ in the following form:

$$\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{V}\cdot\boldsymbol{\Theta}\mathbf{X}_i\epsilon_i^b\big(\mathbf{1}(i\le\lfloor nt\rfloor)-\lfloor nt\rfloor/n\big),\ \text{with } \tau_0\le t\le 1-\tau_0. \qquad \text{(S8.53)}$$

Let $\boldsymbol{Z}_i=\mathbf{V}\cdot\boldsymbol{\Theta}\mathbf{X}_i\epsilon_i$ and $\boldsymbol{G}_i=\mathbf{V}\cdot\boldsymbol{\Theta}\mathbf{X}_i\epsilon_i^b$ for $i=1,\ldots,n$. Note that $\boldsymbol{G}_i$ follows multivariate Gaussian distributions with mean zero and the same covariance matrix as $\boldsymbol{Z}_i$. We aim to use $\max_{\tau_0\le t\le 1-\tau_0}\|\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty}$ to approximate $\max_{\tau_0\le t\le 1-\tau_0}\|\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty}$. Hence, it remains to verify that the conditions of Lemma 1 hold. In fact, by Assumptions (**A.1**) and (**A.2**), we can show that Assumptions (**M1**) - (**M3**) hold for $\mathbf{V}\cdot\boldsymbol{\Theta}\mathbf{X}_i\epsilon_i$ with $1\le i\le n$. Hence, by Lemma 1, we have

$$\sup_{z\in(0,\infty)}\left|\mathbb{P}\big(\max_{\tau_0\le t\le 1-\tau_0}\|\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty}\le z\big)-\mathbb{P}\big(\max_{\tau_0\le t\le 1-\tau_0}\|\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty}\le z\big)\right|\le Cn^{-\zeta_0}. \tag{S8.54}$$

Step 4. In this step, we aim to combine the previous results to prove

$$\sup_{z\in(0,\infty)}\left|\mathbb{P}(T_{\mathcal{G}}\le z)-\mathbb{P}(T_{\mathcal{G}}^b\le z|\mathcal{X})\right|=o_p(1),\ \text{as } n,p\to\infty. \tag{S8.55}$$

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

$$\rho_0:=\mathbb{P}(T_{\mathcal{G}}>z)-\mathbb{P}(T_{\mathcal{G}}^b>z|\mathcal{X}). \tag{S8.56}$$

We first consider the upper bound. Note that $T_{\mathcal{G}}=\max\limits_{t\in[\tau_0,1-\tau_0]}\|\boldsymbol{C}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty}$. By plugging $\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)$ in $T_{\mathcal{G}}$ and using the triangle inequality of $\|\cdot\|_{\mathcal{G},\infty}$, we have

$$\mathbb{P}(T_{\mathcal{G}}>z)\le\mathbb{P}\big(\max_{t\in[\tau_0,1-\tau_0]}\|\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty}>z-\epsilon\big)+\rho_1, \tag{S8.57}$$

where $\rho_1:=\mathbb{P}\big(\max\limits_{t\in[\tau_0,1-\tau_0]}\|\boldsymbol{C}(\lfloor nt\rfloor)-\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty}>\epsilon\big)$. By Lemma 10, we have $\rho_1=o(1)$. Recall $\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)$ defined in (S8.49). For $\mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]}\|\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty}>$

$z - \epsilon$), we then have

$$\mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]} \|\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} > z-\epsilon) \le \underbrace{\mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]} \|\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} > z-\epsilon|\mathcal{X})}_{\rho_3} + \rho_2, \tag{S8.58}$$

where

$$\rho_2 := \sup_{x\in(0,\infty)} |\mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]} \|\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} > x) - \mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]} \|\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} > x|\mathcal{X})|.$$

By Step 3, we have proved $\rho_2 \le Cn^{-\zeta_0}$ holds. For ρ_3, we have $\rho_3 = \rho_4 + \rho_5$, where

$$\begin{aligned} \rho_4 &:= \mathbb{P}(z-\epsilon < \max_{t\in[\tau_0,1-\tau_0]} \|\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} < z+\epsilon|\mathcal{X}), \\ \rho_5 &:= \mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]} \|\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} > z+\epsilon|\mathcal{X}). \end{aligned} \tag{S8.59}$$

By Lemma 2, we have proved $\rho_4 = o_p(1)$. So far, we have proved that

$$\mathbb{P}(T_{\mathcal{G}} > z) \le \mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]} \|\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} > z+\epsilon|\mathcal{X}) + o_p(1). \tag{S8.60}$$

Note that $T_{\mathcal{G}}^b := \max_{t\in[\tau_0,1-\tau_0]} \|\boldsymbol{C}^b(\lfloor nt\rfloor)\|_{\mathcal{G},\infty}$. By the triangle inequality, we have

$$\mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]} \|\widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} > z+\epsilon|\mathcal{X}) \le \mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]} \|\boldsymbol{C}^{b}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} > z|\mathcal{X}) + \rho_6, \tag{S8.61}$$

where $\rho_6 := \mathbb{P}(\max_{t\in[\tau_0,1-\tau_0]} \|\boldsymbol{C}^b(\lfloor nt\rfloor) - \widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\|_{\mathcal{G},\infty} > \epsilon|\mathcal{X})$. By Lemma 11, we have proved $\rho_6 = o_p(1)$. Combining (S8.60) and (S8.61), we have

$$\mathbb{P}(T_{\mathcal{G}} > z) \le \mathbb{P}(T_{\mathcal{G}}^b > z|\mathcal{X}) + o_p(1). \tag{S8.62}$$

With a similar proof technique, we can also obtain the lower bound and prove

$$|\mathbb{P}(T_{\mathcal{G}} > z) - \mathbb{P}(T_{\mathcal{G}}^b > z|\mathcal{X})| = o_p(1) \tag{S8.63}$$

holds uniformly in $z \in (0, \infty)$, which finishes the proof of Theorem 2. $\square$

S8.4 Proof of Theorem 3

Proof. Without loss of generality, we assume $\delta_j := \beta_j^{(1)} - \beta_j^{(2)} \geq 0$. As a mild technical assumption, throughout this section, we assume $s\sqrt{\log(p)/n\tau_0}\|\boldsymbol{\delta}\|_\infty/\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} = o(1)$. For each $t \in [\tau_0, 1-\tau_0]$, define $\boldsymbol{Z}(\lfloor nt \rfloor) = \big(Z_1(\lfloor nt \rfloor), \ldots, Z_p(\lfloor nt \rfloor)\big)^\top$ with

$$Z_j(\lfloor nt \rfloor) := \sqrt{n}\frac{\lfloor nt \rfloor}{n}\Big(1 - \frac{\lfloor nt \rfloor}{n}\Big)(\breve{\beta}_j^{(0,t)} - \breve{\beta}_j^{(t,1)}), \text{ for } 1 \leq j \leq p. \tag{S8.64}$$

Note that there is no variance estimator in $Z_j(\lfloor nt \rfloor)$. By definition, we have

$$\widehat{t}_{0,\mathcal{G}} := \underset{t \in [\tau_0, 1-\tau_0]}{\arg\max} \|\boldsymbol{Z}(\lfloor nt \rfloor)\|_{\mathcal{G},\infty}.$$

For notational simplicity, we abbreviate $\widehat{t}_{0,\mathcal{G}}$ to $\widehat{t}_0$. Moreover, we assume $\widehat{t}_0 \in [t_0, 1-\tau_0]$. To give the proof, we need to make decompositions on $\boldsymbol{Z}(\lfloor nt \rfloor)$. We first define $\boldsymbol{\delta}(t) = (\delta_1(t), \ldots, \delta_p(t))^\top$:

$$\begin{aligned}\boldsymbol{\delta}(t) := \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt_0 \rfloor^*}{n}\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big)\mathbf{1}\{t \in [\tau_0, t_0]\} \\ + \sqrt{n}\frac{\lfloor nt_0 \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big)\mathbf{1}\{t \in [t_0, 1-\tau_0]\},\end{aligned} \tag{S8.65}$$

and $\boldsymbol{R}^{(0,t)} = (R_1^{(0,t)}, \ldots, R_p^{(0,t)})^\top$ and $\boldsymbol{R}^{(t,1)} = (R_1^{(t,1)}, \ldots, R_p^{(t,1)})^\top$:

$$\boldsymbol{R}^{(0,t)} := \boldsymbol{R}^{(0,t),\mathrm{I}} \mathbf{1}\big\{t \in [\tau_0, t_0]\big\} + \boldsymbol{R}^{(0,t),\mathrm{II}} \mathbf{1}\big\{t \in [t_0, 1-\tau_0]\big\},$$

$$\boldsymbol{R}^{(t,1)} := \boldsymbol{R}^{(t,1),\mathrm{I}} \mathbf{1}\big\{t \in [\tau_0, t_0]\big\} + \boldsymbol{R}^{(t,1),\mathrm{II}} \mathbf{1}\big\{t \in [t_0, 1-\tau_0]\big\}, \tag{S8.66}$$

where $\boldsymbol{R}^{(0,t),\mathrm{I}} - \boldsymbol{R}^{(t,1),\mathrm{II}}$ are defined as

$$\begin{aligned}
\boldsymbol{R}^{(0,t),\mathrm{I}} &:= -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(1)}\big), \\
\boldsymbol{R}^{(0,t),\mathrm{II}} &:= -\frac{\lfloor nt \rfloor - \lfloor nt_0 \rfloor}{\lfloor nt \rfloor}\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t_0,t)} - \mathbf{I}\big)\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big) - \big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(1)}\big), \\
\boldsymbol{R}^{(t,1),\mathrm{I}} &:= -\frac{\lfloor nt_0 \rfloor - \lfloor nt \rfloor}{\lfloor nt \rfloor^*}\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,t_0)} - \mathbf{I}\big)\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big) - \big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,1)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(2)}\big), \\
\boldsymbol{R}^{(t,1),\mathrm{II}} &:= -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t,1)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(t,1)} - \boldsymbol{\beta}^{(2)}\big).
\end{aligned} \tag{S8.67}$$

Then, by the definitions of $\breve{\boldsymbol{\beta}}^{(0,t)}$ and $\breve{\boldsymbol{\beta}}^{(t,1)}$, similar to the analysis of Step 2 in Section S8.3, under $\mathbf{H}_1$, we can write $\boldsymbol{Z}(\lfloor nt \rfloor)$ as follows:

$$\boldsymbol{Z}(\lfloor nt \rfloor) = \boldsymbol{\delta}(t) + \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\Big(\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}\widehat{\boldsymbol{\xi}}_i - \frac{1}{\lfloor nt \rfloor^*}\sum_{i=\lfloor nt \rfloor+1}^{n}\widehat{\boldsymbol{\xi}}_i + \boldsymbol{R}^{(0,t)} - \boldsymbol{R}^{(t,1)}\Big), \tag{S8.68}$$

where $\widehat{\boldsymbol{\xi}}_i := (\widehat{\xi}_{i,1}, \ldots, \widehat{\xi}_{i,p})^\top$ with $\widehat{\xi}_{i,j} = \widehat{\boldsymbol{\Theta}}_j^\top \mathbf{X}_i \epsilon_i$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$.

In addition to the decomposition, let $\boldsymbol{\delta} = \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}$ and we assume

$$\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} \gg \sqrt{\frac{\log(|\mathcal{G}|n)}{n}}.$$

Let $j^* \in \mathcal{G}$ such that $Z_{j^*}(\lfloor n\widehat{t}_0 \rfloor) = \max_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0 \rfloor)$. The following Lemma 12 shows that $\liminf_{n\to\infty} \delta_{j^*}/\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} \geq 1$. The proof of Lemma 12 is given in Section S9.3.

Lemma 12. *Suppose Assumptions* (**A**.**1**) – (**A**.**5**) *hold. Then, with probability tending to one, we have* $\liminf_{n\to\infty} \delta_{j^*}/\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} \geq 1$.

Furthermore, define the event

$$\begin{aligned}
\mathcal{H}_1 &= \Big\{ \max\nolimits_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0 \rfloor) = \max\nolimits_{j\in\mathcal{G}} |Z_j(\lfloor n\widehat{t}_0 \rfloor)| := \|\boldsymbol{Z}(\lfloor n\widehat{t}_0 \rfloor)\|_{\mathcal{G},\infty} \Big\}, \\
\mathcal{H}_2 &= \Big\{ Z_{j^*}(\lfloor nt_0 \rfloor) = |Z_{j^*}(\lfloor nt_0 \rfloor)| \Big\}.
\end{aligned} \tag{S8.69}$$

The following Lemma 13 shows that $\mathcal{H}_1 \cap \mathcal{H}_2$ occurs with high probability. The proof of Lemma 13 is provided in Section S9.4.

Lemma 13. *Suppose Assumptions* (**A**.**1**) – (**A**.**5**) *hold. Then we have*

$$\mathbb{P}(\mathcal{H}_1 \cap \mathcal{H}_2) \geq 1 - C_1(np)^{-C_2}, \tag{S8.70}$$

where C_1 and C_2 are universal positive constants not depending on n or p.

Using Lemmas 12 and 13, we are ready to give the proof. Specifically, by the above two lemmas, we have:

$$\begin{aligned}\|\boldsymbol{Z}(\lfloor nt_0\rfloor)\|_{\mathcal{G},\infty}-\|\boldsymbol{Z}(\lfloor n\widehat{t}_0\rfloor)\|_{\mathcal{G},\infty} &= \|\boldsymbol{Z}(\lfloor nt_0\rfloor)\|_{\mathcal{G},\infty}-Z_{j^*}(\lfloor n\widehat{t}_0\rfloor)\\ &\geq Z_{j^*}(\lfloor nt_0\rfloor)-Z_{j^*}(\lfloor n\widehat{t}_0\rfloor).\end{aligned}$$

Moreover, by the decomposition of $\boldsymbol{Z}(\lfloor nt\rfloor)$ in (S8.68), we have:

$$Z_{j^*}(\lfloor nt_0\rfloor)-Z_{j^*}(\lfloor n\widehat{t}_0\rfloor)\geq\sqrt{n}\frac{\lfloor nt_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor-\lfloor nt_0\rfloor}{n}\delta_{j^*}+I-II, \quad \text{(S8.71)}$$

where

$$\begin{aligned}I &= \frac{1}{\sqrt{n}}\Big(\sum_{i=\lfloor nt_0\rfloor+1}^{\lfloor n\widehat{t}_0\rfloor}\widehat{\xi}_{i,j}-\frac{\lfloor n\widehat{t}_0\rfloor-\lfloor nt_0\rfloor}{n}\sum_{i=1}^{n}\widehat{\xi}_{i,j}\Big),\\ II &= \sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}\big(\|\boldsymbol{R}^{(0,\widehat{t}_0),II}\|_\infty+\|\boldsymbol{R}^{(\widehat{t}_0,1),II}\|_\infty\big)\\ &\quad+\sqrt{n}\frac{\lfloor nt_0\rfloor}{n}\frac{\lfloor nt_0\rfloor^*}{n}\big(\|\boldsymbol{R}^{(0,t_0),II}\|_\infty+\|\boldsymbol{R}^{(t_0,1),II}\|_\infty\big).\end{aligned} \quad \text{(S8.72)}$$

Note that by Assumptions (**A**.**1**) – (**A**.**3**), $\widehat{\xi}_{i,j}$ follows the sub-exponential distribution. Using Bernstein's inequalities, we can prove that:

$$\max_{t\in[t_0,1-\tau_0]}\max_{j\in\mathcal{G}}\frac{|\frac{1}{\sqrt{n}}\big(\sum_{i=\lfloor nt_0\rfloor+1}^{\lfloor nt\rfloor}\widehat{\xi}_{i,j}-\frac{\lfloor nt\rfloor-\lfloor nt_0\rfloor}{n}\sum_{i=1}^{n}\widehat{\xi}_{i,j}\big)|}{(\lfloor nt\rfloor-\lfloor nt_0\rfloor)^{1/2}}=O_p\Big(\sqrt{\frac{\log(|\mathcal{G}|)}{n}}\Big). \quad \text{(S8.73)}$$

Moreover, the following Lemma 14 shows that II can be decomposed into three terms. The proof of Lemma 14 is given in Section S9.5.

Lemma 14. *Suppose Assumptions* (**A**.**1**) – (**A**.**5**) *hold. For II in (S8.71), with probability tending to 1, we have*

$$II\leq C_1\sqrt{\log(|\mathcal{G}|n)\frac{\lfloor n\widehat{t}_0\rfloor-\lfloor nt_0\rfloor}{n}}s\|\boldsymbol{\delta}\|_\infty+C_2\sqrt{n}s\frac{\log(|\mathcal{G}|n)}{n}+o\Big(\sqrt{n}\frac{\lfloor nt_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor-\lfloor nt_0\rfloor}{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\Big).$$

where $C_1, C_2 > 0$ are some constants not depending on n or p.

Considering (S8.71) - (S8.73), by Lemmas 12 and 14, we have:

$$
\begin{aligned}
&\|\boldsymbol{Z}(\lfloor nt_0 \rfloor)\|_{\mathcal{G},\infty} - \|\boldsymbol{Z}(\lfloor n\widehat{t}_0 \rfloor)\|_{\mathcal{G},\infty} \\
&\geq Z_{j^*}(\lfloor nt_0 \rfloor) - Z_{j^*}(\lfloor n\widehat{t}_0 \rfloor) \\
&\geq \sqrt{n}\frac{\lfloor nt_0 \rfloor}{n}\frac{\lfloor n\widehat{t}_0 \rfloor - \lfloor nt_0 \rfloor}{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} - C_1\sqrt{\frac{\lfloor n\widehat{t}_0 \rfloor - \lfloor nt_0 \rfloor \log(|\mathcal{G}|)}{n}} \\
&-C_2\sqrt{\log(|\mathcal{G}|n)\frac{\lfloor n\widehat{t}_0 \rfloor - \lfloor nt_0 \rfloor}{n}}s\|\boldsymbol{\delta}\|_{\infty} - C_3\sqrt{n}s\frac{\log(|\mathcal{G}|n)}{n} - o\big(\sqrt{n}\frac{\lfloor nt_0 \rfloor}{n}\frac{\lfloor n\widehat{t}_0 \rfloor - \lfloor n\tau_0 \rfloor}{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\big).
\end{aligned}
\tag{S8.74}
$$

Note that $\|\boldsymbol{Z}(\lfloor nt_0 \rfloor)\|_{\mathcal{G},\infty} - \|\boldsymbol{Z}(\lfloor n\widehat{t}_0 \rfloor)\|_{\mathcal{G},\infty} \leq 0$. Hence, by (S8.73), we have:

$$
\begin{aligned}
&\frac{1}{2}\sqrt{n}\frac{\lfloor nt_0 \rfloor}{n}\frac{\lfloor n\widehat{t}_0 \rfloor - \lfloor nt_0 \rfloor}{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} \\
&\leq 3\max\Big(C_1\sqrt{\frac{\lfloor n\widehat{t}_0 \rfloor - \lfloor nt_0 \rfloor \log(|\mathcal{G}|)}{n}}, C_2\sqrt{\log(|\mathcal{G}|n)\frac{\lfloor n\widehat{t}_0 \rfloor - \lfloor nt_0 \rfloor}{n}}s\|\boldsymbol{\delta}\|_{\infty}, C_3\sqrt{n}s\frac{\log(|\mathcal{G}|n)}{n}\Big).
\end{aligned}
$$

This implies that with probability tending to 1, we must have

$$
\frac{\lfloor n\widehat{t}_0 \rfloor - \lfloor nt_0 \rfloor}{n} \leq C^*\max\Big(\frac{\log(|\mathcal{G}|)}{n\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}^2}, \frac{\log(|\mathcal{G}|)s^2\|\boldsymbol{\delta}\|_{\infty}^2}{n\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}^2}, \frac{\log(|\mathcal{G}|)s\|\boldsymbol{\delta}\|_{\infty}}{n\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}^2}\Big) \leq C^*\frac{\log(|\mathcal{G}|)}{n\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}^2},
$$

where the second inequality comes from Assumption (**A.6**), and C^* is some big enough constant not depending on n or p, which completes the proof of Theorem 3. □

S8.5 Proof of Theorem 4

Proof. Note that by (S8.25) in Theorem 1 and by Assumption (**A.4**), we have shown that $\max_{1\leq j,k\leq p}|\widehat{\omega}_{j,k} - \omega_{j,k}| = o_p(1)$. Hence, to prove Theorem 4, it remains to prove that $|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| = o_p(1)$, where $\widehat{\sigma}_\epsilon^2$ is the weighted variance estimator as defined in (2.19). Without loss of generality, we assume the change point estimator $\widehat{t}_{0,\mathcal{G}} \in [\tau_0, t_0]$, where $\widehat{t}_{0,\mathcal{G}}$ is obtained by (2.18). To simplify notations, we denote $\widehat{t}_{0,\mathcal{G}}$ by $\widehat{t}_0$. Thoughout this section, we denote

$$\epsilon_n := \frac{\lfloor nt_0 \rfloor - \lfloor n\widehat{t}_0 \rfloor}{n}, \quad \text{and } \boldsymbol{\delta} = \boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}.$$

Furthermore, by definition, we can write $\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2$ as the following 8 parts:

$$\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2 = I + II + III + IV + V + VI + VII + VIII, \tag{S8.75}$$

where $I - VIII$ are defined as

$$\begin{aligned}
&I = \frac{1}{n}\sum_{i=1}^{n}(\epsilon_i^2 - \sigma_\epsilon^2), && II = \frac{1}{n}\Big\|\mathbf{X}_{(0,\widehat{t}_0)}\big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)} - \boldsymbol{\beta}^{(1)}\big)\Big\|_2^2,\\
&III = \frac{2}{n}(\boldsymbol{\epsilon}_{(0,\widehat{t}_0)})^\top \mathbf{X}_{(0,\widehat{t}_0)}\big(\boldsymbol{\beta}^{(1)} - \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)}\big), && IV = \frac{1}{n}\Big\|\mathbf{X}_{(\widehat{t}_0,1)}\big(\widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)} - \boldsymbol{\beta}^{(2)}\big)\Big\|_2^2,\\
&V = \frac{2}{n}(\boldsymbol{\epsilon}_{(\widehat{t}_0,1)})^\top \mathbf{X}_{(\widehat{t}_0,1)}\big(\boldsymbol{\beta}^{(2)} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)}\big), && VI = \frac{2}{n}(\boldsymbol{\epsilon}_{(\widehat{t}_0,t_0)})^\top \mathbf{X}_{(\widehat{t}_0,t_0)}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}),\\
&VII = \frac{1}{n}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})^\top(\mathbf{X}_{(\widehat{t}_0,t_0)})^\top \mathbf{X}_{(\widehat{t}_0,t_0)}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}),\\
&VIII = \frac{1}{n}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})^\top(\mathbf{X}_{(\widehat{t}_0,t_0)})^\top \mathbf{X}_{(\widehat{t}_0,t_0)}(\boldsymbol{\beta}^{(2)} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)}).
\end{aligned} \tag{S8.76}$$

By (S8.75), we need to bound the eight parts on the RHS of (S8.75), respectively. For the rest of the proof, we assume the event $\cap_{t\in[\tau_0,1-\tau_0]}\{\mathcal{A}(t)\cap\mathcal{B}(t)\}$ holds. For I, using (S8.23) in Theorem 1, we have $I = o(1)$ as $n, p \to \infty$. For II, by Lemma 8, we have $II \le Cs^{(1)}\frac{\log(p)}{n} = o(1)$ as $n, p \to \infty$. For

III, by Lemma 8 and Assumption (**A**.**4**), we have

$$
\begin{aligned}
III \quad &\leq C\frac{\lfloor n\widehat{t}_0\rfloor}{n}\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_0\rfloor}}\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)}-\boldsymbol{\beta}^{(1)}\|_1,\\
&\leq Cs^{(1)}\frac{\log(p)}{\lfloor n\widehat{t}_0\rfloor}\leq Cs^{(1)}\frac{\log(p)}{\lfloor n\tau_0\rfloor}=o_p(1).
\end{aligned}
$$

Recall $s=s^{(1)}\vee s^{(2)}$. For IV, by Lemma 8 and Assumption (**A**.**4**), we have

$$
\begin{aligned}
IV \quad &\leq C\Big(\frac{\lfloor nt_0\rfloor-\lfloor n\widehat{t}_0\rfloor}{\lfloor n\widehat{t}_0\rfloor^*}\Big)^2\big\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\big\|_2^2,\\
&\leq C\Big(\frac{\lfloor nt_0\rfloor-\lfloor n\widehat{t}_0\rfloor}{n}\Big)^2\big\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\big\|_2^2=O_p(\epsilon_n^2 s\|\boldsymbol{\delta}\|_\infty^2).
\end{aligned}
$$

For V, by Lemma 8 and Assumption (**A**.**4**), we have

$$
\begin{aligned}
|V| \quad &\leq C\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_0\rfloor^*}}\big\|\widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)}-\boldsymbol{\beta}^{(2)}\big\|_1,\\
&\leq C\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_0\rfloor^*}}\times\frac{\lfloor nt_0\rfloor-\lfloor n\widehat{t}_0\rfloor}{\lfloor n\widehat{t}_0\rfloor^*}\big\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\big\|_1,\\
&\leq C\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_0\rfloor^*}}\frac{\lfloor nt_0\rfloor-\lfloor n\widehat{t}_0\rfloor}{n}\big\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\big\|_1=O_p(\epsilon_n s\sqrt{\frac{\log(p)}{n}}\|\boldsymbol{\delta}\|_\infty).
\end{aligned}
$$

For VI, by Assumptions (**A**.**1**) and (**A**.**3**), we have

$$
\begin{aligned}
|VI| \quad &\leq C\frac{\lfloor nt_0\rfloor-\lfloor n\widehat{t}_0\rfloor}{n}\sqrt{\frac{\log(p)}{\lfloor nt_0\rfloor-\lfloor n\widehat{t}_0\rfloor}}\big\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\big\|_1,\\
&\leq C\sqrt{\frac{\lfloor nt_0\rfloor-\lfloor n\widehat{t}_0\rfloor}{n}}\sqrt{\frac{\log(p)}{n}}\big\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\big\|_1=O_p(s\sqrt{\epsilon_n\frac{\log(p)}{n}}\|\boldsymbol{\delta}\|_\infty).
\end{aligned}
$$

For VII, using the fact that $\boldsymbol{x}^\top \mathbf{A}\boldsymbol{x} \leq \|\mathbf{A}\|_\infty \|\boldsymbol{x}\|_1^2$, we have

$$\begin{aligned}
VII \quad &=_{(1)} \frac{\lfloor nt_0 \rfloor - \lfloor n\widehat{t}_0 \rfloor}{n} (\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})^\top \widehat{\boldsymbol{\Sigma}}_{(\widehat{t}_0, t_0)} (\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}) \\
&=_{(2)} \frac{\lfloor nt_0 \rfloor - \lfloor n\widehat{t}_0 \rfloor}{n} (\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})^\top (\widehat{\boldsymbol{\Sigma}}_{(\widehat{t}_0, t_0)} - \boldsymbol{\Sigma}) (\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}) \\
&\quad + \frac{\lfloor nt_0 \rfloor - \lfloor n\widehat{t}_0 \rfloor}{n} (\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)})^\top \boldsymbol{\Sigma} (\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}) \\
&\leq_{(3)} C_1 \sqrt{\epsilon_n \frac{\log(p)}{n}} \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_1^2 + C_2 \epsilon_n \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2^2 \\
&=_{(4)} O_p(s^2 \sqrt{\epsilon_n \frac{\log(p)}{n}} \|\boldsymbol{\delta}\|_\infty^2 + \epsilon_n s \|\boldsymbol{\delta}\|_\infty^2),
\end{aligned}$$

where (3) comes from the concentration inequality for $\|\widehat{\boldsymbol{\Sigma}}_{(\widehat{t}_0, t_0)} - \boldsymbol{\Sigma}\|_\infty$ and by Assumption (**A**.**3**) that $\Sigma_{j,j} = O(1)$. Lastly, for $VIII$, by Lemma 8, and similar to VII, we have

$$VIII = O_p(\epsilon_n^2 s^2 \|\boldsymbol{\delta}\|_\infty^2 + s^2 \sqrt{\epsilon_n^3 \frac{\log(p)}{n}} \|\boldsymbol{\delta}\|_\infty^2).$$

Combining the obtained upper bounds of $I, \ldots, VIII$, we have

$$\begin{aligned}
|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| &= o_p(1) + O_p(\epsilon_n^2 s \|\boldsymbol{\delta}\|_\infty^2) + O_p(\epsilon_n s \sqrt{\frac{\log(p)}{n}} \|\boldsymbol{\delta}\|_\infty) + O_p(s \sqrt{\epsilon_n \frac{\log(p)}{n}} \|\boldsymbol{\delta}\|_\infty) \\
&+ O_p(\epsilon_n s \|\boldsymbol{\delta}\|_\infty^2 + s^2 \sqrt{\epsilon_n \frac{\log(p)}{n}} \|\boldsymbol{\delta}\|_\infty^2) + O_p(\epsilon_n^2 s^2 \|\boldsymbol{\delta}\|_\infty^2 + s^2 \sqrt{\epsilon_n^3 \frac{\log(p)}{n}} \|\boldsymbol{\delta}\|_\infty^2).
\end{aligned} \tag{S8.77}$$

By (S8.77), to bound $|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2|$, we consider the following two cases:

Case1 : The signal satisfies $\|\boldsymbol{\delta}\|_\infty \gg \sqrt{\log(p)/n}$. In this case, by Theorem 3, we have $\epsilon_n = o_p(1)$. Moreover, by Assumption (**A**.**6**), we have $s\|\boldsymbol{\delta}\|_\infty = O(1)$ and $\|\boldsymbol{\delta}\|_\infty = o(1)$. Combining (S8.77), we have $|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| = o_p(1)$.

Case2 : The signal satisfies $\|\boldsymbol{\delta}\|_\infty = O(\sqrt{\log(p)/n})$. In this case, we can not obtain a consistent change point estimator. In other words, we only have $\epsilon_n = O_p(1)$. Moreover, we can show that

$$|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| = O_p(s^2\|\boldsymbol{\delta}\|_\infty^2). \tag{S8.78}$$

Considering (S8.78), and by the assumption that $s\sqrt{\log(p)/n} = o(1)$, we have $|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| = o_p(1)$, which finishes the proof. $\square$

S8.6 Proof of Theorem 5

Proof. As a very mild technical assumption, throughout this section, we assume

$$s\sqrt{\log(p)/n\tau_0}\|\boldsymbol{\delta}\|_\infty/\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} = o(1).$$

The proof of Theorem 5 proceeds in two steps. In Step 1, we obtain the upper bound of $c_{T_\mathcal{G}^b}(1-\alpha)$, where $c_{T_\mathcal{G}^b}(1-\alpha)$ is the $1-\alpha$ quantile of $T_\mathcal{G}^b$, which is defined as

$$c_{T_\mathcal{G}^b}(1-\alpha) := \inf\left\{t : \mathbb{P}(T_\mathcal{G}^b \le t|\mathcal{X}) \ge 1-\alpha\right\}. \tag{S8.79}$$

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

$$\mathbb{P}\big(T_\mathcal{G} \ge c_{T_\mathcal{G}^b}(1-\alpha)\big) \to 1, \text{ as } n, p \to \infty. \tag{S8.80}$$

Note that $\{\Phi_{\mathcal{G},\alpha} = 1\} \Leftrightarrow \{T_\mathcal{G} \ge \widehat{c}_{T_\mathcal{G}^b}(1-\alpha)\}$, where

$$\widehat{c}_{T_\mathcal{G}^b}(1-\alpha) := \inf\left\{t : \frac{1}{B+1}\sum_{b=1}^{B}\mathbf{1}\{T_\mathcal{G}^b \le t|\mathcal{X}\} \ge 1-\alpha\right\}. \tag{S8.81}$$

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

Step 1: In this step, we aim to obtain the upper bound for $c_{T_{\mathcal{G}}^b}(1-\alpha)$. Define $\widehat{\xi}_{i,j}^b = \widehat{\boldsymbol{\Theta}}_j^\top \boldsymbol{X}_i \epsilon_i^b$ for $1 \le i \le n$ and $1 \le j \le p$. Recall $C_j^b(\lfloor nt \rfloor)$ in (S8.45) and the decomposition in (S8.47). By the definition of $T_{\mathcal{G}}^b$ and using the fact that $\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n} \le 1$ with $t \in [\tau_0, 1-\tau_0]$, we have

$$
\begin{aligned}
T_{\mathcal{G}}^b &\le \max_{t\in[\tau_0,1-\tau_0]} \max_{j\in\mathcal{G}} |C_j^{b,\mathrm{I}}(\lfloor nt \rfloor)| + \max_{t\in[\tau_0,1-\tau_0]} \max_{j\in\mathcal{G}} |C_j^{b,\mathrm{II}}(\lfloor nt \rfloor)| \\
&\le W_{\mathcal{G}}^b + \max_{t\in[\tau_0,1-\tau_0]} \max_{j\in\mathcal{G}} |C_j^{b,\mathrm{II}}(\lfloor nt \rfloor)|,
\end{aligned}
\tag{S8.82}
$$

where

$$
W_{\mathcal{G}}^b := \max_{t\in[\tau_0,1-\tau_0]} \max_{j\in\mathcal{G}} \underbrace{\sqrt{n}\sqrt{\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}} \frac{\left|\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor} \widehat{\xi}_{i,j}^b - \frac{1}{\lfloor nt \rfloor^*}\sum_{i=\lfloor nt \rfloor+1}^{n} \widehat{\xi}_{i,j}^b\right|}{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}}}_{D_j^b(\lfloor nt \rfloor)}.
\tag{S8.83}
$$

By (S8.82), we have $c_{T_{\mathcal{G}}^b}(1-\alpha) \le c_{W_{\mathcal{G}}^b}(1-\alpha) + \max_{t\in[\tau_0,1-\tau_0]} \max_{j\in\mathcal{G}} |C_j^{b,\mathrm{II}}(\lfloor nt \rfloor)|$, where $c_{W_{\mathcal{G}}^b}(1-\alpha)$ is the $1-\alpha$ quantile of $W_{\mathcal{G}}^b$. Hence, to obtain the upper bound of $c_{T_{\mathcal{G}}^b}(1-\alpha)$, it is sufficient to get the upper bound of $c_{W_{\mathcal{G}}^b}(1-\alpha)$ and $\max_{t\in[\tau_0,1-\tau_0]} \max_{j\in\mathcal{G}} |C_j^{b,\mathrm{II}}(\lfloor nt \rfloor)|$, respectively.

We first consider $c_{W_{\mathcal{G}}^b}(1-\alpha)$. By the definition of $D_j^b(\lfloor nt \rfloor)$ in (S8.83), conditional on $\mathcal{X}$, some basic calculations show that

$$D_j^b(\lfloor nt \rfloor) \sim N(0, \sigma_j^2(t)), \text{ with } t \in [\tau_0, 1-\tau_0] \text{ and } 1 \leq j \leq p, \tag{S8.84}$$

where

$$\sigma_j^2(t) := \frac{\widehat{\boldsymbol{\Theta}}_j^\top \Big(\frac{\lfloor nt \rfloor^*}{n} \sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{X}_i \boldsymbol{X}_i^\top + \frac{\lfloor nt \rfloor}{n} \sum_{i=\lfloor nt \rfloor+1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top \Big) \widehat{\boldsymbol{\Theta}}_j}{\widehat{\boldsymbol{\Theta}}_j^\top \big(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top \big) \widehat{\boldsymbol{\Theta}}_j}. \tag{S8.85}$$

Under Assumptions (**A**.**1**) - (**A**.**5**), we can prove that as $n, p \to \infty$

$$\max_{t \in [\tau_0, 1-\tau_0]} \max_{1 \leq j \leq p} |\sigma_j^2(t) - 1| = o_p(1). \tag{S8.86}$$

Let $q' = |\mathcal{G}|(n - 2\lfloor n\tau_0 \rfloor + 1)$. Combining (S8.84) and (S8.86), and using Lemma 3, for any $t > 0$, we have

$$\mathbb{E}\Big(\max_{t \in [\tau_0, 1-\tau_0]} \max_{j \in \mathcal{G}} |D_j^b(\lfloor nt \rfloor)| \Big) \leq \frac{\log(2p')}{t} + \frac{tA_0^2}{2}, \text{ with } A_0^2 := \frac{3}{2}. \tag{S8.87}$$

Furthermore, taking $t = A_0^{-1}\sqrt{2\log(q')}$ in (S8.87), we have

$$\mathbb{E}\Big(\max_{t \in [\tau_0, 1-\tau_0]} \max_{j \in \mathcal{G}} |D_j^b(\lfloor nt \rfloor)| \Big) \leq A_0 \sqrt{2\log(q')}\big(1 + \frac{1}{2\log q'}\big). \tag{S8.88}$$

By Theorem 5.8 in Boucheron et al. (2013), we have

$$\mathbb{P}\Big(\max_{\substack{\tau_0 \leq t \leq 1-\tau_0 \\ j \in \mathcal{G}}} |D_j^b(\lfloor nt \rfloor)| \geq \mathbb{E}\Big[\max_{\substack{\tau_0 \leq t \leq 1-\tau_0 \\ j \in \mathcal{G}}} |D_j^b(\lfloor nt \rfloor)| \Big] + z \Big| \mathcal{X} \Big) \leq \exp\Big(- \frac{z^2}{2A_0^2} \Big). \tag{S8.89}$$

Based on (S8.88), and taking $z = A_0\sqrt{2\log(\alpha^{-1})}$ in (S8.89), we have

$$c_{W_{\mathcal{G}}^b}(1-\alpha) \leq A_0 \sqrt{2\log(q')}\big(1 + \frac{1}{2\log q'}\big) + A_0\sqrt{2\log(\alpha^{-1})}. \tag{S8.90}$$

After obtaining the upper bound of $c_{W^b_{\mathcal{G}}}(1-\alpha)$ in (S8.90), we next consider the upper bound of $\max_{t\in[\tau_0,1-\tau_0]}\max_{j\in\mathcal{G}}|C_j^{b,\mathrm{II}}(\lfloor nt\rfloor)|$. To this end, we define

$$\mathcal{E}'=\Big\{\min_{1\le j\le p}\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j}\ge c_\epsilon\kappa_1^{-1}/2,\quad \max_{1\le j\le p}\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j}\le 2C_\epsilon\kappa_2\Big\}. \tag{S8.91}$$

By Theorem 4 and Assumptions (**A.2**) and (**A.3**), we have $\mathbb{P}(\mathcal{E}')\to 1$ as $n,p\to\infty$. Under $\mathcal{E}'$, by the definition of $C_j^{b,\mathrm{II}}(\lfloor nt\rfloor)$ in (S8.47), we have

$$\begin{aligned}
&\max_{t\in[\tau_0,1-\tau_0]}\max_{j\in\mathcal{G}}|C_j^{b,\mathrm{II}}(\lfloor nt\rfloor)|\\
&\le \underbrace{C_1\max_{t\in[\tau_0,1-\tau_0]}\sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\|\boldsymbol{\Delta}^{b,(0,t)}\|_{\mathcal{G},\infty}}_{\boldsymbol{\Delta}_1}\\
&\qquad+\underbrace{C_1\max_{t\in[\tau_0,1-\tau_0]}\sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\|\boldsymbol{\Delta}^{b,(t,1)}\|_{\mathcal{G},\infty}}_{\boldsymbol{\Delta}_2},
\end{aligned}\tag{S8.92}$$

where $C_1:=\sqrt{C_\epsilon\kappa_1^{-1}/2}$, $\boldsymbol{\Delta}^{b,(0,t)}$ and $\boldsymbol{\Delta}^{b,(t,1)}$ are defined in (S8.44). Next, we consider $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$, respectively. Without loss of generality, we assume $\widehat{t}_{0,\mathcal{G}}\in[\tau_0,t_0]$.

Control of $\boldsymbol{\Delta}_1$. For $\boldsymbol{\Delta}_1$, by the definition of $\boldsymbol{\Delta}^{b,(0,t)}$ in (S8.44), we have

$$\boldsymbol{\Delta}_1\le C_1\Big(\underbrace{\max_{t\in[\tau_0,\widehat{t}_{0,\mathcal{G}}]}\sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\|\boldsymbol{\Delta}^{b,(0,t),\mathrm{I}}\|_{\mathcal{G},\infty}}_{\boldsymbol{\Delta}_{1,1}}\vee\underbrace{\max_{t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0]}\sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\|\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}}\|_{\mathcal{G},\infty}}_{\boldsymbol{\Delta}_{1,2}}\Big). \tag{S8.93}$$

Control of $\boldsymbol{\Delta}_{1,1}$. For $\boldsymbol{\Delta}_{1,1}$, consider $\boldsymbol{\Delta}^{b,(0,t),\mathrm{I}}$ in (S8.38) with $t\in[\tau_0,\widehat{t}_{0,\mathcal{G}}]$. Conditional on $\mathcal{X}$, using concentration inequalities and by Lemma

8, we have

$$
\begin{aligned}
&\|\boldsymbol{\Delta}^{b,(0,t),\mathrm{I}}\|_{\mathcal{G},\infty} \\
&\leq C\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}\big\|\widehat{\boldsymbol{\beta}}^{b,(0,t)}-\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big\|_1 \\
&\leq Cs\big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big)\frac{\log(pn)}{\lfloor nt\rfloor}, \\
&\leq Cs^{(1)}\frac{\log(pn)}{\lfloor nt\rfloor}(\text{by Lemma 8}).
\end{aligned}
\tag{S8.94}
$$

Hence, by (S8.94), we have

$$
\begin{aligned}
\boldsymbol{\Delta}_{1,1} &= \max_{t\in[\tau_0,\widehat{t}_{0,\mathcal{G}}]}\sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\|\boldsymbol{\Delta}^{b,(0,t),\mathrm{I}}\|_{\mathcal{G},\infty}, \\
&\leq Cs\frac{\log(pn)}{\sqrt{n}} = o\big(\sqrt{\log(|\mathcal{G}|n)}\big),
\end{aligned}
\tag{S8.95}
$$

where the last equation of (S8.95) comes from the assumption that $s\sqrt{\log(pn)/n}=o(1)$ with $s:=s^{(1)}\vee s^{(2)}$ and $|\mathcal{G}|=p^\gamma$ for $\gamma\in(0,1]$.

Control of $\boldsymbol{\Delta}_{1,2}$. For $\boldsymbol{\Delta}_{1,2}$, considering $\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}}$ in (S8.42) with $t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0]$, we have

$$
\|\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}}\|_{\mathcal{G},\infty}\leq\|\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}\|_{\mathcal{G},\infty}+\|\boldsymbol{\Delta}_2^{b,(0,t),\mathrm{II}}\|_{\mathcal{G},\infty},
\tag{S8.96}
$$

where

$$
\begin{aligned}
\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}} &= -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)}-\mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{b,(0,t)}-\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big), \\
\boldsymbol{\Delta}_2^{b,(0,t),\mathrm{II}} &= -\frac{\lfloor nt\rfloor-\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor nt\rfloor}\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(\widehat{t}_{0,\mathcal{G}},t)}-\mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}-\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\big).
\end{aligned}
\tag{S8.97}
$$

Hence, by (S8.97), we need to consider $\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}$ and $\boldsymbol{\Delta}_2^{b,(0,t),\mathrm{II}}$, respectively.

Control of $\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}$. For $\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}$, using Lemma 8 for the bootstrap based samples, we have

$$
\begin{aligned}
&\|\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}\|_{\mathcal{G},\infty} \\
&\leq C\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}\big\|\widehat{\boldsymbol{\beta}}^{b,(0,t)}-\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big\|_1, \\
&\leq C\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}\frac{\lfloor nt\rfloor-\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor nt\rfloor}\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}-\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\|_1, \\
&\leq C\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}\Big(\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}-\boldsymbol{\beta}^{(1)}\|_1+\|\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}-\boldsymbol{\beta}^{(2)}\|_1+\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\|_1\Big).
\end{aligned}
\tag{S8.98}
$$

Note that we assume $\widehat{t}_{0,\mathcal{G}}\in[\tau_0,t_0]$. Using Lemma 8, we have

$$
\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}-\boldsymbol{\beta}^{(1)}\|_1\leq Cs\sqrt{\frac{\log(p)}{n\tau_0}},\ \ \|\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}-\boldsymbol{\beta}^{(2)}\|_1\leq C\frac{\lfloor nt_0\rfloor-\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor^*}\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\|_1.
\tag{S8.99}
$$

Combining (S8.98) and (S8.99), and using the fact that $\|\boldsymbol{\delta}\|_1\leq s\|\boldsymbol{\delta}\|_\infty$, we have

$$
\|\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}\|_{\mathcal{G},\infty}\leq C_1s\frac{\log(pn)}{\lfloor n\tau_0\rfloor}+C_2s\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}\|\boldsymbol{\delta}\|_\infty,\ \text{with}\ t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0].
\tag{S8.100}
$$

Control of $\boldsymbol{\Delta}_2^{b,(0,t),\mathrm{II}}$. After bounding $\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}$ in (S8.100), we next consider $\boldsymbol{\Delta}_2^{b,(0,t),\mathrm{II}}$. Using concentration inequalities and the trianlge inequality and by Lemma 8, we have

$$
\begin{aligned}
&\|\boldsymbol{\Delta}_2^{b,(0,t),\mathrm{II}}\|_{\mathcal{G},\infty} \\
&\leq C\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}\big\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}-\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\big\|_1, \\
&\leq C\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}\Big(\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}-\boldsymbol{\beta}^{(1)}\|_1+\|\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}-\boldsymbol{\beta}^{(2)}\|_1+\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\|_1\Big), \\
&\leq C\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}\Big(s\sqrt{\frac{\log(p)}{n\tau_0}}+\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\|_1+\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\|_1\Big)(\text{Lemma 8}), \\
&\leq C_1 s\frac{\log(pn)}{\lfloor n\tau_0\rfloor}+C_2 s\sqrt{\frac{\log(p)}{\lfloor nt\rfloor}}\|\boldsymbol{\delta}\|_\infty.
\end{aligned}
\tag{S8.101}
$$

Combining (S8.96), (S8.100), and (S8.101), by Assumption (**A.4**), we have

$$
\begin{aligned}
\boldsymbol{\Delta}_{1,2} &= \max_{t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0]}\sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\|\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}}\|_{\mathcal{G},\infty} \\
&\leq C_1\sqrt{n}s\frac{\log(pn)}{n\tau_0}+C_2\sqrt{n}\big(s\sqrt{\frac{\log(p)}{n\tau_0}}\|\boldsymbol{\delta}\|_\infty\big), \\
&\leq o\big(\sqrt{\log(|\mathcal{G}|n)}\big)+C_2\sqrt{n}\big(s\sqrt{\frac{\log(p)}{n\tau_0}}\|\boldsymbol{\delta}\|_\infty\big).
\end{aligned}
\tag{S8.102}
$$

Combining (S8.93), (S8.95), and (S9.178), we have

$$
\boldsymbol{\Delta}_1 \leq o\big(\sqrt{\log(|\mathcal{G}|n)}\big)+C_1\sqrt{n}\big(s\sqrt{\frac{\log(p)}{n\tau_0}}\|\boldsymbol{\delta}\|_\infty\big). \tag{S8.103}
$$

Control of $\boldsymbol{\Delta}_2$. Similarly, we can obtain the upper bound for $\boldsymbol{\Delta}_2$ as

$$
\boldsymbol{\Delta}_2 \leq o\big(\sqrt{\log(|\mathcal{G}|n)}\big)+C_2\sqrt{n}\big(s\sqrt{\frac{\log(p)}{n\tau_0}}\|\boldsymbol{\delta}\|_\infty\big). \tag{S8.104}
$$

Combining (S8.92), (S8.103), and (S8.104), we have

$$
\max_{t\in[\tau_0,1-\tau_0]}\max_{j\in\mathcal{G}}|C_j^{b,\mathrm{II}}(\lfloor nt\rfloor)| \leq o\big(\sqrt{\log(|\mathcal{G}|n)}\big)+C_1\sqrt{n}\big(s\sqrt{\frac{\log(p)}{n\tau_0}}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\big). \tag{S8.105}
$$

Finally, using (S8.82), (S8.90), and (S8.105), we obtain an upper bound of $c_{T^b_{\mathcal{G}}}(1-\alpha)$ as

$$\begin{aligned} c_{T^b_{\mathcal{G}}}(1-\alpha) \quad &\leq A_0\sqrt{2\log(q')}\big(1+\frac{1}{2\log q'}\big)+A_0\sqrt{2\log(\alpha^{-1})}+o\big(\sqrt{\log(|\mathcal{G}|n)}\big) \\ &\quad +C_1\sqrt{n}\big(s\sqrt{\frac{\log(p)}{n\tau_0}}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\big). \end{aligned} \tag{S8.106}$$

Step 2: In this step, we aim to prove that $\mathbb{P}\big(T_{\mathcal{G}} \geq c_{T^b_{\mathcal{G}}}(1-\alpha)\big) \to 1$ as $n, p \to \infty$. Let

$$\begin{aligned} c^{\mathrm{u}}_{T^b_{\mathcal{G}}}(1-\alpha) \quad &= A_0\sqrt{2\log(q')}\big(1+\frac{1}{2\log q'}\big)+A_0\sqrt{2\log(\alpha^{-1})} \\ &\quad +o\big(\sqrt{\log(|\mathcal{G}|n)}\big)+C_1\sqrt{n}\big(s\sqrt{\frac{\log(p)}{n\tau_0}}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\big). \end{aligned} \tag{S8.107}$$

Considering the upper bound obtained in (S8.106), it is sufficient to prove $H_1 \to 1$, where

$$H_1 = \mathbb{P}\big(T_{\mathcal{G}} \geq c^{\mathrm{u}}_{T^b_{\mathcal{G}}}(1-\alpha)\big). \tag{S8.108}$$

By replacing $\widehat{\sigma}_\epsilon\widehat{\omega}_{j,j}$ by its true values, we define the oracle testing statistics as

$$\widetilde{T}_{\mathcal{G}} = \max_{t\in[\tau_0,1-\tau_0]}\max_{j\in\mathcal{G}}\sqrt{n}\frac{\lfloor nt\rfloor}{n}\Big(1-\frac{\lfloor nt\rfloor}{n}\Big)\Big|\frac{\breve{\beta}^{(0,t)}_j-\breve{\beta}^{(t,1)}_j}{\sqrt{\sigma^2_\epsilon\omega_{j,j}}}\Big|. \tag{S8.109}$$

Considering (S8.108) and (S8.109), it is sufficient to prove $H_2 \to 1$ as $n, p \to \infty$, where

$$H_2 = \mathbb{P}\big(\widetilde{T}_{\mathcal{G}} \geq c^{\mathrm{u}}_{T^b_{\mathcal{G}}}(1-\alpha)+|T_{\mathcal{G}}-\widetilde{T}_{\mathcal{G}}|\big). \tag{S8.110}$$

Recall $\{Z_j(\lfloor nt \rfloor), \tau_0 \le t \le 1-\tau_0, 1 \le j \le p\}$ defined in (S8.64). By definition, we have

$$\widetilde{T}_{\mathcal{G}} = \max_{t \in [\tau_0, 1-\tau_0]} \max_{j \in \mathcal{G}} \frac{|Z_j(\lfloor nt \rfloor)|}{\sqrt{\sigma_\epsilon^2 \omega_{j,j}}}. \tag{S8.111}$$

Let $\boldsymbol{Z}(\lfloor nt \rfloor) = \big(Z_1(\lfloor nt \rfloor), \ldots, Z_p(\lfloor nt \rfloor)\big)^\top$. Under $\mathbf{H}_1$, we have the following decomposition:

$$\boldsymbol{Z}(\lfloor nt \rfloor) = \boldsymbol{\delta}(t) + \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\Big(\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}\widehat{\boldsymbol{\xi}}_i - \frac{1}{\lfloor nt \rfloor^*}\sum_{i=\lfloor nt \rfloor+1}^{n}\widehat{\boldsymbol{\xi}}_i + \boldsymbol{R}^{(0,t)} - \boldsymbol{R}^{(t,1)}\Big), \tag{S8.112}$$

where $\widehat{\boldsymbol{\xi}}_i := (\widehat{\xi}_{i,1}, \ldots, \widehat{\xi}_{i,p})^\top$ with $\widehat{\xi}_{i,j} = \widehat{\boldsymbol{\Theta}}_j^\top \mathbf{X}_i \epsilon_i$, $\boldsymbol{\delta}(t) = (\delta_1(t), \ldots, \delta_p(t))^\top$ is defined in (S8.65), $\boldsymbol{R}^{(0,t)}$ and $\boldsymbol{R}^{(t,1)}$ are defined in (S8.66). Using (S8.112), under the event $\mathcal{E}'$, we have

$$\widetilde{T}_{\mathcal{G}} \ge \max_{t \in [\tau_0, 1-\tau_0]} \max_{j \in \mathcal{G}} \frac{\delta_j(t)}{\sqrt{\sigma_\epsilon^2 \omega_{j,j}}} - (c_\epsilon \kappa_2^{-1}/2)^{-1/2}(\boldsymbol{R}_1 + \boldsymbol{R}_2 + \boldsymbol{R}_3), \tag{S8.113}$$

with

$$\begin{aligned}
\boldsymbol{R}_1 &= \max_{t \in [\tau_0, 1-\tau_0]} \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\Big\|\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}\widehat{\boldsymbol{\xi}}_i - \frac{1}{\lfloor nt \rfloor^*}\sum_{i=\lfloor nt \rfloor+1}^{n}\widehat{\boldsymbol{\xi}}_i\Big\|_{\mathcal{G},\infty}, \\
\boldsymbol{R}_2 &= \max_{t \in [\tau_0, 1-\tau_0]} \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\|\boldsymbol{R}^{(0,t)}\|_{\mathcal{G},\infty}, \\
\boldsymbol{R}_3 &= \max_{t \in [\tau_0, 1-\tau_0]} \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\|\boldsymbol{R}^{(t,1)}\|_{\mathcal{G},\infty}.
\end{aligned} \tag{S8.114}$$

By (S8.110) and (S8.113), to prove $H_2 \to 1$, it is sufficient to prove $H_3 \to 1$,

where

$$
\begin{aligned}
H_3 \quad &= \mathbb{P}\Big(\max_{t \in [\tau_0, 1-\tau_0]} \max_{j \in \mathcal{G}} \frac{\delta_j(t)}{\sqrt{\sigma_\epsilon^2 \omega_{j,j}}} \geq (c_\epsilon \kappa_2^{-1}/2)^{-1/2} (\boldsymbol{R}_1 + \boldsymbol{R}_2 + \boldsymbol{R}_3) \\
&\qquad + c^{\mathrm{u}}_{T^b_{\mathcal{G}}}(1-\alpha) + |T_{\mathcal{G}} - \widetilde{T}_{\mathcal{G}}| \Big).
\end{aligned}
\tag{S8.115}
$$

Next, we prove $H_3 \to 1$. To this end, we need to obtain the upper bound of $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, and $\boldsymbol{R}_3$, and $|T_{\mathcal{G}} - \widetilde{T}_{\mathcal{G}}|$, respectively.

Control of $\boldsymbol{R}_1$. We first consider $\boldsymbol{R}_1$. By Assumptions (**A.1**) – (**A.5**), using basic concentration inequalities, we can prove that with probability at least $1 - C_1(np)^{-C_2}$,

$$
\boldsymbol{R}_1 \leq C_2 \sqrt{\log(|\mathcal{G}|n)}. \tag{S8.116}
$$

Control of $\boldsymbol{R}_2$. We next bound $\boldsymbol{R}_2$. Considering $\boldsymbol{R}^{(0,t)}$ in (S8.66), we have

$$
\boldsymbol{R}_2 \leq \underbrace{\max_{t \in [\tau_0, t_0]} \sqrt{n} \frac{\lfloor nt \rfloor}{n} \frac{\lfloor nt \rfloor^*}{n} \big\| \boldsymbol{R}^{(0,t),\mathrm{I}} \big\|_{\mathcal{G},\infty}}_{\boldsymbol{R}_{2,1}} \vee \underbrace{\max_{t \in [t_0, 1-\tau_0]} \sqrt{n} \frac{\lfloor nt \rfloor}{n} \frac{\lfloor nt \rfloor^*}{n} \big\| \boldsymbol{R}^{(0,t),\mathrm{II}} \big\|_{\mathcal{G},\infty}}_{\boldsymbol{R}_{2,2}}, \tag{S8.117}
$$

where $\boldsymbol{R}^{(0,t),\mathrm{I}}$ and $\boldsymbol{R}^{(0,t),\mathrm{II}}$ are defined in (S8.67). Next, we bound $\boldsymbol{R}_{2,1}$ and $\boldsymbol{R}_{2,2}$, respectively.

Control of $\boldsymbol{R}_{2,1}$. For $\boldsymbol{R}_{2,1}$, using concentration inequalities and Lemma

8, we have

$$\begin{aligned}\boldsymbol{R}_{2,1} &\le C_1 \max_{t\in[\tau_0,t_0]} \sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}\big\|\widehat{\boldsymbol{\beta}}^{(0,t)}-\boldsymbol{\beta}^{(1)}\big\|_1,\\ &\le C_1 s\frac{\log(pn)}{\sqrt{n}} = o\big(\log(|\mathcal{G}|n)\big),\end{aligned}\tag{S8.118}$$

where the last equation of (S8.118) comes from the assumption that $s\sqrt{\log(pn)/n} = o(1)$ and $|\mathcal{G}| = p^{\gamma}$ with $\gamma \in (0, 1]$.

Control of $\boldsymbol{R}_{2,2}$. For $\boldsymbol{R}_{2,2}$, by the decomposition in (S8.67), we have

$$\boldsymbol{R}_{2,2} \le \boldsymbol{R}_{2,2,1} + \boldsymbol{R}_{2,2,2},\tag{S8.119}$$

where $\boldsymbol{R}_{2,2,1}$ and $\boldsymbol{R}_{2,2,2}$ are defined as

$$\begin{aligned}\boldsymbol{R}_{2,2,1} &= \max_{t\in[t_0,1-\tau_0]} \sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\frac{\lfloor nt_0\rfloor-\lfloor nt\rfloor}{\lfloor nt\rfloor}\big\|\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t_0,t)}-\mathbf{I}\big)\big(\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\big)\|_{\mathcal{G},\infty},\\ &\le C_1\sqrt{n}\sqrt{\frac{\log(pn)}{n\tau_0}}\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(2)}\|_1,\\ &\le C_1\sqrt{n}s\sqrt{\frac{\log(pn)}{n\tau_0}}\big\|\boldsymbol{\delta}\big\|_\infty,\end{aligned}\tag{S8.120}$$

and

$$\begin{aligned}\boldsymbol{R}_{2,2,2} &= \max_{t\in[t_0,1-\tau_0]} \sqrt{n}\frac{\lfloor nt\rfloor}{n}\frac{\lfloor nt\rfloor^*}{n}\big\|\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)}-\mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,t)}-\boldsymbol{\beta}^{(1)}\big)\big\|_{\mathcal{G},\infty},\\ &\le \max_{t\in[t_0,1-\tau_0]} C_1\sqrt{n}\sqrt{\frac{\log(pn)}{n\tau_0}}\|\widehat{\boldsymbol{\beta}}^{(0,t)}-\boldsymbol{\beta}^{(1)}\|_1,\\ &\le \max_{t\in[t_0,1-\tau_0]} C_1\sqrt{n}\sqrt{\frac{\log(pn)}{n\tau_0}}\frac{\lfloor nt\rfloor-\lfloor nt_0\rfloor}{\lfloor nt\rfloor}\|\boldsymbol{\beta}^{(2)}-\boldsymbol{\beta}^{(1)}\|_1,\ \text{(by Lemma 8)}\\ &\le C_1\sqrt{n}s\sqrt{\frac{\log(pn)}{n\tau_0}}\big\|\boldsymbol{\delta}\big\|_\infty.\end{aligned}\tag{S8.121}$$

Combining (S8.117), (S8.118), (S8.119), (S8.120), (S8.121), we have

$$\boldsymbol{R}_2 \leq o\big(\log(|\mathcal{G}|n)\big) + C_1\sqrt{n}s\sqrt{\frac{\log(pn)}{n\tau_0}}\big\|\boldsymbol{\delta}\big\|_\infty. \tag{S8.122}$$

Control of $\boldsymbol{R}_3$. With a similar proof, we can obtain the upper bound of $\boldsymbol{R}_3$ as

$$\boldsymbol{R}_3 \leq o\big(\log(|\mathcal{G}|n)\big) + C_1\sqrt{n}s\sqrt{\frac{\log(pn)}{n\tau_0}}\big\|\boldsymbol{\delta}\big\|_\infty. \tag{S8.123}$$

Control of $|T_\mathcal{G} - \widetilde{T}_\mathcal{G}|$. After bounding $\boldsymbol{R}_1$, $\boldsymbol{R}_2$, and $\boldsymbol{R}_3$ in (S8.116), (S8.122), and (S8.123), we next bound $|T_\mathcal{G} - \widetilde{T}_\mathcal{G}|$. Using the fact that $|\max_i|a_i| - \max_i|b_i|| \leq \max_i|a_i - b_i|$, we have

$$\begin{aligned}
&|T_\mathcal{G} - \widetilde{T}_\mathcal{G}| \\
&= \Big|\max_{t\in[\tau_0,1-\tau_0]}\max_{j\in\mathcal{G}}\frac{|Z_j(\lfloor nt\rfloor)|}{\sqrt{\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j}}} - \max_{t\in[\tau_0,1-\tau_0]}\max_{j\in\mathcal{G}}\frac{|Z_j(\lfloor nt\rfloor)|}{\sqrt{\sigma_\epsilon^2\omega_{j,j}}}\Big|, \\
&\leq \max_{t\in[\tau_0,1-\tau_0]}\max_{j\in\mathcal{G}}\Big|\frac{Z_j(\lfloor nt\rfloor)}{\sqrt{\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j}}} - \frac{Z_j(\lfloor nt\rfloor)}{\sqrt{\sigma_\epsilon^2\omega_{j,j}}}\Big|, \\
&\leq \widetilde{T}_\mathcal{G}\max_{j\in\mathcal{G}}\Big|\frac{\sqrt{\sigma_\epsilon^2\omega_{j,j}}}{\sqrt{\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j}}} - 1\Big|.
\end{aligned} \tag{S8.124}$$

Note that conditional on the event $\mathcal{E}'$, using Theorem 4, we have

$$\max_{j\in\mathcal{G}}\Big|\frac{\sqrt{\sigma_\epsilon^2\omega_{j,j}}}{\sqrt{\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j}}} - 1\Big| \leq C_1\underbrace{\max_{1\leq j\leq p}|\widehat{\sigma}_\epsilon^2\widehat{\omega}_{j,j} - \sigma_\epsilon^2\omega_{j,j}|}_{\epsilon_n'} = o_p(1). \tag{S8.125}$$

Considering (S8.124) and (S8.125), using the decomposition for $T_\mathcal{G}$ in (S8.112), we have

$$|T_\mathcal{G} - \widetilde{T}_\mathcal{G}| \leq C_1\epsilon_n'\max_{t\in[\tau_0,1-\tau_0]}\max_{j\in\mathcal{G}}\frac{\delta_j(t)}{\sqrt{\sigma_\epsilon^2\omega_{j,j}}} + C_2\epsilon_n'(\boldsymbol{R}_1 + \boldsymbol{R}_2 + \boldsymbol{R}_3). \tag{S8.126}$$

Let $\epsilon''_n = s\sqrt{\log(pn)/n\tau_0}\|\boldsymbol{\delta}\|_\infty/\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}$. By the definition of $\boldsymbol{\delta}(t)$ in (S8.65), we have

$$
\begin{aligned}
\max_{t\in[\tau_0,1-\tau_0]}\max_{j\in\mathcal{G}}\frac{\delta_j(t)}{\sqrt{\sigma_\epsilon^2\omega_{j,j}}} &= \sqrt{n}\frac{\lfloor nt_0\rfloor}{n}\frac{\lfloor nt_0\rfloor^*}{n}\max_{j\in\mathcal{G}}\frac{|\beta_j^{(1)}-\beta_j^{(2)}|}{\sqrt{\sigma_\epsilon^2\omega_{j,j}}},\\
&= \sqrt{n}\max_{j\in\mathcal{G}}\Big|\frac{t_0(1-t_0)(\beta_j^{(2)}-\beta_j^{(1)})}{(\sigma_\epsilon^2\omega_{j,j})^{1/2}}\Big| + O(\frac{1}{\sqrt{n}}),
\end{aligned}
\tag{S8.127}
$$

where the last equation comes from the fact that $|\lfloor nt\rfloor/n - t| = O(1/n)$ as $n\to\infty$.

Finally, for H_3 in (S8.115), considering the upper bounds in (S8.107), (S8.116), (S8.122), (S8.123), (S8.126), we have

$$
\begin{aligned}
H_3 &\geq \mathbb{P}\Big(\sqrt{n}\max_{j\in\mathcal{G}}\Big|\frac{t_0(1-t_0)(\beta_j^{(2)}-\beta_j^{(1)})}{(\sigma_\epsilon^2\omega_{j,j})^{1/2}}\Big| \geq C_1\sqrt{2\log(|\mathcal{G}|n)} + C_2\sqrt{2\log(\alpha^{-1})}\\
&\qquad + C_3(\epsilon'_n\vee\epsilon''_n)(\sqrt{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty})\Big),\\
&\geq \mathbb{P}\Big(\sqrt{n}\max_{j\in\mathcal{G}}|D_j| \geq \frac{C_4}{(1-\epsilon'_n\vee\epsilon''_n)}\big(\sqrt{2\log(|\mathcal{G}|n)} + \sqrt{2\log(\alpha^{-1})}\big)\Big).
\end{aligned}
\tag{S8.128}
$$

Considering (S8.128), by choosing a large enough constant in (3.29), we have $H_3 \to 1$, which completes the proof of Theorem 5.

□

S9 Proofs of lemmas in Section S7

S9.1 Proof of Lemma 10

Proof. In this section, we aim to prove

$$\mathbb{P}\big(\max_{\tau_0 \le t \le 1-\tau_0} \big\|\boldsymbol{C}(\lfloor nt \rfloor) - \widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt \rfloor)\big\|_{\mathcal{G},\infty} \ge \epsilon\big) = o(1). \tag{S9.129}$$

Without loss of generality, we assume $\mathcal{G} = \{1, \ldots, p\}$. Using the triangle inequality, we have

$$\mathbb{P}\big(\max_{\tau_0 \le t \le 1-\tau_0} \big\|\boldsymbol{C}(\lfloor nt \rfloor) - \widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt \rfloor)\big\|_{\infty} \ge \epsilon\big) \le D_1 + D_2, \tag{S9.130}$$

where

$$\begin{aligned} D_1 &:= \mathbb{P}\big(\max_{\tau_0 \le t \le 1-\tau_0} \big\|\boldsymbol{C}(\lfloor nt \rfloor) - \boldsymbol{C}^{\mathrm{I}}(\lfloor nt \rfloor)\big\|_{\infty} \ge \epsilon/2\big), \\ D_2 &:= \mathbb{P}\big(\max_{\tau_0 \le t \le 1-\tau_0} \big\|\boldsymbol{C}^{\mathrm{I}}(\lfloor nt \rfloor) - \widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt \rfloor)\big\|_{\infty} \ge \epsilon/2\big). \end{aligned} \tag{S9.131}$$

Control of D_1. By (S9.130), to prove (S9.129), we need to bound D_1 and D_2, respectively. We first consider D_1. To this end, we define

$$\mathcal{E} = \big\{\min_{1 \le j \le p} \widehat{\sigma}_{\epsilon}^2 \widehat{\omega}_{j,j} > c_{\epsilon}\kappa_1^{-1}/2\big\}, \tag{S9.132}$$

where κ_2 and c_ϵ are defined in Assumptions (**A**.**2**) and (**A**.**3**). By introuducing $\mathcal{E}$, we have

$$D_1 \le \mathbb{P}\big(\max_{\tau_0 \le t \le 1-\tau_0} \big\|\boldsymbol{C}(\lfloor nt \rfloor) - \boldsymbol{C}^{\mathrm{I}}(\lfloor nt \rfloor)\big\|_{\infty} \ge \epsilon/2 \cap \mathcal{E}\big) + \mathbb{P}(\mathcal{E}^c). \tag{S9.133}$$

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

$$
\begin{aligned}
&\mathbb{P}\big(\max_{\tau_0 \le t \le 1-\tau_0} \big\|\boldsymbol{C}(\lfloor nt \rfloor) - \boldsymbol{C}^{\mathrm{I}}(\lfloor nt \rfloor)\big\|_\infty \ge \epsilon/2 \cap \mathcal{E}\big) \\
&\le \mathbb{P}\big(\max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \sqrt{n}\frac{\lfloor nt \rfloor}{n}\frac{\lfloor nt \rfloor^*}{n}\big(\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}\big)^{-1/2} |\Delta_j^{(0,t)} - \Delta_j^{(t,1)}| \ge \epsilon/2\big), \\
&\le \mathbb{P}\big(\max_{t\in[\tau_0,1-\tau_0]} \|\boldsymbol{\Delta}^{(0,t)} - \boldsymbol{\Delta}^{(t,1)}\|_\infty \ge \tfrac{1}{2}\sqrt{c_\epsilon \kappa_2^{-1}/2}\epsilon n^{-1/2}\big), \\
&\le \mathbb{P}\big(\max_{t\in[\tau_0,1-\tau_0]} \|\boldsymbol{\Delta}^{(0,t)}\|_\infty \ge C_1 \epsilon n^{-1/2}\big) + \mathbb{P}\big(\max_{t\in[\tau_0,1-\tau_0]} \|\boldsymbol{\Delta}^{(t,1)}\|_\infty \ge C_2 \epsilon n^{-1/2}\big).
\end{aligned}
\tag{S9.134}
$$

By the definitions of $\boldsymbol{\Delta}^{(0,t)}$ and $\boldsymbol{\Delta}^{(t,1)}$ in (S8.29), we have

$$
\|\boldsymbol{\Delta}^{(0,t)}\|_\infty \le \big\|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big\|_\infty \big\|\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0)}\big\|_1. \tag{S9.135}
$$

To bound $\|\boldsymbol{\Delta}^{(0,t)}\|_\infty$, we need to consider $\big\|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big\|_\infty$ and $\big\|\widehat{\boldsymbol{\beta}}^{(0,t)} - \boldsymbol{\beta}^{(0)}\big\|_1$, respectively. For $\big\|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big\|_\infty$, by the triangle inequality, we have

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big\|_\infty \\
&\le \|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}^n - \mathbf{I}\big\|_\infty + \|(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\widehat{\boldsymbol{\Sigma}}^n - \widehat{\boldsymbol{\Sigma}}_{(0,t)})\big\|_\infty + \|\boldsymbol{\Theta}(\widehat{\boldsymbol{\Sigma}}^n - \widehat{\boldsymbol{\Sigma}}_{(0,t)})\big\|_\infty, \\
&\le \|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}^n - \mathbf{I}\big\|_\infty + \|(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\widehat{\boldsymbol{\Sigma}}^n - \widehat{\boldsymbol{\Sigma}}_{(0,t)})\big\|_\infty + \|\boldsymbol{\Theta}\widehat{\boldsymbol{\Sigma}}^n - \mathbf{I}\big\|_\infty + \|\boldsymbol{\Theta}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big\|_\infty, \\
&\le \|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}^n - \mathbf{I}\big\|_\infty + \max_{1\le j\le p} \|\widehat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\|_1 \|\widehat{\boldsymbol{\Sigma}}^n - \widehat{\boldsymbol{\Sigma}}_{(0,t)}\|_\infty + \|\boldsymbol{\Theta}\widehat{\boldsymbol{\Sigma}}^n - \mathbf{I}\big\|_\infty + \|\boldsymbol{\Theta}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big\|_\infty.
\end{aligned}
\tag{S9.136}
$$

To bound $\|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big\|_\infty$, we consider the four parts on the RHS of (S9.136), respectively.

For $\|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}^n - \mathbf{I}\big\|_\infty$, by Van de Geer et al. (2014) and Assumption (**A**.**5**), we have

$$\|\widehat{\mathbf{\Theta}}\widehat{\mathbf{\Sigma}}^n - \mathbf{I}\|_\infty \leq O_p\Big(\max_{1\leq j\leq p}\lambda_{(j)}\Big) = O_p\Big(\sqrt{\frac{\log(p)}{n}}\Big). \tag{S9.137}$$

For $\max_{1\leq j\leq p}\|\widehat{\mathbf{\Theta}}_j - \mathbf{\Theta}_j\|_1\|\widehat{\mathbf{\Sigma}}^n - \widehat{\mathbf{\Sigma}}_{(0,t)}\|_\infty$, by Lemma 4, we have

$$\max_{1\leq j\leq p}\|\widehat{\mathbf{\Theta}}_j - \mathbf{\Theta}_j\|_1 = O_p\Big(s_j\sqrt{\frac{\log(p)}{n}}\Big). \tag{S9.138}$$

Note that we can write $\|\widehat{\mathbf{\Sigma}}^n - \widehat{\mathbf{\Sigma}}_{(0,t)}\|_\infty$ into

$$\|\widehat{\mathbf{\Sigma}}_{(0,t)}-\widehat{\mathbf{\Sigma}}^n\|_\infty = \max_{1\leq j,k\leq p}\Big|\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\big(X_{i,j}X_{i,k}-\mathbb{E}(X_{i,j}X_{i,k})\big)-\frac{1}{n}\sum_{i=1}^{n}\big(X_{i,j}X_{i,k}-\mathbb{E}(X_{i,j}X_{i,k})\big)\Big|. \tag{S9.139}$$

Under Assumption (**A.1**), $X_{i,j}X_{i,k} - \mathbb{E}X_{i,j}X_{i,k}$ follows sub-exponential distributions for $1\leq j,k\leq p$ and $1\leq i\leq n$. By Bernstein's inequality, with probabilty tending to 1, we have

$$\|\widehat{\mathbf{\Sigma}}_{(0,t)} - \widehat{\mathbf{\Sigma}}^n\|_\infty \leq C_3\sqrt{\frac{\log(pn)}{\lfloor nt\rfloor}}. \tag{S9.140}$$

For $\|\mathbf{\Theta}\widehat{\mathbf{\Sigma}}^n - \mathbf{I}\|_\infty + \|\mathbf{\Theta}\widehat{\mathbf{\Sigma}}_{(0,t)} - \mathbf{I}\|_\infty$, using concentration inequalities again, we have

$$\|\mathbf{\Theta}\widehat{\mathbf{\Sigma}}^n - \mathbf{I}\|_\infty + \|\mathbf{\Theta}\widehat{\mathbf{\Sigma}}_{(0,t)} - \mathbf{I}\|_\infty = O_p\Big(\sqrt{\frac{\log pn}{\lfloor nt\rfloor}}\Big). \tag{S9.141}$$

Combining the results in (S9.137) - (S9.141), we have

$$\max_{t\in[\tau_0,1-\tau_0]}\|\widehat{\mathbf{\Theta}}\widehat{\mathbf{\Sigma}}_{(0,t)} - \mathbf{I}\|_\infty \leq O_p\Big(\sqrt{\frac{\log(pn)}{\lfloor n\tau_0\rfloor}}\Big). \tag{S9.142}$$

Note that by Lemma 8, under $\mathbf{H}_0$, $\sup_{t\in[\tau_0,1-\tau_0]} \left\|\widehat{\boldsymbol{\beta}}^{(0,t)}-\boldsymbol{\beta}^{(0)}\right\|_1 \leq s^{(0)}\sqrt{\log(p)/\lfloor n\tau_0\rfloor}$ holds. Considering (S9.142), we have

$$\max_{t\in[\tau_0,1-\tau_0]} \|\boldsymbol{\Delta}^{(0,t)}\|_\infty \leq O_p\Big(s^{(0)}\frac{\log(pn)}{\lfloor n\tau_0\rfloor}\Big). \tag{S9.143}$$

With a similar proof technique , for $\max_{t\in[\tau_0,1-\tau_0]} \|\boldsymbol{\Delta}^{(t,1)}\|_\infty$, we can obtain

$$\max_{t\in[\tau_0,1-\tau_0]} \|\boldsymbol{\Delta}^{(t,1)}\|_\infty \leq O_p\Big(s^{(0)}\frac{\log(pn)}{\lfloor n\tau_0\rfloor}\Big). \tag{S9.144}$$

Note that

$$\epsilon = C\max(\max_{1\leq j\leq p} s_j\frac{\log(pn)}{\sqrt{n}}, s\sqrt{n}\frac{\log(pn)}{\lfloor n\tau_0\rfloor}) \tag{S9.145}$$

holds for some large enough constant $C>0$. Considering (S9.134), (S9.143), and (S9.144), as $n,p\to\infty$, we have $D_1=o(1)$.

Control of D_2. After bounding D_1, we next consider D_2. By the definitions of $\boldsymbol{C}^{\mathrm{I}}(\lfloor nt\rfloor)$ and $\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)$ in (S8.33) and (S8.34), and using the triangle inequality, we have

$$\max_{t\in[\tau_0,1-\tau_0]} \left\|\boldsymbol{C}^{\mathrm{I}}(\lfloor nt\rfloor)-\widetilde{\boldsymbol{C}}^{\mathrm{I}}(\lfloor nt\rfloor)\right\|_\infty \leq I+II+III, \tag{S9.146}$$

where $I - III$ are defined as

$$
\begin{aligned}
I &= \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \Big| \frac{\sqrt{\sigma_\epsilon^2 \omega_{j,j}}}{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}} - 1 \Big| \Big| \frac{\Big(\sum_{i=1}^{\lfloor nt \rfloor} (\widehat{\boldsymbol{\Theta}}_j^\top - \boldsymbol{\Theta}_j^\top) \mathbf{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} (\widehat{\boldsymbol{\Theta}}_j^\top - \boldsymbol{\Theta}_j^\top) \mathbf{X}_i \epsilon_i \Big)}{\sqrt{n \sigma_\epsilon^2 \omega_{j,j}}} \Big|, \\
II &= \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \Big| \frac{\sqrt{\sigma_\epsilon^2 \omega_{j,j}}}{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}} - 1 \Big| \Big| \frac{\Big(\sum_{i=1}^{\lfloor nt \rfloor} \boldsymbol{\Theta}_j^\top \mathbf{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \boldsymbol{\Theta}_j^\top \mathbf{X}_i \epsilon_i \Big)}{\sqrt{n \sigma_\epsilon^2 \omega_{j,j}}} \Big|, \\
III &= \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \Big| \frac{\Big(\sum_{i=1}^{\lfloor nt \rfloor} (\widehat{\boldsymbol{\Theta}}_j^\top - \boldsymbol{\Theta}_j^\top) \mathbf{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} (\widehat{\boldsymbol{\Theta}}_j^\top - \boldsymbol{\Theta}_j^\top) \mathbf{X}_i \epsilon_i \Big)}{\sqrt{n \sigma_\epsilon^2 \omega_{j,j}}} \Big|.
\end{aligned}
\tag{S9.147}
$$

We next consider I, II, and III, respectively. For I, we have $I \le I^{(1)} + I^{(2)}$, where

$$
\begin{aligned}
I^{(1)} &= \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \Big| \frac{\sqrt{\sigma_\epsilon^2 \omega_{j,j}}}{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}} - 1 \Big|, \\
I^{(2)} &= \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \Big| \frac{\Big(\sum_{i=1}^{\lfloor nt \rfloor} (\widehat{\boldsymbol{\Theta}}_j^\top - \boldsymbol{\Theta}_j^\top) \mathbf{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} (\widehat{\boldsymbol{\Theta}}_j^\top - \boldsymbol{\Theta}_j^\top) \mathbf{X}_i \epsilon_i \Big)}{\sqrt{n \sigma_\epsilon^2 \omega_{j,j}}} \Big|.
\end{aligned}
\tag{S9.148}
$$

To bound $I^{(1)}$, define

$$
\widetilde{I}^{(1)} = \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \Big| \frac{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}}}{\sqrt{\sigma_\epsilon^2 \omega_{j,j}}} - 1 \Big| = \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \Big| \frac{\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}} - \sqrt{\sigma_\epsilon^2 \omega_{j,j}}}{\sqrt{\sigma_\epsilon^2 \omega_{j,j}}} \Big|.
\tag{S9.149}
$$

Using the fact that $a^2 - b^2 = (a-b)(a+b)$, we have

$$\begin{aligned}
\widetilde{I}^{(1)} &= \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \left| \frac{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j} - \sigma_\epsilon^2 \omega_{j,j}}{\sqrt{\sigma_\epsilon^2 \omega_{j,j}}\left(\sqrt{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}} + \sqrt{\sigma_\epsilon^2 \omega_{j,j}}\right)} \right|, \\
&\le \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \left| \frac{\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j} - \sigma_\epsilon^2 \omega_{j,j}}{\sigma_\epsilon^2 \omega_{j,j}} \right|, \\
&\le C \max_{t\in[\tau_0,1-\tau_0]} \max_{1\le j\le p} \left| \widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j} - \sigma_\epsilon^2 \omega_{j,j} \right| (\text{Assumptions } (\mathbf{A.2}) \ \text{ and } (\mathbf{A.3})), \\
&\le O_p\Big(\sqrt{\frac{\log(n)}{n}} + \max\nolimits_j \lambda_{(j)} \sqrt{s_j} \Big),
\end{aligned} \tag{S9.150}$$

where the last inequality comes from Theorem 1. Using (S9.150), and by Lemma C.1 in Zhou et al. (2018), we have

$$I^{(1)} \le O_p\Big(\sqrt{\frac{\log(n)}{n}} + \max_j \lambda_{(j)} \sqrt{s_j} \Big). \tag{S9.151}$$

For $I^{(2)}$, note that for two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, we have $\|\boldsymbol{x}^\top \boldsymbol{y}\|_\infty \le \|\boldsymbol{x}\|_1 \|\boldsymbol{y}\|_\infty$. By Assumptions $(\mathbf{A.2})$ and $(\mathbf{A.3})$, there exists a universal positive constant C such that

$$I^{(2)} \le C \max_{1\le j\le p} \|\widehat{\mathbf{\Theta}}_j^\top - \mathbf{\Theta}_j^\top\|_1 \max_{t\in[\tau_0,1-\tau_0]} \Big\| \frac{1}{\sqrt{n}} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \mathbf{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \mathbf{X}_i \epsilon_i \Big) \Big\|_\infty. \tag{S9.152}$$

By Lemma 4, we have $\max_{1\le j\le p} \|\widehat{\mathbf{\Theta}}_j^\top - \mathbf{\Theta}_j^\top\|_1 \le O_p\Big(\max_{1\le j\le p} s_j \sqrt{\frac{\log(p)}{n}} \Big) = o_p(1)$. Note that Assumptions $(\mathbf{A.1})$ and $(\mathbf{A.2})$ imply that $\mathbf{X}_i \epsilon_i$ follows the sub-exponential distribution. Using Lemma 5, we have

$$\max_{t\in[\tau_0,1-\tau_0]} \Big\| \frac{1}{\sqrt{n}} \Big(\sum_{i=1}^{\lfloor nt \rfloor} \mathbf{X}_i \epsilon_i - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} \mathbf{X}_i \epsilon_i \Big) \Big\|_\infty \le O_p\big(\sqrt{\log(pn)} \big). \tag{S9.153}$$

Combining (S9.151) and (S9.152), we have

$$I \leq O_p\Big(\max_{1\leq j\leq p} s_j \frac{\log^{3/2}(pn)}{n} + \max_{1\leq j\leq p} \frac{(s_j \log(pn))^{3/2}}{n}\Big). \tag{S9.154}$$

Similarly, for II and III, we can obtain their upper bounds as follows:

$$\begin{aligned} II &\leq \times O_p\Big(\frac{\log(pn)}{\sqrt{n}} + \max_{1\leq j\leq p} \frac{\log(pn)\sqrt{s_j}}{\sqrt{n}}\Big), \\ III &\leq O_p\Big(\max_{1\leq j\leq p} s_j \frac{\log(pn)}{\sqrt{n}}\Big). \end{aligned} \tag{S9.155}$$

Considering (S9.145), (S9.146), (S9.147), (S9.154), and (S9.155), as $n, p \to \infty$, we have $D_2 \to 0$.

Finally, combining (S9.130), $D_1 \to 0$, and $D_2 \to 0$, we complete the proof of Lemma 10. $\square$

S9.2 Proof of Lemma 11

Proof. In this section, we aim to prove

$$\mathbb{P}\Big(\max_{\tau_0\leq t\leq 1-\tau_0} \big\|\boldsymbol{C}^b(\lfloor nt\rfloor) - \widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\big\|_{\mathcal{G},\infty} \geq \epsilon|\mathcal{X}\Big) = o(1). \tag{S9.156}$$

Without loss of generality, we assume $\mathcal{G} = \{1, \ldots, p\}$. Using the triangle inequality, we have

$$\mathbb{P}\Big(\max_{\tau_0\leq t\leq 1-\tau_0} \big\|\boldsymbol{C}^b(\lfloor nt\rfloor) - \widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\big\|_{\infty} \geq \epsilon\Big) \leq E_1 + E_2, \tag{S9.157}$$

where

$$E_1 := \mathbb{P}\big(\max_{\tau_0\le t\le 1-\tau_0}\big\|\boldsymbol{C}^b(\lfloor nt\rfloor) - \boldsymbol{C}^{b,\mathrm{I}}(\lfloor nt\rfloor)\big\|_\infty \ge \epsilon/2|\mathcal{X}\big),$$

$$E_2 := \mathbb{P}\big(\max_{\tau_0\le t\le 1-\tau_0}\big\|\boldsymbol{C}^{b,\mathrm{I}}(\lfloor nt\rfloor) - \widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\big\|_\infty \ge \epsilon/2|\mathcal{X}\big). \tag{S9.158}$$

Hence, to prove (S9.156), we need to prove $E_1 \to 0$ and $E_2 \to 0$, respectively.

Control of E_2. We first consider E_2. Similar to the analysis in Section S9.1, we can show that

$$\max_{\tau_0\le t\le 1-\tau_0}\big\|\boldsymbol{C}^{b,\mathrm{I}}(\lfloor nt\rfloor) - \widetilde{\boldsymbol{C}}^{b,\mathrm{I}}(\lfloor nt\rfloor)\big\|_\infty \le O_p\Big(\max_{1\le j\le p} s_j\frac{\log(pn)}{\sqrt{n}}\Big). \tag{S9.159}$$

Note that

$$\epsilon := Cs^{(0)}\max_{1\le j\le p} s_j\frac{\log(pn)}{\sqrt{\lfloor n\tau_0\rfloor}}. \tag{S9.160}$$

By choosing a large enough constant C in ϵ, we have $E_2 \to 0$ as $n, p \to \infty$.

Control of E_1. Next, we consider E_1. Recall $\mathcal{E}$ defined in (S9.132). For E_1, we have

$$E_1 \le \mathbb{P}\big(\max_{\tau_0\le t\le 1-\tau_0}\big\|\boldsymbol{C}^b(\lfloor nt\rfloor) - \boldsymbol{C}^{b,\mathrm{I}}(\lfloor nt\rfloor)\big\|_\infty \ge \epsilon/2\cap\mathcal{E}\big) + \mathbb{P}(\mathcal{E}^c). \tag{S9.161}$$

By Theorem 1, we have $\mathbb{P}(\mathcal{E}^c) = o(1)$ as $n, p \to \infty$. By the definitions of $\boldsymbol{C}^b(\lfloor nt \rfloor)$ and $\boldsymbol{C}^{b,\mathrm{I}}(\lfloor nt \rfloor)$ in (S8.49), under the event $\mathcal{E}$, we have

$$\begin{aligned}
&\mathbb{P}\big(\max_{\tau_0 \le t \le 1-\tau_0} \big\|\boldsymbol{C}^b(\lfloor nt \rfloor) - \boldsymbol{C}^{\mathrm{I}}(\lfloor nt \rfloor)\big\|_{\mathcal{G},\infty} \ge \epsilon/2 \cap \mathcal{E} | \mathcal{X}\big) \\
&\le \mathbb{P}\big(\max_{t \in [\tau_0, 1-\tau_0]} \max_{1 \le j \le p} \sqrt{n} \frac{\lfloor nt \rfloor}{n} \frac{\lfloor nt \rfloor^*}{n} \big(\widehat{\sigma}_\epsilon^2 \widehat{\omega}_{j,j}\big)^{-1/2} |\Delta_j^{b,\lfloor nt \rfloor} - \Delta_j^{b,\lfloor nt \rfloor^*}| \ge \epsilon/2 | \mathcal{X}\big) \\
&\le \mathbb{P}\big(\max_{t \in [\tau_0, 1-\tau_0]} \|\boldsymbol{\Delta}^{b,(0,t)} - \boldsymbol{\Delta}^{b,(t,1)}\|_\infty \ge \tfrac{1}{2}\sqrt{c_\epsilon \kappa_2^{-1}/2}\,\epsilon n^{-1/2}\big) \\
&\le \mathbb{P}\big(\max_{t \in [\tau_0, 1-\tau_0]} \|\boldsymbol{\Delta}^{b,(0,t)}\|_\infty \ge C_1 \epsilon n^{-1/2}\big) + \mathbb{P}\big(\max_{t \in [\tau_0, 1-\tau_0]} \|\boldsymbol{\Delta}^{b,(t,1)}\|_\infty \ge C_2 \epsilon n^{-1/2}\big).
\end{aligned} \tag{S9.162}$$

To bound (S9.162), we need to obtain the upper bounds of $\max_{t \in [\tau_0, 1-\tau_0]} \|\boldsymbol{\Delta}^{b,(0,t)}\|_\infty$ and $\max_{t \in [\tau_0, 1-\tau_0]} \|\boldsymbol{\Delta}^{b,(t,1)}\|_\infty$.

Control of $\max_{t \in [\tau_0, 1-\tau_0]} \|\boldsymbol{\Delta}^{b,(0,t)}\|_\infty$. We first obtain the upper bound of $\max_{t \in [\tau_0, 1-\tau_0]} \|\boldsymbol{\Delta}^{b,(0,t)}\|_\infty$. To this end, we consider two cases:

Case 1 : $t \in [\tau_0, \widehat{t}_{0,\mathcal{G}}]$. In this case, by the definition of $\boldsymbol{\Delta}^{b,(0,t)}$ in (S8.44), it reduces to

$$\boldsymbol{\Delta}^{b,(0,t),\mathrm{I}} = -\big(\widehat{\boldsymbol{\Theta}} \widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{b,(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big). \tag{S9.163}$$

Using the fact that $\|\mathbf{A}\boldsymbol{x}\|_\infty \le \|\mathbf{A}\|_\infty \|\boldsymbol{x}\|_1$, we have

$$\|\boldsymbol{\Delta}^{b,(0,t),\mathrm{I}}\|_\infty \le \big\|\big(\widehat{\boldsymbol{\Theta}} \widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big\|_\infty \big\|\big(\widehat{\boldsymbol{\beta}}^{b,(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big)\big\|_1. \tag{S9.164}$$

By (S9.142), we have

$$\max_{t \in [\tau_0, \widehat{t}_{0,\mathcal{G}}]} \big\|\big(\widehat{\boldsymbol{\Theta}} \widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big\|_\infty \le O_p\Big(\sqrt{\frac{\log(pn)}{\lfloor n\tau_0 \rfloor}}\Big). \tag{S9.165}$$

Note that under $\mathbf{H}_0$, by Lemma 8, the lasso estimator $\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}$ has the following properties:

$$\begin{aligned}
&\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}-\boldsymbol{\beta}^{(0)}\|_q \le O_p\Big((s^{(0)})^{\frac{1}{q}}\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}}\Big),\ \text{for } q=1,2,\\
&\|\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}-\boldsymbol{\beta}^{(0)}\|_q \le O_p\Big((s^{(0)})^{\frac{1}{q}}\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor^*}}\Big),\ \text{for } q=1,2,\\
&\qquad \text{and } \widehat{s}^{(1)},\widehat{s}^{(2)} \le O_p(s^{(0)}),
\end{aligned} \tag{S9.166}$$

where $\widehat{s}^{(1)} := |\widehat{\mathcal{S}}^{(1)}|$ with $\widehat{\mathcal{S}}^{(1)} := \{1 \le j \le p : \widehat{\beta}_j^{(0,\widehat{t}_{0,\mathcal{G}})} \ne 0\}$ and $\widehat{s}^{(2)} = |\widehat{\mathcal{S}}^{(2)}|$ with $\widehat{\mathcal{S}}^{(2)} = \{1 \le j \le p : \widehat{\beta}_j^{(\widehat{t}_{0,\mathcal{G}},1)} \ne 0\}$. Given $\mathcal{X}$, using Lemma 8 again, for $q=1,2$, we have

$$\|\widehat{\boldsymbol{\beta}}^{b,(0,t)}-\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\|_q \quad \le O_p\Big((\widehat{s}^{(1)})^{\frac{1}{q}}\sqrt{\frac{\log(p)}{\lfloor nt\rfloor}}\Big) \le O_p\Big((s^{(0)})^{\frac{1}{q}}\sqrt{\frac{\log(p)}{\lfloor n\tau_0\rfloor}}\Big). \tag{S9.167}$$

Combining (S9.165) and (S9.167), conditional on $\mathcal{X}$, for the case of $t \in [\tau_0, \widehat{t}_{0,\mathcal{G}}]$, we have

$$\max_{t\in[\tau_0,\widehat{t}_{0,\mathcal{G}}]} \|\boldsymbol{\Delta}^{b,(0,t),\mathrm{I}}\|_\infty \le O_p\Big(s^{(0)}\frac{\log(pn)}{\lfloor n\tau_0\rfloor}\Big). \tag{S9.168}$$

Case 2 : $t \in [\widehat{t}_{0,\mathcal{G}}, 1-\tau_0]$. In this case, $\boldsymbol{\Delta}^{b,(0,t)}$ reduces to $\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}}$ in (S8.42). By its definition, we can decompose $\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}}$ into the following two terms:

$$\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}} = \boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}} + \boldsymbol{\Delta}_2^{b,(0,t),\mathrm{II}}, \tag{S9.169}$$

where

$$\begin{aligned}
\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}} &= -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{b,(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\big), \\
\boldsymbol{\Delta}_2^{b,(0,t),\mathrm{II}} &= -\frac{\lfloor nt\rfloor - \lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor nt\rfloor}\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(\widehat{t}_{0,\mathcal{G}},t)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\big), \\
\widehat{\boldsymbol{\Sigma}}_{(\widehat{t}_{0,\mathcal{G}},t)} &:= \frac{(\mathbf{X}_{(\widehat{t}_{0,\mathcal{G}},t)})^\top \mathbf{X}_{(\widehat{t}_{0,\mathcal{G}},t)}}{\lfloor nt\rfloor - \lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor + 1}.
\end{aligned} \tag{S9.170}$$

By (S9.169), for controlling $\boldsymbol{\Delta}^{b,(0,t),\mathrm{II}}$, we need to consider $\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}$ and $\boldsymbol{\Delta}_2^{b,(0,t),\mathrm{II}}$, respectively.

Control of $\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}$. We first consider $\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}$. Similar to the analysis of (S9.136) - (S9.142), we can prove that

$$\max_{t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0]} \Big\|\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,t)} - \mathbf{I}\big)\Big\|_\infty \le O_p\Big(\sqrt{\frac{\log(pn)}{\lfloor n\widehat{\tau}_{0,\mathcal{G}}\rfloor}}\Big) \le O_p\Big(\sqrt{\frac{\log(pn)}{\lfloor n\tau_0\rfloor}}\Big). \tag{S9.171}$$

Note that $\widehat{\boldsymbol{\beta}}^{b,(0,t)}$ is constructed using data both before $\lfloor n\widehat{\tau}_{0,\mathcal{G}}\rfloor$ and after $\lfloor n\widehat{\tau}_{0,\mathcal{G}}\rfloor$. By Lemma 8, conditional on $\mathcal{X}$, we have

$$\begin{aligned}
&\|\widehat{\boldsymbol{\beta}}^{b,(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}\|_1 \\
&\le C_1 \max\Big(\widehat{s}^{(1)}\sqrt{\frac{\log p}{n}}, \frac{\lfloor nt\rfloor - \lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor nt\rfloor}\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})} - \widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\|_1\Big), \\
&\le C_2 s^{(0)} \max\Big(\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}}, \sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor^*}}\Big)\big(\text{by}\quad (\text{S9.166})\big), \\
&\le C_3 s^{(0)}\sqrt{\frac{\log(p)}{\lfloor n\tau_0\rfloor}}.
\end{aligned} \tag{S9.172}$$

Combining (S9.171) and (S9.172), we have

$$\max_{t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0]} \|\boldsymbol{\Delta}_1^{b,(0,t),\mathrm{II}}\|_\infty \le O_p\Big(s^{(0)}\frac{\log(p)}{\lfloor n\tau_0\rfloor}\Big). \tag{S9.173}$$

Control of $\mathbf{\Delta}_2^{b,(0,t),\mathrm{II}}$. After bounding $\max_{t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0]}\|\mathbf{\Delta}_1^{b,(0,t),\mathrm{II}}\|_\infty$, we next consider $\max_{t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0]}\|\mathbf{\Delta}_2^{b,(0,t),\mathrm{II}}\|_\infty$. Similar to the analysis of (S9.136) - (S9.142), we have

$$\max_{t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0]}\Big\|\frac{\lfloor nt\rfloor-\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}{\lfloor nt\rfloor}\big(\widehat{\mathbf{\Theta}}\widehat{\mathbf{\Sigma}}_{(\widehat{t}_{0,\mathcal{G}},t)}-\mathbf{I}\big)\Big\|_\infty\le O_p\Big(\sqrt{\frac{\log(pn)}{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}}\Big).\quad\text{(S9.174)}$$

By (S9.166), we have

$$\begin{aligned}
&\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}-\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}\|_1\\
&\le\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_{0,\mathcal{G}})}-\boldsymbol{\beta}^{(0)}\|_1+\|\widehat{\boldsymbol{\beta}}^{(\widehat{t}_{0,\mathcal{G}},1)}-\boldsymbol{\beta}^{(0)}\|_1,\\
&\le C_1s^{(0)}\max\Big(\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor}},\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_{0,\mathcal{G}}\rfloor^*}}\Big),\\
&\le C_2s^{(0)}\sqrt{\log(p)/\lfloor n\tau_0\rfloor}.
\end{aligned}\quad\text{(S9.175)}$$

Combining (S9.174) and (S9.175), we have

$$\max_{t\in[\widehat{t}_{0,\mathcal{G}},1-\tau_0]}\|\mathbf{\Delta}_2^{b,(0,t),\mathrm{II}}\|_\infty\le O_p\Big(s^{(0)}\frac{\log(p)}{\lfloor n\tau_0\rfloor}\Big).\quad\text{(S9.176)}$$

Considering (S9.168), (S9.169), (S9.173), and (S9.176), we obtain

$$\max_{t\in[\tau_0,1-\tau_0]}\|\mathbf{\Delta}^{b,(0,t)}\|_\infty\le O_p\Big(s^{(0)}\frac{\log(pn)}{\lfloor n\tau_0\rfloor}\Big).\quad\text{(S9.177)}$$

Control of $\max_{t\in[\tau_0,1-\tau_0]}\|\mathbf{\Delta}^{b,(t,1)}\|_\infty$. After bounding $\max_{t\in[\tau_0,1-\tau_0]}\|\mathbf{\Delta}^{b,(0,t)}\|_\infty$, we next consider $\max_{t\in[\tau_0,1-\tau_0]}\|\mathbf{\Delta}^{b,(t,1)}\|_\infty$. Using a similar proof technique, we can obtain

$$\max_{t\in[\tau_0,1-\tau_0]}\|\mathbf{\Delta}^{b,(t,1)}\|_\infty\le O_p\Big(s^{(0)}\frac{\log(pn)}{\lfloor n\tau_0\rfloor}\Big).\quad\text{(S9.178)}$$

Finally, considering (S9.160), (S9.162), (S9.177), and (S9.178), by choosing a large enough constant C in ϵ, we have $E_1 \to 0$, which completes the proof of Lemma 11. □

S9.3 Proof of Lemma 12

Proof. We give the proof by contradiction. Suppose there is a constant $c < 1$ such that

$$\delta_{j^*} \leq c\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}.$$

On one hand, by the decomposition of $\boldsymbol{Z}(\lfloor nt \rfloor)$ in (S8.68), at time point $\widehat{t}_0$, we have:

$$\begin{aligned}
\|\boldsymbol{Z}(\lfloor n\widehat{t}_0 \rfloor)\|_{\mathcal{G},\infty} &:= Z_{j^*}(\lfloor n\widehat{t}_0 \rfloor) \\
&\leq \sqrt{n}\frac{\lfloor nt_0 \rfloor}{n}\frac{\lfloor n\widehat{t}_0 \rfloor^*}{n}\delta_{j^*} + C_2\sqrt{\log(|\mathcal{G}|\lfloor n\tau_0 \rfloor)} + o_p\big(\sqrt{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\big) \\
&\leq \sqrt{n}\frac{\lfloor nt_0 \rfloor}{n}\frac{\lfloor n\widehat{t}_0 \rfloor^*}{n}c(1 + o_p(1))\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}.
\end{aligned}$$

On the other hand, at time point t_0, we have:

$$\begin{aligned}
\|\boldsymbol{Z}(\lfloor nt_0 \rfloor)\|_{\mathcal{G},\infty} &= \max_{j\in\mathcal{G}} |Z_j(\lfloor nt_0 \rfloor)| \\
&\geq \sqrt{n}\frac{\lfloor nt_0 \rfloor}{n}\frac{\lfloor nt_0 \rfloor^*}{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} - C_2\sqrt{\log(|\mathcal{G}|\lfloor n\tau_0 \rfloor)} - o_p\big(\sqrt{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\big) \\
&\geq \sqrt{n}\frac{\lfloor nt_0 \rfloor}{n}\frac{\lfloor n\widehat{t}_0 \rfloor^*}{n}(1 - o_p(1))\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}.
\end{aligned}$$

Considering the above results, we have: $\mathbb{P}(\|\boldsymbol{Z}(\lfloor nt_0 \rfloor)\|_{\mathcal{G},\infty} > \|\boldsymbol{Z}(\lfloor n\widehat{t}_0 \rfloor)\|_{\mathcal{G},\infty}) \to 1$, which is contradicted to the fact that $\widehat{t}_0$ is the maximizer of $\|\boldsymbol{Z}(\lfloor nt \rfloor)\|_{\mathcal{G},\infty}$.

□

S9.4 Proof of Lemma 13

Proof of $\mathcal{H}_1$. Without loss of generality, we assume $\widehat{t}_0 \in [t_0, 1-\tau_0]$. The proof proceeds in two steps. In Step 1, we prove that

$$\big|\max_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor)\big| \geq \big|\min_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor)\big|. \tag{S9.179}$$

By noting that

$$\max_{j\in\mathcal{G}} |Z_j(\lfloor n\widehat{t}_0\rfloor)| = |\max_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor)| \vee |\min_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor)|,$$

we have $\max_{j\in\mathcal{G}} |Z_j(\lfloor n\widehat{t}_0\rfloor)| = |\max_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor)\big|$. In Step 2, we prove $\max_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor) \geq 0$. Note that $Z_{j^*}(\lfloor n\widehat{t}_0\rfloor) = \max_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor)$. Combining Steps 1 and 2, we complete the proof.

Now, we consider the two steps, respectively. By the decomposition of $\boldsymbol{Z}(\lfloor nt\rfloor)$ in (S8.68), at time point $\widehat{t}_0$, we have

$$\begin{aligned}
&\boldsymbol{Z}(\lfloor n\widehat{t}_0\rfloor) - \boldsymbol{\delta}(\widehat{t}_0)\\
&= \sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}\Big(\frac{1}{\lfloor n\widehat{t}_0\rfloor}\sum_{i=1}^{\lfloor n\widehat{t}_0\rfloor}\widehat{\boldsymbol{\xi}}_i - \frac{1}{\lfloor n\widehat{t}_0\rfloor^*}\sum_{i=\lfloor n\widehat{t}_0\rfloor+1}^{n}\widehat{\boldsymbol{\xi}}_i + \boldsymbol{R}^{(0,\widehat{t}_0),II} - \boldsymbol{R}^{(\widehat{t}_0,1),II}\Big).
\end{aligned} \tag{S9.180}$$

By Assumptions (**A.1**) – (**A.3**), using concentration inequalities, we can prove that with probability at least $1-(np)^{-C_1}$,

$$\Big\|\sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}\Big(\frac{1}{\lfloor n\widehat{t}_0\rfloor}\sum_{i=1}^{\lfloor n\widehat{t}_0\rfloor}\widehat{\boldsymbol{\xi}}_i - \frac{1}{\lfloor n\widehat{t}_0\rfloor^*}\sum_{i=\lfloor n\widehat{t}_0\rfloor+1}^{n}\widehat{\boldsymbol{\xi}}_i\Big)\Big\|_{\mathcal{G},\infty} \leq C_2\sqrt{\log(|\mathcal{G}|\lfloor n\tau_0\rfloor)}. \tag{S9.181}$$

Next, we consider the control of $\|\boldsymbol{R}^{(0,\widehat{t}_0),II}\|_{\mathcal{G},\infty}$ and $\|\boldsymbol{R}^{(\widehat{t}_0,1),II}\|_{\mathcal{G},\infty}$.

Control of $\|\boldsymbol{R}^{(0,\widehat{t}_0),II}\|_{\mathcal{G},\infty}$. By the definition of $\boldsymbol{R}^{(0,\widehat{t}_0),II}$ in (S8.67), using the triangle inequality, we have

$$\big\|\boldsymbol{R}^{(0,\widehat{t}_0),\mathrm{II}}\big\|_{\mathcal{G},\infty} \le \big\|\boldsymbol{R}_1^{(0,\widehat{t}_0),\mathrm{II}}\big\|_\infty + \big\|\boldsymbol{R}_2^{(0,\widehat{t}_0),\mathrm{II}}\big\|_\infty, \tag{S9.182}$$

where $\boldsymbol{R}_1^{(0,\widehat{t}_0),\mathrm{II}}$ and $\boldsymbol{R}_2^{(0,\widehat{t}_0),\mathrm{II}}$ are defined as

$$\begin{aligned}
\boldsymbol{R}_1^{(0,\widehat{t}_0),\mathrm{II}} &:= -\frac{\lfloor n\widehat{t}_0\rfloor - \lfloor nt_0\rfloor}{\lfloor n\widehat{t}_0\rfloor}\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t_0,\widehat{t}_0)} - \mathbf{I}\big)\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big),\\
\boldsymbol{R}_2^{(0,\widehat{t}_0),\mathrm{II}} &:= -\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,\widehat{t}_0)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)} - \boldsymbol{\beta}^{(1)}\big).
\end{aligned} \tag{S9.183}$$

Using the fact that $\|\mathbf{A}\boldsymbol{x}\|_\infty \le \|\mathbf{A}\|_\infty\|\boldsymbol{x}\|_1$ and by concentration inequalities, we have,

$$\|\boldsymbol{R}_1^{(0,\widehat{t}_0),\mathrm{II}}\|_\infty \quad \le C_1\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_0\rfloor}}\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_1 \le C_1 s\sqrt{\frac{\log(p)}{\lfloor n\widehat{t}_0\rfloor}}\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\|_\infty = o(\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}),$$

where the last equation comes from the assumption that $s\sqrt{\log(pn)/n\tau_0}\|\boldsymbol{\delta}\|_\infty/\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} = o(1)$. For $\|\boldsymbol{R}_2^{(0,\widehat{t}_0),\mathrm{II}}\|_\infty$, using Lemma 8 and concentration inequalities, we have

$$\begin{aligned}
\|\boldsymbol{R}_2^{(0,\widehat{t}_0),\mathrm{II}}\|_\infty \quad &\le C_1\sqrt{\frac{\log(pn)}{\lfloor n\widehat{t}_0\rfloor}}\Big(\frac{\lfloor n\widehat{t}_0\rfloor - \lfloor nt_0\rfloor}{\lfloor n\widehat{t}_0\rfloor}\big\|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\big\|_1\Big),\\
&\le C_2\sqrt{\frac{\log(pn)}{n\tau_0}}s\big\|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big\|_\infty := o(\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}).
\end{aligned} \tag{S9.184}$$

Control of $\|\boldsymbol{R}^{(\widehat{t}_0,1),II}\|_{\mathcal{G},\infty}$. By the definition of $\boldsymbol{R}^{(\widehat{t}_0,1),II}$ in (S8.67), and using Lemma 8, we have

$$\|\boldsymbol{R}^{(\widehat{t}_0,1)}\|_{\mathcal{G},\infty} \le \|\boldsymbol{R}^{(\widehat{t}_0,1)}\|_\infty \le O_p\Big(\sqrt{\frac{\log(pn)}{\lfloor n\widehat{t}_0\rfloor^*}}\Big)\Big\|\widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)} - \boldsymbol{\beta}^{(2)}\Big\|_1 \le O_p\big(s^{(2)}\frac{\log(pn)}{\lfloor n\widehat{t}_0\rfloor^*}\big). \tag{S9.185}$$

Note that we assume $s\sqrt{\log(pn)/n\tau_0} = o(1)$ with $s := s^{(1)} \vee s^{(2)}$ and $|\mathcal{G}| = p^{\gamma}$ with $\gamma \in (0,1]$. Considering the above results, we have

$$\sqrt{n}\|\boldsymbol{R}^{(0,\widehat{t}_0)}\|_{\mathcal{G},\infty} \leq o_p\big(\sqrt{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\big), \quad \sqrt{n}\|\boldsymbol{R}^{(\widehat{t}_0,1)}\|_{\mathcal{G},\infty} \leq o_p\big(\log(|\mathcal{G}|n)\big). \tag{S9.186}$$

Combining (S9.180) – (S9.186), with probability at least $1-(np)^{-C_1}$, we have

$$\|\boldsymbol{Z}(\lfloor n\widehat{t}_0\rfloor) - \boldsymbol{\delta}(\widehat{t}_0)\|_{\mathcal{G},\infty} := \max_{j\in\mathcal{G}} |Z_j(\lfloor n\widehat{t}_0\rfloor) - \delta_j(\widehat{t}_0)| \leq K^*, \tag{S9.187}$$

where $K^* := C_2\sqrt{\log(|\mathcal{G}|n)} + o\big(\sqrt{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\big)$. Note that

$$\delta_j(\widehat{t}_0) := \sqrt{n}\frac{\lfloor nt_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}(\beta_j^{(1)} - \beta_j^{(2)}) \geq 0.$$

By (S9.187), using the fact $|\max_i a_i - \max_i b_i| \leq \max_i |a_i - b_i|$ for two sequences $\{a_i\}$ and $\{b_i\}$, we have

$$\begin{aligned} &\min_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor) \geq -K^*, \\ &\max_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor) \geq \max_{j\in\mathcal{G}} \delta_j(\lfloor n\widehat{t}_0\rfloor) - K^* \geq K^*, \end{aligned} \tag{S9.188}$$

where the last inequality in (S9.188) comes from the assumption $\|\boldsymbol{\delta}\|_{\mathcal{G},\infty} \gg \sqrt{\log(pn)/n\tau_0}$. By (S9.188), we have $\big|\max_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor)\big| \geq \big|\min_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor)\big|$ and $\max_{j\in\mathcal{G}} Z_j(\lfloor n\widehat{t}_0\rfloor) \geq 0$, which finishes the proof of $\mathcal{H}_1$ in Lemma 13.

Proof of $\mathcal{H}_2$. Note that the proof of $\mathcal{H}_2$ is similar and easier, to save space, we omit the details.

S9.5 Proof of Lemma 14

Proof. We first bound $\boldsymbol{R}^{(0,\widehat{t}_0),\mathrm{II}}$. By its definition in (S8.67), we have

$$\sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}\|\boldsymbol{R}^{(0,\widehat{t}_0),\mathrm{II}}\|_{\mathcal{G},\infty} \leq \underbrace{\sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor - \lfloor nt_0\rfloor}{n}\big\|\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(t_0,\widehat{t}_0)} - \mathbf{I}\big)\big(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(2)}\big)\big\|_\infty}_{I}$$
$$+ \underbrace{\sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}\big\|\big(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{(0,\widehat{t}_0)} - \mathbf{I}\big)\big(\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)} - \boldsymbol{\beta}^{(1)}\big)\big\|_\infty}_{II}. \tag{S9.189}$$

For I, using concentration inequalities, we have

$$\begin{aligned} I &\leq C_1\sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor - \lfloor nt_0\rfloor}{n}\sqrt{\frac{\log(pn)}{\lfloor n\widehat{t}_0\rfloor - \lfloor nt_0\rfloor}}\|\boldsymbol{\delta}\|_1, \\ &\leq C_1\sqrt{\log(|\mathcal{G}|n)\frac{\lfloor n\widehat{t}_0\rfloor - \lfloor nt_0\rfloor}{n}}s\|\boldsymbol{\delta}\|_\infty. \end{aligned} \tag{S9.190}$$

For II, using concentration inequalities and by Lemma 8, we have

$$\begin{aligned} II &\leq C_2\sqrt{n}\sqrt{\frac{\log(pn)}{\lfloor n\widehat{t}_0\rfloor}}\big\|\widehat{\boldsymbol{\beta}}^{(0,\widehat{t}_0)} - \boldsymbol{\beta}^{(1)}\big\|_1 \\ &\leq C_2\sqrt{n}\sqrt{\frac{\log(pn)}{\lfloor n\widehat{t}_0\rfloor}}\frac{\lfloor n\widehat{t}_0\rfloor - \lfloor nt_0\rfloor}{n}s\big\|\boldsymbol{\delta}\big\|_\infty, \\ &\leq C_2\sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor - \lfloor nt_0\rfloor}{n}\sqrt{\frac{\log(pn)}{n\tau_0}}s\|\boldsymbol{\delta}\|_\infty \\ &= o\Big(\sqrt{n}\frac{\lfloor nt_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor - \lfloor nt_0\rfloor}{n}\|\boldsymbol{\delta}\|_{\mathcal{G},\infty}\Big). \end{aligned} \tag{S9.191}$$

After bounding $\boldsymbol{R}^{(0,\widehat{t}_0),\mathrm{II}}$ in (S9.190) and (S9.191), we next consider $\boldsymbol{R}^{(\widehat{t}_0,1),\mathrm{II}}$.

Using concentration inequalities and the upper bound of estimation error of $\widehat{\boldsymbol{\beta}}^{(\widehat{t}_0,1)}$ for $\boldsymbol{\beta}^{(2)}$, we have

$$\sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}\|\boldsymbol{R}^{(\widehat{t}_0,1),\mathrm{II}}\|_{\mathcal{G},\infty} \leq \sqrt{n}\frac{\lfloor n\widehat{t}_0\rfloor}{n}\frac{\lfloor n\widehat{t}_0\rfloor^*}{n}\|\boldsymbol{R}^{(\widehat{t}_0,1),\mathrm{II}}\|_\infty \leq C\sqrt{n}s\frac{\log(|\mathcal{G}|n)}{n}, \tag{S9.192}$$

where the last inequality comes from the assumption that $|\mathcal{G}| = p^{\gamma}$ with $\gamma \in (0, 1]$.

We next consider $\boldsymbol{R}^{(0,t_0),\mathrm{II}}$ and $\boldsymbol{R}^{(t_0,1),\mathrm{II}}$. Using concentration inequalities and the upper bounds of estimation errors of $\widehat{\boldsymbol{\beta}}^{(0,t_0)}$ and $\widehat{\boldsymbol{\beta}}^{(t_0,1)}$ (see Lemma 8), we have

$$\sqrt{n}\frac{\lfloor nt_0 \rfloor}{n}\frac{\lfloor nt_0 \rfloor^*}{n}\left(\|\boldsymbol{R}^{(0,t_0),\mathrm{II}}\|_{\mathcal{G},\infty} \vee \|\boldsymbol{R}^{(t_0,1),\mathrm{II}}\|_{\mathcal{G},\infty}\right) \leq C\sqrt{n}s\frac{\log(|\mathcal{G}|n)}{n}. \tag{S9.193}$$

Finally, combining (S9.190) – (S9.193), we complete the proof. □

S10 Proofs of useful lemmas

S10.1 Proof of Lemma 5

Proof. The result of Lemma 5 can be obtained by using Bernstein's inequality for sub-exponential distributions. To save space, we omit the details here. □

S10.2 Proof of Lemma 6

Proof. We only consider the proof of (S7.14). The proof of (S7.15) is similar.

Using some straightforward calculations, we have

$$\begin{aligned}
&\frac{1}{2\lfloor nt\rfloor}\big\|\boldsymbol{Y}_{(0,t)}-\mathbf{X}_{(0,t)}\widehat{\boldsymbol{\beta}}^{(0,t)}\big\|_2^2-\frac{1}{2\lfloor nt\rfloor}\big\|\boldsymbol{Y}_{(0,t)}-\mathbf{X}_{(0,t)}\boldsymbol{\beta}^{(0,t)}\big\|_2^2,\\
&\quad=\frac{1}{2\lfloor nt\rfloor}\big\|\mathbf{X}_{(0,t)}(\boldsymbol{\beta}^{(0,t)}-\widehat{\boldsymbol{\beta}}^{(0,t)})\big\|_2^2-\frac{1}{\lfloor nt\rfloor}(\widehat{\boldsymbol{\beta}}^{(0,t)}-\boldsymbol{\beta}^{(0,t)})^\top\mathbf{X}_{(0,t)}^\top(\boldsymbol{Y}_{(0,t)}-\mathbf{X}_{(0,t)}\boldsymbol{\beta}^{(0,t)}).
\end{aligned} \tag{S10.194}$$

By noting that $\widehat{\boldsymbol{\beta}}^{(0,t)}$ is the minimizer of (2.9), we have

$$\frac{1}{2\lfloor nt\rfloor}\big\|\boldsymbol{Y}_{(0,t)}-\mathbf{X}_{(0,t)}\widehat{\boldsymbol{\beta}}^{(0,t)}\big\|_2^2-\frac{1}{2\lfloor nt\rfloor}\big\|\boldsymbol{Y}_{(0,t)}-\mathbf{X}_{(0,t)}\boldsymbol{\beta}^{(0,t)}\big\|_2^2\le\lambda_1(t)\|\boldsymbol{\beta}^{(0,t)}\|_1-\lambda_1(t)\|\widehat{\boldsymbol{\beta}}^{(0,t)}\|_1. \tag{S10.195}$$

Note that $|\boldsymbol{x}^\top\boldsymbol{y}|\le\|\boldsymbol{x}\|_\infty\|\boldsymbol{y}\|_1$ for two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Combining (S10.194) and (S10.195), under the event $\mathcal{A}^{(1)}(t)$, taking $\lambda_1(t)=2\lambda^{(1)}$ as defined in (S7.13) , we have

$$\begin{aligned}
&\frac{1}{2\lfloor nt\rfloor}\big\|\mathbf{X}_{(0,t)}(\boldsymbol{\beta}^{(0,t)}-\widehat{\boldsymbol{\beta}}^{(0,t)})\big\|_2^2\\
&\le\lambda_1(t)\big\|\boldsymbol{\beta}^{(0,t)}\big\|_1-\lambda_1(t)\big\|\widehat{\boldsymbol{\beta}}^{(0,t)}\big\|_1+\frac{\lambda_1(t)}{2}\big\|\boldsymbol{\beta}^{(0,t)}-\widehat{\boldsymbol{\beta}}^{(0,t)}\big\|_1,\\
&\le\lambda_1(t)\big\|\boldsymbol{\beta}^{(0,t)}\big\|_1-\lambda_1(t)\big\|\widehat{\boldsymbol{\beta}}^{(0,t)}\big\|_1+\frac{\lambda_1(t)}{2}\big\|\boldsymbol{\beta}^{(0,t)}-\widehat{\boldsymbol{\beta}}^{(0,t)}\big\|_1.
\end{aligned} \tag{S10.196}$$

Note that

$$|\beta_j^{(0,t)}|-|\widehat{\beta}_j^{(0,t)}|+|\beta_j^{(0,t)}-\widehat{\beta}_j^{(0,t)}|=0,\quad\text{if}\ \ j\in J^c(\boldsymbol{\beta}^{(0,t)}). \tag{S10.197}$$

Adding $2^{-1}\|\boldsymbol{\beta}^{(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,t)}\|_1$ on both sides of (S10.196), and considering (S10.197), we have

$$\begin{aligned}
&\frac{1}{2\lfloor nt \rfloor}\big\|\mathbf{X}_{(0,t)}(\boldsymbol{\beta}^{(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,t)})\big\|_2^2 + \frac{\lambda_1(t)}{2}\big\|\boldsymbol{\beta}^{(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,t)}\big\|_1, \\
&\leq \lambda_1(t)\big\|\boldsymbol{\beta}^{(1)}_{J(\boldsymbol{\beta}^{(0,t)})}\big\|_1 - \lambda_1(t)\big\|\widehat{\boldsymbol{\beta}}^{(0,t)}_{J(\boldsymbol{\beta}^{(1)})}\big\|_1 + \lambda_1(t)\big\|(\boldsymbol{\beta}^{(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,t)})_{J(\boldsymbol{\beta}^{(0,t)})}\big\|_1, \\
&\leq 2\lambda_1(t)\big\|(\boldsymbol{\beta}^{(0,t)} - \widehat{\boldsymbol{\beta}}^{(0,t)})_{J(\boldsymbol{\beta}^{(0,t)})}\big\|_1,
\end{aligned} \tag{S10.198}$$

which completes the proof of (S7.14). □

S10.3 Proof of Lemma 7

Proof. Note that by Lemma 6, and the URE conditions in Assumption (A.3), under the event $\{\mathcal{A}(t) \cap \mathcal{B}(t)\}$, using standard analysis of lasso estimation (see Pages 1728 – 1729 in Bickel et al. (2009)), one can prove that (S7.17) holds. To save space, we omit the details. □

S10.4 Proof of Lemma 9

Proof. By definitions of $\mathcal{A}(t)$ and $\mathcal{B}(t)$, to prove (S7.18), we need to bound $\mathbb{P}(\mathcal{A}^c(t))$ and $\mathbb{P}(\mathcal{B}^c(t))$, respectively. We first consider $\mathbb{P}(\mathcal{A}^c(t))$. Before that, we need some notations. We denote $\boldsymbol{\beta}_i$ as the regression coefficients for the i-th observation. By definition, we have

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}^{(1)}\mathbf{1}\{i \leq \lfloor nt_0 \rfloor\} + \boldsymbol{\beta}^{(2)}\mathbf{1}\{i > \lfloor nt_0 \rfloor\}.$$

Recall $\boldsymbol{\beta}^{(0,t)}$ as $\boldsymbol{\beta}^{(0,t)} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \mathbb{E}\big\|\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\boldsymbol{\beta}\big\|_2^2$. By the first order condition, we have:

$$\mathbb{E}\big[\mathbf{X}_{(0,t)}^\top(\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\boldsymbol{\beta}^{(0,t)})\big] = \mathbb{E}\big[\sum_{i=1}^{\lfloor nt\rfloor}\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^{(0,t)})\big] = \mathbf{0}.$$

Hence, by the first order condition, we have:

$$\Big\|\frac{1}{\lfloor nt\rfloor}\big(\mathbf{X}_{(0,t)}\big)^\top(\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\boldsymbol{\beta}^{(0,t)}\big)\Big\|_\infty = \Big\|\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\big(\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^{(0,t)}) - \mathbb{E}[\boldsymbol{X}_i(Y_i - \boldsymbol{X}_i^\top\boldsymbol{\beta}^{(0,t)})]\big)\Big\|_\infty.$$

By noting that $Y_i = \boldsymbol{X}_i^\top\boldsymbol{\beta}_i + \epsilon_i$, $\mathbb{E}\epsilon_i = 0$, and the independence between ϵ_i and $\boldsymbol{X}_i$, we have:

$$\begin{aligned}&\frac{1}{\lfloor nt\rfloor}\big(\mathbf{X}_{(0,t)}\big)^\top(\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\boldsymbol{\beta}^{(0,t)})\\ &= \frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\boldsymbol{X}_i\epsilon_i + \frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\big(\boldsymbol{X}_i\boldsymbol{X}_i^\top(\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)}) - \mathbb{E}[\boldsymbol{X}_i\boldsymbol{X}_i^\top(\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)})]\big).\end{aligned}$$

Based on the above decomposition, we have:

$$\Big\|\frac{1}{\lfloor nt\rfloor}\big(\mathbf{X}_{(0,t)}\big)^\top(\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)}\boldsymbol{\beta}^{(0,t)}\big)\Big\|_\infty \le I + II,$$

where

$$I = \Big\|\frac{1}{\lfloor nt\rfloor}\boldsymbol{\epsilon}_{(0,t)}^\top\mathbf{X}_{(0,t)}\Big\|_\infty, II = \Big\|\frac{1}{\lfloor nt\rfloor}\sum_{i=1}^{\lfloor nt\rfloor}\big(\boldsymbol{X}_i\boldsymbol{X}_i^\top(\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)}) - \mathbb{E}[\boldsymbol{X}_i\boldsymbol{X}_i^\top(\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)})]\big)\Big\|_\infty.$$

Control of I. We first consider I. By Assumptions (**A.1**) and (**A.2**), for $1 \le i \le n$ and $1 \le j \le p$, $\epsilon_i X_{i,j}$ follows the sub-exponential distribution.

By Bernstein's inequality, for each $x > 0$, we have

$$\begin{aligned}
&\mathbb{P}\big(\big\|\frac{1}{\lfloor nt \rfloor}\boldsymbol{\epsilon}_{(0,t)}^{\top}\mathbf{X}_{(0,t)}\big\|_{\infty} \geq x\big) \\
&= \mathbb{P}\Big(\bigcup_{1\leq j\leq p}\Big|\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}\epsilon_i X_{i,j}\Big| \geq x\Big), \\
&\leq p \max_{1\leq j\leq p}\mathbb{P}\Big(\Big|\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}\epsilon_i X_{i,j}\Big| \geq x\Big), \\
&\leq C_1 p \exp(-C_2\lfloor nt \rfloor x^2).
\end{aligned} \tag{S10.199}$$

By (S10.199), taking $\lambda^{(1)} = K_1\sqrt{\log(pn)/\lfloor nt \rfloor}$ for some big enough constant $K_1 > 0$, we have

$$\mathbb{P}\Big(\big\|\frac{1}{\lfloor nt \rfloor}\boldsymbol{\epsilon}_{(0,t)})^{\top}\mathbf{X}_{(0,t)}\big\|_{\infty} \geq \lambda^{(1)}\Big) \leq C_3(pn)^{-C_4}, \tag{S10.200}$$

where C_3 and C_4 are some big enough constants.

Control of II. Next, we consider II. Note that for $t \in [\tau_0, t_0]$, $II = 0$. Hence, in what follows, we consider the non-trivial case that $t \in [t_0, 1-\tau_0]$. Let

$$Z_i = \boldsymbol{X}_i^{\top}(\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)})/\|\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)}\|_2, \quad W_i = \|\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)}\|_2.$$

By Assumption (**A**.**1**), Z_i follows the sub-Gaussian distributions. Moreover, By Assumptions (**A**.**1**) and (**A**.**2**), for $1 \leq i \leq n$ and $1 \leq j \leq p$, $Z_i X_{i,j}$ follows the sub-exponential distribution. Hence, for II, we have

$$\begin{aligned}
&\Big\|\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}\big(\boldsymbol{X}_i\boldsymbol{X}_i^{\top}(\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)}) - \mathbb{E}[\boldsymbol{X}_i\boldsymbol{X}_i^{\top}(\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)})]\big)\Big\|_{\infty} \\
&= \max_j\Big|\frac{1}{\lfloor nt \rfloor}\sum_{i=1}^{\lfloor nt \rfloor}W_i\big(X_{i,j}Z_i - \mathbb{E}[X_{i,j}Z_i]\big)\Big|.
\end{aligned}$$

For each fixed j, using concentration inequality for weighted sub-exponential sums, we have:

$$\begin{aligned}
&\mathbb{P}(\max_j |\frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} W_i \big(X_{i,j} Z_i - \mathbb{E}[X_{i,j} Z_i]\big)| \geq x) \\
&\quad \leq p \max_j \mathbb{P}(|\frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} W_i \big(X_{i,j} Z_i - \mathbb{E}[X_{i,j} Z_i]\big)| \geq x) \\
&\quad \leq 2p \exp(\frac{-C_1 \lfloor nt \rfloor^2 x^2}{\|\boldsymbol{W}\|^2}).
\end{aligned}$$

Note that by definition, we have $\|\boldsymbol{W}\| = \sqrt{\lfloor nt_0 \rfloor (1 - \lfloor nt_0 \rfloor / \lfloor nt \rfloor)} \|\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}\|_2 \leq \sqrt{\lfloor nt_0 \rfloor} C_{\boldsymbol{\Delta}}$. Hence, taking $\lambda^{(1)} = K_2 \sqrt{\log(p)} \|\boldsymbol{W}\| / \lfloor nt \rfloor = K_2' \sqrt{\log(pn) / \lfloor nt \rfloor}$, we have:

$$\mathbb{P}\Big\{ \Big\| \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} \big(\boldsymbol{X}_i \boldsymbol{X}_i^\top (\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)}) - \mathbb{E}[\boldsymbol{X}_i \boldsymbol{X}_i^\top (\boldsymbol{\beta}_i - \boldsymbol{\beta}^{(0,t)})] \big) \Big\|_\infty \geq \lambda^{(1)} \Big\} \leq C_3 (p)^{-C_4}.$$

for some big enough $C_3, C_4 > 0$. Combining the above two cases, taking $\lambda^{(1)} = K_1 \sqrt{\log(pn) / \lfloor nt \rfloor}$ for some big enough constant $K_1 > 0$, we have:

$$\mathbb{P}\Big\{ \big\| \frac{1}{\lfloor nt \rfloor} \big(\mathbf{X}_{(0,t)} \big)^\top (\boldsymbol{Y}_{(0,t)} - \mathbf{X}_{(0,t)} \boldsymbol{\beta}^{(0,t)} \big\|_\infty \geq \lambda^{(1)} \Big\} \leq C_3 (pn)^{-C_4}.$$

With similar arguments as above, we can also prove that

$$\mathbb{P}\Big(\mathcal{B}^c(t) \Big) \leq C_3 (np)^{-C_4}. \tag{S10.201}$$

Finally, combining (S10.200) and (S10.201), and noting that

$$\begin{aligned}
&\mathbb{P}\Big(\bigcap_{t\in[\tau_0,1-\tau_0]}\big\{\mathcal{A}(t)\cap\mathcal{B}(t)\big\}\\
&= 1-\mathbb{P}\Big(\bigcup_{t\in[\tau_0,1-\tau_0]}\big\{\mathcal{A}^c(t)\cup\mathcal{B}^c(t)\big\},\\
&\geq 1-\sum_{t\in[\tau_0,1-\tau_0]}\big(\mathbb{P}(\mathcal{A}^c(t))+\mathbb{P}(\mathcal{B}^c(t))\big),\\
&\geq 1-C_1(np)^{-C_2},
\end{aligned} \tag{S10.202}$$

we complete the proof of Lemma 9. □

References

Bai, Y. and A. Safikhani (2023). A unified framework for change point detection in high-dimensional linear models. *Statistica Sinica 33*, 1–28.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics 37*(4), 1705–1732.

Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Cho, H. and D. Owens (2022). High-dimensional data segmentation in regression settings permitting heavy tails and temporal dependence. *arXiv preprint arXiv:2209.08892*.

Folstein, M. F., S. E. Folstein, and P. R. Mchugh (1975). ”Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research 12*(3), 0–198.

He, Z., D. Cheng, and Y. Zhao (2023). Multiple testing of local extrema for detection of structural breaks in piecewise linear models. *arXiv preprint arXiv:2308.04368*.

Kaul, A., S. B. Fotopoulos, V. K. Jandhyala, and A. Safikhani (2021). Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics 15*, 71–134.

Kaul, A., V. K. Jandhyala, and S. B. Fotopoulos (2019). An efficient two step algorithm for high dimensional change point regression models without grid search. *The Journal of Machine Learning Research 20*, 1–40.

Liu, B., C. Zhou, X.-S. Zhang, and Y. Liu (2020). A unified data-adaptive framework for high dimensional change point detection. *Journal of Royal Statistical Society, Series B (Statistical Methodology) 82*(4), 933–963.

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.

Raskutti, G., M. J. Wainwright, and B. Yu (2010). Restricted eigenvalue properties for correlated Gaussian designs. *The Journal of Machine Learning Research 11*, 2241–2259.

Van De Geer, S. and P. Bühlmann (2009). On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics 3*, 1360–1392.

Van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics 42*(3), 1166–1202.

Vostrikova, L. Y. (1981). Detecting disorder in multidimensional random process. *Soviet Math. Dokl 24*, 55–59.

Wang, D., Z. Zhao, K. Z. Lin, and R. Willett (2021). Statistically and computationally efficient change point localization in regression settings. *The Journal of Machine Learning Research 22*(248), 1–46.

Wang, F., O. Madrid, Y. Yu, and A. Rinaldo (2022). Denoising and change point localisation in piecewise-constant high-dimensional regression coefficients. In *International Conference on Artificial Intelligence and Statistics*, pp. 4309–4338. PMLR.

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*.

Yu, G., Y. Liang, S. Lu, and Y. Liu (2020). Confidence intervals for sparse penalized regression. *Journal of the American Statistical Association 115*, 794–809.

Yu, G. and Y. Liu (2016). Sparse regression incorporating graphical structure among predictors. *Journal of the American Statistical Association 111*, 707–720.

Zhang, D. and D. Shen (2012). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage 59*, 895–907.

Zhou, C., W.-X. Zhou, X.-S. Zhang, and H. Liu (2018). A unified framework for testing high dimensional parameters: a data-adaptive approach. *Preprint arXiv:1808.02648*.