CHANGE POINT INFERENCE IN HIGH-DIMENSIONAL REGRESSION MODELS UNDER TEMPORAL DEPENDENCE

Haotian Xu, Daren Wang, Zifeng Zhao, Yi Yu

LIDAM Discussion Paper ISBA
2022 / 27

ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

Change point inference in high-dimensional regression models under temporal dependence

Haotian Xu[1], Daren Wang[2], Zifeng Zhao[3], and Yi Yu[4]

[1]Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain
[2]Department of Statistics, University of Notre Dame
[3]Mendoza College of Business, University of Notre Dame
[4]Department of Statistics, University of Warwick

September 1, 2022

Abstract

This paper concerns about the limiting distributions of change point estimators, in a high-dimensional linear regression time series context, where a regression object $(y_t, X_t) \in \mathbb{R} \times \mathbb{R}^p$ is observed at every time point $t \in \{1, \ldots, n\}$. At unknown time points, called change points, the regression coefficients change, with the jump sizes measured in ℓ_2-norm. We provide limiting distributions of the change point estimators in the regimes where the minimal jump size vanishes and where it remains a constant. We allow for both the covariate and noise sequences to be temporally dependent, in the functional dependence framework, which is the first time seen in the change point inference literature. We show that a block-type long-run variance estimator is consistent under the functional dependence, which facilitates the practical implementation of our derived limiting distributions. We also present a few important byproducts of our analysis, which are of their own interest. These include a novel variant of the dynamic programming algorithm to boost the computational efficiency, consistent change point localisation rates under temporal dependence and a new Bernstein inequality for data possessing functional dependence. Extensive numerical results are provided to support our theoretical results. The proposed methods are implemented in the R package `changepoints` (Xu et al., 2021).

Keywords. High-dimensional linear regression; Change point inference; Functional dependence; Long-run variance; Confidence interval.

1 Introduction

Given a sequence of data $\{(y_t, X_t)\}_{t=1}^n \subset \mathbb{R} \times \mathbb{R}^p$, with the dimensionality p being a function of the sample size n, assume that

$$y_t = X_t^\top \beta_t^* + \epsilon_t, \quad t = 1, \ldots, n, \tag{1}$$

where $\{y_t\}_{t=1}^n$ are the responses, $\{X_t\}_{t=1}^n$ are the covariates, $\{\epsilon_t\}_{t=1}^n$ are the error terms and $\{\beta_t^*\}_{t=1}^n$ are true regression coefficients. Across time, assume that the coefficients sequence $\{\beta_t^*\}_{t=1}^n$ possesses a piecewise-constant pattern, i.e.

$$\beta_t^* \neq \beta_{t-1}^*, \quad \text{if and only if} \quad t \in \{\eta_1, \ldots, \eta_K\}, \tag{2}$$

where $\{\eta_k\}_{k=1}^K \subset \{1, \ldots, n\}$ are called change points, satisfying that

$$1 = \eta_0 < \eta_1 < \cdots < \eta_K < \eta_{K+1} = n + 1. \tag{3}$$

We refer to the model specified in (1), (2) and (3) as the high-dimensional linear regression model with change points. Data of this pattern commonly arise from various application areas, including biology, climatology, economics, finance and neuroscience. As a concrete example, Rinaldo et al. (2021) studied an air quality data set (Chu, 2016), with the daily index data of Particular Matter 10 in 2015 in Banqiao as responses, other climate and air quality indices of Banqiao and other nearby cities as covariates. Rinaldo et al. (2021) localised two change points of the regression coefficients, corresponding to the first strong typhoon in 2015 in the area and the beginning of the severe air-pollution season starting in fall/winter.

The considered high-dimensional linear regression model with change points falls into the vast body of change point analysis, which has been identified as one of the major challenges for modern data applications (National Research Council, 2013). The primary interest of change point analysis is to study the possibly piecewise-constant pattern in various data types. The best studied data type is fixed-dimensional time series, see e.g. Yao (1988), Yao and Au (1989), Shao and Zhang (2010), Cao and Wu (2015), Wang et al. (2020), Verzelen et al. (2020) and Wang et al. (2021c). We refer to Csörgő and Horváth (1997), Brodsky and Darkhovsky (2013) for book-length treatments, and to Aue and Horváth (2013), Casini and Perron (2019) for survey articles. Due to the increasing availability of big data, high-dimensional data have recently received great attention in change point analysis, including analysis in high-dimensional vector sequences (e.g. Wang et al., 2021a; Pilliat et al., 2020), matrices/networks sequences (e.g. Bhattacharjee et al., 2018; Wang et al., 2021a), functional curves sequences (e.g. Berkes et al., 2009a; Aue et al., 2018; Madrid Padilla et al., 2022), among many others.

Depending on the specific statistical task, work on change point analysis (in the high-dimensional context) can be further grouped into several branches: (i) Chen et al. (2021), Yu and Chen (2021), Wang et al. (2022), Zhang et al. (2021) and Wang and Zhao (2022) conducted hypothesis testing on the existence of change points. This task is referred to as testing. (ii) Lee et al. (2016), Kaul et al. (2019), Wang et al. (2021b), Rinaldo et al. (2021), Chen et al. (2021), Yu and Chen (2021) and Kaul and Michailidis (2021) provided estimation error controls on the estimated change point locations. This task is referred to as localisation/estimation. (iii) Bai (2010), Wang and Shao (2020), Kaul and Michailidis (2021) and Chen et al. (2021) derived limiting distributions of the change point estimators. This task is referred to as inference.

For all three statistical tasks we summarised above, a handle on the unknown variance is always desirable, and in fact vital for the testing and inference tasks. However, most of the existing change point literature in high-dimension provides only heuristic variance estimators justified through numerical experiments (e.g. Kaul and Michailidis, 2021; Wang and Zhao, 2022). For the problem of mean change in a high-dimensional time series, Chen et al. (2021) proposed a robust block-type estimator for the long-run covariance matrix, with theoretical justification, and Zhang et al. (2021) and Wang et al. (2022) cleverly avoided variance estimation via the use of U-statistics and self-normalization.

Besides different data types and different tasks, different temporal dependence notions have also been seen in the change point analysis literature. Temporal dependence assumptions include independence assumptions, linear process assumptions, different types of mixing conditions (α, β, τ, ρ, etc.) and the functional dependence assumption (Wu, 2005). In particular, the functional dependence measure (see Section 1.2 for more details) covers a large class of dependent processes, including linear and nonlinear processes, and enjoys mathematical convenience when dealing with more complex data types. Incorporating temporal dependence in high-dimensional data is in general challenging but is attracting increasing attention due to its practical importance. In the existing literature, the high-dimensional linear regression model with change points is studied under independence assumptions (Lee et al., 2016; Kaul et al., 2019; Wang et al., 2021b; Rinaldo et al., 2021) for estimation and β-mixing conditions (Wang and Zhao, 2022) for testing. To the best of our knowledge, this model is yet to be studied under functional dependence.

In this paper, we consider the inference task under the high-dimensional linear regression model with change points specified in (1)-(3). To our best knowledge, in the existing literature, such model has only been studied in the testing (e.g. Wang and Zhao, 2022) or estimation (e.g. Lee et al., 2016; Kaul et al., 2019; Wang et al., 2021b) aspects, while the inference task has yet to be explored. More discussions on connections with existing literature are deferred to Section 3.3 after we present our main results. To achieve broad applicability, we allow for both the covariate and noise sequences to possess temporal dependence, detailed in Section 1.2. This is, arguably, the most general setting considered in the literature. Under temporal dependence and high-dimensionality, we further derive and justify a block-type long-run variance estimator, which facilitates the construction of confidence intervals for the estimated change points.

1.1 List of contributions

We summarise the contributions of this paper as follows.

- We conduct inference for the high-dimensional linear regression model with change points, allowing for temporal dependence and potentially multiple change points. See Section 3.2. Depending on whether the jump size vanishes as the sample size diverges, the limiting distributions are shown to have two different non-degenerate regimes. This two-regime phenomenon has previously been observed for mean change under the fixed-dimensional time series (e.g. Yao, 1987; Yao and Au, 1989; Bai, 1994), high-dimensional vector time series (e.g. Kaul and Michailidis, 2021) or functional time series setting (e.g. Aue et al., 2009). We believe such result is the first time seen under the high-dimensional regression setting, which requires substantial theoretical effort due to high-dimensionality and temporal dependence.

- To facilitate the practical implementation of our derived limiting distributions, it is vital to provide a theoretically-justified long-run variance estimator. In Section 4, we provide a consistent, block-type estimator for the long-run variance. Such result is again, the first time seen in the literature for high-dimensional regression under functional dependence. Based on the long-run variance estimator, we provide a completely data-driven procedure for constructing confidence intervals for change points in Section 4.1.

- In addition, there are three notable byproducts of our analysis that are of independent interest.
  - To improve computational efficiency, we propose a new dynamic programming algorithm to solve the optimal partition problem (e.g. Friedrich et al., 2008; Killick et al., 2012) with a Lasso-estimation sub-routine, which lowers the computational complexity from $O(n^3p^2)$ to $O(n^2p^2)$. See Section 2.1.

– In addition to the limiting distributions of the change point estimators, we also provide its localisation properties under functional dependence and more general tail behaviours. See Section 3.4. The final result is jointly characterised by the tail-behaviour and the strength of temporal dependence of covariates and errors. When the model assumes sub-Gaussian covariates and errors without temporal independence, our results recover the optimal localisation error rate derived in Rinaldo et al. (2021).

– Last but not least, we provide a novel Bernstein inequality under the functional dependence. See Section 3.4. This is an exciting new addition to the literature for solving a wide range of high-dimensional problems under functional dependence. For instance, in this paper, we derive the estimation error bounds on the Lasso estimator under functional dependence, as an application of this new Bernstein inequality.

Notation and organisation

For any vector $v \in \mathbb{R}^p$ and any $q \in (0, \infty]$, let $|v|_q$ be its ℓ_q-norm. For any random variable $Z \in \mathbb{R}$ and any $q > 0$, let $\|Z\|_q = \{\mathbb{E}[|Z|^q]\}^{1/q}$, if $\mathbb{E}[|Z|^q] < \infty$. For any matrix A, let $|A|_\infty$ denote its entry-wise max norm. For any set S, let $|S|$ denote its cardinality. For two deterministic or random $\mathbb{R}$-valued sequences $a_n, b_n > 0$, write $a_n \gg b_n$ if $a_n/b_n \to \infty$ as n diverges. Write $a_n \lesssim b_n$ if $a_n/b_n \leq C$, for some absolute constant $C > 0$ and for all $n \geq 1$. For a deterministic or random $\mathbb{R}$-valued sequence a_n, write that a sequence of random variable $X_n = O_p(a_n)$ if $\lim_{M\to\infty} \limsup_n \mathbb{P}(|X_n| \geq M a_n) = 0$. Write $X_n = o_p(a_n)$ if $\limsup_n \mathbb{P}(|X_n| \geq M a_n) = 0$ for all $M > 0$. The convergences in distribution and probability are respectively denoted by $\xrightarrow{\mathcal{D}}$ and $\xrightarrow{P.}$.

The rest of the paper is organised as follows. Functional dependence is introduced in Section 1.2. Section 2 consists of our estimation procedure, which starts with a new variant of the dynamic programming algorithm tailored for the Lasso estimation sub-routine in Section 2.1. The limiting distributions of the change point estimators are collected in Section 3, accompanied by the new results on Bernstein's inequality, the Lasso estimator and change point localisation rates. In Section 4, the long-run variance estimator along with its consistency are presented. Numerical studies including a real data application to macroeconomics are conducted in Section 5. All the proofs and technical details are relegated to the Appendix.

1.2 Background: Functional dependence

Before formally introducing functional dependence, we remark that the functional dependence framework is attractive in two aspects: (1) It is general enough to include a large class of dependence assumptions, including ARMA models, linear processes, GARCH models, iterated random functions (Wu, 2005, 2011) and their high-dimensional variants. (2) It quantifies general nonlinear dependence in terms of moments and is thus easy to be computed explicitly for nonlinear processes. Many important mathematical tools which are widely used for independent sequences or martingale difference sequences are not available for general nonlinear processes. Functional dependence enables us to import some of the tools from the independence territory to the dependence one, while accommodating a large class of dependence assumptions. In contrast, different mixing type dependence assumptions assert the mixing coefficients to be based on σ-algebras, thus can be difficult to compute and transfer to useful mathematical tools.

Functional dependence of the covariate sequence. For each $t \in \mathbb{Z}$, let

$$X_t = (X_{t1}, \ldots, X_{tp})^\top = G(\mathcal{F}_t^X), \tag{4}$$

where $G(\cdot) = (g_1(\cdot), \ldots, g_p(\cdot))^\top$ is an $\mathbb{R}^p$-valued measurable function with input $\mathcal{F}_t^X = \{\mathcal{X}_s\}_{s \leq t}$, and $\{\mathcal{X}_t\}_{t\in\mathbb{Z}}$ is a collection of independent and identically distributed (i.i.d.) random elements. Note that (4) leads to the strict stationarity of $\{X_t\}_{t\in\mathbb{Z}}$.

For any $t \in \mathbb{Z}$ and any integer pair $s_2 \leq s_1$, let $\mathcal{X}_t^*$ be an independent copy of $\mathcal{X}_t$ and $X_{t,\{s_1,s_2\}} = G(\mathcal{F}_{t,\{s_1,s_2\}}^X)$ be a coupled random variable, where

$$\mathcal{F}_{t,\{s_1,s_2\}}^X = \begin{cases} \{\ldots, \mathcal{X}_{s_2-1}, \mathcal{X}_{s_2}^*, \ldots, \mathcal{X}_{s_1}^*, \mathcal{X}_{s_1+1}, \ldots, \mathcal{X}_t\}, & s_1 < t, \\ \{\ldots, \mathcal{X}_{s_2-1}, \mathcal{X}_{s_2}^*, \ldots, \mathcal{X}_t^*\}, & s_2 \leq t \leq s_1, \\ \{\ldots, \mathcal{X}_t\}, & t < s_2. \end{cases} \tag{5}$$

For any $s, t \in \mathbb{Z}$, let $\mathcal{F}_{t,\{s\}}^X = \mathcal{F}_{t,\{s,s\}}^X$, satisfying that when $t < s$, $\mathcal{F}_{t,\{s\}}^X = \mathcal{F}_t^X$ and $X_{t,\{s\}} = X_t$.

For any $v \in \mathbb{R}^p$, any $q > 0$ satisfying that $\|v^\top X_1\|_q < \infty$ and any $s \in \mathbb{Z}$, let

$$\delta_{s,q}^X(v) = \|v^\top X_1 - v^\top X_{1,\{1-s\}}\|_q = \|v^\top X_s - v^\top X_{s,\{0\}}\|_q, \tag{6}$$

due to the strict stationarity implied by (4). Define the uniform functional dependence measure and its cumulative version as

$$\delta_{s,q}^X = \sup_{|v|_2=1} \delta_{s,q}^X(v) \quad \text{and} \quad \Delta_{m,q}^X = \sum_{s=m}^{\infty} \delta_{s,q}^X, \quad m \in \mathbb{Z}, \tag{7}$$

respectively. In this paper, functional dependence assumption is imposed on the decay rate of the cumulative uniform function dependence measure $\Delta_{m,q}^X$, see, for example (20).

Functional dependence of the noise sequence. For each $t \in \mathbb{Z}$, let

$$\epsilon_t = g(\mathcal{F}_t^\epsilon), \tag{8}$$

where $g(\cdot)$ is an $\mathbb{R}$-valued measurable function with input $\mathcal{F}_t^\epsilon = \{\varepsilon_s\}_{s\leq t}$, and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a collection of i.i.d. random elements. Note that (8) leads to the strict stationarity of $\{\epsilon_t\}_{t\in\mathbb{Z}}$.

For any $t \in \mathbb{Z}$ and any integer pair $s_2 \leq s_1$, let ε_t^* be an independent copy of ε_t and $\epsilon_{t,\{s_1,s_2\}} = g(\mathcal{F}_{t,\{s_1,s_2\}}^\epsilon)$ be a coupled random variable, defined in the same way as that in (5). For any $q > 0$ satisfying that $\|\epsilon_1\|_q < \infty$ and any $s \in \mathbb{Z}$, let the functional dependence measure and its cumulative version be

$$\delta_{s,q}^\epsilon = \|\epsilon_1 - \epsilon_{1,\{1-s\}}\|_q = \|\epsilon_s - \epsilon_{s,\{0\}}\|_q \quad \text{and} \quad \Delta_{m,q}^\epsilon = \sum_{s=m}^{\infty} \delta_{s,q}^\epsilon, \quad m \in \mathbb{Z}, \tag{9}$$

respectively. Same as above, functional dependence assumption is imposed on the decay rate of the cumulative function dependence measure $\Delta_{m,q}^\epsilon$, see, for example (22).

We remark that although $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ are assumed to be strictly stationary, which is a direct consequence of the time-invariance of $G(\cdot)$ and $g(\cdot)$, our theoretical analysis in this paper extends to nonstationary cases, which is indeed needed for change point analysis.

Functional dependence of nonstationary sequences. Despite that $\{X_t\}_{t\in\mathbb{Z}}$ and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ are assumed to be strictly stationary, we still need to deal with nonstationary sequences, for instance $\{X_{tj}X_t^\top\beta_t^*\}_{t\in\mathbb{Z}}$, $j\in\{1,\ldots,p\}$ and $\{\epsilon_t X_t^\top\beta_t^*\}_{t\in\mathbb{Z}}$, since the coefficient sequence $\{\beta_t^*\}$ may possess change points. We therefore introduce the functional dependence measure for a (possibly) nonstationary process $\{Z_t\}_{t\in\mathbb{Z}}\subset\mathbb{R}$. Suppose that, for each $t\in\mathbb{Z}$,

$$Z_t = g_t(\mathcal{F}_t^\zeta), \tag{10}$$

where $g_t(\cdot)$ is a time-dependent $\mathbb{R}$-valued measurable functions with input $\mathcal{F}_t^\zeta=\{\zeta_s\}_{s\le t}$, and $\{\zeta_t\}_{t\in\mathbb{Z}}$ is a collection of i.i.d. random elements. For any $t\in\mathbb{Z}$ and any integer pair $s_2\le s_1$, let ζ_t^* be an independent copy of ζ_t and $Z_{t,\{s_1,s_2\}}=g_t(\mathcal{F}^\zeta_{t,\{s_1,s_2\}})$ be a coupled random variable, defined in the same way as that in (5). For any $q>0$ satisfying that $\sup_t\|Z_t\|_q<\infty$ and any $s\in\mathbb{Z}$, let the functional dependence measure and its cumulative version be

$$\delta^Z_{s,q}=\sup_{t\in\mathbb{Z}}\|Z_t-Z_{t,\{t-s\}}\|_q \quad\text{and}\quad \Delta^Z_{m,q}=\sum_{s=m}^{\infty}\delta^Z_{s,q},\quad m\in\mathbb{Z}, \tag{11}$$

respectively, and we again impose functional dependence assumptions on the decay rate of $\Delta^Z_{m,q}$.

2 Methodology

To provide a change point estimator with tractable limiting distributions, we exploit a two-step procedure, where we obtain a preliminary estimator in the first step and further refine it in the second step. This two-step strategy has been summoned regularly in recent literature to handle the multiple change point scenario. We study the two steps in Sections 2.1 and 2.2, respectively.

2.1 Preliminary estimators

To obtain preliminary estimators, we exploit an ℓ_0-penalised estimation procedure (e.g. Friedrich et al., 2008; Killick et al., 2012; Wang et al., 2020). Let

$$\widehat{\mathcal{P}}=\widehat{\mathcal{P}}(\zeta)\in\arg\min_{\mathcal{P}}\left\{\sum_{\mathcal{I}\in\mathcal{P}}\mathcal{G}(\mathcal{I})+\zeta|\mathcal{P}|\right\}, \tag{12}$$

where the minimisation is over all possible integer partitions of $\{1,\ldots,n\}$, $\mathcal{P}$ denotes an integer partition and $\zeta>0$ is a penalisation parameter. For any integer interval $\mathcal{I}\subset(0,n]$, the loss function $\mathcal{G}(\cdot)$ is defined as

$$\mathcal{G}(\mathcal{I})=\begin{cases}\sum_{t\in\mathcal{I}}\{-2y_tX_t^\top\widehat{\beta}_{\mathcal{I}}+\widehat{\beta}_{\mathcal{I}}^\top X_tX_t^\top\widehat{\beta}_{\mathcal{I}}\}, & |\mathcal{I}|\ge\zeta,\\ 0, & \text{otherwise},\end{cases} \tag{13}$$

with a to-be-specified $\zeta>0$. The estimator $\widehat{\beta}_{\mathcal{I}}$ in (13) is defined as

$$\widehat{\beta}_{\mathcal{I}}=\arg\min_{\beta\in\mathbb{R}^p}\left\{\sum_{t\in\mathcal{I}}(y_t-X_t^\top\beta)^2+\lambda|\mathcal{I}|^{1/2}|\beta|_1\right\}, \tag{14}$$

where $\lambda > 0$ is the Lasso tuning parameter to be specified. Note that there is a one-to-one relationship between a subset of integers and an integer partition, using the left endpoints of all the intervals. An estimator $\widehat{\mathcal{P}}$ therefore corresponds to a unique set of change point estimators denoted as $\widehat{\mathcal{B}}$ in Algorithm 1.

Remark 1 (Tuning parameter ζ in (12) and (13))**.** The same tuning parameter ζ is used in both (12) and (13) for different purposes. In (12), ζ penalises over-partitioning and is a high-probability upper bound controlling the loss function $\mathcal{G}(\mathcal{I})$, for desirable $\mathcal{I}$. In (13), ζ restricts the attention to the intervals of lengths at least ζ. To see why these two parameters coincides, we discuss from two aspects. Denote κ as the minimum jump size (see Assumption 1**b** for more details). Firstly, (12) implies that the localisation errors are of order at least ζ/κ^2 and (13) implies that the localisation errors are of order at least ζ. Since $\kappa = O(1)$, $\zeta/\kappa^2 \gtrsim \zeta$, which implies that the tuning parameter in (12) should be no smaller than that in (13). Secondly, the success of ζ serving as a high-probability upper bound in (12) relies on a restricted eigenvalue condition, holding only when the sample size is large enough, i.e. the sample size is at least larger than the fluctuation. This means that the tuning parameter in (12) should be no larger than that in (13). Based on these two, we see that these two tuning parameters can take the same value. More discussions on ζ can be found in Section 3.2.

Despite the different form, we remark that the optimisation problem specified in (12) and (13) is equivalent to that in Rinaldo et al. (2021). An equivalent formulation in (13) enables us to apply a more computationally-efficient optimisation in Algorithm 1. To solve the minimisation problem, we propose a variant of the dynamic programming (DP) algorithm (e.g. Algorithm 1 in Friedrich et al., 2008), based on the observation that the loss function $\mathcal{G}(\mathcal{I})$ is the sum of partial sums of $X_t X_t^\top$ and $y_t X_t^\top$, which can be updated iteratively. This formulation suggests that we can save computational costs by storing the partial sums in the memory. We coin this variant as dynamic programming with dynamic update (DPDU), detailed in Algorithm 1.

Denote $\mathbf{Lasso}(p)$ as the computational cost of the lasso procedure given quantities $\sum_{t=1}^n y_t X_t \in \mathbb{R}^p$ and $\sum_{t=1}^n X_t X_t^\top \in \mathbb{R}^{p\times p}$. The exact value of $\mathbf{Lasso}(p)$ depends on the choice of algorithms. For example, for coordinate descent algorithms, $\mathbf{Lasso}(p) = O(p^2)$, see Lemma 11. For the LARS algorithm, $\mathbf{Lasso}(p) = O(p^3)$, see Efron et al. (2004). The memory and computational costs of the DPDU are collected below.

Lemma 1. *For* DPDU$(\{y_t, X_t\}_{t=1}^n, \lambda, \zeta)$ *with* $\{X_t\}_{t=1}^n \subset \mathbb{R}^p$ *and any* $\lambda, \zeta > 0$*, the memory cost is* $O(np^2)$ *and the computational cost is* $O(n^2p^2 + n^2\mathbf{Lasso}(p))$*.*

Note that the computational cost of a standard DP (e.g. Algorithm 1 in Friedrich et al., 2008) is $O(n^3p^2 + n^2\mathbf{Lasso}(p))$. DPDU improves the computational efficiency by decreasing the term n^3 to n^2, with the help of storing the partial sums in the memory. To achieve this improvement on the computational efficiency, we pay a higher memory cost, since the memory cost of DP for such problem is $O(p^2)$.

2.2 Final estimators

Denote the preliminary estimators obtained from the optimisation problem specified in (12), (13) and (14), or equivalently from Algorithm 1 as

$$\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}} = \widehat{\mathcal{B}}. \tag{16}$$

Algorithm 1 Dynamic Programming with Dynamic Update. DPDU $(\{(y_t, X_t)\}_{t=1}^n, \lambda, \zeta)$

INPUT: Data $\{(y_t, X_t)\}_{t=1}^n$, tuning parameters $\lambda, \zeta > 0$.

$\widehat{\mathcal{B}} \leftarrow \varnothing$, $\mathfrak{p} \leftarrow (-1, \dots, -1)^\top \in \mathbb{R}^n$, $B \leftarrow (-\zeta, \infty, \dots, \infty)^\top \in \overline{\mathbb{R}}^{n+1}$

for $t \in \{1, \dots, n\}$ **do**

$\mathcal{M}(t,t) \leftarrow (0) \in \mathbb{R}^{p \times p}$, $\mathcal{V}(t,t) \leftarrow (0) \in \mathbb{R}^p$

end for

for $r \in \{2, \dots, n+1\}$ **do**

for $l \in \{1, \dots, r-1\}$ **do**

$$\mathcal{M}(r,l) \leftarrow \mathcal{M}(r-1,l) + X_{r-1}X_{r-1}^\top \quad \text{and} \quad \mathcal{V}(r,l) \leftarrow \mathcal{V}(r-1,l) + y_{r-1}X_{r-1}^\top \tag{15}$$

$\mathcal{I} \leftarrow [l, r) \cap \mathbb{Z}$.

if $|\mathcal{I}| \geq \zeta$ **then**

$\widehat{\beta}_\mathcal{I} \leftarrow \arg\min_{\beta \in \mathbb{R}^p} |\mathcal{I}|^{-1}\{-2\mathcal{V}(r,l)\beta + \beta^\top \mathcal{M}(r,l)\beta\} + \lambda/\sqrt{|\mathcal{I}|}|\beta|_1$

$\mathcal{G}(\mathcal{I}) \leftarrow -2\mathcal{V}(r,l)\widehat{\beta}_\mathcal{I} + \widehat{\beta}_\mathcal{I}^\top \mathcal{M}(r,l)\widehat{\beta}_\mathcal{I}$

else

$\mathcal{G}(\mathcal{I}) \leftarrow 0$

end if

$b \leftarrow B_l + \zeta + \mathcal{G}(\mathcal{I})$

if $b < B_r$ **then**

$B_r \leftarrow b$, $\mathfrak{p}_r \leftarrow l$

end if

end for

Replace $\{\mathcal{M}(r-1,1), \dots, \mathcal{M}(r-1,r-2)\}$ with $\{\mathcal{M}(r,1), \dots, \mathcal{M}(r,r-2)\}$ in the memory.

Replace $\{\mathcal{V}(r-1,1), \dots, \mathcal{V}(r-1,r-2)\}$ with $\{\mathcal{V}(r,1), \dots, \mathcal{V}(r,r-2)\}$ in the memory.

end for

$k \leftarrow n$

while $k > 1$ **do**

$h \leftarrow \mathfrak{p}_k$, $\widehat{\mathcal{B}} \leftarrow \widehat{\mathcal{B}} \cup h$, $k \leftarrow h$

end while

OUTPUT: $\widehat{\mathcal{B}}$, $\widehat{K} = |\widehat{\mathcal{B}}|$, $\{\widehat{\beta}_k\}_{k=0}^{\widehat{K}}$

Denote the corresponding regression coefficient estimators (14) as $\{\widehat{\beta}_k\}_{k=0}^{\widehat{K}}$. Our final estimators are obtained via a local refinement step, which is considered in Rinaldo et al. (2021), Yu et al. (2022) and others. The final estimators are defined as

$$\widetilde{\eta}_k = \underset{s_k < \eta < e_k}{\arg\min}\, Q_k(\eta) = \underset{s_k < \eta < e_k}{\arg\min} \left\{ \sum_{t=s_k}^{\eta-1} (y_t - X_t^\top \widehat{\beta}_{k-1})^2 + \sum_{t=\eta}^{e_k-1} (y_t - X_t^\top \widehat{\beta}_k)^2 \right\}, \tag{17}$$

where $k \in \{1, \dots \widehat{K}\}$ and (s_k, e_k) is defined as

$$s_k = 9\widehat{\eta}_{k-1}/10 + \widehat{\eta}_k/10 \quad \text{and} \quad e_k = \widehat{\eta}_k/10 + 9\widehat{\eta}_{k+1}/10. \tag{18}$$

The rationale of the final estimators $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{K}}$ is that, provided the preliminary estimators $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ are good enough, each interval (s_k, e_k) contains one and only one true change point η_k.

The choice of the constants 9/10 and 1/10 in (18) is arbitrary. To be specific, Lemma 5 shows that, under certain conditions, with high probability, the output of DPDU satisfies that

$$\widehat{K} = K \quad \text{and} \quad \max_{k=1,\ldots,K} \frac{|\widehat{\eta}_k - \eta_k|}{\min\{\eta_k - \eta_{k-1},\, \eta_{k+1} - \eta_k\}} \to 0, \quad n \to \infty,$$

which guarantees the success of the $\{(s_k, e_k)\}_{k=1}^K$ construction above.

3 Theory

We kick off by a list of assumptions in Section 3.1, followed by the main results on the limiting distributions in Section 3.2. The connection of our main results with relevant literature is discussed in Section 3.3. We conclude this section by discussing a few important byproducts of our analysis in Section 3.4.

3.1 Assumptions

Assumption 1 formalises some of the model assumptions we have mentioned so far.

Assumption 1 (Model)**.** *Consider the high-dimensional linear regression model with change points specified in* (1), (2) *and* (3)*, with the covariate sequence* $\{X_t\}_{t=1}^n \subset \mathbb{R}^p$ *and noise sequence* $\{\epsilon_t\}_{t=1}^n \subset \mathbb{R}$ *independent of each other. For any* $k \in \{1, \ldots, K\}$*, where* $K \geq 1$ *is an absolute constant, let the* k*th jump vector and its normalised version be* $\Psi_k = \beta^*_{\eta_k} - \beta^*_{\eta_k - 1}$ *and* $v_k = \Psi_k/|\Psi_k|_2 = \Psi_k/\kappa_k$*, respectively.*
a. *(Sparsity) For any* $t \in \{1, \ldots, n\}$*, assume that there exists support* $S_t \subset \{1, \ldots, p\}$*, with* $|S_t| \leq \mathfrak{s}$*, such that* $\beta^*_{t,j} = 0$*, if* $j \notin S_t$*.*
b. *(Change points) Let the minimal jump size be*

$$\kappa = \min_{1 \leq k \leq K} \kappa_k = \min_{1 \leq k \leq K} |\Psi_k|_2. \tag{19}$$

Assume that there exists an absolute constant $C_\kappa > 0$ *such that for any* $k \in \{1, \ldots, K\}$*,* $\kappa_k \leq C_\kappa$*. Let the minimal spacing be* $\Delta = \min_{1 \leq k \leq K+1}(\eta_k - \eta_{k-1})$*.*

In Assumption 1, we impose the ℓ_0-sparsity on the regression coefficients, following the suit of standard high-dimensional linear regression literature. We also upper bound the jump size κ_k, i.e. the ℓ_2-norm of the difference between the true regression coefficients before and after the kth change point for $k \in \{1, \ldots, K\}$. Since the larger the jump size, the more prominent the change points, Assumption 1**b** focuses our analysis in the more challenging territory. In Assumption 1, although the number of change points K is assumed to be of order $O(1)$, we do allow the minimum spacing Δ to be a function of n and to vanish as n diverges, as specified later in Assumption 4. We assume that K to be independent of n for simplicity of presentation, as our main result Theorem 3 focuses on the limiting distribution of every change point estimator.

We explicitly state the independence between the covariate and noise sequences in Assumption 1. In Assumptions 2 and 3 below, we will further exploit the flexibility of our model by specifying the assumptions on the covariate and noise sequences, respectively.

Assumption 2 (Covariate sequence)**.** *Assume that the covariate sequence $\{X_t\}_{t=1}^n \subset \mathbb{R}^p$ is a consecutive subsequence of an infinite sequence $\{X_t\}_{t\in\mathbb{Z}} \subset \mathbb{R}^p$, which has marginal mean zero and is of the form (4), with covariance matrix $\Sigma = \mathrm{Cov}(X_1)$.*
a. *(Functional dependence) For the cumulative uniform functional dependence measure $\Delta^X_{\cdot,\cdot}$ defined in (6) and (7), for some absolute constants $\gamma_1, c > 0$, assume that there exists an absolute constant $D_X > 0$ such that*

$$\sup_{m\geq 0} \exp(cm^{\gamma_1})\Delta^X_{m,4} \leq D_X. \tag{20}$$

b. *(Tail behaviour) Assume that there exist absolute constants $\gamma_2 \in (0, 2]$ and $C_X > 0$, such that for any $\tau > 0$,*

$$\sup_{|v|_2=1} \mathbb{P}(|v^\top X_1| > \tau) \leq 2\exp\{-(\tau/C_X)^{\gamma_2}\}. \tag{21}$$

c. *(Marginal covariance matrix) Assume that the smallest and largest eigenvalues of Σ satisfy $0 < c_{\min} \leq \Lambda_{\min}(\Sigma) \leq \Lambda_{\max}(\Sigma) < \infty$, where $c_{\min} > 0$ is an absolute constant. For any $k \in \{1, \ldots, K\}$, assume that there exists an absolute constant $\varpi_k > 0$ such that $v_k^\top \Sigma v_k \to \varpi_k$, as $n \to \infty$.*

As mentioned in Section 1.2, the functional dependence is measured through the quantity $\Delta^X_{\cdot,\cdot}$ and their decay rate. Assumption 2**a** specifies that $\Delta^X_{m,4}$ decays exponentially in m at the rate of at least $\exp(-cm^{\gamma_1})$. Recall the definition of $\Delta^X_{\cdot,\cdot}$ in (6) and (7), the term 4 means that the functional dependence is measured in the 4th moment. This moment condition is required to unlock a functional central limit theorem under functional dependence. The term γ_1 characterises the exponential decay rate. In the extreme case that $\gamma_1 = \infty$, Assumption 2**a** essentially reduces to the i.i.d. case.

Assumption 2**b** regulates the tail behaviour of the covariate random vectors' marginal distribution. The specific rate is characterised by the parameter $\gamma_2 \in (0, 2]$, indicating that the marginal distribution falls in the sub-Weibull family, with sub-Exponential ($\gamma_2 = 1$) and sub-Gaussian ($\gamma_2 = 2$) as its special cases.

Temporal dependence and moment assumptions are two indispensable assumptions when studying the theoretical properties of high-dimensional time series. In the existing literature (e.g. Wu and Wu, 2016; Zhang and Wu, 2017; Kuchibhotla et al., 2021), these two assumptions are often entangled via a single dependence-adjusted moment assumption. Instead, Assumption 2 provides separate assumptions on the temporal dependence and tail behaviour, which is more intuitive. We specify the temporal dependence and tail behaviour of the noise sequence $\{\epsilon_t\}_{t=1}^n$ in Assumption 3.

Assumption 3 (Noise sequence)**.** *Assume that the noise sequence $\{\epsilon_t\}_{t=1}^n \subset \mathbb{R}$ is a consecutive subsequence of an infinite sequence $\{\epsilon_t\}_{t\in\mathbb{Z}} \subset \mathbb{R}$, which has marginal mean zero and is of the form (8), with variance $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_1) > 0$ being an absolute constant.*
a. *(Functional dependence) For the cumulative functional dependence measure $\Delta^\epsilon_{\cdot,\cdot}$, defined in (9), for some absolute constants $\gamma_1, c > 0$, assume there exists an absolute constant $D_\epsilon > 0$ such that*

$$\sup_{m\geq 0} \exp(cm^{\gamma_1})\Delta^\epsilon_{m,4} \leq D_\epsilon. \tag{22}$$

b. *(Tail behaviour) Assume that there exist absolute constants $\gamma_2 \in (0, 2]$ and $C_\epsilon > 0$, such that for any $\tau > 0$, $\mathbb{P}(|\epsilon_0| > \tau) \leq 2\exp\{-(\tau/C_\epsilon)^{\gamma_2}\}$.*

In spite of the same notation γ_1 and γ_2 used in Assumptions 2 and 3, there is no reason to force these two sequences to have the same functional dependence measure decay rates or the same tail behaviour. We adopt the same notation for simplicity and they can be seen as worse cases between two. To be specific, let $\gamma_1(X)$ and $\gamma_1(\epsilon)$ be the functional dependence measure decay rate indicator for the covariate and noise sequences, respectively. In Assumptions 2 and 3, we in fact let $\gamma_1 = \min\{\gamma_1(X), \gamma_1(\epsilon)\}$, and similarly $\gamma_2 = \min\{\gamma_2(X), \gamma_2(\epsilon)\}$.

With the temporal dependence and tail behaviour specified for both the covariate and noise sequences, we introduce the parameter

$$\gamma = \left(\gamma_1^{-1} + 2\gamma_2^{-1}\right)^{-1}. \tag{23}$$

The coefficient 2 in the term γ_2^{-1} accounts for the tail behaviour of the cross-term sequence $\{\epsilon_t X_t\}_{t=1}^n$. In Assumptions 2**b** and 3**b**, we allow $\gamma_1 \in (0, \infty)$ provided that $\gamma < 1$. We restrict our analysis to $\gamma_2 \leq 2$ in this paper, as $\gamma_2 > 2$ corresponds to tails lighter than sub-Gaussian, which is less interesting for practical purposes.

We state the required signal-to-noise ratio condition below.

Assumption 4 (Signal-to-noise ratio)**.**
a. *Assume that there exists a sufficiently large absolute constant $C_{\mathrm{snr}} > 0$ such that*

$$\Delta\kappa^2 \geq C_{\mathrm{snr}}\left\{\mathfrak{s}\log(pn)\right\}^{2/\gamma-1}\alpha_n, \tag{24}$$

where $\alpha_n > 0$ is any diverging sequence.
b. *Assume that there exists a sufficiently large absolute constant $C_{\mathrm{snr}} > 0$ such that*

$$\Delta\kappa^2 \geq C_{\mathrm{snr}}\mathfrak{s}^{2/\gamma}\{\log(pn)\}^{4/\gamma-1}\alpha_n,$$

where $\alpha_n > 0$ is any diverging sequence.

Assumption 4**a** and **b** impose lower bounds on the quantity $\Delta\kappa^2$ to ensure theoretical guarantees of the change point estimation (Lemma 5) and inference (Theorem 3) procedures, respectively. Note that Assumption 4**b** is strictly stronger than Assumption 4**a**. This means that deriving the limiting distributions is strictly harder than providing a high-probability estimation error upper bound. Such signal-to-noise ratio conditions are widely seen in different change point analysis problems. In Rinaldo et al. (2021), it is shown that for change point estimation of the model specified in Assumption 1, when the data considered are temporally independent and sub-Gaussian ($\gamma = 1$), the minimax optimal condition is that $\Delta\kappa^2 \asymp \mathfrak{s}$, saving a logarithmic factor, matching our Assumption 4**a**. We extend Rinaldo et al. (2021) by allowing for temporal dependence, indexed by γ_1, and potentially heavier tails, indexed by γ_2. Such effects are captured by γ, which not only inflates the logarithmic term but also affects the term regarding the sparsity parameter $\mathfrak{s}$.

Similar effects of γ on the sparsity level are seen in the existing literature, on the Lasso estimation in high-dimensional linear regression problems with temporal dependence, but without the presence of change points (e.g. Wu and Wu, 2016; Wong et al., 2020; Kuchibhotla et al., 2021; Han and Tsay, 2020). In the presence of change points, Chen et al. (2021) characterise the jump in terms of vector ℓ_∞-norm, which turns a high-dimensional problem to a univariate one. The parameter γ's counterpart therein hence does not affect the sparsity parameter therein.

3.2 Limiting distributions

For the true change points $\{\eta_k\}_{k=1}^K$, in Section 2 we proposed a two-step estimator $\{\widetilde{\eta}_k\}_{k=1,\ldots,\widehat{K}}$. Under the assumptions listed in Section 3.1, we present the limiting distributions of $\{\widetilde{\eta}_k\}_{k=1,\ldots,\widehat{K}}$ in two regimes, regarding the jump sizes $\{\kappa_k\}_{k=1}^K$, defined in (19): (i) the non-vanishing regime where $\kappa_k \to \varrho_k$, as $n \to \infty$, with $\varrho_k > 0$ being an absolute constant; and (ii) the vanishing regime where $\kappa_k \to 0$ as $n \to \infty$.

Since the limiting distribution in the vanishing regime involves an unknown long-run variance, we first define the long-run variance and prove its existence.

Lemma 2. *Suppose that Assumptions 1, 2 and 3 hold. For $k \in \{1, \ldots, K\}$, the long-run variance*

$$\sigma_\infty^2(k) = 4 \lim_{n\to\infty} \mathrm{Var}\left(n^{-1/2}\sum_{t=1}^n \epsilon_t v_k^\top X_t\right) \tag{25}$$

exists and satisfies that $\sigma_\infty^2(k) < \infty$.

We are now ready to present the main results.

Theorem 3. *Given data $\{(y_t, X_t)\}_{t=1}^n$, suppose that Assumptions 1, 2, 3 and 4**b** hold. Let $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{K}}$ be the change point estimators defined in (17), with*

- *the intervals $\{(s_k, e_k)\}_{k=1}^{\widehat{K}}$ defined in (18);*
- *the preliminary estimators $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ from $\mathrm{DPDU}(\{(y_t, X_t)\}_{t=1}^n, \lambda, \zeta)$ detailed in Algorithm 1;*
- *the DPDU tuning parameters $\lambda = C_\lambda\sqrt{\log(pn)}$ and $\zeta = C_\zeta\{\mathfrak{s}\log(pn)\}^{2/\gamma-1}$, with $C_\lambda, C_\zeta > 0$ being absolute constants; and*
- *the Lasso estimators $\{\widehat{\beta}_k\}_{k=1}^{\widehat{K}}$ defined in (14) with the intervals constructed based on $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$.*

a. *(Non-vanishing regime) For $k \in \{1, \ldots, K\}$, if $\kappa_k \to \varrho_k$, as $n \to \infty$, with $\varrho_k > 0$ being an absolute constant, then the following results hold.*

a.1. *The estimation error satisfies that $|\widetilde{\eta}_k - \eta_k| = O_p(1)$, as $n \to \infty$.*

a.2. *When $n \to \infty$,*

$$\widetilde{\eta}_k - \eta_k \xrightarrow{\mathcal{D}} \mathop{\arg\min}_{r\in\mathbb{Z}} P_k(r),$$

where, for $r \in \mathbb{Z}$, $P_k(r)$ is a two-sided random walk defined as

$$P_k(r) = \begin{cases} \sum_{t=r}^{-1}\{-2\varrho_k\epsilon_t v_k^\top X_t + \varrho_k^2(v_k^\top X_t)^2\}, & r < 0, \\ 0, & r = 0, \\ \sum_{t=1}^{r}\{2\varrho_k\epsilon_t v_k^\top X_t + \varrho_k^2(v_k^\top X_t)^2\}, & r > 0. \end{cases}$$

b. *(Vanishing regime) For $k \in \{1, \ldots, K\}$, if $\kappa_k \to 0$, as $n \to \infty$, then the following results hold.*

b.1. *The estimation error satisfies that $|\widetilde{\eta}_k - \eta_k| = O_p(\kappa_k^{-2})$, as $n \to \infty$.*

b.2. *When $n \to \infty$,*

$$\kappa_k^2(\widetilde{\eta}_k - \eta_k) \xrightarrow{\mathcal{D}} \arg\min_{r \in \mathbb{R}} \{\varpi_k |r| + \sigma_\infty(k)\mathbb{W}(r)\},$$

where $\mathbb{W}(r)$ is a two-sided standard Brownian motion defined as

$$\mathbb{W}(r) = \begin{cases} \mathbb{B}_1(-r), & r < 0, \\ 0, & r = 0, \\ \mathbb{B}_2(r), & r > 0, \end{cases}$$

with $\mathbb{B}_1(r)$ and $\mathbb{B}_2(r)$ being two independent standard Brownian motions.

Theorem 3 provides not only estimation error controls of the change point estimators $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{K}}$ under two scenarios of jump sizes, but more importantly their limiting distributions. We remark that the localisation rate in Theorem 3 is an exact match of the minimax lower bounds derived under the temporal independence assumption (e.g. Rinaldo et al., 2021) for the high-dimensional regression model with change points, while all previous works are off by a logarithmic factor. As far as we are aware, this exact match is the first time seen in the change point literature in high-dimensional regression setting.

In the following, we further discuss Theorem 3 in more details from the aspects of tuning parameters, the sketch of proof and the intuition of the two different forms of limiting distributions.

Choices of the tuning parameters. Recall that the estimators $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{K}}$ are obtained based on the preliminary estimators $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$, which are outputs of the DPDU algorithm detailed in Algorithm 1. There are two tuning parameters involved in the DPDU step: the Lasso penalisation parameter λ and the interval length cutoff ζ, see (12), (13) and (14).

- Lasso parameter λ. It is interesting to see that despite the temporal dependence and more general tail behaviour, the Lasso parameter λ is chosen to be of order $\log^{1/2}(np)$, which is the same as that in the independence and sub-Gaussian tail case (e.g. Bühlmann and van de Geer, 2011). This is because, the Lasso procedures are only conducted on the intervals of length at least ζ. To elaborate, we investigate a large-probability upper bound on the term $|\mathcal{I}|^{-1}|\sum_{t\in\mathcal{I}} X_t\epsilon_t|_\infty$. Due to our new Bernstein's inequality in Theorem 4, it is shown in Lemma 30 that, with large probability,

$$|\mathcal{I}|^{-1}|\sum_{t\in\mathcal{I}} X_t\epsilon_t|_\infty \lesssim \sqrt{\frac{\log(np)}{|\mathcal{I}|}} + \frac{\log^{1/\gamma}(np)}{|\mathcal{I}|}.$$

  Provided that $|\mathcal{I}| \geq \zeta \gtrsim \{\log(pn)\}^{2/\gamma-1}$, the term $\sqrt{\log(np)/|\mathcal{I}|}$ dominates $\log^{1/\gamma}(np)/|\mathcal{I}|$. This explains the choice of λ.

- Interval length cutoff ζ. In addition to facilitate the choice of λ, the choice of ζ is also to enable the application of a restricted eigenvalue condition we provide for data with functional dependence, see Theorem 33. We would also like to point out, since this paper focuses on a more challenging case where $\max_{k=1}^K \kappa_k \lesssim 1$, Assumption 4**a** guarantees that

$$\Delta \gtrsim \kappa^{-2}\{\mathfrak{s}\log(np)\}^{2/\gamma-1}\alpha_n \gtrsim \zeta. \tag{26}$$

Our Lasso procedure only considers intervals of length at least ζ, implying that the change point localisation error is at least of order ζ. This price is acceptable by noticing that $\Delta \gtrsim \zeta$, as demonstrated in (26).

Sketch of the proof. The proof can be roughly modularised into three steps.

- We first show that with large probability, the preliminary estimators are good enough such that for each $k \in \{1, \ldots, K\}$, the following holds. (i) The interval (s_k, e_k) defined in (18) contains one and only one true change point η_k, which is sufficiently far away from both s_k and e_k; and (ii) the Lasso estimators $\widehat{\beta}_k$ is a good enough estimator of the true coefficient β_k. These two statements have been shown previously in Rinaldo et al. (2021) for the temporal independence and sub-Gaussian case. In Lemmas 5 and 6, we provide the functional dependence and general tail behaviour counterparts for this set of results, based on our new Bernstein inequality in Theorem 4.

- Based on the preliminary estimators, we then show that $\{\kappa_k^2 |\widetilde{\eta}_k - \eta_k|\}_{k=1}^{\widehat{K}}$ are uniformly tight. This is a stepping stone towards our final limiting distributions, but this result is interesting *per se*. The preliminary estimators satisfy that, with large probability,

$$\max_{k=1}^{K} \kappa_k^2 |\widehat{\eta}_k - \eta_k| \lesssim \{\mathfrak{s} \log(np)\}^{2/\gamma - 1},$$

  the upper bound of which is diverging. The final estimators, however, refine the preliminary results by constraining the localisation error to $O_p(\kappa_k^{-2})$.

- The limiting distributions are thus derived, by exploiting the argmax continuous mapping theorem (e.g. Theorem 3.2.2 in van der Vaart and Wellner, 1996) and the functional central limit theorem under temporal dependence (e.g. Theorem 3 in Wu, 2011).

Limiting distributions. The limiting distributions differ based on the limits of the jump sizes $\{\kappa_k\}_{k=1}^K$. To explain this, we provide more insights in the proofs. For $k \in \{1, \ldots, K\}$, recall that $\widetilde{\eta}_k$ is the minimiser to

$$Q_k(\eta) = \sum_{t=s_k}^{\eta-1} (y_t - X_t^\top \widehat{\beta}_{k-1})^2 + \sum_{t=\eta}^{e_k-1} (y_t - X_t^\top \widehat{\beta}_k)^2.$$

This is equivalent to finding the minimiser of

$$\widetilde{Q}_k(r) = Q_k(\eta_k + r) - Q_k(\eta_k) = \begin{cases} \sum_{t=\eta_k}^{\eta_k+r-1} \left\{(y_t - X_t^\top \widehat{\beta}_{k-1})^2 - (y_t - X_t^\top \widehat{\beta}_k)^2\right\}, & r \in \mathbb{Z}_+, \\ 0, & r = 0, \\ \sum_{t=\eta_k+r+1}^{\eta_k} \left\{(y_t - X_t^\top \widehat{\beta}_k)^2 - (y_t - X_t^\top \widehat{\beta}_{k-1})^2\right\}, & r \in \mathbb{Z}_-. \end{cases}$$

Considering only the $r > 0$ case, provided that $\widehat{\beta}_{k-1}$ and $\widehat{\beta}_k$ are good enough, $\widetilde{Q}_k(r)$ can be thus written as the sum of

$$Q_k^*(\eta_k + r) - Q_k^*(\eta_k) = \sum_{t=\eta_k}^{\eta_k+r-1} \left\{(y_t - X_t^\top \beta_{\eta_{k-1}}^*)^2 - (y_t - X_t^\top \beta_{\eta_k}^*)^2\right\}$$

$$= \kappa_k \sum_{t=\eta_k}^{\eta_k+r-1} (2\epsilon_t v_k^\top X_t) + \kappa_k^2 \sum_{t=\eta_k}^{\eta_k+r-1} (v_k^\top X_t)^2 \quad (27)$$

and the remainder term, which is of order $o_p(r^{1/2}\kappa_k \vee r\kappa_k^2 \vee 1)$.

- In the non-vanishing case, all the choices of r considered are finite due to the uniform tightness of $|\widehat{\eta}_k - \eta_k|$ and κ_k is also finite due to Assumption 1. Term (27) thus can be approximated by a two-sided random walk distribution.
- In the vanishing case, the remainder term vanishes. Regarding the terms in (27), all r, such that $r\kappa_k^2$ are finite, are considered. Since κ_k vanishes, the functional central limit theorem leads to the two-sided Brownian motion distribution in the limit.

3.3 Connections with relevant literature

Theorem 3 presents limiting distributions of change point estimators, under both the non-vanishing and vanishing jump size regimes, in a high-dimensional linear regression model, with nonlinear temporal dependence in both covariate and noise sequences. Such results are novel at a few fronts, especially that this is a new addition to high-dimensional linear regression settings. Nevertheless, limiting distributions of change point estimators have been studied previously in other models. At a high level, all are built on different versions of argmax continuous mapping theorem, but specific work faces specific challenges brought in by specific models and temporal dependence assumptions.

The pioneer work by Yao and Au (1989) studies a mean change problem in an independent univariate sequence. They present a two-sided random walk-type limiting distribution in the non-vanishing regime, as a univariate mean counterpart of Theorem 3**a**. Despite being univariate and independent, when the distribution of noise is unknown, there is no practical guidance on utilising the limiting distribution.

Fully-observed functional mean change point problems are studied in Aue et al. (2009) and Aue et al. (2018). Both works study one change point scenario with the minimal spacing $\Delta = cn$, where $c \in (0, 1)$ being a constant. In both these two papers, limiting distributions of change point estimators are derived in non-vanishing and vanishing regimes. Although this is a different model, the two limiting distributions are also in the form of two-sided random walks and Brownian motions, respectively. Aue et al. (2009) consider a temporally-independent case and the noise sequence in Aue et al. (2018) is assumed to be a linear process.

High-dimensional mean change point problems have been recently studied in a stream of papers, including Bai (2010), Wang and Shao (2020), Chen et al. (2021), Kaul et al. (2021) and Kaul and Michailidis (2021). This is again a fundamentally different model, where at every time point, a high-dimensional vector is observed and the means of the vectors change at certain points. For high-dimensional mean change point problems, in terms of the jump size limits regimes, Bai (2010), Wang and Shao (2020) and Chen et al. (2021) consider the vanishing jump size regime, providing two-sided Brownian motion type limiting distributions; while Kaul et al. (2021) and Kaul and Michailidis (2021) consider both regimes, providing two-sided random walks and Brownian motions type limiting distributions. In terms of the number of change points, Bai (2010), Wang and Shao (2020) and Kaul et al. (2021) study one change point scenario with the minimal spacing $\Delta = cn$, where $c \in (0, 1)$ being a constant; while Chen et al. (2021) and Kaul and Michailidis (2021) allow for multiple change points. In terms of the temporal dependence assumption, Wang and Shao

(2020), Kaul et al. (2021) and Kaul and Michailidis (2021) assume temporal independence; while Bai (2010) and Chen et al. (2021) assume the noise sequence is a linear process.

In view of these aforementioned works, we see that our framework is by far the most general setting. Since we allow both the high-dimensional covariate and univariate noise sequences to be temporally dependent in a nonlinear way, new techniques handling the large deviation behaviours of stationary and non-stationary sequences are developed. This is unseen in the existing work. Moreover, compared to the mean change problems, we need to conduct high-dimensional regression based on non-stationary and temporally-dependent sequences. This relies on a set of new techniques, including a new restricted eigenvalue-type result, which is, among other key technical ingredients, not dealt with in the existing change point inference literature.

3.4 Intermediate results

We collect some important intermediate results of our analysis, which are interesting byproducts *per se*. The one at the core of the analysis is a new Bernstein inequality, controlling the large deviation bound under temporal dependence and more general tail behaviour.

Theorem 4. *Let $\{Z_t\}_{t\in\mathbb{Z}}$ be an $\mathbb{R}$-valued, mean-zero, possibly nonstationary process of the form (10). Let the cumulative functional dependence measure $\Delta^Z_{m,2}$ be defined in (11). Assume that there exist absolute constants $\gamma_1(Z), \gamma_2(Z), c, C_{\mathrm{FDM}} > 0$ such that*

$$\sup_{m\geq 0} \exp(cm^{\gamma_1(Z)})\Delta^Z_{m,2} \leq C_{\mathrm{FDM}} \tag{28}$$

and $\sup_{t\in\mathbb{Z}} \mathbb{P}(|Z_t| > x) \leq \exp(1 - x^{\gamma_2(Z)})$, for any $x > 0$. Let $\gamma(Z) = \{1/\gamma_1(Z) + 1/\gamma_2(Z)\}^{-1} < 1$. There exist absolute constants $c_1, c_2 > 0$ such that for any $x \geq 1$ and integer $m \geq 3$,

$$\mathbb{P}\left\{m^{-1/2}\left|\sum_{t=1}^m Z_t\right| \geq x\right\} \leq m\exp\{-c_1 x^{\gamma(Z)} m^{\gamma(Z)/2}\} + 2\exp\{-c_2 x^2\}. \tag{29}$$

Theorem 4 is repeatedly used in the proof of Theorem 3. Specific forms of $\{Z_t\}_{t\in\mathbb{Z}}$ used include strictly stationary series $\{\epsilon_t X_{tj}\}$ and nonstationary series $\{\epsilon_t X_t^\top \beta_t^*\}_{t\in\mathbb{Z}}$. Note that the moment condition in the cumulative functional dependence measure $\Delta^Z_{m,2}$ used in Theorem 4 is 2. This further explains why the fourth moment condition is assumed in Assumptions 2 and 3. In (29), we see that the decay rate for a sum of m observations is governed by the interplay of two terms. When the sample size is large enough, the decay rate is governed by the same rate as that in the temporal independence and sub-Gaussian case (e.g. Vershynin, 2018). This coincides with the conventional wisdom in understanding standard Bernstein's inequality.

With the help of Theorem 4, we are able to show the properties of preliminary estimators under temporal dependence and the general tail behaviour. The localisation error, the error bounds on Lasso estimators and the error bounds on the jump size estimators are presented in Lemmas 5, 6 and 7, respectively.

Lemma 5 (Preliminary change point estimators)**.** *Given data $\{(y_t, X_t)\}_{t=1}^n$, suppose that Assumptions 1, 2, 3 and 4**a** hold. Let $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ be the change point estimators output by $\mathrm{DPDU}(\{y_t, X_t\}_{t=1}^n,$ λ, $\zeta)$, with $\lambda = C_\lambda\sqrt{\log(np)}$ and $\zeta = C_\zeta\{\mathfrak{s}\log(pn)\}^{2/\gamma-1}$, $C_\lambda, C_\zeta > 0$ being absolute constants. It then holds with probability at least $1-(n\vee p)^{-3}$ that, with an absolute constant $C_{\mathfrak{e}} > 0$,*

$$\widehat{K} = K \quad \textit{and} \quad \max_{k=1}^{K} \kappa_k^2 |\widehat{\eta}_k - \eta_k| \leq C_{\mathfrak{e}}\{\mathfrak{s}\log(pn)\}^{2/\gamma-1}.$$

Lemma 6 (Lasso estimators)**.** *Under the same conditions and with the same notation as that in Lemma 5, let $\{\widehat{\beta}_k\}_{k=1}^{\widehat{K}}$ be the Lasso estimators defined in (14) with the intervals constructed based on $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$. It then holds with probability at least $1-(n\vee p)^{-3}$ that, with an absolute constant $C>0$,*

$$\max\left\{\max_{k=1}^{K}\kappa_k^{-1}|\widehat{\beta}_k-\beta^*_{\eta_k}|_2,\quad \max_{k=0}^{K-1}\kappa_{k+1}^{-1}|\widehat{\beta}_k-\beta^*_{\eta_k}|_2,\right.$$
$$\left.\mathfrak{s}^{-1/2}\max_{k=1}^{K}\kappa_k^{-1}|\widehat{\beta}_k-\beta^*_{\eta_k}|_1,\quad \mathfrak{s}^{-1/2}\max_{k=0}^{K-1}\kappa_{k+1}^{-1}|\widehat{\beta}_k-\beta^*_{\eta_k}|_1\right\}\leq C\alpha_n^{-1/2},$$

*where $\alpha_n>0$ is any diverging sequence specified in Assumption 4**a**.*

Lemma 7 (Jump size estimators)**.** *Under the same conditions and with the same notation as that in Lemma 5, for each $k\in\{1,\ldots,\widehat{K}\}$, let*

$$\widehat{\kappa}_k=|\widehat{\beta}_k-\widehat{\beta}_{k-1}|_2 \tag{30}$$

be the kth jump size estimator. It then holds with probability at least $1-(n\vee p)^{-3}$ that, with an absolute constant $C>0$,

$$\max_{k=1}^{K}|\widehat{\kappa}_k/\kappa_k-1|\leq C\alpha_n^{-1/2},$$

*where $\alpha_n>0$ is any diverging sequence specified in Assumption 4**a**.*

4 Consistent long-run variance estimation

As pointed out in existing change point inference literature, e.g. Bai (2010), in the non-vanishing jump size regime, change points can be estimated within a constant error rate; while in the vanishing jump size regime, the estimators have diverging localisation error rates. In practice, it is therefore more interesting to study the limiting distributions in the vanishing regime.

To practically perform inference on change point locations based on change point estimators $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{K}}$ in the vanishing jump size regime, it is crucial to have access to consistent estimators of the long-run variances $\{\sigma^2_\infty(k)\}_{k=1}^{K}$ and the drift coefficients $\{\varpi_k\}_{k=1}^{K}$, involved in the limiting distribution in Theorem 3**b.2**. In this section, we first propose a block-type long-run variance estimator and derive its consistency. With the help of this long-run variance estimator, we are able to provide a complete inference procedure of the vanishing regime in Section 4.1.

As we have mentioned in Section 1, most of the existing literature for change point inference in high dimension lacks theoretical discussions on the long-run variance estimator. To the best of our knowledge, Chen et al. (2021) is one of the few existing works showing the consistency of the long-run variance estimators. Chen et al. (2021) studied the consistency of a block-type estimator, for linear processes. In the following, we are to show that a block-type estimator, detailed in Algorithm 2, is also consistent for the general, nonlinear, temporal dependence sequences.

Theorem 8. *Under all the assumptions and with all the notation in Theorem 3, let $\{\widehat{\sigma}^2_\infty(k)\}_{k=1}^{\widehat{K}}$ be the output of Algorithm 2 with $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ output by* DPDU *in Algorithm 1, $\{\widehat{\beta}_k\}_{k=1}^{\widehat{K}}$ constructed from (14) based on the intervals with endpoints $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$, $\{\widehat{\kappa}_k\}_{k=1}^{\widehat{K}}$ defined in (30), $\{(s_k,e_k)\}_{k=1}^{\widehat{K}}$ defined in (18) and $R\in\mathbb{N}$ satisfying $(e_k-s_k)\gg R\gg\sqrt{e_k-s_k}$, as n diverges. We have that*

$$\max_{k=1}^{K}|\widehat{\sigma}^2_\infty(k)-\sigma^2_\infty(k)|\xrightarrow{P.}0,\quad n\to\infty.$$

Algorithm 2 Long-run variance estimators

INPUT: $\{(y_t, X_t)\}_{t=1}^n$, $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$, $\{\widehat{\beta}_k\}_{k=1}^{\widehat{K}}$, $\{\widehat{\kappa}_k\}_{k=1}^{\widehat{K}}$, $\{(s_k, e_k)\}_{k=1}^{\widehat{K}}$ and tuning parameter $R \in \mathbb{N}$

  for $k \in \{1, \dots, \widehat{K}\}$ **do**

    for $t \in \{s_k, \dots, e_k - 1\}$ **do**

      $Z_t^{(k)} \leftarrow (y_t - X_t^\top \widehat{\beta}_k) X_t^\top (\widehat{\beta}_{k+1} - \widehat{\beta}_k)$, $Z_t^{(k+1)} \leftarrow (y_t - X_t^\top \widehat{\beta}_{k+1}) X_t^\top (\widehat{\beta}_{k+1} - \widehat{\beta}_k)$

      $Z_t \leftarrow Z_t^{(k)} + Z_t^{(k+1)}$

    end for

    $S \leftarrow \lfloor (e_k - s_k)/(2R) \rfloor$

    for $r \in \{1, \dots, 2R\}$ **do**

      $\mathcal{S}_r \leftarrow \{s_k + (r-1)S, \dots, s_k + rS - 1\}$

    end for

    for $r \in \{1, \dots, R\}$ **do**

      $D_r \leftarrow (2S)^{-1/2}(\sum_{t \in \mathcal{S}_{2r-1}} Z_t - \sum_{t \in \mathcal{S}_{2r}} Z_t)$

    end for

    $\widehat{\sigma}_\infty^2(k) \leftarrow R^{-1} \widehat{\kappa}_k^{-2} \sum_{r=1}^R D_r^2$

  end for

OUTPUT: $\{\widehat{\sigma}_\infty^2(k)\}_{k=1}^{\widehat{K}}$

Theorem 8 demonstrates the consistency of the long-run variance estimators $\{\widehat{\sigma}_\infty^2(k)\}_{k=1}^{\widehat{K}}$, which are functions of $\{Z_t\}_{t=1}^n$ defined in Algorithm 2. The sequence $\{Z_t\}_{t=1}^n$ is a sample version of the unobservable sequence $\{Z_t^*\}_{t=1}^n$, defined as

$$Z_t^* = \begin{cases} 2\epsilon_t \Psi_k^\top X_t - (\Psi_k^\top X_t)^2, & t = s_k, \dots, \eta_k - 1, \\ 2\epsilon_t \Psi_k^\top X_t + (\Psi_k^\top X_t)^2, & t = \eta_k, \dots, e_k - 1. \end{cases}$$

As shown from Lemma 12 and the proof of Theorem 3, $\sigma_\infty^2(k)$ is, in fact, the long-run variance of the process $\{\kappa_k^{-1} Z_t^*\}_{t \in \mathbb{Z}}$. Note that $\mathrm{Var}(Z_t^*)$ maintains the same for all $t \in (s_k, e_k]$, but $\mathbb{E}[Z_t^*]$ changes at η_k from $-\Psi_k^\top \Sigma \Psi_k$ to $\Psi_k^\top \Sigma \Psi_k$. This requires a robust estimation method against the mean change at the unknown change point η_k and is materialised via the block-type estimator. It is also possible to construct a long-run variance estimator based on the stationary process $\{2\epsilon_t v_k^\top X_t\}_{t \in \mathbb{Z}}$. We consider the estimator based on $\{\kappa_k^{-1} Z_t^*\}_{t \in \mathbb{Z}}$ due to consideration of finite sample performances.

For any $k \in \{1, \dots, \widehat{K}\}$, the estimator $\widehat{\sigma}_\infty^2(k)$ is constructed based on data from the interval (s_k, e_k). When the data are assumed to be independent, the sample variance usually serves as a good estimator of the population variance. To capture the temporal dependence, we proceed by constructing sample variances of means of blocks of data, with block width $S \to \infty$, as n diverges. This is done by dividing the interval (s_k, e_k) into $2R$ width-S intervals, with $2R = \lfloor (e_k - s_k)/S \rfloor$. To subtract the mean when estimating the variance, we define D_r, $r \in \{1, \dots, R\}$, by taking the difference of the sample means in two consecutive blocks. As long as there is no change point in $\mathcal{S}_{2r-1} \cup \mathcal{S}_{2r}$, it roughly holds that $\mathbb{E}\{D_r\} = 0$. With large probability, there is one and only one true change point in (s_k, e_k). This means, only one D_r, $r \in \{1, \dots, R\}$, does not have mean zero, namely D_{r^*}. The long-run variance can thus be estimated by the sample mean of $\{D_r^2\}_{r=1}^R$, hedging out the effect of D_{r^*}, provided that $R \to \infty$, as n diverges.

4.1 Confidence interval construction

An important application of the consistent long-run variance estimator derived above is to construct confidence intervals for the change points. In view of the limiting distributions

$$\kappa_k^2(\widetilde{\eta}_k - \eta_k) \xrightarrow{\mathcal{D}} \mathop{\arg\min}_{r\in\mathbb{R}} \{\varpi_k|r| + \sigma_\infty(k)\mathbb{W}(r)\}, \quad n\to\infty, \forall k \in \{1,\ldots,K\},$$

derived in Theorem 3**b**, aside from the long-run variance, the drift coefficients $\varpi_k = v_k^\top \Sigma v_k$, $k \in \{1,\ldots,K\}$, are also unknown. In the rest of this subsection, we first propose consistent estimators of $\{\varpi_k\}_{k=1}^K$ in Proposition 9 and then provide a confidence interval construction strategy.

Proposition 9 (Consistent estimators of $\{\varpi_k\}_{k=1}^K$)**.** *Given data $\{(y_t, X_t)\}_{t=1}^n$, suppose that Assumptions 1, 2, 3 and 4***a** *hold. Let $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ be the change point estimators output by* $\mathrm{DPDU}(\{y_t, X_t\}_{t=1}^n, \lambda, \zeta)$*, with $\lambda = C_\lambda\sqrt{\log(np)}$ and $\zeta = C_\zeta\{\mathfrak{s}\log(pn)\}^{2/\gamma-1}$, $C_\lambda, C_\zeta > 0$ being absolute constants. Let $\{\widehat{\beta}_k\}_{k=1}^{\widehat{K}}$ be the Lasso estimators defined in (14) with the intervals constructed based on $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$. Let $\{\widehat{\kappa}_k\}_{k=1}^{\widehat{K}}$ be the jump size estimators defined in (30). For $k \in \{1,\ldots,\widehat{K}\}$, define the drift estimators*

$$\widehat{\varpi}_k = \frac{1}{n\widehat{\kappa}_k^2}\sum_{t=1}^n (\widehat{\beta}_k - \widehat{\beta}_{k-1})^\top X_t X_t^\top (\widehat{\beta}_k - \widehat{\beta}_{k-1}). \tag{31}$$

It then holds that

$$\max_{k=1}^K |\widehat{\varpi}_k - \varpi_k| \xrightarrow{P.} 0, \quad n\to\infty.$$

Comparing the conditions required in Theorem 8 and Proposition 9, we see that the slightly stronger signal-to-noise ratio condition Assumption 4**b** is in fact needed when estimating the long-run variance. The weaker condition Assumption 4**a** is sufficient for Proposition 9, which only concerns the properties of the initial change point estimators.

For $k \in \{1,\ldots,\widehat{K}\}$, with the ingredients: change point estimator $\widetilde{\eta}_k$ defined in (17), local refinement interval (s_k, e_k) defined in (18), long-run variance estimator $\widehat{\sigma}_\infty(k)$ output by Algorithm 2 and drift estimator $\widehat{\varpi}_k$ defined in (31), we streamline the construction of a numerical $(1-\alpha)$-confidence interval of η_k, with a user-specified $\alpha \in (0,1)$.

Step 1. Let $B \in \mathbb{Z}_+$ and $M \in \mathbb{R}_+$. For $b \in \{1,\ldots,B\}$, let

$$\widehat{u}^{(b)} = \mathop{\arg\min}_{r\in(-M,M)} \{\widehat{\varpi}_k|r| + \widehat{\sigma}_\infty(k)\mathbb{W}^{(b)}(r)\}, \tag{32}$$

where

$$\mathbb{W}^{(b)}(r) = \begin{cases} \frac{1}{\sqrt{n}}\sum_{i=\lceil nr\rceil}^{-1} z_i^{(b)}, & r<0,\\ 0, & r=0,\\ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nr\rfloor} z_i^{(b)}, & r>0,\end{cases}$$

the quantity n is the sample size and $\{z_i^{(b)}\}_{i=-\lfloor nM\rfloor}^{\lceil nM\rceil}$ are independent standard Gaussian random variables.

Step 2. Let $\widehat{q}_u(1-\alpha/2)$ and $\widehat{q}_u(\alpha/2)$ be the $(1-\alpha/2)$- and $\alpha/2$-quantiles of the empirical distribution of $\{\widehat{u}^{(b)}\}_{b=1}^B$. Provided that $\widehat{\kappa}_k \neq 0$, a numerical $(1-\alpha)$ confidence interval of η_k is

$$\left[\widetilde{\eta}_k + \frac{\widehat{q}_u(\alpha/2)}{\widehat{\kappa}_k^2}\mathbb{1}\{\widehat{\kappa}_k \neq 0\}, \widetilde{\eta}_k + \frac{\widehat{q}_u(1-\alpha/2)}{\widehat{\kappa}_k^2}\mathbb{1}\{\widehat{\kappa}_k \neq 0\}\right]. \tag{33}$$

The rationale of this procedure is that, following Theorem 3, each $\widehat{u}^{(b)}$ is a simulated $\widehat{\kappa}_k^2(\widetilde{\eta}_k - \eta_k)$. Due to the uniform tightness, it is sufficient to restrict the optimisation (32) in the interval $(-M, M)$, with a sufficiently large $M > 0$. When B is large enough, the empirical distribution of $\{\widehat{u}^{(b)}\}_{b=1}^B$ serves as a reasonable approximation of the limiting distribution of $\widehat{\kappa}_k^2(\widetilde{\eta}_k - \eta_k)$, which justifies that (33) is a numerical $(1-\alpha)$ confidence interval of η_k. This is shown in Theorem 10 below, that the set $\{\widehat{u}^{(b)}\}_{b=1}^B$ approximates the sampling distribution of $\widehat{\kappa}_k^2(\widetilde{\eta}_k - \eta_k)$.

Theorem 10. *Under all the assumptions in Theorem 8, for any $k \in \{1, \ldots, K\}$, let $\widehat{u}^{(1)}$ be defined in (32) with $M = \infty$ and $\widehat{\kappa}_k^2$ be the kth jump size estimator defined in (30). It holds that*

$$\frac{\kappa_k^2}{\widehat{\kappa}_k^2}\widehat{u}^{(1)}\mathbb{1}\{\widehat{\kappa}_k \neq 0\} \xrightarrow{\mathcal{D}} \operatorname*{arg\,min}_{r \in \mathbb{R}} \left\{\varpi_k|r| + \sigma_\infty(k)\mathbb{W}(r)\right\}, \quad n \to \infty.$$

5 Numerical experiments

In this section, we conduct numerical studies to investigate the computational efficiency of the proposed DPDU algorithm (Algorithm 1) in Section 5.1 and the numerical performance of change point confidence intervals (33) in Section 5.2. This section is concluded with a real data application in Section 5.3 to showcase the practicability of our proposed methodology.

5.1 Computational efficiency

Recall that we have shown in Lemma 1 that DPDU improves the computational efficiency of the standard dynamic programming (DP). We refer readers to Friedrich et al. (2008) for a thorough discussion on a generic dynamic programming algorithm (a.k.a. optimal partition algorithm, Algorithm 1 therein) and Rinaldo et al. (2021) for the deployment of DP in the high-dimensional linear regression change point localisation context. The implementation of DP and DPDU with the same coordinate descent for the Lasso is available in the R package `changepoints`[1] (Xu et al., 2021) as `DP.regression` and `DPDU.regression` functions, respectively.

In this subsection, we reinforce the computational efficiency of DPDU over DP through simulations. Given the data $\{(y_t, X_t)\}_{t=1}^n \subset \mathbb{R} \times \mathbb{R}^p$ and fixed tuning parameters, we compare DUDP and DP in the settings where $n \in \{200, 300, 400\}$ and $p \in \{30, 50, 70\}$. All the simulations are run on a machine equipped with an Apple M2 chip (8-core CPU). Figure 1 depicts the boxplots of computational cost (in seconds) over 100 repetitions with the x-axis being p. For each fixed p, as n linearly increases, it shows that the computational cost of DP increases much faster compared to that of DPDU, justifying the implication of Lemma 1. For each fixed n, as p linearly increases, it shows that the increasing rates of the computational cost of DP are similar to that of DPDU. In all considered settings, DPDU is nearly twice as fast as DP. All findings align with Lemma 1.

[1]The development version of `changepoints` is available at https://github.com/HaotianXu/changepoints.

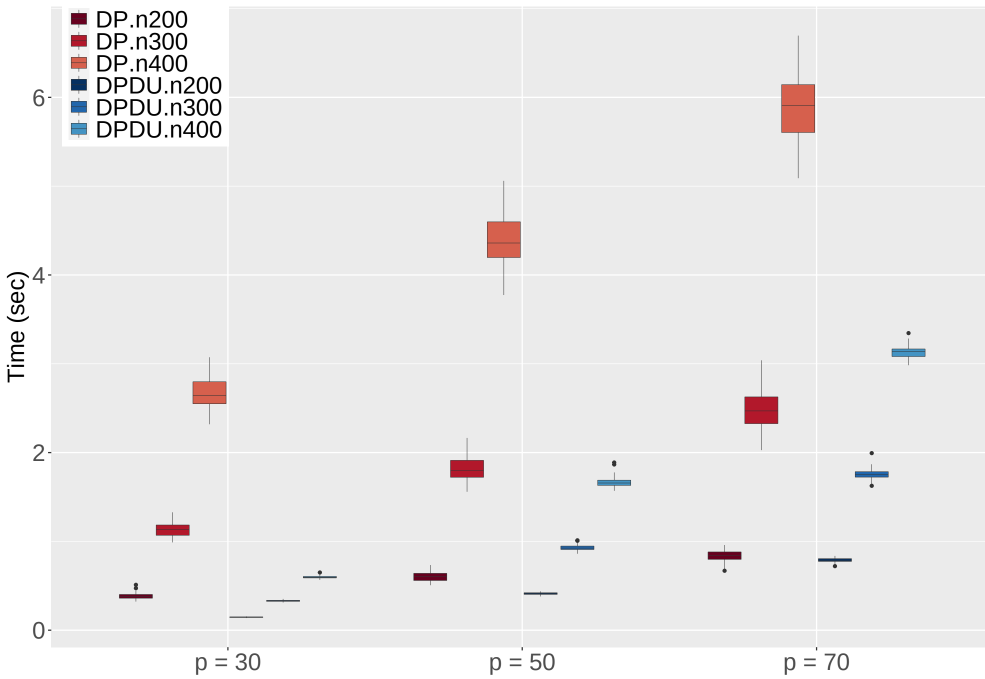


Figure 1: Boxplots of the computational cost (in seconds) of DP and DPDU over 100 repetitions.

5.2 Simulation studies

Our change point inference procedure contains three steps: (i) the preliminary estimation (see Section 2.1), (ii) the local refinement (see Section 2.2) and (iii) the confidence interval construction (see Section 4.1). To the best of our knowledge, no competitor exists for change point inference in high-dimensional regression settings. We therefore focus on investigating the performance of our proposed procedure for change point inference in different scenarios. As an important byproduct, the localisation results are also reported.

Settings. We simulate data from the model described in (1), i.e.

$$y_t = X_t^\top \beta_t^* + \epsilon_t, \quad t = 1, \ldots, n.$$

We consider temporal dependence in both $\{X_t\}_{t=1}^n$ and $\{\epsilon_t\}_{t=1}^n$. The covariate process $\{X_t\}_{t=1}^n$ is generated from an autoregressive model, i.e. $X_t = 0.3X_{t-1} + \sqrt{1-0.3^2}e_t$ with $e_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_p)$ and I_p being the p-dimensional identity matrix. The noise process $\{\epsilon_t\}_{t=1}^n$ is generated by a moving average model, i.e. $\epsilon_t = (e'_t + 0.3e'_{t-1})/(2\sqrt{1+0.3^2})$ with $e'_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Let $\beta_0 = (\beta_{0i}, i = 1, \ldots, p)^\top$ with $\beta_{0i} = 2^{-1}\mathfrak{s}^{-1/2}\kappa$ for $i = 1, \ldots, \mathfrak{s}$ and zero otherwise. Let $\{\eta_k\}_{k=1}^K$ be the true change points. We consider the regression coefficients

$$\beta_t^* = \begin{cases} (-1)^0\beta_0, & t \in \{1, \ldots, \eta_1 - 1\}, \\ (-1)^1\beta_0, & t \in \{\eta_1, \ldots, \eta_2 - 1\}, \\ \ldots \\ (-1)^K\beta_0, & t \in \{\eta_K, \ldots, n\}. \end{cases}$$

In each scenario below, we simulate 500 repetitions and fix $\mathfrak{s} = 5$.

Evaluation measurements. Let $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ and $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{K}}$ be the preliminary and final change point estimators defined in (16) and (17), respectively. To assess the performance of localisation, in Tables 1, 3 and 5, we report (i) the proportions of the repetitions with $\widehat{K} < K$ (i.e. underestimating the number of change points) and $\widehat{K} > K$ (i.e. overestimating the number of change points), and (ii) means and standard deviations (in parentheses) of scaled Hausdorff distances of the preliminary and final estimators. The scaled Hausdorff distance of $\{\widehat{\eta}_k\}_{k=1}^{K}$ is defined as

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

where we set $\widehat{\eta}_0 = 1$ and $\widehat{\eta}_{\widehat{K}+1} = n+1$. We define similarly $d_{\mathrm{H}}^{\mathrm{fin}}$ for $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{K}}$.

The performance of change point inference is measured by the coverage rate of the constructed confidence interval, i.e.

$$\mathrm{cover}_k(1-\alpha) = \mathbb{1}\Bigg\{\eta_k \in \Bigg[\Bigg\lfloor \widetilde{\eta}_k + \frac{\widehat{q}_u(\alpha/2)}{\widehat{\kappa}_k^2} 1\{\widehat{\kappa}_k \neq 0\}\Bigg\rfloor, \Bigg\lceil \widetilde{\eta}_k + \frac{\widehat{q}_u(1-\alpha/2)}{\widehat{\kappa}_k^2} 1\{\widehat{\kappa}_k \neq 0\}\Bigg\rceil\Bigg]\Bigg\},$$

where for $k = 1, \ldots, K$, $\widehat{\kappa}_k$ is defined in (30), $\widehat{q}_u(\alpha/2)$ and $\widehat{q}_u(1-\alpha/2)$ are defined in (33) with $M = n$ and $B = 1000$. We remark that the floor and ceiling functions are applied respectively to the lower and upper endpoint of the confidence interval given in (33), considering that change points are integers. We report the average coverage among all the repetitions with $\widehat{K} = K$ (out of 500 repetitions), along with the widths of numerical $(1-\alpha)$ confidence intervals, i.e.

$$\mathrm{width}_k(1-\alpha) = \Bigg\lceil \widetilde{\eta}_k + \frac{\widehat{q}_u(1-\alpha/2)}{\widehat{\kappa}_k^2} 1\{\widehat{\kappa}_k \neq 0\}\Bigg\rceil - \Bigg\lfloor \widetilde{\eta}_k + \frac{\widehat{q}_u(\alpha/2)}{\widehat{\kappa}_k^2} 1\{\widehat{\kappa}_k \neq 0\}\Bigg\rfloor.$$

In Tables 2, 4, 6 and 7, we report the average $\mathrm{cover}_k(1-\alpha)$ and means and standard deviations (in parentheses) of $\mathrm{width}_k(1-\alpha)$ with $\alpha \in \{1\%, 5\%\}$.

Tuning parameters selection and other unknown quantities estimation. Three tuning parameters are involved in the proposed change point inference procedure: λ and ζ for the DPDU algorithm (see Algorithm 1) in the preliminary estimation and the number of pairs R for long-run variance estimation (see Algorithm 2) in the confidence interval construction. We adopt the cross-validation method proposed by Rinaldo et al. (2021) to select λ and ζ. Specifically, we first divide $\{(y_t, X_t)\}_{t=1}^n$ into training and validation sets according to odd and even indices. For each possible value of λ and ζ considered, we obtain the DPDU outputs ($\widehat{\mathcal{B}}$, $\widehat{K}$ and $\{\widehat{\beta}_k\}_{k=0}^{\widehat{K}}$) based on the training set. Using the DPDU outputs and the validation set, we compute the least squared prediction error as the validation loss. We select the values of λ and ζ, which minimise the validation loss. The selection of R is guided by Theorem 8, i.e. $R = \Big\lfloor \Big(\max_{k=1}^{\widehat{K}} \{e_k - s_k\} \Big)^{3/5} \Big\rfloor$ with $\{(s_k, e_k)\}_{k=1}^{\widehat{K}}$ given in (18). In addition, we use $\{\widehat{\kappa}_k\}_{k=1}^{\widehat{K}}$, $\{\widehat{\sigma}_\infty^2(k)\}_{k=1}^{\widehat{K}}$ and $\{\widehat{\varpi}_k\}_{k=1}^{\widehat{K}}$, defined respectively in (30), Algorithm 2 and (31), to estimate unknown quantities in constructing confidence intervals.

Scenario 1: single change point. Let $K = 1$, $\eta = n/2$ and the jump size $\kappa = 2$. Vary $n \in \{100, 200, 300, 400\}$ and $p \in \{100, 200, 300\}$. Tables 1 and 2 respectively summarise the localisation and inference performances of η. We can see that both the preliminary and final estimators perform better as n increases or p decreases. Our proposed inference procedure produces good coverage in

Table 1: Localisation results of Scenario 1.

n	$\widehat{K} < K$	$\widehat{K} > K$	$d_{\mathrm{H}}^{\mathrm{pre}}$	$d_{\mathrm{H}}^{\mathrm{fin}}$
		$K = 1,\ \kappa = 2$		
		$p = 100$		
100	0.188	0	0.119 (0.183)	0.111 (0.186)
200	0	0	0.007 (0.013)	0.003 (0.005)
300	0	0	0.003 (0.006)	0.001 (0.002)
400	0	0	0.002 (0.005)	0.001 (0.002)
		$p = 200$		
100	0.428	0	0.234 (0.225)	0.230 (0.228)
200	0	0	0.009 (0.017)	0.004 (0.007)
300	0	0	0.004 (0.006)	0.002 (0.003)
400	0	0	0.002 (0.004)	0.001 (0.002)
		$p = 300$		
100	0.610	0	0.319 (0.218)	0.315 (0.221)
200	0.004	0	0.013 (0.035)	0.007 (0.032)
300	0	0	0.004 (0.008)	0.002 (0.003)
400	0	0	0.003 (0.005)	0.001 (0.003)

Table 2: Inference results of Scenario 1.

		$K = 1,\ \kappa = 2$		
	$\alpha = 0.01$		$\alpha = 0.05$	
n	cover$(1-\alpha)$	width$(1-\alpha)$	cover$(1-\alpha)$	width$(1-\alpha)$
		$p = 100$		
100	0.938	16.721 (12.351)	0.897	11.867 (8.193)
200	0.980	7.152 (1.913)	0.956	5.254 (1.416)
300	0.988	6.072 (1.255)	0.980	4.420 (0.879)
400	0.974	5.270 (1.060)	0.958	4.022 (0.553)
		$p = 200$		
100	0.920	28.972 (23.469)	0.867	20.549 (16.514)
200	0.978	8.222 (2.655)	0.962	5.926 (1.894)
300	0.994	6.284 (1.201)	0.968	4.556 (0.921)
400	0.992	5.366 (1.065)	0.982	4.080 (0.599)
		$p = 300$		
100	0.913	33.431 (22.318)	0.877	23.754 (16.469)
200	0.984	10.490 (5.637)	0.946	7.616 (4.165)
300	0.984	6.562 (1.406)	0.966	4.794 (1.055)
400	0.966	5.562 (1.053)	0.944	4.158 (0.680)

the considered settings. As n increases or p decreases, the average length of our constructed confidence intervals decreases.

Scenario 2: varying jump size κ. Consider the same model as that in Scenario 1. Fix $(n, p) = (200, 300)$ and vary $\kappa \in \{1, 2, 3, 4\}$. The localisation performances are summarised in Table 3, which shows that all the estimators achieve better localisation performances as κ increases. Table 4 summarises the inference performances. It shows that the best performance is achieved when κ is between 2 and 3. When κ is too small – it leads to a bad localisation performance, or too big – it may fall in the non-vanishing jump size regime, the inference performance may deteriorate.

Table 3: Localisation results of Scenario 2.

$K = 1$, $n = 200$ $p = 300$				
κ	$\widehat{K} < K$	$\widehat{K} > K$	$d_{\mathrm{H}}^{\mathrm{pre}}$	$d_{\mathrm{H}}^{\mathrm{fin}}$
1	0.532	0	0.279 (0.232)	0.274 (0.237)
2	0.004	0	0.013 (0.035)	0.007 (0.032)
3	0	0	0.007 (0.011)	0.002 (0.004)
4	0	0	0.005 (0.008)	0.001 (0.003)

Table 4: Inference results of Scenario 2.

$K = 1$, $n = 200$, $p = 300$				
	$\alpha = 0.01$		$\alpha = 0.05$	
κ	cover$(1-\alpha)$	width$(1-\alpha)$	cover$(1-\alpha)$	width$(1-\alpha)$
1	0.953	37.427 (16.649)	0.902	27.487 (11.563)
2	0.984	10.490 (5.637)	0.946	7.616 (4.165)
3	0.978	4.546 (1.367)	0.960	3.422 (1.183)
4	0.976	2.938 (1.064)	0.966	2.188 (0.594)

Scenario 3: unequally-spaced multiple change points. Let $K = 2$ and the unequally-spaced change points $\{\eta_1, \eta_2\} = \{n/4, 5n/8\}$. We fix $\kappa_1 = \kappa_2 = 2$, varying $n \in \{200, 400, 800\}$ and $p \in \{100, 200, 300\}$. Table 5 shows that the localisation performances of both preliminary and final estimators improve as n increases or p decreases. For change point inference, the comparison between Tables 6 and 7 suggests that our inference procedure applied on η_2, the one with a larger sample size, performs better.

5.3 Real data application

We apply the proposed procedure to the Federal Reserve Economic Database (FRED)[2] (McCracken and Ng, 2016). Wang and Zhao (2022) recently conducted a hypothesis testing on the existence of change points on the same data set. Their results suggest that there exist change points in the relationship between the monthly growth rate of the US industrial production (IP) index – an important indicator of macroeconomic activity, and other macroeconomic variables. Consider the response variable (monthly growth rate of IP, in percentage scale) $y_t = \log(\mathrm{IP}_t/\mathrm{IP}_{t-1})$, and the

[2]The data set is publicly available at `https://research.stlouisfed.org/econ/mccracken/fred-databases`.

Table 5: Localisation results of Scenario 3.

		$K = 1,\ \kappa_1 = \kappa_2 = 2$		
n	$\widehat{K} < K$	$\widehat{K} > K$	$d_{\mathrm{H}}^{\mathrm{pre}}$	$d_{\mathrm{H}}^{\mathrm{fin}}$
		$p = 100$		
200	0.036	0	0.034 (0.050)	0.026 (0.070)
400	0	0	0.005 (0.005)	0.003 (0.003)
800	0	0	0.002 (0.002)	0.001 (0.001)
		$p = 200$		
200	0.208	0.004	0.084 (0.110)	0.090 (0.142)
400	0	0	0.007 (0.009)	0.002 (0.003)
800	0	0	0.002 (0.002)	0.001 (0.001)
		$p = 300$		
200	0.392	0.004	0.136 (0.133)	0.158 (0.170)
400	0	0	0.007 (0.009)	0.002 (0.003)
800	0	0	0.002 (0.003)	0.001 (0.001)

Table 6: Inference results for the first change point (η_1) in Scenario 3.

		$K = 1,\ \kappa_1 = \kappa_2 = 2$		
	$\alpha = 0.01$		$\alpha = 0.05$	
n	$\mathrm{cover}_1(1-\alpha)$	$\mathrm{width}_1(1-\alpha)$	$\mathrm{cover}_1(1-\alpha)$	$\mathrm{width}_1(1-\alpha)$
		$p = 100$		
200	0.927	12.110 (7.063)	0.876	8.654 (4.886)
400	0.984	6.844 (1.641)	0.960	4.954 (1.258)
800	0.974	5.290 (0.923)	0.962	4.026 (0.312)
		$p = 200$		
200	0.914	19.043 (13.639)	0.865	13.668 (9.382)
400	0.978	7.278 (1.798)	0.964	5.286 (1.356)
800	0.984	5.348 (0.908)	0.972	4.062 (0.377)
		$p = 300$		
200	0.924	27.298 (22.191)	0.884	19.308 (15.590)
400	0.994	7.792 (2.090)	0.982	5.690 (1.531)
800	0.972	5.352 (0.983)	0.966	4.044 (0.388)

Table 7: Inference results for the second change point (η_2) in Scenario 3.

$K = 1$, $\kappa_1 = \kappa_2 = 2$				
	$\alpha = 0.01$		$\alpha = 0.05$	
n	$\text{cover}_2(1-\alpha)$	$\text{width}_2(1-\alpha)$	$\text{cover}_2(1-\alpha)$	$\text{width}_2(1-\alpha)$
		$p = 100$		
200	0.967	10.697 (5.545)	0.948	7.701 (3.830)
400	0.972	6.112 (1.181)	0.950	4.400 (0.816)
800	0.980	4.926 (0.931)	0.976	4.034 (0.247)
		$p = 200$		
200	0.982	17.244 (14.101)	0.954	12.345 (9.833)
400	0.980	6.254 (1.202)	0.966	4.572 (0.922)
800	0.986	4.994 (0.916)	0.976	3.992 (0.237)
		$p = 300$		
200	0.954	24.891 (17.430)	0.917	17.709 (12.593)
400	0.984	6.640 (1.295)	0.976	4.818 (1.011)
800	0.982	5.066 (0.914)	0.972	4.006 (0.224)

covariates with 116 macroeconomic variables recorded for the month $t-1$ (after removing variables containing missing values).

Following the recommendation given on the FRED website, we transform the raw data into stationary time series and remove outliers using the R package `fbi` (Chen et al., 2022). We consider the period from January 2000 to December 2019 with sample size $n = 240$. Tuning parameters selection and unknown quantities estimation follow the same method specified in Section 5.2. We first perform DPDU, which outputs July 2008 as the initial change point estimator. We then refine the initial estimator and obtain December 2007 as the final estimator, with a 95% confidence interval [February 2007, October 2008] and a 99% confidence interval [September 2006, March 2009]. This change point coincides with the financial crisis of 2008. The final estimator closely relates to the beginning of the avalanche of financial industry bankruptcies. The confidence intervals, especially the 99% confidence interval, well include the milestone events, e.g. the bankruptcy of Lehman Brothers, the acquisition of Merrill Lynch by Bank of America, the bailout and sell of Bear Stearns, United States bear market of 2007–2009, among many others.

Note that, we have also implemented three other change point localisation methods, BSA (Leonardi and Bühlmann, 2016), SGL (Zhang et al., 2015) and VPC (Wang et al., 2021b), on the same data set. BSA and SGL report no change point. VPC estimates a change point at August 2008, which lies in our 99% confidence interval.

6 Discussion

In this paper, we derived limiting distributions of change point estimators, in the context of high-dimensional linear regression model, with temporal dependence in both covariate and noise sequences. We believe such limiting distributions are first time seen in the high-dimensional linear regression change point literature, including those temporally-independent scenarios. In addition to the limiting distributions, this paper is also a host of various interesting byproducts.

One potential extension is the implementation of the limiting distribution when the jump size is non-vanishing. Although from a change point detection point of view, the vanishing jump size scenario is much more challenging, the limiting distribution of change point estimators when the jump size is non-vanishing lacks universality. This is to say, as detailed in Theorem 3**a**, the limiting distribution is a function of the data generating mechanism. This is understandable, since in this case the localisation error of order $O(1)$, and such phenomenon is observed and exists in simpler settings (e.g. Yao, 1988; Aue et al., 2009; Mukherjee et al., 2022). It would be challenging and interesting to devise a practical way to distinguish the vanishing and non-vanishing jump size regimes, and further utilise the derived limiting distributions in the non-vanishing regime.

References

Aue, A., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16.

Aue, A., Rice, G., and Sönmez, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):509–529.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, 15(5):453–472.

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, 157(1):78–92.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567.

Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. (2009a). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):927–946.

Berkes, I., Hörmann, S., and Schauer, J. (2009b). Asymptotic results for the empirical process of stationary sequences. *Stochastic processes and their applications*, 119(4):1298–1324.

Bhattacharjee, M., Banerjee, M., and Michailidis, G. (2018). Change point estimation in a dynamic stochastic block model. *arXiv preprint arXiv:1812.03090*.

Brodsky, E. and Darkhovsky, B. S. (2013). *Nonparametric Methods in Change Point Problems*. Springer Netherlands.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Burkholder, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. *Astérisque*, 157(158):75–94.

Cao, H. and Wu, W. B. (2015). Changepoint estimation: another look at multiple testing problems. *Biometrika*, 102(4):974–980.

Casini, A. and Perron, P. (2019). Structural breaks in time series. In *Oxford Research Encyclopedia of Economics and Finance (forthcoming)*. Oxford University Press.

Chen, L., Wang, W., and Wu, W. B. (2021). Inference of breakpoints in high-dimensional time series. *Journal of the American Statistical Association*, pages 1–13.

Chen, Y. B., Ng, S., and Bai, J. (2022). *fbi: Factor-Based Imputation and FRED-MD/QD Data Set*. R package version 0.6.0.

Chu, N. (2016). Air quality in northern taiwan.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley Series in Probability and Statistics. Wiley.

Dedecker, J. and Prieur, C. (2004). Coupling for τ-dependent sequences and applications. *Journal of Theoretical Probability*, 17(4):861–885.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, 32(2):407–499.

Friedrich, F., Kempe, A., Liebscher, V., and Winkler, G. (2008). Complexity penalized m-estimation: fast computation. *Journal of Computational and Graphical Statistics*, 17(1):201–224.

Han, Y. and Tsay, R. S. (2020). High-dimensional linear regression for dependent data with applications to nowcasting. *Statistica Sinica*, 30(4):1797–1827.

Kaul, A., Fotopoulos, S. B., Jandhyala, V. K., and Safikhani, A. (2021). Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics*, 15(1):71–134.

Kaul, A., Jandhyala, V. K., and Fotopoulos, S. B. (2019). An efficient two step algorithm for high dimensional change point regression models without grid search. *J. Mach. Learn. Res.*, 20:111–1.

Kaul, A. and Michailidis, G. (2021). Inference for change points in high dimensional mean shift models. *arXiv preprint arXiv:2107.09150*.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598.

Kuchibhotla, A. K., Brown, L. D., Buja, A., George, E. I., and Zhao, L. (2021). Uniform-in-submodel bounds for linear regression in a model-free framework. *Econometric Theory*, pages 1–47.

Lee, S., Seo, M. H., and Shin, Y. (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 Bühlmann, P. (2016). Computationally efficient change point detection for high-dimensional regression. *arXiv preprint arXiv:1601.03704*.

Liu, W., Xiao, H., and Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statistica Sinica*, pages 1257–1272.

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, pages 1637–1664.

Madrid Padilla, C. M., Wang, D., Zhao, Z., and Yu, Y. (2022). Change-point detection for sparse and dense functional data in general dimensions. *arXiv e-prints*, pages arXiv–2205.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3):435–474.

Mukherjee, D., Banerjee, M., and Ritov, Y. (2022). On robust learning in the canonical change point problem under heavy tailed errors in finite and growing dimensions. *Electronic Journal of Statistics*, 16(1):1153–1252.

National Research Council (2013). *Frontiers in Massive Data Analysis*. The National Academies Press, Washington, DC.

Pilliat, E., Carpentier, A., and Verzelen, N. (2020). Optimal multiple change-point detection for high-dimensional data. *arXiv preprint arXiv:2011.07818*.

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

Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *Journal of Theoretical Probability*, 22(1):146–163.

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

van der Vaart, A. and Wellner, J. (1996). *Weak convergence and empirical processes: with applications to statistics*. Springer.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.

Verzelen, N., Fromont, M., Lerasle, M., and Reynaud-Bouret, P. (2020). Optimal change-point detection and localization. *arXiv preprint arXiv:2010.11470*.

Wang, D., Yu, Y., and Rinaldo, A. (2020). Univariate mean change point detection: Penalization, cusum and optimality. *Electronic Journal of Statistics*, 14(1):1917–1961.

Wang, D., Yu, Y., and Rinaldo, A. (2021a). Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics*, 49(1):203–232.

Wang, D. and Zhao, Z. (2022). Optimal change-point testing for high-dimensional linear models with temporal dependence. *arXiv preprint arXiv:2205.03880*.

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

Wang, R. and Shao, X. (2020). Dating the break in high-dimensional data. *arXiv preprint arXiv:2002.04115*.

Wang, R., Zhu, C., Volgushev, S., and Shao, X. (2022). Inference for change points in high-dimensional data via selfnormalization. *The Annals of Statistics*, 50(2):781–806.

Wang, W., Wu, W. B., and Richter, S. (2021c). Testing for parameter change epochs in garch time series. *Available at SSRN 3791556*.

Wong, K. C., Li, Z., and Tewari, A. (2020). Lasso guarantees for beta-mixing heavy-tailed time series. *The Annals of Statistics*, 48(2):1124–1142.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and its Interface*, 4(2):207–226.

Wu, W.-B. and Wu, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, 10(1):352–379.

Xu, H., Padilla, O., Wang, D., and Li, M. (2021). *changepoints: A Collection of Change-Point Detection Methods*. R package version 1.0.0.

Yao, Y.-C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *The Annals of Statistics*, pages 1321–1328.

Yao, Y.-C. (1988). Estimating the number of change-points via schwarz'criterion. *Statistics & Probability Letters*, 6(3):181–189.

Yao, Y.-C. and Au, S.-T. (1989). Least-squares estimation of a step function. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 370–381.

Yu, M. and Chen, X. (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.

Yu, Y., Chatterjee, S., and Xu, H. (2022). Localising change points in piecewise polynomials of general degrees. *Electronic Journal of Statistics*, 16(1):1855–1890.

Zhang, B., Geng, J., and Lai, L. (2015). Multiple change-points estimation in linear regression models via sparse group lasso. *IEEE Transactions on Signal Processing*, 63(9):2209–2224.

Zhang, D. and Wu, W. B. (2017). Gaussian approximation for high dimensional time series. *The Annals of Statistics*, 45(5):1895–1919.

Zhang, Y., Wang, R., and Shao, X. (2021). Adaptive inference for change points in high-dimensional data. *Journal of the American Statistical Association*, pages 1–12.

Appendices

We collect all the technical details in the Appendices. Appendix A collects all the proofs of the results in Section 2. Appendix B collects the proofs of the main results, which are in Section 3.2. Appendix C collects the proofs of the intermediate results, which are in Section 3.4, except the proof of Theorem 4. Appendix D collects the proofs of the results regarding the long-run variance estimator and its application, which are in Section 4. Some additional technical details are collected in Appendix E. Appendix F is of its own interest, providing the proof of Theorem 4. Appendix F is a self-contained section and the notation therein only applies within the section.

A Proofs of results in Section 2

Given data $\{y_t, X_t\}_{t=1}^n \subset \mathbb{R} \times \mathbb{R}^p$ and tuning parameter $\lambda > 0$, let

$$\widehat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{n} \sum_{t=1}^n (y_t - X_t^\top \beta)^2 + \frac{\lambda}{\sqrt{n}} |\beta|_1 \right\}.$$

Assume for simplicity that there is no intercept. Consider the coordinate descent algorithm. Given $\widetilde{\beta} \in \mathbb{R}^p$, the coordinate-wise update of $\widetilde{\beta}$ has the form

$$\widetilde{\beta}_j \leftarrow S\left(\frac{1}{n} \sum_{t=1}^n X_{tj}(y_t - \widetilde{y}_t^{(j)}),\ \lambda \right), \quad j \in \{1, \ldots, p\},$$

where $\widetilde{y}_t^{(j)} = \sum_{k \neq j} X_{tk} \widetilde{\beta}_k$ and $S(z, \lambda)$ is the soft-thresholding operator $S(z, \lambda) = \mathrm{sign}(z)(|z| - \lambda)_+$.

Lemma 11. *Given the quantities $\sum_{t=1}^n y_t X_t \in \mathbb{R}^p$ and $\sum_{t=1}^n X_t X_t^\top \in \mathbb{R}^{p \times p}$. Each complete iteration of the coordinate descent has computational cost of order $O(p^2)$.*

Proof. Let $\widetilde{\beta}$ be given. Then for each fixed $j \in \{1, \ldots, p\}$, note that

$$\sum_{t=1}^n X_{tj}(y_t - \widetilde{y}_t^{(j)}) = \sum_{t=1}^n X_{tj} y_t - \sum_{k \neq j} \left(\sum_{t=1}^n X_{tj} X_{tk} \right) \widetilde{\beta}_k.$$

Therefore, with $\sum_{t=1}^n y_t X_t \in \mathbb{R}^p$ and $\sum_{t=1}^n X_t X_t^\top \in \mathbb{R}^{p \times p}$ stored in the memory, it takes $O(p)$ operations to compute $\sum_{t=1}^n X_{tj}(y_t - \widetilde{y}_t^{(j)})$ and update $\widetilde{\beta}_j$. So it takes $O(p^2)$ to update each of the p coordinates once, which is a complete iteration of the coordinate descent. □

Proof of Lemma 1. Note that $\mathcal{M}(r, l) = \sum_{t=l}^{r-1} X_t X_t^\top$ and $\mathcal{V}(r, l) = \sum_{t=l}^{r-1} y_t X_t^\top$. The claim on the memory cost follows from the observation that only $\{\mathcal{M}(r, l)\}_{l=1,\ldots,r-1}$ and $\{\mathcal{V}(r, l)\}_{l=1,\ldots,r-1}$ are temporally stored in the memories.

As for the computational cost, from Algorithm 1, the major computational cost comes from the three sources: computing $\mathcal{M}(r, l)$, $\mathcal{V}(r, l)$ and $\mathcal{F}(\mathcal{I})$.

Step 1. For any fixed $r \in \{2, \ldots, n+1\}$, computing $\{\mathcal{M}(r, 1), \ldots, \mathcal{M}(r, r-1)\}$ based on $\{\mathcal{M}(r-1, 1), \ldots, \mathcal{M}(r-1, r-2)\}$ requires at most $O(np^2)$ operations. Computing $\{\mathcal{V}(r, 1), \ldots, \mathcal{V}(r, r-1)\}$ based on $\{\mathcal{V}(r-1, 1), \ldots, \mathcal{V}(r-1, r-2)\}$ requires at most $O(np)$ operations. It therefore requires $O(n^2p^2)$ operations to compute $\{\mathcal{M}(r, l)\}_{2 \leq r \leq n+1, 1 \leq l \leq r-1}$ and $\{\mathcal{V}(r, l)\}_{2 \leq r \leq n+1, 1 \leq l \leq r-1}$.

Step 2. For any $\mathcal{I} = [l, r)$, given $\mathcal{M}(r, l)$ and $\mathcal{V}(r, l)$, it requires $\mathbf{Lasso}(p)$ to compute $\widehat{\beta}_{\mathcal{I}}$. Given $\widehat{\beta}_{\mathcal{I}}$, it requires $O(p^2)$ to compute $\mathcal{F}(\mathcal{I})$. There are at most n^2 many intervals in $(0, n]$. So the cost of computing $\{\mathcal{F}(\mathcal{I})\}_{\mathcal{I} \subset (0,n]}$ via DPDU is $O(n^2p^2 + n^2\mathbf{Lasso}(p))$. □

B Proofs for Section 3.2

B.1 Proof of Lemma 2

Proof of Lemma 2. For any fixed $n \in \mathbb{Z}_+$, we consider the following quantity

$$\operatorname{Var}\left(n^{-1/2}\sum_{t=1}^{n} 2\epsilon_t v_k^\top X_t\right).$$

Note that $\{\epsilon_t\}$ is independent of $\{X_t\}$.

Upper bound. We first consider the upper bounds using the martingale decomposition and Burkholder's inequality (Lemma 24). For any t, we decompose the random variable $\epsilon_t v_k^\top X_t$ as

$$\epsilon_t v_k^\top X_t = \epsilon_t v_k^\top X_t - \mathbb{E}[\epsilon_t v_k^\top X_t] = \sum_{i=0}^{\infty} \mathcal{P}_{t-i}(\epsilon_t v_k^\top X_t),$$

where $\{\mathcal{P}_{t-i}(\epsilon_t v_k^\top X_t)\}_{t\in\mathbb{Z}}$ is a martingale difference sequence. It holds for any fixed $n \in \mathbb{Z}_+$ that

$$\begin{aligned}
&\sqrt{\operatorname{Var}\left(n^{-1/2}\sum_{t=1}^{n}\epsilon_t v_k^\top X_t\right)} = \left\|n^{-1/2}\sum_{t=1}^{n}\left\{\epsilon_t v_k^\top X_t - \mathbb{E}[\epsilon_t v_k^\top X_t]\right\}\right\|_2 \\
=&\left\|n^{-1/2}\sum_{t=1}^{n}\sum_{i=0}^{\infty}\mathcal{P}_{t-i}(\epsilon_t v_k^\top X_t)\right\|_2 \le n^{-1/2}\sum_{i=0}^{\infty}\left\|\sum_{t=1}^{n}\mathcal{P}_{t-i}(\epsilon_t v_k^\top X_t)\right\|_2 \\
\le& n^{-1/2}\sum_{i=0}^{\infty}\left(\sum_{t=1}^{n}\left\|\mathcal{P}_{t-i}(\epsilon_t v_k^\top X_t)\right\|_2^2\right)^{1/2} \le \sum_{i=0}^{\infty}\left\|\epsilon_i v_k^\top X_i - \epsilon_{i,\{0\}} v_k^\top X_{i,\{0\}}\right\|_2 \\
=&\sum_{i=0}^{\infty}\left\|\epsilon_i v_k^\top X_i - \epsilon_{i,\{0\}} v_k^\top X_i + \epsilon_{i,\{0\}} v_k^\top X_i - \epsilon_{i,\{0\}} v_k^\top X_{i,\{0\}}\right\|_2 \\
\le&\sum_{i=0}^{\infty}\left\|(\epsilon_i - \epsilon_{i,\{0\}}) v_k^\top X_i\right\|_2 + \sum_{i=0}^{\infty}\left\|\epsilon_{i,\{0\}} v_k^\top (X_i - X_{i,\{0\}})\right\|_2 \\
\le&\left\|v_k^\top X_0\right\|_4\sum_{i=0}^{\infty}\delta_{i,4}^{\epsilon} + \|\epsilon_0\|_4\sum_{i=0}^{\infty}\left\|v_k^\top (X_i - X_{i,\{0\}})\right\|_4 \\
\le& D_\epsilon \sup_{|v|_2=1}\left\|v^\top X_0\right\|_4 + D_X\|\epsilon_0\|_4,
\end{aligned} \tag{34}$$

where the first inequality follows from Fubini's theorem and the triangle inequality, the second inequality follows from Burkholder's inequality, the fourth inequality follows from the triangle inequality, the fifth inequality follows from Hölder's inequality, and the sixth inequality follows from Assumptions 2 and 3. The third inequality follows from the stationarity and the fact that for any $i \ge 0$ and any random variable Z_t of the form (10) that

$$\begin{aligned}
&\|\mathcal{P}_{t-i}Z_t\|_2 = \|\mathbb{E}[Z_t|\mathcal{F}_{t-i}] - \mathbb{E}[Z_t|\mathcal{F}_{t-i-1}]\|_2 \\
=&\|\mathbb{E}[Z_t|\mathcal{F}_{t-i}] - \mathbb{E}[Z_{t,\{t-i\}}|\mathcal{F}_{t-i-1}]\|_2 \\
=&\|\mathbb{E}[Z_t - Z_{t,\{t-i\}}|\mathcal{F}_{t-i}]\|_2 \le \|Z_t - Z_{t,\{t-i\}}\|_2,
\end{aligned}$$

where the second and the third equalities follow from the definition of the coupled random variables $Z_{t,\{t-i\}}$, and the first inequality follows from Jensen's inequality.

Letting $n \to \infty$, in the vanishing regime, it satisfies that

$$\sigma_\infty^2 \le 4\Big(D_\epsilon \sup_{|v|_2=1} \big\|v^\top X_0\big\|_4 + D_X\|\epsilon_0\|_4\Big)^2 < \infty.$$

Convergence. Next, we show that $\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^n 2\epsilon_t v_k^\top X_t\right)$ has a limit. It is suffice to show that $\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^n 2\epsilon_t v_k^\top X_t\right)$ is a Cauchy sequence. Note that $\{2\epsilon_t v_k^\top X_t\}_{t\in\mathbb{Z}}$ is a stationary process. Denote the lag-ℓ covariance as

$$\gamma(\ell) = \mathrm{Cov}\left(a\epsilon_{t-\ell}v_k^\top X_{t-\ell}, a\epsilon_t v_k^\top X_t\right) = 4\mathbb{E}[\epsilon_{t-\ell}\epsilon_t]\mathbb{E}\big[v_k^\top X_{t-\ell}X_t^\top v_k\big].$$

For $\mathbb{E}[\epsilon_{t-\ell}\epsilon_t]$, we have that for $\ell \ge 0$

$$\begin{aligned}
\big|\mathbb{E}[\epsilon_{t-\ell}\epsilon_t]\big| &= \left|\mathbb{E}\Big[\Big(\sum_{k=0}^\infty \mathcal{P}_{-k}\epsilon_t\Big)\Big(\sum_{k=0}^\infty \mathcal{P}_{\ell-k}\epsilon_{t+l}\Big)\Big]\right| = \left|\sum_{k=0}^\infty \mathbb{E}\big[(\mathcal{P}_{-k}\epsilon_t)(\mathcal{P}_{-k}\epsilon_{t+\ell})\big]\right| \\
&\le \sum_{k=0}^\infty \left|\mathbb{E}\big[(\mathcal{P}_{-k}\epsilon_t)(\mathcal{P}_{-k}\epsilon_{t+\ell})\big]\right| \le \sum_{k=0}^\infty \|\mathcal{P}_{-k}\epsilon_t\|_2\|\mathcal{P}_{-k}\epsilon_{t+\ell}\|_2 \le \sum_{k=0}^\infty \delta_{k,2}^\epsilon\delta_{k+\ell,2}^\epsilon \\
&\le \sqrt{\sum_{k=0}^\infty (\delta_{k,2}^\epsilon)^2}\sqrt{\sum_{k=0}^\infty (\delta_{k+\ell,2}^\epsilon)^2} \le \Big(\sum_{k=0}^\infty \delta_{k,2}^\epsilon\Big)\Big(\sum_{k=\ell}^\infty \delta_{k,2}^\epsilon\Big) = \Delta_{0,2}^\epsilon\Delta_{\ell,2}^\epsilon,
\end{aligned}$$

where the first inequality follows the triangle inequality and the second and fourth inequalities follow Hölder's inequality. The second equality also follows the orthogonality of $\mathcal{P}_j\cdot$, i.e. for $i < j$

$$\mathbb{E}[(\mathcal{P}_i\epsilon_r)(\mathcal{P}_j\epsilon_s)] = \mathbb{E}[\mathbb{E}[(\mathcal{P}_i\epsilon_r)(\mathcal{P}_j\epsilon_s)|\mathcal{F}_i]] = \mathbb{E}[(\mathcal{P}_i\epsilon_r)\mathbb{E}[\epsilon_s - \epsilon_s|\mathcal{F}_i]] = 0,$$

and the orthogonality also holds for $i > j$ by symmetry. The third inequality is due to the fact that

$$\begin{aligned}
\|\mathcal{P}_j\epsilon_i\|_2 =& \|\mathbb{E}[\epsilon_i|\mathcal{F}_j] - \mathbb{E}[\epsilon_i|\mathcal{F}_{j-1}]\|_2 = \|\mathbb{E}[\epsilon_i|\mathcal{F}_j] - \mathbb{E}[\epsilon_{i,\{j\}}|\mathcal{F}_{j-1}]\|_2 \\
=& \|\mathbb{E}[\epsilon_i - \epsilon_{i,\{j\}}|\mathcal{F}_j]\|_2 \le \|\epsilon_i - \epsilon_{i,\{j\}}\|_2 \le \delta_{i-j,2}^\epsilon,
\end{aligned}$$

where the second and the third equality follows the definition of the coupled random variables $\epsilon_{i,\{j\}}$, the first inequality follows Jensen's inequality, and the second inequality follows the definition of the functional dependence measure. By the same arguments, we have for any $\ell < 0$ that

$$\big|\mathbb{E}[\epsilon_{t-\ell}, \epsilon_t]\big| \le \Delta_{0,2}^\epsilon\Delta_{-\ell,2}^\epsilon.$$

It follows from Assumption 3 that

$$\big|\mathbb{E}[\epsilon_{t-\ell}, \epsilon_t]\big| \le D_\epsilon^2 \exp(-c|\ell|^{\gamma_1}).$$

By the same arguments, it follows from Assumption 2 that

$$\Big|\mathbb{E}\big[v_k^\top X_{t-\ell}X_t^\top v_k\big]\Big| \le D_X^2 \exp(-c|\ell|^{\gamma_1}).$$

Therefore, we have

$$\left|\gamma(\ell)\right| \le 4D_\epsilon^2 D_X^2 \exp(-2c|\ell|^{\gamma_1}). \tag{35}$$

By some calculation, we have that

$$\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n} 2\epsilon_t v_k^\top X_t\right) = \sum_{-n<\ell<n}\Big(1-\frac{|\ell|}{n}\Big)\gamma(\ell).$$

It follows from (35) that

$$\begin{aligned}
&\left|\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n} 2\epsilon_t v_k^\top X_t\right) - \mathrm{Var}\left(m^{-1/2}\sum_{t=1}^{m} 2\epsilon_t v_k^\top X_t\right)\right| \\
\le & \frac{2}{\min\{n,m\}}\sum_{\ell=0}^{\min\{n,m\}-1} \ell|\gamma(\ell)| + \sum_{\ell=\min\{n,m\}}^{\max\{n,m\}} |\gamma(\ell)| \\
\le & \frac{8D_\epsilon^2 D_X^2}{\min\{n,m\}}\sum_{\ell=0}^{\min\{n,m\}-1} \ell\exp(-2c|\ell|^{\gamma_1}) + 4D_\epsilon^2 D_X^2 \sum_{\ell=\min\{n,m\}}^{\max\{n,m\}} \exp(-2c|\ell|^{\gamma_1}),
\end{aligned}$$

which shows that $\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n} 2\epsilon_t v_k^\top X_t\right)$ is Cauchy sequence since the sum converges. It further leads to that $\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n} 2\epsilon_t v_k^\top X_t\right)$ has a limit. Combining the upper bounds and the existence of a limit completes the proof. ◻

Lemma 12. *Under the same assumptions in Lemma 2. We have the equivalence that*

$$\begin{aligned}
\sigma_\infty^2(k) =& 4\lim_{n\to\infty}\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n}\epsilon_t v_k^\top X_t\right) \\
=& \lim_{n\to\infty}\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n}\kappa_k^{-1}\big\{a\epsilon_t\Psi_k^\top X_t + (\Psi_k^\top X_t)^2\big\}\right)
\end{aligned} \tag{36}$$

for any $a\in\{-2,2\}$.

Proof. For any fixed $n\in\mathbb{Z}_+$, we have the following decomposition that

$$\begin{aligned}
&\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n}\kappa_k^{-1}\big\{a\epsilon_t\Psi_k^\top X_t + (\Psi_k^\top X_t)^2\big\}\right) \\
=& 4\,\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n}\epsilon_t v_k^\top X_t\right) + \kappa_k^{-2}\,\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n}(\Psi_k^\top X_t)^2\right),
\end{aligned}$$

which follows from the independence between $\{\epsilon_t\}$ and $\{X_t\}$ and $\mathbb{E}[\epsilon_t]=0$.

Note that the upper bound and the convergence of $4\,\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n}\epsilon_t v_k^\top X_t\right)$ have been studied in the proof of Lemma 2. It is suffice to show that

$$\lim_{n\to\infty}\kappa_k^{-2}\,\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n}(\Psi_k^\top X_t)^2\right)=0. \tag{37}$$

Consider the process $\{\kappa_k^{-1}(\Psi_k^\top X_t)^2\}_{t\in\mathbb{Z}}$. Using the similar arguments in (34) of the proof of Lemma 2, it holds for any fixed $n \in \mathbb{Z}_+$ that

$$\begin{aligned}\sqrt{\mathrm{Var}\left(n^{-1/2}\sum_{t=1}^{n}(\Psi_k^\top X_t)^2\right)} \leq & 2\sum_{i=0}^{\infty}\left\|\Psi_k^\top(X_i+X_{i,\{0\}})\right\|_4\left\|\Psi_k^\top(X_i-X_{i,\{0\}})\right\|_4 \\ \leq & 4\kappa_k^2 \sup_{|v|_2=1}\left\|v^\top X_0\right\|_4 D_X,\end{aligned}$$

which further leads to (37). $\square$

B.2 Proof of Theorem 3

Define the event

$$\mathcal{E}_{\mathrm{DP}} = \left\{\widehat{K} = K \text{ and } \max_{1\leq k\leq K}\kappa_k^2|\eta_k - \widehat{\eta}_k| \leq C_{\mathfrak{e}}\big(\mathfrak{s}\log(np)+\zeta\big)\right\}. \tag{38}$$

By Lemma 5, under Assumptions 1, 2 and 3, we have that $\mathcal{E}_{\mathrm{DP}}$ holds with probability at least $1-cn^{-3}$.

Proof of Theorem 3.
Preliminary. Since $\mathcal{E}_{\mathrm{DP}}$ holds with probability tending to 1, we condition on the $\mathcal{E}_{\mathrm{DP}}$ in the following proof. For $k \in \{0,1,\ldots,K\}$, let $\mathcal{I}_k = [l_k, r_k)$, where $l_k = \widehat{\eta}_k$, $r_k = \widehat{\eta}_{k+1}$, $\widehat{\eta}_0 = 1$ and $\widehat{\eta}_{K+1} = n+1$. We have that

$$|l_k - \eta_k| \leq C_{\mathfrak{e}}\frac{\mathfrak{s}\log(p)+\zeta}{\kappa_k^2} \text{ and } |r_k - \eta_{k+1}| \leq C_{\mathfrak{e}}\frac{\mathfrak{s}\log(p)+\zeta}{\kappa_{k+1}^2}.$$

It follows from Assumption 4 that

$$|\mathcal{I}_k| = r_k - l_k \geq \eta_{k+1} - \eta_k - C_{\mathfrak{e}}\frac{\mathfrak{s}\log(np)+\zeta}{\kappa_k^2} - C_{\mathfrak{e}}\frac{\mathfrak{s}\log(np)+\zeta}{\kappa_{k+1}^2} \geq \frac{\Delta}{2} \geq \zeta,$$

and

$$|\mathcal{I}_k| \leq \eta_{k+1} - \eta_k + C_{\mathfrak{e}}\frac{\mathfrak{s}\log(np)+\zeta}{\kappa_k^2} + C_{\mathfrak{e}}\frac{\mathfrak{s}\log(np)+\zeta}{\kappa_{k+1}^2} \leq C_I(\eta_{k+1}-\eta_k),$$

where $C_I > 1$ is some absolute constant. Therefore, by Lemma 19, we have that for any $k \in \{0,1,\ldots,K\}$

$$|\widehat{\beta}_k - \beta^*_{\mathcal{I}_k}|_2^2 \leq \frac{C_1\mathfrak{s}\log(p)}{\Delta}, \tag{39}$$

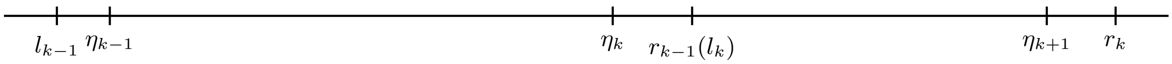
$$|\widehat{\beta}_k - \beta^*_{\mathcal{I}_k}|_1 \leq C_2\mathfrak{s}\sqrt{\frac{\log(p)}{\Delta}}, \tag{40}$$

$$|(\widehat{\beta}_k)_{S^c}|_1 \leq 3|(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}_k})_S|_1, \tag{41}$$

Moreover, by Assumption 4, we have that

$$\Delta \geq C_{snr}\alpha_n(\mathfrak{s}\log(np))^{2/\gamma-1}\max\left\{\kappa_k^{-2}, \kappa_{k-1}^{-1}\kappa_k^{-1}, \kappa_k^{-1}\kappa_{k+1}^{-1}\right\}. \tag{42}$$

By Proposition 13, $[l_{k-1}, r_k) = [l_{k-1}, r_{k-1}) \cup [l_k, r_k)$ can contain 1, 2 or 3 change points. One of the cases that $[l_{k-1}, r_k)$ contains 3 change points is illustrated in the following figure, and the biases will be analysed. The other cases are similar and simpler, and therefore omitted.



In the following, we analyse three types of biases. It follows that

$$\begin{aligned}
&\beta^*_{\eta_{k-1}} - \beta^*_{\mathcal{I}_{k-1}} = \beta^*_{\eta_{k-1}} - \frac{\eta_{k-1} - l_{k-1}}{|\mathcal{I}_{k-1}|}\beta^*_{\eta_{k-2}} - \frac{\eta_k - \eta_{k-1}}{|\mathcal{I}_{k-1}|}\beta^*_{\eta_{k-1}} - \frac{r_{k-1} - \eta_k}{|\mathcal{I}_{k-1}|}\beta^*_{\eta_k} \\
\leq &\frac{\eta_{k-1} - l_{k-1}}{|\mathcal{I}_{k-1}|}(\beta^*_{\eta_{k-1}} - \beta^*_{\eta_{k-2}}) + \frac{r_{k-1} - \eta_k}{|\mathcal{I}_{k-1}|}(\beta^*_{\eta_{k-1}} - \beta^*_{\eta_k}).
\end{aligned}$$

Similarly,

$$\beta^*_{\eta_k} - \beta^*_{\mathcal{I}_k} = \beta^*_{\eta_k} - \frac{\eta_{k+1} - l_k}{|\mathcal{I}_k|}\beta^*_{\eta_k} - \frac{r_k - \eta_{k+1}}{|\mathcal{I}_k|}\beta^*_{\eta_{k+1}} \leq \frac{r_k - \eta_{k+1}}{|\mathcal{I}_k|}(\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}),$$

and

$$\begin{aligned}
&\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_k} \\
= &\frac{\eta_{k-1} - l_{k-1}}{|\mathcal{I}_{k-1}|}\beta^*_{\eta_{k-2}} + \frac{\eta_k - \eta_{k-1}}{|\mathcal{I}_{k-1}|}\beta^*_{\eta_{k-1}} + \frac{r_{k-1} - \eta_k}{|\mathcal{I}_{k-1}|}\beta^*_{\eta_k} - \beta^*_{\eta_k} \\
= &\frac{\eta_{k-1} - l_{k-1}}{|\mathcal{I}_{k-1}|}\left(\beta^*_{\eta_{k-2}} - \beta^*_{\eta_{k-1}} + \beta^*_{\eta_{k-1}} - \beta^*_{\eta_k}\right) + \frac{\eta_k - \eta_{k-1}}{|\mathcal{I}_{k-1}|}\left(\beta^*_{\eta_{k-1}} - \beta^*_{\eta_k}\right).
\end{aligned}$$

Therefore, we have that

$$\left|\beta^*_{\eta_{k-1}} - \beta^*_{\mathcal{I}_{k-1}}\right|_2 \leq 2C_{\mathfrak{e}}\frac{\mathfrak{s}\log(p) + \zeta}{\Delta\kappa^2_{k-1}}\kappa_{k-1} + 2C_{\mathfrak{e}}\frac{\mathfrak{s}\log(p) + \zeta}{\Delta\kappa^2_k}\kappa_k \leq C_3\kappa_k\alpha_n^{-1}, \tag{43}$$

$$\left|\beta^*_{\eta_k} - \beta^*_{\mathcal{I}_k}\right|_2 \leq 2C_{\mathfrak{e}}\frac{\mathfrak{s}\log(p) + \zeta}{\Delta\kappa^2_{k+1}}\kappa_{k+1} \leq C_4\kappa_k\alpha_n^{-1}, \tag{44}$$

and

$$\left|\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_k}\right|_2 \leq 2C_{\mathfrak{e}}\frac{\mathfrak{s}\log(p) + \zeta}{\Delta\kappa^2_{k-1}}(\kappa_{k-1} + \kappa_k) + \kappa_k \leq C_5\kappa_k. \tag{45}$$

Uniform tightness of $\kappa_k^2|\widetilde{\eta}_k - \eta_k|$. Denote $r = \widetilde{\eta}_k - \eta_k$. Without loss of generality, suppose $r \geq 0$. Since $\widetilde{\eta}_k = \eta_k + r$, defined in (17), is the minimizer of $Q_k(\eta)$, it follows that

$$Q_k(\eta_k + r) - Q_k(\eta_k) \leq 0.$$

Observe that

$$\begin{aligned}
Q_k(\eta_k + r) - Q_k(\eta_k) = &\sum_{t=\eta_k}^{\eta_k+r-1}\left\{(y_t - X_t^\top\widehat{\beta}_{k-1})^2 - (y_t - X_t^\top\widehat{\beta}_k)^2\right\} \\
= \sum_{t=\eta_k}^{\eta_k+r-1}\left\{(y_t - X_t^\top\widehat{\beta}_{k-1})^2 - (y_t - X_t^\top\beta^*_{\mathcal{I}_{k-1}})^2\right\} &- \sum_{t=\eta_k}^{\eta_k+r-1}\left\{(y_t - X_t^\top\widehat{\beta}_k)^2 - (y_t - X_t^\top\beta^*_{\mathcal{I}_k})^2\right\}
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{t=\eta_k}^{\eta_k+r-1} \left\{(y_t - X_t^\top \beta^*_{\mathcal{I}_{k-1}})^2 - (y_t - X_t^\top \beta^*_{\eta_{k-1}})^2\right\} - \sum_{t=\eta_k}^{\eta_k+r-1} \left\{(y_t - X_t^\top \beta^*_{\mathcal{I}_k})^2 - (y_t - X_t^\top \beta^*_{\eta_k})^2\right\} \\
&+ \sum_{t=\eta_k}^{\eta_k+r-1} \left\{(y_t - X_t^\top \beta^*_{\eta_{k-1}})^2 - (y_t - X_t^\top \beta^*_{\eta_k})^2\right\} \\
=& I - II + III - IV + V.
\end{aligned}
$$

Therefore, we have that

$$
V \leq -I + II - III + IV \leq |I| + |II| + |III| + |IV|. \tag{46}
$$

If $r \leq 1/\kappa_k^2$ or $r \leq 2$, then there is nothing to show. So for the rest of the argument, for contradiction, assume that

$$
r \geq \frac{1}{\kappa_k^2} \quad \text{and} \quad r \geq 3.
$$

Step 1: the order of magnitude of I. We have that

$$
\begin{aligned}
I =& \sum_{t=\eta_k}^{\eta_k+r-1} (X_t^\top \widehat{\beta}_{k-1} - X_t^\top \beta^*_{\mathcal{I}_{k-1}})^2 - 2(\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} X_t(y_t - X_t^\top \beta^*_{\mathcal{I}_{k-1}}) \\
=& \sum_{t=\eta_k}^{\eta_k+r-1} (X_t^\top \widehat{\beta}_{k-1} - X_t^\top \beta^*_{\mathcal{I}_{k-1}})^2 - 2(\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} X_t X_t^\top (\beta^*_{\eta_k} - \beta^*_{\mathcal{I}_{k-1}}) \\
&- 2(\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} X_t \epsilon_t \\
=& I_1 - 2I_2 - 2I_3.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&|I_1| = \sum_{t=\eta_k}^{\eta_k+r-1} (X_t^\top \widehat{\beta}_{k-1} - X_t^\top \beta^*_{\mathcal{I}_{k-1}})^2 \\
=& (\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}})^\top \left\{ \sum_{t=\eta_k}^{\eta_k+r-1} (X_t X_t^\top - \Sigma) \right\} (\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}}) + r(\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}})^\top \Sigma (\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}}) \\
=& O_p\left(\frac{\mathfrak{s}\log(p)}{\Delta} \left(\sqrt{r\mathfrak{s}\log(p)} + \{\mathfrak{s}\log(p)\}^{1/\gamma} + r \right) \right) \\
=& O_p\left(\alpha_n^{-1} \left\{ \sqrt{r\kappa_k^2} (\mathfrak{s}\log(p))^{-2/\gamma+5/2} + (\mathfrak{s}\log(p))^{-1/\gamma+2} + r\kappa_k^2 (\mathfrak{s}\log(p))^{-2/\gamma+2} \right\} \right),
\end{aligned}
$$

where the third equality follows from Lemma 37 and (41) and (39), and the fourth equality follows from (42) and $\kappa_k < C_\kappa$. Similarly

$$
|I_2| = \left| (\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} X_t X_t^\top (\beta^*_{\eta_k} - \beta^*_{\mathcal{I}_{k-1}}) \right|
$$

$$
\begin{aligned}
\leq & \big|\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}}\big|_1 \Big|(\beta^*_{\eta_k} - \beta^*_{\mathcal{I}_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} (X_t X_t^\top - \Sigma)\Big|_\infty + r\Lambda_{\max}(\Sigma)\big|\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}}\big|_2 \big|\beta^*_{\eta_k} - \beta^*_{\mathcal{I}_{k-1}}\big|_2 \\
= & O_p\Bigg(\big|\beta^*_{\eta_k} - \beta^*_{\mathcal{I}_{k-1}}\big|_2 \mathfrak{s}\sqrt{\frac{\log(p)}{\Delta}}\Big(\sqrt{r\log(p)} + \log^{1/\gamma}(p)\Big) + r\Lambda_{\max}(\Sigma)\sqrt{\frac{\mathfrak{s}\log(p)}{\Delta}}\big|\beta^*_{\eta_k} - \beta^*_{\mathcal{I}_{k-1}}\big|_2\Bigg) \\
= & O_p\Bigg(\alpha_n^{-1/2}\Big\{\sqrt{r\kappa_k^2}(\mathfrak{s}\log(p))^{-1/\gamma+3/2} + \mathfrak{s}^{-1/\gamma+3/2}\log(p) + r\kappa_k^2(\mathfrak{s}\log(p))^{-1/\gamma+1}\Big\}\Bigg),
\end{aligned}
$$

where the second equality follows from Lemma 36 and the last equality follows from (43) and $\kappa_k < C_\kappa$. Moreover,

$$
\begin{aligned}
|I_3| = & \Big|(\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} X_t\epsilon_t\Big| \leq \big|\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}}\big|_1 \Big|\sum_{t=\eta_k}^{\eta_k+r-1} X_t\epsilon_t\Big|_\infty \\
= & O_p\Bigg(\mathfrak{s}\sqrt{\frac{\log(p)}{\Delta}}\Big(\sqrt{r\log(p)} + \log^{1/\gamma}(p)\Big)\Bigg) = O_p\Bigg(\alpha_n^{-1/2}\Big\{\sqrt{r\kappa_k^2}(\mathfrak{s}\log(p))^{-1/\gamma+3/2} + \mathfrak{s}^{-1/\gamma+3/2}\log(p)\Big\}\Bigg),
\end{aligned}
$$

where the second equality follows from Lemma 36 and $\kappa_k < C_\kappa$. Therefore,

$$
\begin{aligned}
|I| = & O_p\Bigg(\alpha_n^{-1}\Big\{\sqrt{r\kappa_k^2}(\mathfrak{s}\log(p))^{-2/\gamma+5/2} + (\mathfrak{s}\log(p))^{-1/\gamma+2} + r\kappa_k^2(\mathfrak{s}\log(p))^{-2/\gamma+2}\Big\} \\
& + \alpha_n^{-1/2}\Big\{\sqrt{r\kappa_k^2}(\mathfrak{s}\log(p))^{-1/\gamma+3/2} + \mathfrak{s}^{-1/\gamma+3/2}\log(p) + r\kappa_k^2(\mathfrak{s}\log(p))^{-1/\gamma+1}\Big\}\Bigg) \\
= & o_p\Big(\sqrt{r\kappa_k^2}\Big) + o_p(1) + o_p(r\kappa_k^2),
\end{aligned} \tag{47}
$$

where the second inequality follows from Assumption 4**b.**

Step 2: the order of magnitude of II. The same arguments for the term I lead to

$$
|II| = o_p\Big(\sqrt{r\kappa_k^2}\Big) + o_p(1) + o_p(r\kappa_k^2). \tag{48}
$$

Step 3: the order of magnitude of III. We have that

$$
\begin{aligned}
III = & \sum_{t=\eta_k}^{\eta_k+r-1} (X_t^\top\beta^*_{\mathcal{I}_{k-1}} - X_t^\top\beta^*_{\eta_{k-1}})^2 - 2(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} X_t(y_t - X_t^\top\beta^*_{\eta_{k-1}}) \\
= & \sum_{t=\eta_k}^{\eta_k+r-1} (X_t^\top\beta^*_{\mathcal{I}_{k-1}} - X_t^\top\beta^*_{\eta_{k-1}})^2 - 2(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} X_tX_t^\top(\beta^*_{\eta_k} - \beta^*_{\eta_{k-1}}) \\
& - 2(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} X_t\epsilon_t \\
= & r(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top\Sigma(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}) + (\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} (X_tX_t^\top - \Sigma)(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}) \\
& - 2(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} X_t\epsilon_t - 2(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k+r-1} (X_tX_t^\top - \Sigma)(\beta^*_{\eta_k} - \beta^*_{\eta_{k-1}})
\end{aligned}
$$

$$
\begin{aligned}
& - 2r(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \Sigma (\beta^*_{\eta_k} - \beta^*_{\eta_{k-1}}) \\
= & III_1 + III_2 - 2III_3 - 2III_4 - 2III_5.
\end{aligned}
$$

For the term III_1, (43) leads to

$$
|III_1| \leq r\Lambda_{\max}(\Sigma) \big|\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}\big|_2^2 \leq C_1 r \kappa_k^2 \alpha_n^{-2}.
$$

For the term III_2, let

$$
z_t = \frac{1}{|\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}|_2^2} (\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \big(X_t X_t^\top - \Sigma\big)(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}).
$$

Note that for some $q > 2$, we have that

$$
\begin{aligned}
& \big\|z_t - z_{t,\{0\}}\big\|_q \\
\leq & \Bigg\| \frac{1}{|\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}|_2^2} (\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \Big\{ X_t (X_t - X_{t,\{0\}})^\top + (X_t - X_{t,\{0\}}) X_{t,\{0\}}^\top \Big\} (\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}) \Bigg\|_q \\
\leq & 2 \sup_{|v|_2 = 1} \big\|v^\top X_1\big\|_{2q} \delta^X_{t,2q},
\end{aligned}
$$

and

$$
\sum_{t=1}^\infty \|z_t - z_{t,\{0\}}\|_q \leq 2 \sup_{|v|_2=1} \big\|v^\top X_1\big\|_{2q} \Delta^X_{0,2q} < \infty.
$$

It holds that

$$
|III_2| = (\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \sum_{t=\eta_k}^{\eta_k + r - 1} \big(X_t X_t^\top - \Sigma\big)(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}) = O_p\big(\alpha_n^{-2} \sqrt{r\kappa_k^2} \{\log(r\kappa_k^2) + 1\}\big),
$$

where the second equality follows from Lemma 27 with $\nu = \kappa_k^2$, (43) and $\kappa_k \leq C_\kappa$.

Next, we consider the term III_3. Let

$$
z_t = \frac{1}{|\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}|_2} (\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top X_t \epsilon_t.
$$

Note that for some $q > 2$, we have that

$$
\begin{aligned}
& \big\|z_t - z_{t,\{0\}}\big\|_q \\
\leq & \Bigg\| \frac{1}{|\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}|_2} (\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}})^\top \Big\{ X_t(\epsilon_t - \epsilon_{t,\{0\}}) + (X_t - X_{t,\{0\}})\epsilon_{t,\{0\}} \Big\} \Bigg\|_q \\
\leq & \sup_{|v|_2=1} \big\|v^\top X_1\big\|_{2q} \delta^\epsilon_{t,2q} + \sup_{|v|_2=1} \big\|\epsilon_1\big\|_{2q} \delta^X_{t,2q},
\end{aligned}
$$

and

$$
\sum_{t=1}^\infty \|z_t - z_{t,\{0\}}\|_q \leq \sup_{|v|_2=1} \big\|v^\top X_1\big\|_{2q} \Delta^\epsilon_{0,2q} + \sup_{|v|_2=1} \big\|\epsilon_1\big\|_{2q} \Delta^X_{0,2q} < \infty.
$$

By (43) and Lemma 27 and letting $\nu = \kappa_k^2$, we have that

$$|III_3| = O_p\big(\alpha_n^{-1}\sqrt{r\kappa_k^2}\{\log(r\kappa_k^2)+1\}\big).$$

For the term III_4, similar arguments for the term III_2 lead to

$$|III_4| = O_p\big(\alpha_n^{-1}\sqrt{r\kappa_k^2}\{\log(r\kappa_k^2)+1\}\big).$$

For the term III_5, we have that

$$|III_5| \leq r\Lambda_{\max}(\Sigma)\big|\beta^*_{\eta_k} - \beta^*_{\eta_{k-1}}\big|_2\big|\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}\big|_2 \leq C\alpha_n^{-1}r\kappa_k^2,$$

where the second inequality follows from (43). Therefore, we have that

$$|III| = o_p\big(\sqrt{r\kappa_k^2}\{\log(r\kappa_k^2)+1\}\big) + o_p(r\kappa_k^2). \tag{49}$$

Step 4: the order of magnitude of IV. The same arguments for the term III lead to

$$|IV| = o_p\big(\sqrt{r\kappa_k^2}\{\log(r\kappa_k^2)+1\}\big) + o_p(r\kappa_k^2). \tag{50}$$

Step 5: lower bound of V. Observe that

$$\begin{aligned}
V =& \sum_{t=\eta_k}^{\eta_k+r-1}\big\{(y_t - X_t^\top\beta^*_{\eta_{k-1}})^2 - (y_t - X_t^\top\beta^*_{\eta_k})^2\big\} \\
=& \sum_{t=\eta_k}^{\eta_k+r-1}(X_t^\top\beta^*_{\eta_{k-1}} - X_t^\top\beta^*_{\eta_k})^2 - 2\sum_{t=\eta_k}^{\eta_k+r-1}(y_t - X_t^\top\beta^*_{\eta_k})(X_t^\top\beta^*_{\eta_{k-1}} - X_t^\top\beta^*_{\eta_k}) \\
=& r(\beta^*_{\eta_{k-1}} - \beta^*_{\eta_k})^\top\Sigma(\beta^*_{\eta_{k-1}} - \beta^*_{\eta_k}) + (\beta^*_{\eta_{k-1}} - \beta^*_{\eta_k})^\top\sum_{t=\eta_k}^{\eta_k+r-1}\big(X_tX_t^\top - \Sigma\big)(\beta^*_{\eta_{k-1}} - \beta^*_{\eta_k}) \\
&- 2(\beta^*_{\eta_{k-1}} - \beta^*_{\eta_k})^\top\sum_{t=\eta_k}^{\eta_k+r-1}\epsilon_tX_t \\
=& V_1 + V_2 - 2V_3.
\end{aligned}$$

Thus, we have that

$$V \geq V_1 - |V_2| - 2|V_3|.$$

For the term V_1, we have that

$$V_1 \geq r\Lambda_{\min}(\Sigma)\big|\beta^*_{\eta_{k-1}} - \beta^*_{\eta_k}\big|_2^2 \geq c_{\min}r\kappa_k^2,$$

where the second inequality follows from Assumption 2.

The analysis of the term V_2 and V_3 are similar to those of III_2 and III_3, and thus omitted. It holds that

$$|V_2| = O_p\big(\sqrt{r\kappa_k^2}\{\log(r\kappa_k^2)+1\}\big) \text{ and } |V_3| = O_p\big(\sqrt{r\kappa_k^2}\{\log(r\kappa_k^2)+1\}\big),$$

Therefore, we have that

$$V \geq Cr\kappa_k^2 + O_p\big(\sqrt{r\kappa_k^2}\{\log(r\kappa_k^2) + 1\}\big). \tag{51}$$

Combining (46), (47), (48), (49), (50) and (51), we have uniformly for all $r \geq 1/\kappa_k^2$ that

$$Cr\kappa_k^2 + O_p\big(\sqrt{r\kappa_k^2}\{\log(r\kappa_k^2) + 1\}\big) \leq o_p\Big(\sqrt{r\kappa_k^2}\Big) + o_p(1) + o_p(r\kappa_k^2),$$

which implies that

$$r\kappa_k^2 = O_p(1),$$

and complete the proofs of **a.1** and **b.1**.

Limiting distributions. For any $k \in \{1, \dots, K\}$, given the end points s_k and e_k and the true coefficients $\beta^*_{\eta_{k-1}}$ and $\beta^*_{\eta_k}$ before and after the change point η_k, we define the function on $\eta \in (s_k, e_k)$ as

$$Q_k^*(\eta) = \sum_{t=s_k}^{\eta-1} (y_t - X_t^\top \beta^*_{\eta_{k-1}})^2 + \sum_{t=\eta}^{e_k-1} (y_t - X_t^\top \beta^*_{\eta_k})^2. \tag{52}$$

Note that

$$V = \sum_{t=\eta_k}^{\eta_k+r-1} \big\{(y_t - X_t^\top \beta^*_{\eta_{k-1}})^2 - (y_t - X_t^\top \beta^*_{\eta_k})^2\big\} = Q^*(\eta_k + r) - Q^*(\eta_k).$$

Due to the uniform tightness of $r\kappa_k^2$, we have that as $n \to \infty$

$$\big|Q(\eta_k + r) - Q(\eta_k) - (Q^*(\eta_k + r) - Q^*(\eta_k))\big| \leq |I| + |II| + |III| + |IV| \xrightarrow{p} 0.$$

Therefore, it is sufficient to show the limiting distributions of $Q^*(\eta_k + r) - Q^*(\eta_k)$ when $n \to \infty$.

Non-vanishing regime. Observe that for $r > 0$, we have that when $n \to \infty$

$$\begin{aligned}
Q_k^*(\eta_k + r) - Q_k^*(\eta_k) &= \sum_{t=\eta_k}^{\eta_k+r-1} \big\{(y_t - X_t^\top \beta^*_{\eta_{k-1}})^2 - (y_t - X_t^\top \beta^*_{\eta_k})^2\big\} \\
&= \sum_{t=\eta_k}^{\eta_k+r-1} \big\{(y_t - X_t^\top \beta^*_{\eta_k} + \Psi_k^\top X_t)^2 - (y_t - X_t^\top \beta^*_{\eta_k})^2\big\} \\
&= \sum_{t=\eta_k}^{\eta_k+r-1} \big\{2\epsilon_t \Psi_k^\top X_t + (\Psi_k^\top X_t)^2\big\} \xrightarrow{\mathcal{D}} \sum_{t=\eta_k}^{\eta_k+r-1} \big\{2\varrho_k \epsilon_t v_k^\top X_t + \varrho_k^2 (v_k^\top X_t)^2\big\}.
\end{aligned}$$

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

$$Q_k^*(\eta_k + r) - Q_k^*(\eta_k) = \sum_{t=\eta_k+r}^{\eta_k-1} \big\{(y_t - X_t^\top \beta^*_{\eta_k})^2 - (y_t - X_t^\top \beta^*_{\eta_{k-1}})^2\big\}$$

$$= \sum_{t=\eta_k+r}^{\eta_k-1} \{(y_t - X_t^\top \beta^*_{\eta_{k-1}} - \Psi_k^\top X_t)^2 - (y_t - X_t^\top \beta^*_{\eta_{k-1}})^2\}$$

$$= \sum_{t=\eta_k+r}^{\eta_k-1} \big\{ -2\epsilon_t \Psi_k^\top X_t + (\Psi_k^\top X_t)^2 \big\} \xrightarrow{\mathcal{D}} \sum_{t=\eta_k+r}^{\eta_k-1} \big\{ -2\varrho_k \epsilon_t v_k^\top X_t + \varrho_k^2 (v_k^\top X_t)^2 \big\}.$$

So Slutsky's theorem and the Argmax (or Argmin) continuous mapping theorem (see 3.2.2 Theorem van der Vaart and Wellner, 1996) lead to

$$\widetilde{\eta}_k - \eta_k \xrightarrow{\mathcal{D}} \arg\min_r P(r),$$

which completes the proof of **a.2**.

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

$$Q_k^*(\eta_k + rm) - Q_k^*(\eta_k) = \sum_{t=\eta_k}^{\eta_k+rm-1} \big\{(y_t - X_t^\top \beta^*_{\eta_{k-1}})^2 - (y_t - X_t^\top \beta^*_{\eta_k})^2\big\}$$

$$= \sum_{t=\eta_k}^{\eta_k+rm-1} \big\{(y_t - X_t^\top \beta^*_{\eta_k} + \Psi_k^\top X_t)^2 - (y_t - X_t^\top \beta^*_{\eta_k})^2\big\}$$

$$= \frac{1}{\sqrt{m}} \sum_{t=\eta_k}^{\eta_k+rm-1} \big\{2\epsilon_t v_k^\top X_t\big\} + \frac{1}{m} \sum_{t=\eta_k}^{\eta_k+rm-1} \big\{(v_k^\top X_t)^2 - v_k^\top \Sigma v_k\big\} + r v_k^\top \Sigma v_k.$$

Note that condition of the functional CLT under functional dependence (Theorem 3 in Wu, 2011) can be verified easily under Assumptions 2 and 3, By the functional CLT, we have that when $n \to \infty$

$$\frac{1}{\sqrt{m}} \sum_{t=\eta_k}^{\eta_k+rm-1} \big\{2\epsilon_t v_k^\top X_t\big\} \xrightarrow{\mathcal{D}} \sigma_\infty(k)\mathbb{B}(r) \text{ and } \frac{1}{m} \sum_{t=\eta_k}^{\eta_k+rm-1} \big\{(v_k^\top X_t)^2 - v_k^\top \Sigma v_k\big\} \xrightarrow{p} 0,$$

where $\mathbb{B}(r)$ is a standard Brownian motion and $\sigma^2_\infty(k)$ is the long-run variance given in (25). Therefore, it holds that when $n \to \infty$

$$Q_k^*(\eta_k + rm) - Q_k^*(\eta_k) \xrightarrow{\mathcal{D}} r\varpi_k + \sigma_\infty(k)\mathbb{B}_1(r).$$

Similarly, for $r < 0$, we have that when $n \to \infty$

$$Q_k^*(\eta_k + rm) - Q_k^*(\eta_k) \xrightarrow{\mathcal{D}} -r\varpi_k + \sigma_\infty(k)\mathbb{B}_2(-r).$$

So Slutsky's theorem and the Argmax (or Argmin) continuous mapping theorem (see 3.2.2 Theorem van der Vaart and Wellner, 1996) lead to

$$\kappa_k^2(\widetilde{\eta}_k - \eta_k) \xrightarrow{\mathcal{D}} \arg\min_r \big\{\varpi_k |r| + \sigma_\infty(k)\mathbb{W}(r)\big\},$$

which completes the proof of **b.2**. $\square$

C Proofs for Section 3.4

C.1 Proof of Lemma 5

Proof of Lemma 5. It follows from Proposition 13 that, $K \leq |\widehat{\mathcal{P}}| \leq 3K$. This combined with Proposition 14 completes the proof. $\square$

Proposition 13. *Under the same conditions in Lemma 5 and letting $\widehat{\mathcal{P}}$ being the solution to Algorithm 1, the following events uniformly hold with probability at least $1 - cn^{-3}$.*

(i) For all intervals $\widehat{\mathcal{I}} = [s, e) \in \widehat{\mathcal{P}}$ containing one and only one true change point η_k, it must be the case that for sufficiently large constant $C_{\mathfrak{e}}$,

$$\min\{e - \eta_k, \eta_k - s\} \leq C_{\mathfrak{e}} \left(\frac{\mathfrak{s} \log(pn) + \zeta}{\kappa_k^2} \right);$$

(ii) for all intervals $\widehat{\mathcal{I}} = [s, e) \in \widehat{\mathcal{P}}$ containing exactly two true change points, say $\eta_k < \eta_{k+1}$, it must be the case that for sufficiently large constant $C_{\mathfrak{e}}$,

$$e - \eta_{k+1} \leq C_{\mathfrak{e}} \left(\frac{\mathfrak{s} \log(pn) + \zeta}{\kappa_{k+1}^2} \right) \quad \text{and} \quad \eta_k - s \leq C_{\mathfrak{e}} \left(\frac{\mathfrak{s} \log(pn) + \zeta}{\kappa_k^2} \right);$$

(iii) for all consecutive intervals $\widehat{\mathcal{I}}$ and $\widehat{\mathcal{J}}$ in $\widehat{\mathcal{P}}$, the interval $\widehat{\mathcal{I}} \cup \widehat{\mathcal{J}}$ contains at least one true change point; and

(iv) no interval $\widehat{\mathcal{I}} \in \widehat{\mathcal{P}}$ contains strictly more than two true change points.

Proposition 14. *Under the same conditions in Lemma 5, with $\widehat{\mathcal{P}}$ being the solution to (1), satisfying $K \leq |\widehat{\mathcal{P}}| \leq 3K$, then with probability at least $1 - cn^{-3}$, it holds that $|\widehat{\mathcal{P}}| = K$.*

Throughout the rest of this section, denote

$$S = \bigcup_{t=1}^{n} S_t.$$

Since there are K change points in $\{\beta_t^*\}_{t=1}^n$, it follows that $|S| \leq \sum_{k=0}^{K} |S_{\eta_k+1}| \leq (K+1)\mathfrak{s}$.

C.1.1 Proof of Proposition 13

In this subsection, we present the four cases of Proposition 13. Throughout this subsection, we assume that all the conditions in Lemma 5 hold.

Lemma 15. *Let $\mathcal{I} = [s, e) \in \widehat{\mathcal{P}}$ be such that $\mathcal{I}$ contains exactly one change point η_k. Then with probability at least $1 - cn^{-5}$, it holds that for sufficiently large constant $C_{\mathfrak{e}}$,*

$$\min\{\eta_k - s, e - \eta_k\} \leq C_{\mathfrak{e}} \left(\frac{\mathfrak{s} \log(pn) + \zeta}{\kappa_k^2} \right).$$

Proof. If either $\eta_k - s \le C_{\mathfrak{e}}\left(\frac{\mathfrak{s}\log(pn)+\zeta}{\kappa_k^2}\right)$ or $e-\eta_k \le C_{\mathfrak{e}}\left(\frac{\mathfrak{s}\log(pn)+\zeta}{\kappa_k^2}\right)$, then there is nothing to show. So assume that

$$\eta_k - s > C_{\mathfrak{e}}\left(\frac{\mathfrak{s}\log(pn)+\zeta}{\kappa_k^2}\right) \quad \text{and} \quad e-\eta_k > C_{\mathfrak{e}}\left(\frac{\mathfrak{s}\log(pn)+\zeta}{\kappa_k^2}\right). \tag{53}$$

For sufficiently large $C_{\mathfrak{e}}$, this implies that

$$\eta_k - s > \zeta \quad \text{and} \quad e-\eta_k > \zeta.$$

Step 1. Denote

$$\mathcal{J}_1 = [s, \eta_k) \quad \text{and} \quad \mathcal{J}_2 = [\eta_k, e).$$

So $|\mathcal{J}_1| \ge \zeta$ and $|\mathcal{J}_2| \ge \zeta$. Thus for any $\ell \in \{1,2\}$

$$\mathcal{H}(\mathcal{J}_\ell) = \sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top \widehat{\beta}_{\mathcal{J}_\ell})^2 \le \sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top \beta^*_{\mathcal{J}_\ell})^2 + C_1\mathfrak{s}\log(pn),$$

where the inequality follows from Lemma 20.

Since $\mathcal{I} \in \widehat{\mathcal{P}}$, it follows that

$$\begin{aligned}\sum_{t\in\mathcal{I}}(y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 &\le \sum_{t\in\mathcal{J}_1}(y_t - X_t^\top \widehat{\beta}_{\mathcal{J}_1})^2 + \sum_{t\in\mathcal{J}_2}(y_t - X_t^\top \widehat{\beta}_{\mathcal{J}_2})^2 + \zeta \\ &\le \sum_{t\in\mathcal{J}_1}(y_t - X_t^\top \beta^*_{\mathcal{J}_1})^2 + \sum_{t\in\mathcal{J}_2}(y_t - X_t^\top \beta^*_{\mathcal{J}_2})^2 + 2C_1\mathfrak{s}\log(pn) + \zeta \end{aligned} \tag{54}$$

Equation (54) gives

$$\begin{aligned}&\sum_{t\in\mathcal{J}_1}(y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 + \sum_{t\in\mathcal{J}_2}(y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 \\ \le &\sum_{t\in\mathcal{J}_1}(y_t - X_t^\top \beta^*_{\mathcal{J}_1})^2 + \sum_{t\in\mathcal{J}_2}(y_t - X_t^\top \beta^*_{\mathcal{J}_2})^2 + 2C_1\mathfrak{s}\log(pn) + \zeta. \end{aligned} \tag{55}$$

Step 2. Equation (55) leads to

$$\begin{aligned}&\sum_{t\in\mathcal{J}_1}\left\{X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_1})\right\}^2 + \sum_{t\in\mathcal{J}_2}\left\{X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_2})\right\}^2 \\ \le &2\sum_{t\in\mathcal{J}_1}\epsilon_t X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_1}) + 2\sum_{t\in\mathcal{J}_2}\epsilon_t X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_2}) + 2C_1\mathfrak{s}\log(pn) + \zeta. \end{aligned} \tag{56}$$

Note that for $\ell \in \{1,2\}$,

$$|(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell})_{S^c}|_1 = |(\widehat{\beta}_{\mathcal{I}})_{S^c}|_1 \le 3|(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}})_S|_1 \le C_3\mathfrak{s}\sqrt{\frac{\log(np)}{|\mathcal{I}|}},$$

where the last two inequalities follows from Lemma 19. So

$$|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell}|_1 = |(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell})_S|_1 + |(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell})_{S^c}|_1 \le \sqrt{\mathfrak{s}}|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell}|_2 + C_3\mathfrak{s}\sqrt{\frac{\log(np)}{|\mathcal{I}|}}. \tag{57}$$

Therefore, for any $\ell = 1, 2$, it holds with probability at least $1 - n^{-5}$ that

$$\begin{aligned}\sum_{t\in\mathcal{J}_\ell} \epsilon_t X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}) \leq & \left|\sum_{t\in\mathcal{J}_\ell} \epsilon_t X_t^\top\right|_\infty |\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_1 \leq C_2\sqrt{\log(np)|\mathcal{J}_\ell|}|\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_1 \\ \leq & C_2\sqrt{\log(np)|\mathcal{J}_\ell|}\left(\sqrt{\mathfrak{s}}|\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_2 + C_3\mathfrak{s}\sqrt{\frac{\log(np)}{|\mathcal{I}|}}\right) \\ \leq & \frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{64}|\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_2^2 + C_4\mathfrak{s}\log(pn), \end{aligned} \tag{58}$$

where the second inequality follows from Lemma 30 and that $|\mathcal{J}_\ell| \geq \zeta$, the third inequality follows from Equation (57), and the last inequality follows from the observation that $|\mathcal{J}_\ell| \leq |\mathcal{I}|$ and the inequality $2ab \leq a^2 + b^2$.

Step 3. For any $\ell = 1, 2$, it holds with probability at least $1 - 2n^{-5}$ that

$$\begin{aligned}&\sum_{t\in\mathcal{J}_\ell} \left\{X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell})\right\}^2 \\ \geq & \frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{2}|\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_2^2 - C_5|\mathcal{J}_\ell|^{1-\gamma}\log(pn)|\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_1^2 \\ \geq & \frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{2}|\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_2^2 - C_6\mathfrak{s}|\mathcal{J}_\ell|^{1-\gamma}\log(pn)|\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_2^2 - C_6|\mathcal{J}_\ell|^{1-\gamma}\log(np)\mathfrak{s}^2\frac{\log(np)}{|\mathcal{I}|} \\ \geq & \frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{4}|\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_2^2 - C_6|\mathcal{J}_\ell|^{1-\gamma}\log(np)\mathfrak{s}^2\frac{\log(np)}{|\mathcal{I}|} \\ \geq & \frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{4}|\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_\ell}|_2^2 - C_7\mathfrak{s}\log(np) \end{aligned} \tag{59}$$

where the first inequality follows from Theorem 32, the second inequality follows from Equation (57), the third inequality follows from the observation that

$$|\mathcal{J}_\ell| \geq \zeta = C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1} \geq C_\zeta(\mathfrak{s}\log(pn))^{1/\gamma},$$

and the last inequality follows from the observation that

$$\frac{|\mathcal{J}_\ell|^{1-\gamma}}{|\mathcal{I}|} \leq |\mathcal{J}_\ell|^{-\gamma} \leq C_\zeta^{-\gamma}(\mathfrak{s}\log(pn))^{-1}.$$

Step 4. Equation (56), Equation (58) and Equation (59) together imply that

$$|\mathcal{J}_1||\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_1}|_2^2 + |\mathcal{J}_2||\widehat{\beta}_\mathcal{I} - \beta^*_{\mathcal{J}_2}|_2^2 \leq C_8\mathfrak{s}\log(np) + \zeta.$$

Observe that

$$\inf_{\beta\in\mathbb{R}^p} |\mathcal{J}_1||\beta - \beta^*_{\mathcal{J}_1}|_2^2 + |\mathcal{J}_2||\beta - \beta^*_{\mathcal{J}_2}|_2^2 = \kappa_k^2\frac{|\mathcal{J}_1||\mathcal{J}_2|}{|\mathcal{I}|} \geq \frac{\kappa_k^2}{2}\min\{|\mathcal{J}_1|, |\mathcal{J}_2|\}.$$

This leads to

$$\frac{\kappa_k^2}{2}\min\{|\mathcal{J}_1|, |\mathcal{J}_2|\} \leq C_8\mathfrak{s}\log(np) + \zeta.$$

Therefore

$$\min\{|\mathcal{J}_1|, |\mathcal{J}_2|\} \leq C_\mathfrak{e}\left(\frac{\mathfrak{s}\log(pn) + \zeta}{\kappa_k^2}\right),$$

which is a contradiction with Equation (53). This immediate gives the desired result. □

Lemma 16. *Let $\mathcal{I} = [s, e) \in \widehat{\mathcal{P}}$ be such that $\mathcal{I}$ contains exactly two change points η_k, η_{k+1}. Then with probability at least $1 - cn^{-5}$, it holds that for sufficiently large constant $C_{\mathfrak{e}}$,*

$$\eta_k - s \leq C_{\mathfrak{e}}\bigg(\frac{\mathfrak{s}\log(np) + \zeta}{\kappa_k^2}\bigg) \quad \text{and} \quad e - \eta_{k+1} \leq C_{\mathfrak{e}}\bigg(\frac{\mathfrak{s}\log(np) + \zeta}{\kappa_{k+1}^2}\bigg).$$

Proof. Denote

$$\mathcal{J}_1 = [s, \eta_k), \quad \mathcal{J}_2 = [\eta_k, \eta_{k+1}) \quad \text{and} \quad \mathcal{J}_3 = [\eta_{k+1}, e).$$

By symmetry, it suffices to show that

$$|\mathcal{J}_1| \leq C_{\mathfrak{e}}\bigg(\frac{\mathfrak{s}\log(np) + \zeta}{\kappa_k^2}\bigg).$$

If $|\mathcal{J}_1| \leq \zeta$, then for sufficiently large $C_{\mathfrak{e}}$, it holds that

$$|\mathcal{J}_1| \leq \zeta \leq C_{\mathfrak{e}}\bigg(\frac{\mathfrak{s}\log(np) + \zeta}{\kappa_k^2}\bigg).$$

So it suffices to assume

$$|\mathcal{J}_1| \geq \zeta.$$

Step 1. Since $\mathcal{I}$ contains two change points, it follows that $|\mathcal{I}| \geq \eta_{k+1} - \eta_k \geq \zeta$. So

$$\mathcal{H}(\mathcal{I}) = \sum_{t \in \mathcal{I}} (y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2.$$

Since for any $\ell \in \{1, 2\}$, $|\mathcal{J}_\ell| \geq \zeta$, it follows with probability at least $1 - n^{-5}$ that

$$\mathcal{H}(\mathcal{J}_\ell) = \sum_{t \in \mathcal{J}_\ell} (y_t - X_t^\top \widehat{\beta}_{\mathcal{J}_\ell})^2 \leq \sum_{t \in \mathcal{J}_\ell} (y_t - X_t^\top \beta^*_{\mathcal{J}_\ell})^2 + C_1 \mathfrak{s}\log(pn)$$

where the inequality follows from Lemma 20. If $|\mathcal{J}_3| \geq \zeta$, then

$$\mathcal{H}(\mathcal{J}_3) = \sum_{t \in \mathcal{J}_3} (y_t - X_t^\top \widehat{\beta}_{\mathcal{J}_3})^2 \leq \sum_{t \in \mathcal{J}_3} (y_t - X_t^\top \beta^*_{\mathcal{J}_3})^2 + C_1 \mathfrak{s}\log(pn),$$

where the inequality follows from Lemma 20. If $|\mathcal{J}_3| \leq \zeta$, then

$$\mathcal{H}(\mathcal{J}_3) = 0 \leq \sum_{t \in \mathcal{J}_3} (y_t - X_t^\top \beta^*_{\mathcal{J}_3})^2 + C_1 \mathfrak{s}\log(pn).$$

So

$$\mathcal{H}(\mathcal{J}_3) \leq \sum_{t \in \mathcal{J}_3} (y_t - X_t^\top \beta^*_{\mathcal{J}_3})^2 + C_1 \mathfrak{s}\log(pn).$$

Step 2. Since $\mathcal{I} \in \widehat{\mathcal{P}}$, we have

$$\mathcal{H}(\mathcal{I}) \leq \mathcal{H}(\mathcal{J}_1) + \mathcal{H}(\mathcal{J}_2) + \mathcal{H}(\mathcal{J}_3) + 2\zeta. \tag{60}$$

The above display and the calculations in **Step 1** imply that

$$\sum_{t\in\mathcal{I}}(y_t - X_t^\top\widehat{\beta}_{\mathcal{I}})^2 \le \sum_{\ell=1}^{3}\sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top\beta^*_{\mathcal{J}_\ell})^2 + 3C_1\mathfrak{s}\log(pn) + 2\zeta. \tag{61}$$

Equation (61) gives

$$\sum_{\ell=1}^{3}\sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top\widehat{\beta}_{\mathcal{I}})^2 \le \sum_{\ell=1}^{3}\sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top\beta^*_{\mathcal{J}_\ell})^2 + 3C_1\mathfrak{s}\log(pn) + 2\zeta. \tag{62}$$

Step 3. Note that for any $\ell \in \{1, 2\}$, $|\mathcal{J}_\ell| \ge \zeta$. So it holds with probability at least $1 - 3n^{-5}$ that

$$\begin{aligned}
&\sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top\widehat{\beta}_{\mathcal{I}})^2 - \sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top\beta^*_{\mathcal{J}_\ell})^2\\
=&\sum_{t\in\mathcal{J}_\ell}\{X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell})\}^2 - 2\sum_{t\in\mathcal{J}_\ell}\epsilon_t X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell})\\
\ge&\frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{4}|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell}|_2^2 - C_2\mathfrak{s}\log(np) - \frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{64}|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell}|_2^2 - C_2'\mathfrak{s}\log(pn)\\
\ge&\frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{8}|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell}|_2^2 - C_3\mathfrak{s}\log(np),
\end{aligned} \tag{63}$$

where the first inequality follows from the same arguments that lead to Equation (59) and Equation (58).

Step 4. If $|\mathcal{J}_3| \ge \zeta$, then the same calculations in **Step 3** give

$$\sum_{t\in\mathcal{J}_3}(y_t - X_t^\top\widehat{\beta}_{\mathcal{I}})^2 - \sum_{t\in\mathcal{J}_3}(y_t - X_t^\top\beta^*_{\mathcal{J}_3})^2 \ge \frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_3|}{8}|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_3}|_2^2 - C_3\mathfrak{s}\log(np). \tag{64}$$

Putting Equation (63) and Equation (64) into Equation (62), it follows that

$$\sum_{\ell=1}^{3}\frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{8}|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell}|_2^2 \le C_4\mathfrak{s}\log(np) + 2\zeta.$$

This leads to

$$|\mathcal{J}_1||\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_1}|_2^2 + |\mathcal{J}_2||\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_2}|_2^2 \le C_4\mathfrak{s}\log(np) + 2\zeta. \tag{65}$$

Observe that

$$\inf_{\beta\in\mathbb{R}^p}|\mathcal{J}_1||\beta - \beta^*_{\mathcal{J}_1}|_2^2 + |\mathcal{J}_2||\beta - \beta^*_{\mathcal{J}_2}|_2^2 = \kappa_k^2\frac{|\mathcal{J}_1||\mathcal{J}_2|}{|\mathcal{J}_1| + |\mathcal{J}_2|} \ge \frac{\kappa_k^2}{2}\min\{|\mathcal{J}_1|, |\mathcal{J}_2|\}.$$

Thus

$$\kappa_k^2\min\{|\mathcal{J}_1|, |\mathcal{J}_2|\} \le C_5\mathfrak{s}\log(np) + 2\zeta,$$

which is

$$\min\{|\mathcal{J}_1|, |\mathcal{J}_2|\} \le \frac{C_5\mathfrak{s}\log(np) + 2\zeta}{\kappa_k^2}.$$

Since $|\mathcal{J}_2| \geq \Delta \geq \frac{C\mathfrak{s}\log(np)+2\alpha_n\zeta}{\kappa_k^2}$ for sufficiently large constant C, it follows that

$$|\mathcal{J}_1| \leq \frac{C_5\mathfrak{s}\log(np)+2\zeta}{\kappa_k^2}.$$

Step 5. If $|\mathcal{J}_3| \leq \zeta$, note that a similar and simpler argument of Lemma 30 gives

$$\mathbb{P}\left(\left|\sum_{t\in\mathcal{J}_3}\epsilon_t^2 - |\mathcal{J}_3|\mathbb{E}(\epsilon_t^2)\right| \leq C_6\left\{\sqrt{|\mathcal{J}_3|\log(n)} + \log^{1/\gamma}(n)\right\}\right) \leq n^{-5}.$$

So with probability at least $1-n^{-5}$, it follows that

$$\sum_{t\in\mathcal{J}_3}\epsilon_t^2 \leq |\mathcal{J}_3|\sigma_\epsilon^2 + C_6\left\{\sqrt{|\mathcal{J}_3|\log(n)} + \log^{1/\gamma}(n)\right\} \leq \frac{\zeta}{4}, \tag{66}$$

where the last inequality follows from that $|\mathcal{J}_3| \leq \zeta$ and $\zeta \geq C_\zeta(\mathfrak{s}\log(pn))^{1/\gamma}$ for sufficient large C_ζ. So

$$\begin{aligned}
&\sum_{t\in\mathcal{J}_3}(y_t - X_t^\top\widehat{\beta}_\mathcal{I})^2 - \sum_{t\in\mathcal{J}_3}(y_t - X_t^\top\beta^*_{\mathcal{J}_3})^2 = \sum_{t\in\mathcal{J}_3}\left\{X_t^\top(\widehat{\beta}_\mathcal{I}-\beta^*_{\mathcal{J}_3})\right\}^2 - 2\sum_{t\in\mathcal{J}_3}\epsilon_t X_t^\top(\widehat{\beta}_\mathcal{I}-\beta^*_{\mathcal{J}_3})\\
\geq&\sum_{i\in\mathcal{J}_3}\left\{X_t^\top(\widehat{\beta}_\mathcal{I}-\beta^*_{\mathcal{J}_3})\right\}^2 - \frac{1}{2}\sum_{t\in\mathcal{J}_3}\left\{X_t^\top(\widehat{\beta}_\mathcal{I}-\beta^*_{\mathcal{J}_3})\right\}^2 - 2\sum_{t\in\mathcal{J}_3}\epsilon_t^2\\
\geq&\frac{1}{2}\sum_{t\in\mathcal{J}_3}\left\{X_t^\top(\widehat{\beta}_\mathcal{I}-\beta^*_{\mathcal{J}_3})\right\}^2 - \frac{\zeta}{2} \geq -\frac{\zeta}{2},
\end{aligned} \tag{67}$$

where the second inequality follows from Equation (66). Putting Equation (63) and Equation (67) into Equation (62), it follows that

$$\sum_{\ell=1}^{2}\frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_\ell|}{8}|\widehat{\beta}_\mathcal{I}-\beta^*_{\mathcal{J}_\ell}|_2^2 \leq C_7\mathfrak{s}\log(np)+C_8\zeta.$$

Similar argument for Equation (65) leads to

$$|\mathcal{J}_1| \leq \frac{C_9\big(\mathfrak{s}\log(np)+\zeta\big)}{\kappa_k^2}.$$

□

Lemma 17. *With probability at least $1-cn^{-3}$, there is no intervals in $\widehat{\mathcal{P}}$ containing three or more true change points.*

Proof. For contradiction, suppose $\mathcal{I} = [s,e) \in \widehat{\mathcal{P}}$ be such that $\{\eta_1,\ldots,\eta_M\} \subset \mathcal{I}$ with $M \geq 3$.

Step 1. Since $\mathcal{I}$ contains three change points, $|\mathcal{I}| \geq \zeta$ and therefore

$$\mathcal{H}(\mathcal{I}) = \sum_{t\in\mathcal{I}}(y_t - X_t^\top\widehat{\beta}_\mathcal{I})^2.$$

Denote

$$\mathcal{J}_1 = [s, \eta_1), \ \mathcal{J}_m = [\eta_{m-1}, \eta_m) \ \text{ for } \ 2 \le m \le M, \ \mathcal{J}_{M+1} = [\eta_M, e).$$

Since for $m \in \{2, \ldots, M\}$, $|\mathcal{J}_m| \ge \zeta$, it follows that

$$\mathcal{H}(\mathcal{J}_m) = \sum_{t \in \mathcal{J}_m} (y_t - X_t^\top \widehat{\beta}_{\mathcal{J}_m})^2 \le \sum_{t \in \mathcal{J}_m} (y_t - X_t^\top \beta^*_{\mathcal{J}_m})^2 + C_1 \mathfrak{s} \log(pn)$$

where the inequality follows from Lemma 20. For $\ell \in \{1, M+1\}$, if $|\mathcal{J}_\ell| \ge \zeta$, then

$$\mathcal{H}(\mathcal{J}_\ell) = \sum_{t \in \mathcal{J}_\ell} (y_t - X_t^\top \widehat{\beta}_{\mathcal{J}_\ell})^2 \le \sum_{t \in \mathcal{J}_\ell} (y_t - X_t^\top \beta^*_{\mathcal{J}_\ell})^2 + C_1 \mathfrak{s} \log(pn),$$

where the inequality follows from Lemma 20. If $|\mathcal{J}_\ell| \le \zeta$, then

$$\mathcal{H}(\mathcal{J}_\ell) = 0 \le \sum_{t \in \mathcal{J}_\ell} (y_t - X_t^\top \beta^*_{\mathcal{J}_\ell})^2 + C_1 \mathfrak{s} \log(pn).$$

So

$$\mathcal{H}(\mathcal{J}_\ell) \le \sum_{t \in \mathcal{J}_\ell} (y_t - X_t^\top \beta^*_{\mathcal{J}_\ell})^2 + C_1 \mathfrak{s} \log(pn).$$

Step 2. Since $\mathcal{I} \in \widehat{\mathcal{P}}$, we have

$$\mathcal{H}(\mathcal{I}) \le \sum_{m=1}^{M+1} \mathcal{H}(\mathcal{J}_m) + M\zeta. \tag{68}$$

The above display and the calculations in **Step 1** and **Step 2** implies that

$$\sum_{t \in \mathcal{I}} (y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 \le \sum_{m=1}^{M+1} \sum_{t \in \mathcal{J}_m} (y_t - X_t^\top \beta^*_t)^2 + (M+1) C_1 \mathfrak{s} \log(pn) + M\zeta. \tag{69}$$

Equation (69) gives

$$\sum_{m=1}^{M+1} \sum_{t \in \mathcal{J}_m} (y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 \le \sum_{m=1}^{M+1} \sum_{t \in \mathcal{J}_m} (y_t - X_t^\top \beta^*_{\mathcal{J}_m})^2 + (M+1) C_1 \mathfrak{s} \log(pn) + M\zeta. \tag{70}$$

Step 3. Observe that $|\mathcal{J}_m| \ge \zeta$ for $m \in \{2, \ldots, M\}$. Using the same argument that leads to Equation (63), it follows that

$$\sum_{t \in \mathcal{J}_m} (y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 - \sum_{t \in \mathcal{J}_m} (y_t - X_t^\top \beta^*_{\mathcal{J}_m})^2 \ge \frac{\Lambda_{\min}(\Sigma) |\mathcal{J}_m|}{8} |\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_m}|_2^2 - C_3 \mathfrak{s} \log(np). \tag{71}$$

Step 4. For $\ell \in \{1, M+1\}$, if $|\mathcal{J}_\ell| \ge \zeta$, then similar to Equation (71),

$$\sum_{t \in \mathcal{J}_\ell} (y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 - \sum_{t \in \mathcal{J}_\ell} (y_t - X_t^\top \beta^*_{\mathcal{J}_\ell})^2 \ge \frac{\Lambda_{\min}(\Sigma) |\mathcal{J}_\ell|}{8} |\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_\ell}|_2^2 - C_3 \mathfrak{s} \log(np) \ge -C_3 \mathfrak{s} \log(np).$$

where the first inequality follows from the same arguments that lead to Equation (59) and Equation (58).
If $|\mathcal{J}_\ell| \le \zeta$, then using the same argument that leads to Equation (67), it follows that

$$\sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 - \sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top \beta^*_{\mathcal{J}_\ell})^2 \ge -2C_4\zeta.$$

Putting the two cases together, it follows that for $\ell \in \{1, M+1\}$,

$$\sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 - \sum_{t\in\mathcal{J}_\ell}(y_t - X_t^\top \beta^*_{\mathcal{J}_\ell})^2 \ge -C_3\mathfrak{s}\log(pn) - 2C_4\zeta. \tag{72}$$

Step 5. Equation (70), Equation (71) and Equation (72) together imply that

$$\sum_{m=2}^{M} \frac{\Lambda_{\min}(\Sigma)|\mathcal{J}_m|}{8}|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_m}|_2^2 \le C_5 M(\mathfrak{s}\log(np) + \zeta). \tag{73}$$

For any $m \in \{2, \ldots, M\}$, it holds that

$$\inf_{\beta\in\mathbb{R}^p} |\mathcal{J}_{m-1}||\beta - \beta^*_{\mathcal{J}_{m-1}}|^2 + |\mathcal{J}_m||\beta - \beta^*_{\mathcal{J}_m}|^2 = \frac{|\mathcal{J}_{m-1}||\mathcal{J}_m|}{|\mathcal{J}_{m-1}| + |\mathcal{J}_m|}\kappa_m^2 \ge \frac{1}{2}\Delta\kappa^2, \tag{74}$$

where the last inequality follows from the assumptions that $\eta_k - \eta_{k-1} \ge \Delta$ and $\kappa_k \ge \kappa$ for all $1 \le k \le K$. So

$$\begin{aligned} 2\sum_{m=2}^{M} |\mathcal{J}_m||\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_m}|_2^2 &\ge \sum_{m=3}^{M}\Big(|\mathcal{J}_{m-1}|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_{m-1}}|_2^2 + |\mathcal{J}_m||\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{J}_m}|_2^2\Big) \\ &\ge (M-2)\frac{1}{2}\Delta\kappa^2 \ge \frac{M}{6}\Delta\kappa^2, \end{aligned} \tag{75}$$

where the second inequality follows from Equation (74) and the last inequality follows from $M \ge 3$. Equation (73) and Equation (75) together imply that

$$\frac{M}{12}\Delta\kappa^2 \le C_5 M\big(\mathfrak{s}\log(np) + \zeta\big). \tag{76}$$

Note that by Assumption 1, we have

$$\Delta\kappa^2 \ge C_{snr}\big(\mathfrak{s}\log(np) + \zeta\big). \tag{77}$$

Equation (76) contradicts Equation (77) for sufficiently large C_{snr}.

□

Lemma 18. *With probability at least $1 - cn^{-3}$, there is no two consecutive intervals $\mathcal{I}_1 = [s,t) \in \widehat{\mathcal{P}}$, $\mathcal{I}_2 = [t,e) \in \widehat{\mathcal{P}}$ such that $\mathcal{I}_1 \cup \mathcal{I}_2$ contains no change points.*

Proof. For contradiction, suppose that

$$\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$$

contains no change points. If $|\mathcal{I}| \le \zeta$, then $\mathcal{H}(\mathcal{I}) = \mathcal{H}(\mathcal{I}_1) = \mathcal{H}(\mathcal{I}_2) = 0$. So

$$\mathcal{H}(\mathcal{I}_1) + \mathcal{H}(\mathcal{I}_2) + \zeta \ge \mathcal{H}(\mathcal{I}).$$

This is a contradiction to the assumption that $\mathcal{I}_1 \in \widehat{\mathcal{P}}$ and $\mathcal{I}_2 \in \widehat{\mathcal{P}}$. For the rest of the proof, assume that $|\mathcal{I}| \ge \zeta$. Therefore, by Lemma 20, it follows that

$$\left| \mathcal{H}(\mathcal{I}) - \sum_{t \in \mathcal{I}} (y_t - X_t^\top \beta_t^*)^2 \right| = \left| \sum_{t \in \mathcal{I}} (y_t - X_t^\top \widehat{\beta}_{\mathcal{I}})^2 - \sum_{t \in \mathcal{I}} (y_t - X_t^\top \beta_{\mathcal{I}}^*)^2 \right| \le C_1 \mathfrak{s} \log(np). \tag{78}$$

Case a. Suppose $|\mathcal{I}_1| \ge \zeta$ and $|\mathcal{I}_2| \ge \zeta$. Let $\ell \in \{1, 2\}$. Then by Lemma 20, it follows that

$$\left| \mathcal{H}(\mathcal{I}_\ell) - \sum_{t \in \mathcal{I}_\ell} (y_t - X_t^\top \beta_t^*)^2 \right| = \left| \sum_{t \in \mathcal{I}_\ell} (y_t - X_t^\top \widehat{\beta}_{\mathcal{I}_\ell})^2 - \sum_{t \in \mathcal{I}_\ell} (y_t - X_t^\top \beta_{\mathcal{I}_\ell}^*)^2 \right| \le C_1 \mathfrak{s} \log(np). \tag{79}$$

So by Equation (78) and (79)

$$\mathcal{H}(\mathcal{I}) \le \sum_{t \in \mathcal{I}} (y_t - X_t^\top \beta_t^*)^2 + C_1 \mathfrak{s} \log(pn) \le \mathcal{H}(\mathcal{I}_1) + \mathcal{H}(\mathcal{I}_2) + 3C_1 \mathfrak{s} \log(pn) \le \mathcal{H}(\mathcal{I}_1) + \mathcal{H}(\mathcal{I}_2) + \zeta.$$

This is a contradiction to the assumption that $\mathcal{I}_1 \in \widehat{\mathcal{P}}$ and $\mathcal{I}_2 \in \widehat{\mathcal{P}}$.

Case b. Suppose $|\mathcal{I}_1| < \zeta$ and $|\mathcal{I}_2| < \zeta$. Then by Lemma 20, it follows that

$$\left| \mathcal{H}(\mathcal{I}_\ell) - \sum_{t \in \mathcal{I}_\ell} (y_t - X_t^\top \beta_t^*)^2 \right| = \left| \mathcal{H}(\mathcal{I}_\ell) - 0 \right| = \left| \sum_{t \in \mathcal{I}_\ell} \epsilon_t^2 \right| \le \frac{\zeta}{4}, \tag{80}$$

where the last inequality follows from Equation (66). So by Equation (78) and (80)

$$\mathcal{H}(\mathcal{I}) \le \sum_{t \in \mathcal{I}} (y_t - X_t^\top \beta_t^*)^2 + C_1 \mathfrak{s} \log(pn) \le \mathcal{H}(\mathcal{I}_1) + \mathcal{H}(\mathcal{I}_2) + C_1 \mathfrak{s} \log(pn) + \frac{\zeta}{2} \le \mathcal{H}(\mathcal{I}_1) + \mathcal{H}(\mathcal{I}_2) + \zeta,$$

where $\zeta \ge C_\zeta(\mathfrak{s} \log(pn))$ is used in the last inequality. This is a contradiction to the assumption that $\mathcal{I}_1 \in \widehat{\mathcal{P}}$ and $\mathcal{I}_2 \in \widehat{\mathcal{P}}$.

Case c. Suppose $|\mathcal{I}_1| < \zeta$ and $|\mathcal{I}_2| \ge \zeta$. Then Equation (79) holds for $\ell = 2$ and Equation (80) holds for $\ell = 1$. So

$$\mathcal{H}(\mathcal{I}) \le \sum_{t \in \mathcal{I}} (y_t - X_t^\top \beta_t^*)^2 + C_1 \mathfrak{s} \log(pn) \le \mathcal{H}(\mathcal{I}_1) + \mathcal{H}(\mathcal{I}_2) + 2C_1 \mathfrak{s} \log(pn) + \frac{\zeta}{4} \le \mathcal{H}(\mathcal{I}_1) + \mathcal{H}(\mathcal{I}_2) + \zeta.$$

This is a contradiction to the assumption that $\mathcal{I}_1 \in \widehat{\mathcal{P}}$ and $\mathcal{I}_2 \in \widehat{\mathcal{P}}$.

Case d. Suppose $|\mathcal{I}_1| \ge \zeta$ and $|\mathcal{I}_2| < \zeta$. This is the same as **Case c.** $\square$

C.1.2 Proof of Proposition 14

Proof of Proposition 14. Denote $S_n^* = \sum_{t=1}^n (y_t - X_t^\top \beta_t^*)^2$. Given any collection $\{t_1, \ldots, t_m\}$, where $t_1 < \cdots < t_m$, and $t_1 = 0$, $t_{m+1} = n$, let

$$S_n(t_1, \ldots, t_m) = \sum_{k=1}^m \mathcal{H}([t_k, t_{k+1})). \tag{81}$$

For any collection of time points, when defining (81), we will assume that the time points are sorted in an increasing order.

Let $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ denote the change points induced by $\widehat{\mathcal{P}}$. It suffices to justify that

$$
\begin{aligned}
S_n^* + K\zeta \geq & S_n(\eta_1, \ldots, \eta_K) + K\zeta - C_a(K+1)\mathfrak{s}\log(pn) & (82)\\
\geq & S_n(\widehat{\eta}_1, \ldots, \widehat{\eta}_{\widehat{K}}) + \widehat{K}\zeta - C_a(K+1)\mathfrak{s}\log(pn) & (83)\\
\geq & S_n(\widehat{\eta}_1, \ldots, \widehat{\eta}_{\widehat{K}}, \eta_1, \ldots, \eta_K) - C_a(K+1)\mathfrak{s}\log(pn) + \widehat{K}\zeta - C_a(K+1)\mathfrak{s}\log(pn) & (84)
\end{aligned}
$$

and that

$$
S_n^* - S_n(\widehat{\eta}_1, \ldots, \widehat{\eta}_{\widehat{K}}, \eta_1, \ldots, \eta_K) \leq C_b(K + \widehat{K} + 2)\mathfrak{s}\log(pn). \tag{85}
$$

This is because if $\widehat{K} \geq K+1$, then

$$
\begin{aligned}
C_b(4K+2)\mathfrak{s}\log(pn) \geq & C_b(K + \widehat{K} + 2)\mathfrak{s}\log(pn)\\
\geq & S_n^* - S_n(\widehat{\eta}_1, \ldots, \widehat{\eta}_{\widehat{K}}, \eta_1, \ldots, \eta_K)\\
\geq & -2C_a(K+1)\mathfrak{s}\log(pn) + (\widehat{K} - K)\zeta\\
\geq & -2C_a(K+1)\mathfrak{s}\log(pn) + \zeta,
\end{aligned}
$$

where the first inequality holds because $|\widehat{\mathcal{P}}| = \widehat{K} \leq 3K$,the second inequality follows from Equation (85), the third inequality follows from Equation (84), and the last inequality holds because $\widehat{K} \geq K+1$. This would imply

$$
\zeta \leq C_b(4K+2)\mathfrak{s}\log(pn) + 2C_a(K+1)\mathfrak{s}\log(pn),
$$

which is a contradiction since $\zeta = C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1} \geq C_\zeta\mathfrak{s}\log(pn)$ for sufficient large C_ζ.

Step 1. Note that (82) is implied by

$$
|S_n^* - S_n(\eta_1, \ldots, \eta_K)| \leq C_a(K+1)\mathfrak{s}\log(pn), \tag{86}
$$

which is an immediate consequence of Lemma 20 and the assumption that $\Delta \geq C_{snr}(\mathfrak{s}\log(pn))^{2/\gamma-1}$.

Step 2. Since $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ are the change points induced by $\widehat{\mathcal{P}}$, (83) holds because $\widehat{\mathcal{P}}$ is a minimiser.

Step 3. For any $\widehat{\mathcal{I}} = (s, e] \in \widehat{\mathcal{P}}$, denote

$$
\widehat{\mathcal{I}} = [s, \eta_{p+1}) \cup \ldots \cup [\eta_{p+q}, e) = \mathcal{J}_1 \cup \ldots \cup \mathcal{J}_{q+1},
$$

where $\{\eta_{p+l}\}_{l=1}^{q+1} = \widehat{\mathcal{I}} \cap \{\eta_k\}_{k=1}^{K}$. Then (84) is an immediate consequence of the following inequality

$$
\mathcal{H}(\widehat{\mathcal{I}}) \geq \sum_{l=1}^{q+1} \mathcal{H}(\mathcal{J}_l) - C_2(q+1)\mathfrak{s}\log(pn). \tag{87}
$$

Note that since $\mathcal{J}_l \subset \widehat{\mathcal{I}}$, it follows that if $|\widehat{\mathcal{I}}| \leq \zeta$, then $\mathcal{H}(\widehat{\mathcal{I}}) = \mathcal{H}(\widehat{\mathcal{J}}_l) = 0$ and Equation (87) trivially holds. Therefore it suffice to assume that

$$
|\widehat{\mathcal{I}}| \geq \zeta = C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1}.
$$

In this case, Equation (87) is implied by

$$\sum_{t\in\widehat{\mathcal{I}}}(y_t - X_t^\top \widehat{\beta}_{\widehat{\mathcal{I}}})^2 \geq \sum_{l=1}^{q+1} H(\mathcal{J}_l) - C_2'(q+1)\mathfrak{s}\log(pn). \tag{88}$$

For any $l \in \{1,\dots,q+1\}$, there are two cases.
Case 1. $|\mathcal{J}_l| \leq \zeta$. Then

$$\mathcal{H}(\mathcal{J}_l) = 0 \leq \sum_{t\in\widehat{\mathcal{I}}}(y_t - X_t^\top \widehat{\beta}_{\widehat{\mathcal{I}}})^2 + C_3\mathfrak{s}\log(pn).$$

Case 2. $|\mathcal{J}_l| > \zeta$. Then

$$\begin{aligned}\mathcal{H}(\mathcal{J}_l) = \sum_{t\in\mathcal{J}_l}(y_t - X_t^\top \widehat{\beta}_{\mathcal{J}_l})^2 &\leq \sum_{t\in\mathcal{J}_l}(y_t - X_t^\top \beta_t^*)^2 + C_4\mathfrak{s}\log(pn) \\ &\leq \sum_{t\in\mathcal{J}_l}(y_t - X_t^\top \widehat{\beta}_{\widehat{\mathcal{I}}})^2 + C_5\mathfrak{s}\log(pn),\end{aligned}$$

where the first inequality follows from Lemma 20, and the second inequality follows from Lemma 21. Therefore for any $l \in \{1,\dots,q+1\}$

$$\mathcal{H}(\mathcal{J}_l) \leq \sum_{t\in\widehat{\mathcal{I}}}(y_t - X_t^\top \widehat{\beta}_{\widehat{\mathcal{I}}})^2 + C_6\mathfrak{s}\log(pn),$$

and this immediately implies Equation (88).

Step 4. Finally, to show (85), observe that from (86), it suffices to show that

$$S_n(\eta_1,\dots,\eta_K) - S_n(\widehat{\eta}_1,\dots,\widehat{\eta}_{\widehat{K}},\eta_1,\dots,\eta_K) \leq C_7(K+\widehat{K}+1)\mathfrak{s}\log(pn).$$

This follows from a similar but simpler argument to **Step 3**. $\square$

C.1.3 Additional results related to DUDP

Given regression data $\{y_t, X_t\}_{t=1}^n$, we study the nonasymptotic properties of the lasso estimator $\widehat{\beta}_{\mathcal{I}}$ defined in (14) for a generic interval $\mathcal{I} \subset (0,n]$.

Lemma 19. *Let $\widehat{\beta}_{\mathcal{I}}$ be defined as in (14) with $\lambda = C_\lambda\sqrt{\log(pn)}$ for some sufficiently large constant $C_\lambda > 0$. Under Assumptions 1, 2 and 3 for any $\mathcal{I} \subset (0,n]$ such that $|\mathcal{I}| \geq \zeta$, it holds uniformly with probability at least $1 - cn^{-5}$ that*

$$|\widehat{\beta}_{\mathcal{I}} - \beta_{\mathcal{I}}^*|_2^2 \leq \frac{C_1\mathfrak{s}\log(pn)}{|\mathcal{I}|}, \tag{89}$$

$$|\widehat{\beta}_{\mathcal{I}} - \beta_{\mathcal{I}}^*|_1 \leq C_2\mathfrak{s}\sqrt{\frac{\log(pn)}{|\mathcal{I}|}}, \tag{90}$$

$$|(\widehat{\beta}_{\mathcal{I}})_{S^c}|_1 \leq 3|(\widehat{\beta}_{\mathcal{I}} - \beta_{\mathcal{I}}^*)_S|_1, \tag{91}$$

where $\beta_{\mathcal{I}}^* = \frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}\beta_t^*$, S is union of the supports of β_t^* for $t \in \mathcal{I}$, i.e. $S = \bigcup_{t\in\mathcal{I}} S_t$, and $C_1, C_2, c > 0$ are absolute constants depending on C_ϵ, D_ϵ, C_X, D_X, C_κ and C_λ.

Lemma 20. *Let $\mathcal{I} = [s, e)$ be any generic interval. If $\mathcal{I}$ contains no change points and that $|\mathcal{I}| \geq \zeta = C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1}$, then it holds that*

$$\mathbb{P}\left(\left|\sum_{t\in\mathcal{I}}(y_t - X_t^\top\widehat{\beta}_\mathcal{I})^2 - \sum_{t\in\mathcal{I}}(y_t - X_t^\top\beta^*_\mathcal{I})^2\right| \geq C\mathfrak{s}\log(np)\right) \leq n^{-5},$$

where $\beta^*_\mathcal{I} = \frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}\beta^*_t$ and $C > 0$ is an absolute constant.

Proof. Note that

$$\begin{aligned}
\left|\sum_{t\in\mathcal{I}}(y_t - X_t^\top\widehat{\beta}_\mathcal{I})^2 - \sum_{t\in\mathcal{I}}(y_t - X_t^\top\beta^*_t)^2\right| =& \left|\sum_{t\in\mathcal{I}}\left\{X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_t)\right\}^2 - 2\sum_{i\in\mathcal{I}}\epsilon_t X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_t)\right| \\
\leq& \sum_{t\in\mathcal{I}}\left\{X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})\right\}^2 \qquad (92)\\
&+ 2\left|\sum_{t\in\mathcal{I}}\epsilon_t X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})\right|. \qquad (93)
\end{aligned}$$

Suppose all the good events in Lemma 19 holds.

Step 1. By Lemma 19, $\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I}$ satisfies the cone condition that

$$|(\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})_{S^c}|_1 \leq 3|(\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})_S|_1.$$

It follows from Theorem 33 that with probability at least $1 - n^{-5}$,

$$\left|\frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}\left\{X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})\right\}^2 - (\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})^\top\Sigma(\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})\right| \leq C_1\sqrt{\frac{\mathfrak{s}\log(pn)}{|\mathcal{I}|}}|\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I}|_2^2.$$

The above display gives

$$\begin{aligned}
\left|\frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}\left\{X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})\right\}^2\right| \leq& \Lambda_{\max}(\Sigma)|\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I}|_2^2 + C_1\sqrt{\frac{\mathfrak{s}\log(pn)}{|\mathcal{I}|}}|\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I}|_2^2 \\
\leq& \Lambda_{\max}(\Sigma)|\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I}|_2^2 + C_1\sqrt{\frac{\mathfrak{s}\log(pn)}{C_\zeta\mathfrak{s}\log(pn)}}|\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I}|_2^2 \\
\leq& \frac{C_2\mathfrak{s}\log(pn)}{|\mathcal{I}|},
\end{aligned}$$

where the second inequality follows from the assumption that $|\mathcal{I}| \geq C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1} \geq C_\zeta\mathfrak{s}\log(pn)$ when $\gamma < 1$, and the last inequality follows from Equation (89) in Lemma 19 and $\Lambda_{\max}(\Sigma) < \infty$. This gives

$$\left|\sum_{t\in\mathcal{I}}\left\{X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})\right\}^2\right| \leq 2C_2\mathfrak{s}\log(pn).$$

Step 2. It holds with probability at least $1 - n^{-5}$ that

$$\left|\sum_{t\in\mathcal{I}}\epsilon_t X_t^\top(\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I})\right| \leq \left|\sum_{t\in\mathcal{I}}\epsilon_t X_t^\top\right|_\infty |\widehat{\beta}_\mathcal{I} - \beta^*_\mathcal{I}|_1 \leq C_3\sqrt{\log(np)|\mathcal{I}|}\mathfrak{s}\sqrt{\frac{\log(pn)}{|\mathcal{I}|}}$$

$$=C_3\mathfrak{s}\log(pn),$$

where the second inequality follows from Lemma 30 with $|\mathcal{I}| \geq C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1}$ and Equation (90) in Lemma 19.

The desired result follows by putting **Step 1** into Equation (92) and **Step 2** into Equation (93). $\square$

Lemma 21. *Under the assumptions and notation in Proposition 14, suppose there exists no true change point in the interval $\mathcal{I}$. For any interval $\mathcal{J} \supset \mathcal{I}$ with*

$$|\mathcal{I}| \geq \zeta = C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1},$$

it holds that with probability at least $1-2n^{-5}$,

$$\sum_{t\in\mathcal{I}}(y_t - X_t^\top \beta^*_\mathcal{I})^2 - \sum_{t\in\mathcal{I}}(y_t - X_t^\top \widehat{\beta}_\mathcal{J})^2 \leq C\mathfrak{s}\log(pn).$$

Proof. Let $\Theta = \beta^*_\mathcal{I} - \widehat{\beta}_\mathcal{J}$. Then with probability at least $1-n^{-5}$, we have

$$\begin{aligned}
\sum_{t\in I}(X_t^\top\Theta)^2 \geq & \frac{\Lambda_{\min}(\Sigma)}{2}|\mathcal{I}||\Theta|_2^2 - C_1\log(pn)|\mathcal{I}|^{1-\gamma}|\Theta|_1^2 \\
\geq & \frac{\Lambda_{\min}(\Sigma)}{2}|\mathcal{I}||\Theta|_2^2 - 2C_1\log(pn)|\mathcal{I}|^{1-\gamma}|\Theta_S|_1^2 - 2C_1\log(pn)|\mathcal{I}|^{1-\gamma}|\Theta_{S^c}|_1^2 \\
\geq & \frac{\Lambda_{\min}(\Sigma)}{2}|\mathcal{I}||\Theta|_2^2 - 2C_1\mathfrak{s}\log(pn)|\mathcal{I}|^{1-\gamma}|\Theta_S|_2^2 - 2C_1\log(pn)|\mathcal{I}|^{1-\gamma}|(\widehat{\beta}_\mathcal{J})_{S^c}|_1^2 \\
\geq & \frac{\Lambda_{\min}(\Sigma)}{4}|\mathcal{I}||\Theta|_2^2 - 2C_1\log(pn)|\mathcal{I}|^{1-\gamma}|(\widehat{\beta}_\mathcal{J})_{S^c}|_1^2
\end{aligned} \tag{94}$$

where the first inequality follows from Theorem 32, the second inequality follows from the inequality that $(|a|+|b|)^2 \leq 2|a|^2+2|b|^2$, the third inequality follows from $(\beta^*_\mathcal{I})_{S^c}=0$ and the last inequality follows from $|\mathcal{I}| \geq C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1} \geq C_\zeta(\mathfrak{s}\log(pn))^{1/\gamma}$ because $\gamma<1$ and that $|\Theta_S|_2 \leq |\Theta|_2$. Since

$$|\mathcal{J}| \geq |\mathcal{I}| \geq C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1},$$

it follows from Lemma 19 that,

$$|(\widehat{\beta}_\mathcal{J})_{S^c}|_1 \leq C_2\mathfrak{s}\sqrt{\frac{\log(pn)}{|\mathcal{I}|}}., \tag{95}$$

So Equation (94) and Equation (95) give

$$\begin{aligned}
\sum_{t\in I}(X_t^\top\Theta)^2 \geq & \frac{\Lambda_{\min}(\Sigma)}{4}|\mathcal{I}||\Theta|_2^2 - 2C_1\log(pn)|\mathcal{I}|^{1-\gamma}C_2^2\mathfrak{s}^2\frac{\log(pn)}{|\mathcal{I}|} \\
\geq & \frac{\Lambda_{\min}(\Sigma)}{4}|\mathcal{I}||\Theta|_2^2 - C_3\frac{\mathfrak{s}^2\log^2(pn)}{|\mathcal{I}|^\gamma} \\
\geq & \frac{\Lambda_{\min}(\Sigma)}{4}|\mathcal{I}||\Theta|_2^2 - C_3\mathfrak{s}\log(pn),
\end{aligned} \tag{96}$$

where the third inequality follows from $|\mathcal{I}| \geq C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1} \geq C_\zeta(\mathfrak{s}\log(pn))^{1/\gamma}$. Similarly

$$
\begin{aligned}
\sum_{t\in I}\epsilon_t X_t^\top\Theta \leq & 2\left|\sum_{t\in I} X_t\epsilon_t\right|_\infty\left(\sqrt{\mathfrak{s}}|\Theta_S|_2 + |(\widehat{\beta}_{\mathcal{J}})_{S^c}|_1\right) \\
\leq & C_4\sqrt{|\mathcal{I}|\log(pn)}\left(\sqrt{\mathfrak{s}}|\Theta|_2 + |(\widehat{\beta}_{\mathcal{J}})_{S^c}|_1\right) \\
\leq & \frac{\Lambda_{\min}(\Sigma)}{8}|\mathcal{I}||\Theta|_2^2 + C_5\mathfrak{s}\log(pn) + C_4\sqrt{|\mathcal{I}|\log(pn)}|(\widehat{\beta}_{\mathcal{J}})_{S^c}|_1 \\
\leq & \frac{\Lambda_{\min}(\Sigma)}{16}|\mathcal{I}||\Theta|_2^2 + C_5\mathfrak{s}\log(pn) + C_6\mathfrak{s}\log(pn),
\end{aligned} \tag{97}
$$

where the second inequality follows from Lemma 30 and that $|\Theta_S|_2 \leq |\Theta|_2$, the third inequality follows from the inequality $2ab \leq a^2 + b^2$, and the last inequality follows from Equation (95). We then have from Equation (96) and Equation (97) that

$$
\begin{aligned}
& \sum_{t\in\mathcal{I}}(y_t - X_t^\top\beta_{\mathcal{I}}^*)^2 - \sum_{t\in\mathcal{I}}(y_t - X_t^\top\widehat{\beta}_{\mathcal{J}})^2 \\
= & 2\sum_{t\in\mathcal{I}}\epsilon_t X_t^\top\Theta - \sum_{t\in\mathcal{I}}(X_t^\top\Theta)^2 \\
\leq & \frac{\Lambda_{\min}(\Sigma)}{8}|\mathcal{I}||\Theta|_2^2 + 2C_5\mathfrak{s}\log(pn) + 2C_6\mathfrak{s}\log(pn) - \frac{\Lambda_{\min}(\Sigma)}{4}|\mathcal{I}||\Theta|_2^2 + C_3\mathfrak{s}\log(pn) \\
\leq & C_7\mathfrak{s}\log(pn),
\end{aligned}
$$

as desired. □

C.2 Proof of Lemma 6

Proof of Lemma 6. Since with probability at least $1 - cn^{-3}$, $\mathcal{E}_{\mathrm{DP}}$ defined in (38) holds, we condition on the $\mathcal{E}_{\mathrm{DP}}$ in the following proof. Let $l_k = \widehat{\eta}_k$ and $r_k = \widehat{\eta}_{k+1}$. We have that

$$
|l_k - \eta_k| \leq C_{\mathfrak{e}}\frac{\mathfrak{s}\log(np) + \zeta}{\kappa_k^2} \quad \text{and} \quad |r_k - \eta_{k+1}| \leq C_{\mathfrak{e}}\frac{\mathfrak{s}\log(np) + \zeta}{\kappa_{k+1}^2}. \tag{98}
$$

By the definitions of the preliminary estimators given in Section 2.1, we have

$$
\widehat{\beta}_k = \arg\min_{\beta\in\mathbb{R}^p}\frac{1}{l_k - r_k}\sum_{t=l_k}^{r_k-1}(y_t - X_t^\top\beta)^2 + \frac{\lambda}{\sqrt{l_k - r_k}}|\beta|_1.
$$

By Assumption 1 and (98), it follows that

$$
l_k - r_k \geq \eta_{k+1} - \eta_k - C_{\mathfrak{e}}\frac{\mathfrak{s}\log(np) + \zeta}{\kappa_k^2} - C_{\mathfrak{e}}\frac{\mathfrak{s}\log(np) + \zeta}{\kappa_{k+1}^2} \geq \frac{\Delta}{2} \geq \zeta. \tag{99}
$$

Therefore, by (89) in Lemma 19,

$$
\left|\widehat{\beta}_k - \frac{1}{l_k - r_k}\sum_{t=l_k}^{r_k-1}\beta_t^*\right|_2^2 \leq C_3\frac{\mathfrak{s}\log(pn)}{l_k - r_k} \leq 2C_3\frac{\mathfrak{s}\log(pn)}{\Delta} \leq C_4\alpha_n^{-1}\kappa_{k+1}^2, \tag{100}
$$

and also

$$\left|\widehat{\beta}_k - \frac{1}{r_k - l_k}\sum_{t=l_k}^{r_k-1}\beta_t^*\right|_2^2 \le C_5\alpha_n^{-1}\kappa_k^2. \tag{101}$$

Moreover, by (90) in Lemma 19,

$$\left|\widehat{\beta}_k - \frac{1}{r_k - l_k}\sum_{t=l_k}^{r_k-1}\beta_t^*\right|_1 \le C_6\mathfrak{s}\sqrt{\frac{\log(pn)}{r_k - l_k}} \le \sqrt{2}C_6\mathfrak{s}\sqrt{\frac{\log(pn)}{\Delta}} \le C_7\mathfrak{s}^{1/2}\alpha_n^{-1/2}\kappa_{k+1}, \tag{102}$$

and also

$$\left|\widehat{\beta}_k - \frac{1}{r_k - l_k}\sum_{t=l_k}^{r_k-1}\beta_t^*\right|_1 \le C_8\mathfrak{s}^{1/2}\alpha_n^{-1/2}\kappa_k. \tag{103}$$

Note that $[l_k, r_k)$ can contain 0, 1 or 2 change points. The case that $[l_k, r_k)$ contains 2 change points is analysed. The other cases are similar and simpler, and therefore omitted. Let η_k and η_{k+1} be the two change points contained in $[l_k, r_k)$, it follows that

$$\begin{aligned}\beta_{\eta_k}^* - \frac{1}{r_k - l_k}\sum_{t=l_k}^{r_k-1}\beta_t^* =& \beta_{\eta_k}^* - \frac{1}{r_k - l_k}\left(\sum_{t=l_k}^{\eta_k-1}\beta_{\eta_{k-1}}^* + \sum_{t=\eta_k}^{\eta_{k+1}-1}\beta_{\eta_k}^* + \sum_{t=\eta_{k+1}}^{r_k-1}\beta_{\eta_{k+1}}^*\right)\\ =& \frac{\eta_{k-1} - l_k}{r_k - l_k}(\beta_{\eta_k}^* - \beta_{\eta_{k-1}}^*) + \frac{r_k - \eta_k}{r_k - l_k}(\beta_{\eta_k}^* - \beta_{\eta_{k+1}}^*).\end{aligned}$$

Therefore

$$\begin{aligned}\left|\beta_{\eta_k}^* - \frac{1}{r_k - l_k}\sum_{t=l_k}^{r_k-1}\beta_t^*\right|_2^2 \le& 2\frac{(\eta_{k-1} - l_k)^2}{(r_k - l_k)^2}(\beta_{\eta_k}^* - \beta_{\eta_{k-1}}^*)^2 + 2\frac{(r_k - \eta_k)^2}{(r_k - l_k)^2}(\beta_{\eta_k}^* - \beta_{\eta_{k+1}}^*)^2\\ \le& 8\frac{(\mathfrak{s}\log(np) + \zeta)^2}{\kappa_k^4\Delta^2}\kappa_k^2 + 8\frac{(\mathfrak{s}\log(np) + \zeta)^2}{\kappa_{k+1}^4\Delta^2}\kappa_{k+1}^2\\ \le& 16\frac{(\mathfrak{s}\log(np) + \zeta)^2}{C_{snr}^2\alpha_n^2(\mathfrak{s}\log(np))^{4/\gamma-2}}\kappa_{k+1}^2\\ =& C_9\alpha_n^{-2}\kappa_{k+1}^2,\end{aligned} \tag{104}$$

where the second inequality follows from (98) and (99), and the third inequality follows from Assumption 1. More specifically, by Assumption 1,

$$\Delta \ge C_{snr}\max_{k\in\{1,\ldots,K\}}\frac{\alpha_n(\mathfrak{s}\log(np))^{2/\gamma-1}}{\kappa_k^2},$$

which further leads to that

$$\Delta^2 \ge C_{snr}^2\frac{\alpha_n^2(\mathfrak{s}\log(np))^{4/\gamma-2}}{\kappa_{k+1}^2\kappa_k^2} \ \text{ and } \ \Delta^2 \ge C_{snr}^2\frac{\alpha_n^2(\mathfrak{s}\log(np))^{4/\gamma-2}}{\kappa_{k+1}^4}.$$

By symmetry, we also have

$$\left|\beta^*_{\eta_k} - \frac{1}{r_k - l_k}\sum_{t=l_k}^{r_k-1}\beta^*_t\right|_2^2 \le C_{10}\alpha_n^{-2}\kappa_k^2, \tag{105}$$

Note that combining (100) and (104) and combining (101) and (105) immediately lead to (i).

Similarly, for the bias, we have that

$$\begin{aligned}\left|\beta^*_{\eta_k} - \frac{1}{r_k - l_k}\sum_{t=l_k}^{r_k-1}\beta^*_t\right|_1 \le & \frac{\eta_{k-1}-l_k}{r_k-l_k}|\beta^*_{\eta_k} - \beta^*_{\eta_{k-1}}|_1 + 2\frac{r_k-\eta_k}{r_k-l_k}|\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}|_1 \\ \le & 2\sqrt{2}\frac{\mathfrak{s}\log(np)+\zeta}{\kappa_k^2\Delta}\mathfrak{s}^{1/2}\kappa_k + 2\sqrt{2}\frac{\mathfrak{s}\log(np)+\zeta}{\kappa_{k+1}^2\Delta}\mathfrak{s}^{1/2}\kappa_{k+1} \\ \le & 4\sqrt{2}\frac{\mathfrak{s}\log(np)+\zeta}{C_{snr}\alpha_n(\mathfrak{s}\log(np))^{2/\gamma-1}}\mathfrak{s}^{1/2}\kappa_{k+1} \\ = & C_{11}\alpha_n^{-1}\mathfrak{s}^{1/2}\kappa_{k+1},\end{aligned} \tag{106}$$

By symmetry, we also have

$$\left|\beta^*_{\eta_k} - \frac{1}{r_k - l_k}\sum_{t=l_k}^{r_k-1}\beta^*_t\right|_1 \le C_{12}\alpha_n^{-1}\mathfrak{s}^{1/2}\kappa_k, \tag{107}$$

Note that combining (102) and (106) and combining (103) and (107) immediately lead to (ii). □

C.3 Proof of Lemma 7

Proof of Lemma 7. By definition,

$$\frac{\widehat{\kappa}_k}{\kappa_k} - 1 = \kappa_k^{-1}|\widehat{\beta}_k - \widehat{\beta}_{k-1}|_2 - \kappa_k^{-1}|\beta^*_{\eta_k} - \beta^*_{\eta_{k-1}}|_2 \le \kappa_k^{-1}|\widehat{\beta}_k - \beta^*_{\eta_k}|_2 + \kappa_k^{-1}|\widehat{\beta}_{k-1} - \beta^*_{\eta_{k-1}}|_2,$$

where the inequality follows from the triangle inequality. Applying Lemma 6 immediately lead to the desired result. □

C.4 Proof of Lemma 19

Lemma 22. *Denote $\beta^*_{\mathcal{I}} = \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\beta^*_i$. Suppose Assumption 1 holds. Let $\mathcal{I}\subset(0,n]$. Then*

$$|\beta^*_{\mathcal{I}} - \beta_i|_2 \le CC_\kappa,$$

for some absolute constant C independent of n.

Proof. It suffices to consider $\mathcal{I} = [1,n]$ and $\beta_i = \beta_0$ as the general case is similar. Let $\Delta_k = \eta_k - \eta_{k-1}$. Observe that

$$\begin{aligned}|\beta^*_{[1,n]} - \beta^*_1|_2 = & \left|\frac{1}{n}\sum_{t=1}^n\beta^*_t - \beta^*_1\right|_2 = \left|\frac{1}{n}\sum_{k=0}^K\Delta_k\beta^*_{\eta_k+1} - \frac{1}{n}\sum_{k=0}^K\Delta_k\beta^*_1\right|_2 \\ \le & \frac{1}{n}\sum_{k=0}^K|\Delta_k(\beta^*_{\eta_k+1} - \beta^*_1)|_2 \le \frac{1}{n}\sum_{k=0}^K\Delta_k(K+1)C_\kappa \le (K+1)C_\kappa.\end{aligned}$$

By Assumption 1, both C_κ and K are finite. □

Proof of Lemma 19. It follows from the definition of $\widehat{\beta}_{\mathcal{I}}$ that

$$\frac{1}{|\mathcal{I}|}\sum_{t\in I}(y_t - X_t^\top\widehat{\beta}_{\mathcal{I}})^2 + \frac{\lambda}{\sqrt{|\mathcal{I}|}}|\widehat{\beta}_{\mathcal{I}}|_1 \le \frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}(y_t - X_t^\top\beta^*_{\mathcal{I}})^2 + \frac{\lambda}{\sqrt{|\mathcal{I}|}}|\beta^*_{\mathcal{I}}|_1.$$

This gives

$$\frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}\left\{X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}})\right\}^2 + \frac{2}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}(y_t - X_t^\top\beta^*_{\mathcal{I}})X_t^\top(\beta^*_{\mathcal{I}} - \widehat{\beta}_{\mathcal{I}}) + \frac{\lambda}{\sqrt{|\mathcal{I}|}}\left|\widehat{\beta}_{\mathcal{I}}\right|_1 \le \frac{\lambda}{\sqrt{|\mathcal{I}|}}|\beta^*_{\mathcal{I}}|_1,$$

and therefore

$$\begin{aligned}
&\frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}\left\{X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}})\right\}^2 + \frac{\lambda}{\sqrt{|\mathcal{I}|}}\left|\widehat{\beta}_{\mathcal{I}}\right|_1 \\
\le&\frac{2}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}\epsilon_t X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}}) + 2(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}})^\top\frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}X_tX_t^\top(\beta^*_t - \beta^*_{\mathcal{I}}) + \frac{\lambda}{\sqrt{|\mathcal{I}|}}\left|\beta^*_{\mathcal{I}}\right|_1 \\
\le&2\left|\frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}\epsilon_t X_t\right|_\infty|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}}|_1 + 2\left|\frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}X_tX_t^\top(\beta^*_{\mathcal{I}} - \beta^*_t)\right|_\infty|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}}|_1 + \frac{\lambda}{\sqrt{|\mathcal{I}|}}\left|\beta^*_{\mathcal{I}}\right|_1. \qquad (108)
\end{aligned}$$

By Lemma 30, with probability at least $1-n^{-5}$, it holds that

$$\left|\frac{1}{|\mathcal{I}|}\sum_{t\in|\mathcal{I}|}X_t\epsilon_t\right|_\infty \le C_3\left(\sqrt{\frac{\log(pn)}{|\mathcal{I}|}} + \frac{\log^{1/\gamma}(np)}{|\mathcal{I}|}\right) \le C_3'\sqrt{\frac{\log(pn)}{|\mathcal{I}|}} \le \frac{\lambda}{8\sqrt{|\mathcal{I}|}},$$

where the second inequality follows from $|\mathcal{I}| \ge C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1}$, and the third inequality follows from the choice of λ. Observe that $\sum_{t\in\mathcal{I}}(\beta^*_{\mathcal{I}} - \beta^*_t) = 0$ and that by Lemma 22, $|\beta^*_{\mathcal{I}} - \beta^*_t|_2 \le C_4C_\kappa$. Therefore by Lemma 31, with probability at east $1-n^{-5}$,

$$\left|\frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}X_tX_t^\top(\beta^*_{\mathcal{I}} - \beta^*_t)\right|_\infty \le C_4\left(\sqrt{\frac{\log(pn)}{|\mathcal{I}|}} + \frac{\log^{1/\gamma}(np)}{|\mathcal{I}|}\right) \le C_4'\sqrt{\frac{\log(pn)}{|\mathcal{I}|}} \le \frac{\lambda}{8\sqrt{|\mathcal{I}|}},$$

where the second inequality follows from $|\mathcal{I}| \ge C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1}$, and the third inequality follows from the choice of λ.

So (108) gives

$$\frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}\left\{X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}})\right\}^2 + \frac{\lambda}{\sqrt{|\mathcal{I}|}}\left|\widehat{\beta}_{\mathcal{I}}\right|_1 \le \frac{\lambda}{2\sqrt{|\mathcal{I}|}}|\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}}|_1 + \frac{\lambda}{\sqrt{|\mathcal{I}|}}\left|\beta^*_{\mathcal{I}}\right|_1.$$

The above inequality and the fact that $|\beta|_1 = |\beta_S|_1 + |\beta_{S^c}|_1$ for any $\beta \in \mathbb{R}^p$ imply that

$$\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\left\{X_t^\top(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}})\right\}^2 + \frac{\lambda}{2\sqrt{|\mathcal{I}|}}|(\widehat{\beta}_{\mathcal{I}})_{S^c}|_1 \le \frac{3\lambda}{2\sqrt{|\mathcal{I}|}}|(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}})_S|_1. \qquad (109)$$

Let $\Theta = \widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}}$. Equation (109) and the observation $(\beta^*_t)_{S^c} = 0$ also imply that

$$\frac{\lambda}{2}|\Theta_{S^c}|_1 = \frac{\lambda}{2}|(\widehat{\beta}_{\mathcal{I}})_{S^c}|_1 \le \frac{3\lambda}{2}|(\widehat{\beta}_{\mathcal{I}} - \beta^*_{\mathcal{I}})_S|_1 = \frac{3\lambda}{2}|\Theta_S|_1.$$

The above inequality gives

$$|\Theta_{S^c}|_1 \leq 3|\Theta_S|_1.$$

Thus Theorem 33 gives

$$\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\left(X_t^\top\Theta\right)^2 \geq \frac{\Lambda_{\min}(\Sigma)}{2}|\Theta|_2^2.$$

Therefore

$$\frac{\Lambda_{\min}(\Sigma)}{2}|\Theta|_2^2 + \frac{\lambda}{2\sqrt{|\mathcal{I}|}}|\Theta_{S^c}|_1 = \frac{\Lambda_{\min}(\Sigma)}{2}|\Theta|_2^2 + \frac{\lambda}{2\sqrt{|\mathcal{I}|}}|(\widehat{\beta}_{\mathcal{I}})_{S^c}|_1 \leq \frac{3\lambda}{2\sqrt{|\mathcal{I}|}}|\Theta_S|_1 \leq \frac{3\lambda\sqrt{\mathfrak{s}}}{2\sqrt{|\mathcal{I}|}}|\Theta|_2, \tag{110}$$

where the first equality follows from $\Theta_{S^c} = (\widehat{\beta}_{\mathcal{I}})_{S^c}$, the first inequality follows from Equation (109), and the second inequality follows from the observation that $|S| \leq K\mathfrak{s}$ and $K < \infty$. Note that Equation (110) leads to

$$|\Theta|_2 \leq \frac{C_5\lambda\sqrt{\mathfrak{s}}}{\sqrt{|\mathcal{I}|}}. \tag{111}$$

This and the fact that $|\Theta_S|_1 \leq \sqrt{|S|}|\Theta|_2 \leq \sqrt{K\mathfrak{s}}|\Theta|_2$ also implies that $|\Theta_S|_1 \leq \frac{C_6\lambda\mathfrak{s}}{\sqrt{|\mathcal{I}|}}$. Since $|\Theta_{S^c}|_1 \leq 3|\Theta_S|_1$, it also holds that

$$|\Theta|_1 = |\Theta_S|_1 + |\Theta_{S^c}|_1 \leq 4|\Theta_S|_1 \leq \frac{4C_6\lambda\mathfrak{s}}{\sqrt{|\mathcal{I}|}}.$$

□

D Proof for Section 4

For $t \in [s_k, e_k)$, the corrected interval given in (18), recall that

$$Z_t^* = \begin{cases} 2\epsilon_t\Psi_k^\top X_t - (\Psi_k^\top X_t)^2, & s_k \leq t \leq \eta_k - 1, \\ 2\epsilon_t\Psi_k^\top X_t + (\Psi_k^\top X_t)^2, & \eta_k \leq t \leq e_k - 1. \end{cases}$$

Given the unobserved process $\{Z_t^*\}_{t=s_k}^{e_k-1}$, we define the auxiliary long-run variance estimator as follows.

Step 1. Divide interval $[s_k, e_k)$ into $2R$ integer intervals with a size S, and denote the $2R$ intervals as $\mathcal{S}_1, \ldots, \mathcal{S}_{2R}$.

Step 2. Compute the rescaled mean differences of Z_t^*'s from the odd interval $\mathcal{S}_{2i-1}$ and the even interval $\mathcal{S}_{2i}$, defined by

$$D_i^* = \frac{1}{\sqrt{2S}}\left\{\sum_{t\in\mathcal{S}_{2i-1}} Z_t^* - \sum_{t\in\mathcal{S}_{2i}} Z_t^*\right\}.$$

Step 3. Compute the auxiliary long-run variance estimator based on the unobservable process $\{Z_t^*\}_{s_k+1}^{e_k}$, which is defined as

$$\breve{\sigma}_\infty^2 = \frac{1}{R}\sum_{i=1}^{R}(\kappa_k^{-1}D_i^*)^2. \tag{112}$$

Lemma 23. *Let $\breve{\sigma}_\infty^2$ be the auxiliary long-run variance estimator defined in (112). Suppose assumptions 1, 2, 3 hold, and assume $S, R \to \infty$ and $R \gg S$ as $n \to \infty$. It holds that*

$$\left|\mathbb{E}[\breve{\sigma}_\infty^2 - \sigma_\infty^2]\right| \to 0.$$

Proof of Lemma 23. We consider the rescaled process $\{\kappa_k^{-1}D_i^*\}_{i\in\mathbb{Z}}$. For notational convenience, denote

$$F_t = v_k^\top X_t X_t^\top v_k - v_k^\top \Sigma v_k \ \text{ and } \ G_t = \epsilon_t v_k^\top X_t.$$

Note that $\{F_t\}_{t\in\mathbb{Z}}$ and $\{G_t\}_{t\in\mathbb{Z}}$ are stationary, and $\mathbb{E}[F_t] = 0$, $\mathbb{E}[G_t] = 0$ and $\mathbb{E}[F_{t_1}G_{t_2}] = 0$ for any $t_1, t_2 \in \mathbb{Z}$. For $l \in \mathbb{Z}$, denote the lag-l autocovariances of $\{F_t\}_{t\in\mathbb{Z}}$ and $\{G_t\}_{t\in\mathbb{Z}}$ respectively as

$$\gamma_l^F = \mathbb{E}[F_{t-l}F_t] \ \text{ and } \ \gamma_l^G = \mathbb{E}[G_{t-l}G_t].$$

Note that $\gamma_l^F = \gamma_{-l}^F$ and $\gamma_l^G = \gamma_{-l}^G$.

Case 1. If there is no change point in $\mathcal{S}_{2i-1}$ and $\mathcal{S}_{2i}$, the mean can be fully cancelled out. More specifically, for any i such that $s_k + 2iS < \eta_k$, it follows that

$$D_i^* = -\frac{\kappa_k^2}{\sqrt{2S}}\left\{\sum_{t\in\mathcal{S}_{2i-1}} F_t - \sum_{t\in\mathcal{S}_{2i}} F_t\right\} + \sqrt{\frac{2}{S}}\kappa_k\left\{\sum_{t\in\mathcal{S}_{2i-1}} G_t - \sum_{t\in\mathcal{S}_{2i}} G_t\right\},$$

and for any i such that $s_k + (2i-1)S \geq \eta_k$ that

$$D_i^* = \frac{\kappa_k^2}{\sqrt{2S}}\left\{\sum_{t\in\mathcal{S}_{2i-1}} F_t - \sum_{t\in\mathcal{S}_{2i}} F_t\right\} + \sqrt{\frac{2}{S}}\kappa_k\left\{\sum_{t\in\mathcal{S}_{2i-1}} G_t - \sum_{t\in\mathcal{S}_{2i}} G_t\right\}.$$

For both of the two above settings, we have that

$$\begin{aligned}\mathbb{E}[(D_i^*)^2] =& \kappa_k^4\gamma_0^F + \kappa_k^4\sum_{l=1}^{S}\left(2 - \frac{3l}{S}\right)\gamma_l^F - \kappa_k^4\sum_{l=S+1}^{2S-1}\left(2 - \frac{l}{S}\right)\gamma_l^F \\ &+ 4\kappa_k^2\gamma_0^G + 4\kappa_k^2\sum_{l=1}^{S}\left(2 - \frac{3l}{S}\right)\gamma_l^G - 4\kappa_k^2\sum_{l=S+1}^{2S-1}\left(2 - \frac{l}{S}\right)\gamma_l^G.\end{aligned}$$

Case 2. If there is a change point $\eta_k \in \mathcal{S}_{2i-1}$ or $\eta_k \in \mathcal{S}_{2i}$, the mean cannot be fully cancelled out, but as we required $R \gg S$ as $n \to \infty$, the bias due to change point is negligible. If $\eta_k \in \mathcal{S}_{2i-1}$, we have that

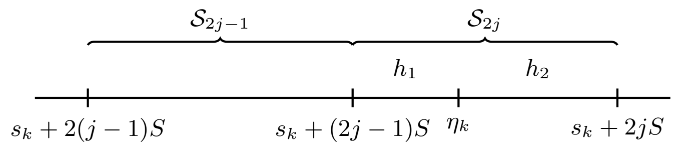
$$D_i^* = -\frac{\kappa_k^2}{\sqrt{2S}}\left\{\sum_{t\in[s_k+\{2i-2\}S,\eta_k)} F_t - \sum_{t\in[\eta_k, s_k+\{2i-1\}S)} F_t + \sum_{t\in\mathcal{S}_{2i}} F_t\right\}$$

$$
+ \sqrt{\frac{2}{S}}\kappa_k\bigg\{\sum_{t\in\mathcal{S}_{2i-1}} G_t - \sum_{t\in\mathcal{S}_{2i}} G_t\bigg\} - \frac{\sqrt{2}\{\eta_k - s_k - (2i-2)S\}}{\sqrt{S}}\kappa_k^2 v_k^\top \Sigma v_k.
$$

If $\eta_k \in \mathcal{S}_{2i}$, we have that

$$
\begin{aligned}
D_i^* = & -\frac{\kappa_k^2}{\sqrt{2S}}\bigg\{\sum_{t\in\mathcal{S}_{2i-1}} F_t - \sum_{t\in[s_k+\{2i-1\}S,\eta_k)} F_t + \sum_{t\in[\eta_k, s_k+2iS)} F_t\bigg\} \\
& + \sqrt{\frac{2}{S}}\kappa_k\bigg\{\sum_{t\in\mathcal{S}_{2i-1}} G_t - \sum_{t\in\mathcal{S}_{2i}} G_t\bigg\} - \frac{\sqrt{2}(\eta_k - s_k - 2iS)}{\sqrt{S}}\kappa_k^2 v_k^\top \Sigma v_k.
\end{aligned}
$$

Without loss of generality, we focus on the **Case 2**, i.e. the case that $\eta_k \in \mathcal{S}_{2j}$ for some $j \in \{1, \ldots, R\}$. Let $h_1 = \eta_k - s_k - (2j-1)S$ and $h_2 = s_k + 2jS - \eta_k$. Assume $h_2 \geq h_1$. The following Figure depicts the definitions of h_1 and h_2.



It follows that

$$
\begin{aligned}
\mathbb{E}[(D_j^*)^2] = & \kappa_k^4\gamma_0^F + \kappa_k^4\sum_{l=1}^{h_1}\Big(2 - \frac{5l}{S}\Big)\gamma_l^F + \kappa_k^4\sum_{l=h_1+1}^{h_2}\Big(2 - \frac{4h_1 + l}{S}\Big)\gamma_l^F + \kappa_k^4\sum_{l=h_2+1}^{S+h_1}\frac{l - 2h_1}{S}\gamma_l^F \\
& + \kappa_k^4\sum_{l=S+h_1+1}^{2S-1}\Big(2 - \frac{l}{S}\Big)\gamma_l^F + 4\kappa_k^2\gamma_0^G + 4\kappa_k^2\sum_{l=1}^{S}\Big(2 - \frac{3l}{S}\Big)\gamma_l^G - 4\kappa_k^2\sum_{l=S+1}^{2S-1}\Big(2 - \frac{l}{S}\Big)\gamma_l^G \\
& + \frac{2h_2^2}{S}\kappa_k^2\big(v_k^\top\Sigma v_k\big)^2 \\
= & \kappa_k^4\gamma_0^F + \kappa_k^4\sum_{l=1}^{S}\Big(2 - \frac{3l}{S}\Big)\gamma_l^F - \kappa_k^4\sum_{l=S+1}^{2S-1}\Big(2 - \frac{l}{S}\Big)\gamma_l^F + 4\kappa_k^2\gamma_0^G + 4\kappa_k^2\sum_{l=1}^{S}\Big(2 - \frac{3l}{S}\Big)\gamma_l^G \\
& - 4\kappa_k^2\sum_{l=S+1}^{2S-1}\Big(2 - \frac{l}{S}\Big)\gamma_l^G - \kappa_k^4\sum_{l=1}^{h_1}\frac{2l}{S}\gamma_l^F - \kappa_k^4\sum_{l=h_1+1}^{h_2}\frac{4h_1 - 2l}{S}\gamma_l^F + \kappa_k^4\sum_{l=h_2+1}^{S}\Big(\frac{4l - 2h_1}{S} - 2\Big)\gamma_l^F \\
& + \kappa_k^4\sum_{l=S+1}^{S+h_1}\Big(2 - \frac{2h_1}{S}\Big)\gamma_l^F + \kappa_k^4\sum_{l=S+h_1+1}^{2S-1}\Big(4 - \frac{2l}{S}\Big)\gamma_l^F + \frac{2h_2^2}{S}\kappa_k^2\big(v_k^\top\Sigma v_k\big)^2.
\end{aligned}
$$

For any $l > 0$, we construct the coupled random variables, which are independence of F_{t-l} and G_{t-l}, as

$$
F_{t,\{t-l,-\infty\}} = v_k^\top X_{t,\{t-l,-\infty\}} X_{t,\{t-l,-\infty\}}^\top v_k - v_k^\top \Sigma v_k
$$

and

$$
G_{t,\{t-l,-\infty\}} = \epsilon_{t,\{t-l,-\infty\}} v_k^\top X_{t,\{t-l,-\infty\}}.
$$

Hence, it follows that

$$
\begin{aligned}
|\gamma_l^F| =& \big|\mathbb{E}[F_{t-l}(F_t - F_{t,\{t-l,-\infty\}})]\big| \le \big\|(v_k^\top X_{t-l})^2\big\|_2 \big\|(v_k^\top X_t)^2 - (v_k^\top X_{t,\{t-l,-\infty\}})^2\big\|_2 \\
\le& 2\big\|v_k^\top X_t\big\|_4^3 \big\|v_k^\top (X_t - X_{t,\{t-l,-\infty\}})\big\|_4 \le 2 \sup_{|v|_2=1} \|v^\top X_t\|_4^3 D_X \exp(-cl^{\gamma_1}) \\
=& K_F \exp(-cl^{\gamma_1}),
\end{aligned}
$$

where the first two inequalities follow from the triangle and Hölder's inequalities and the third inequality follows from Assumption 2. Similarly, by Assumption 2 and Assumption 3, we have that

$$
\begin{aligned}
|\gamma_l^G| =& \big|\mathbb{E}[G_{t-l}(G_t - G_{t,\{t-l,-\infty\}})]\big| \le \big\|\epsilon_t v_k^\top X_t\big\|_2 \big\|\epsilon_t v_k^\top X_t - \epsilon_{t,\{t-l,-\infty\}} v_k^\top X_{t,\{t-l,-\infty\}}\big\|_2 \\
\le& \|\epsilon_t\|_4 \big\|v_k^\top X_t\big\|_4 \big(\|\epsilon_t - \epsilon_{t,\{t-l,-\infty\}}\|_4 \big\|v_k^\top X_t\big\|_4 + \|\epsilon_t\|_4 \big\|v_k^\top (X_t - X_{t,\{t-l,-\infty\}})\big\|_4\big) \\
\le& \|\epsilon_t\|_4 \sup_{|v|_2=1} \|v^\top X_t\|_4 \Big(\|\epsilon_t\|_4 D_X + \sup_{|v|_2=1} \|v^\top X_t\|_4 D_\epsilon\Big) \exp(-cl^{\gamma_1}) \\
=& K_G \exp(-cl^{\gamma_1}).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\kappa_k^2 \big|\mathbb{E}[\check{\sigma}_\infty^2 - \sigma_\infty^2]\big| = \left| \frac{1}{R} \sum_{i \ne j} \mathbb{E}[(D_i^*)^2] + \frac{\mathbb{E}[(D_j^*)^2]}{R} - \kappa_k^4 \sum_{l=-\infty}^{\infty} \gamma_l^F - 4\kappa_k^2 \sum_{l=-\infty}^{\infty} \gamma_l^G \right| \\
=& \Bigg| -\kappa_k^4 \sum_{l=1}^{S} \frac{3l}{S} \gamma_l^F - \kappa_k^4 \sum_{l=S+1}^{2S-1} \Big(4 - \frac{l}{S}\Big) \gamma_l^F - 4\kappa_k^2 \sum_{l=1}^{S} \frac{3l}{S} \gamma_l^G - 4\kappa_k^2 \sum_{l=S+1}^{2S-1} \Big(4 - \frac{l}{S}\Big) \gamma_l^G \\
&- \kappa_k^4 \sum_{l=1}^{h_1} \frac{2l}{RS} \gamma_l^F - \kappa_k^4 \sum_{l=h_1+1}^{h_2} \frac{4h_1 - 2l}{RS} \gamma_l^F + \kappa_k^4 \sum_{l=h_2+1}^{S} \Big(\frac{4l - 2h_1}{RS} - \frac{2}{R}\Big) \gamma_l^F + \kappa_k^4 \sum_{l=S+1}^{S+h_1} \Big(\frac{2}{R} - \frac{2h_1}{RS}\Big) \gamma_l^F \\
&+ \kappa_k^4 \sum_{l=S+h_1+1}^{2S-1} \Big(\frac{4}{R} - \frac{2l}{RS}\Big) \gamma_l^F + \frac{2h_2^2}{RS} \kappa_k^2 \big(v_k^\top \Sigma v_k\big)^2 - 2\kappa_k^4 \sum_{l \ge 2S} \gamma_l^F - 8\kappa_k^2 \sum_{l \ge 2S} \gamma_l^G \Bigg| \\
\le& \frac{3\kappa_k^4}{S} \sum_{l=1}^{S} l|\gamma_l^F| + 3\kappa_k^4 \sum_{l=S+1}^{2S-1} |\gamma_l^F| + \frac{12\kappa_k^2}{S} \sum_{l=1}^{S} l|\gamma_l^G| + 12\kappa_k^2 \sum_{l=S+1}^{2S-1} |\gamma_l^G| + \frac{2\kappa_k^4}{RS} \sum_{l=1}^{h_1} l|\gamma_l^F| + \frac{2\kappa_k^4}{R} \sum_{l=h_1+1}^{2S-1} |\gamma_l^F| \\
&+ \frac{2S}{R} \kappa_k^2 \big(v_k^\top \Sigma v_k\big)^2 + 2\kappa_k^4 \sum_{l \ge 2S} |\gamma_l^F| + 8\kappa_k^2 \sum_{l \ge 2S} |\gamma_l^G| \\
\le& \frac{3}{S} \kappa_k^4 \sum_{l=1}^{S} l|\gamma_l^F| + 3\kappa_k^4 \sum_{l=S+1}^{\infty} |\gamma_l^F| + \frac{2\kappa_k^4}{R} \sum_{l=1}^{2S-1} |\gamma_l^F| + \frac{12\kappa_k^2}{S} \sum_{l=1}^{S} l|\gamma_l^G| + 12\kappa_k^2 \sum_{l=S+1}^{\infty} |\gamma_l^G| + \frac{2S}{R} \kappa_k^2 (\Lambda_{\max}(\Sigma))^2 \\
\le& \frac{3K_F\kappa_k^4 + 12K_G\kappa_k^2}{S} \sum_{l=1}^{S} l \exp(-cl^{\gamma_1}) + (3K_F\kappa_k^4 + 12K_G\kappa_k^2) \sum_{l=S+1}^{\infty} \exp(-cl^{\gamma_1}) + \frac{2K_F\kappa_k^4}{R} \sum_{l=1}^{2S-1} \exp(-cl^{\gamma_1}) \\
&+ \frac{2S}{R} \kappa_k^2 (\Lambda_{\max}(\Sigma))^2 \\
\le& C_1 \frac{K_F\kappa_k^4 + 4K_G\kappa_k^2}{S} + C_2 \frac{K_F\kappa_k^4}{R} + \frac{2S}{R} \kappa_k^2 (\Lambda_{\max}(\Sigma))^2, \qquad (113)
\end{aligned}
$$

where the last inequality follows from the fact that, for any $\nu > 0$,

$$\sup_{S\geq 1} S^{\nu} \sum_{l=S}^{\infty} \exp(-cl^{\gamma_1}) \leq \sum_{l=0}^{\infty} l^{\nu} \exp(-cl^{\gamma_1}) < \infty.$$

Therefore, let $S, R \to \infty$ and $R \gg S$ as $n \to \infty$, it holds that

$$\left|\mathbb{E}[\breve{\sigma}_{\infty}^2 - \sigma_{\infty}^2]\right| \to 0. \tag{114}$$

$\square$

Proof of Theorem 8. Since with probability at least $1-cn^{-3}$, $\mathcal{E}_{\mathrm{DP}}$ defined in (38) holds, we condition on the $\mathcal{E}_{\mathrm{DP}}$ in the following proof. It follows by definition that

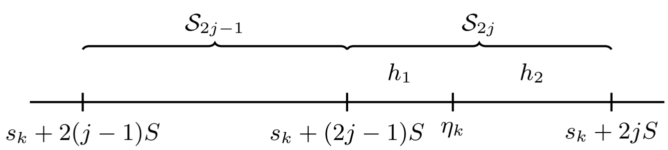
$$\begin{aligned} & Z_t - Z_t^* \\ =& \begin{cases} -2\widehat{\nu}_{k+1}^{\top} X_t X_t^{\top} \Psi_k + (\widehat{\nu}_k - \widehat{\nu}_{k+1})^{\top} X_t X_t^{\top} (\widehat{\nu}_k + \widehat{\nu}_{k+1}) + 2\epsilon_t(\widehat{\nu}_{k+1} - \widehat{\nu}_k)^{\top} X_t, & s_k \leq t \leq \eta_k - 1, \\ -2\widehat{\nu}_{k}^{\top} X_t X_t^{\top} \Psi_k + (\widehat{\nu}_k - \widehat{\nu}_{k+1})^{\top} X_t X_t^{\top} (\widehat{\nu}_k + \widehat{\nu}_{k+1}) + 2\epsilon_t(\widehat{\nu}_{k+1} - \widehat{\nu}_k)^{\top} X_t, & \eta_k \leq t \leq e_k - 1, \end{cases} \end{aligned}$$

where we denote $\widehat{\nu}_k = \widehat{\beta}_k - \beta_{\eta_k}^*$ and $\widehat{\nu}_{k+1} = \widehat{\beta}_{k+1} - \beta_{\eta_{k+1}}^*$.

Recall the estimated jump size $\widehat{\kappa}_k$ defined in (30). Some algebra shows that

$$\begin{aligned} |\widehat{\kappa}_k^2 \widehat{\sigma}_{\infty}^2 - \kappa_k^2 \breve{\sigma}_{\infty}^2| \leq & \left| \frac{1}{R} \sum_{i=1}^{R} \{D_i^2 - (D_i^*)^2\} \right| \leq \frac{1}{R} \sum_{i=1}^{R} (D_i - D_i^*)^2 + \left| \frac{2}{R} \sum_{i=1}^{R} D_i^*(D_i - D_i^*) \right| \\ \leq & \frac{1}{R} \sum_{i=1}^{R} (D_i - D_i^*)^2 + 2\kappa_k \breve{\sigma}_{\infty} \sqrt{\frac{1}{R} \sum_{i=1}^{R} (D_i - D_i^*)^2}. \end{aligned} \tag{115}$$

By Equation (114), we have that $\breve{\sigma}_{\infty}^2 \xrightarrow{P} \sigma_{\infty}^2 \in [0, \infty)$. Next, we will focus on upper bounding $|D_i - D_i^*|$. Without loss of generality, we consider the case that $\eta_k \in \mathcal{S}_{2j}$ for some $j \in \{1, \ldots, R\}$. Let $h_1 = \eta_k - s_k - (2j-1)S$ and $h_2 = s_k + 2jS - \eta_k$. Assume $h_2 \geq h_1$. The following Figure depicts the definitions of h_1 and h_2.



Case 1: $i < j$. It follows from Lemma 6 that

$$\begin{aligned} &\sqrt{2S}|D_i - D_i^*| = \left| \sum_{t\in\mathcal{S}_{2i-1}} (Z_t - Z_t^*) - \sum_{t\in\mathcal{S}_{2i}} (Z_t - Z_t^*) \right| \\ =& \Bigg| \sum_{t\in\mathcal{S}_{2i-1}} \big(-2\widehat{\nu}_{k+1}^{\top} X_t X_t^{\top} \Psi_k + (\widehat{\nu}_k - \widehat{\nu}_{k+1})^{\top} X_t X_t^{\top} (\widehat{\nu}_k + \widehat{\nu}_{k+1}) + 2\epsilon_t(\widehat{\nu}_{k+1} - \widehat{\nu}_k)^{\top} X_t \big) \\ & - \sum_{t\in\mathcal{S}_{2i}} \big(-2\widehat{\nu}_{k+1}^{\top} X_t X_t^{\top} \Psi_k + (\widehat{\nu}_k - \widehat{\nu}_{k+1})^{\top} X_t X_t^{\top} (\widehat{\nu}_k + \widehat{\nu}_{k+1}) + 2\epsilon_t(\widehat{\nu}_{k+1} - \widehat{\nu}_k)^{\top} X_t \big) \Bigg| \end{aligned}$$

$$
\begin{aligned}
\leq & 2\kappa_k|\widehat{\nu}_{k+1}|_1\Big|\sum_{t\in\mathcal{S}_{2i-1}}(X_tX_t^\top-\Sigma)v_k-\sum_{t\in\mathcal{S}_{2i}}(X_tX_t^\top-\Sigma)v_k\Big|_\infty\\
&+2\big|\widehat{\nu}_k\big|_1^2\Big|\sum_{t\in\mathcal{S}_{2i-1}}(X_tX_t^\top-\Sigma)-\sum_{t\in\mathcal{S}_{2i}}(X_tX_t^\top-\Sigma)\Big|_\infty+2|\widehat{\nu}_k|_1\Big|\sum_{t\in\mathcal{S}_{2i-1}}\epsilon_tX_t-\sum_{t\in\mathcal{S}_{2i}}\epsilon_tX_t\Big|_\infty\\
&+2\big|\widehat{\nu}_{k+1}\big|_1^2\Big|\sum_{t\in\mathcal{S}_{2i-1}}(X_tX_t^\top-\Sigma)-\sum_{t\in\mathcal{S}_{2i}}(X_tX_t^\top-\Sigma)\Big|_\infty+2|\widehat{\nu}_{k+1}|_1\Big|\sum_{t\in\mathcal{S}_{2i-1}}\epsilon_tX_t-\sum_{t\in\mathcal{S}_{2i}}\epsilon_tX_t\Big|_\infty.
\end{aligned}
$$

By Lemma 5 and Lemma 6, it follows with probability at least $1-c(n^{-3}+S^{-5})$ that

$$
|D_i-D_i^*|\leq C_1\big(\kappa_k^2\mathfrak{s}^{1/2}\alpha_n^{-1/2}+\kappa_k^2\mathfrak{s}\alpha_n^{-1}+\kappa_k\mathfrak{s}^{1/2}\alpha_n^{-1/2}\big)\bigg(\sqrt{\log(p)}+\frac{\log^{1/\gamma}(p)}{\sqrt{S}}\bigg).
$$

Case 2: $i>j$. Similarly, it follows with probability at least $1-c(n^{-3}+S^{-5})$ that

$$
|D_i-D_i^*|\leq C_2\big(\kappa_k^2\mathfrak{s}^{1/2}\alpha_n^{-1/2}+\kappa_k^2\mathfrak{s}\alpha_n^{-1}+\kappa_k\mathfrak{s}^{1/2}\alpha_n^{-1/2}\big)\bigg(\sqrt{\log(p)}+\frac{\log^{1/\gamma}(p)}{\sqrt{S}}\bigg).
$$

Case 3: $i=j$. We have that

$$
\begin{aligned}
&\sqrt{2S}|D_j-D_j^*|=\Big|\sum_{t\in\mathcal{S}_{2j-1}}(Z_t-Z_t^*)-\sum_{t\in\mathcal{S}_{2j}}(Z_t-Z_t^*)\Big|\\
=&\Big|\sum_{t\in\mathcal{S}_{2j-1}}\big(-2\widehat{\nu}_{k+1}^\top X_tX_t^\top\Psi_k+(\widehat{\nu}_k-\widehat{\nu}_{k+1})^\top X_tX_t^\top(\widehat{\nu}_k+\widehat{\nu}_{k+1})+2\epsilon_t(\widehat{\nu}_{k+1}-\widehat{\nu}_k)^\top X_t\big)\\
&-\sum_{t\in[s_k+\{2j-1\}S,\eta_k)}\big(-2\widehat{\nu}_{k+1}^\top X_tX_t^\top\Psi_k+(\widehat{\nu}_k-\widehat{\nu}_{k+1})^\top X_tX_t^\top(\widehat{\nu}_k+\widehat{\nu}_{k+1})+2\epsilon_t(\widehat{\nu}_{k+1}-\widehat{\nu}_k)^\top X_t\big)\\
&-\sum_{t\in[\eta_k,s_k+2jS)}\big(-2\widehat{\nu}_k^\top X_tX_t^\top\Psi_k+(\widehat{\nu}_k-\widehat{\nu}_{k+1})^\top X_tX_t^\top(\widehat{\nu}_k+\widehat{\nu}_{k+1})+2\epsilon_t(\widehat{\nu}_{k+1}-\widehat{\nu}_k)^\top X_t\big)\Big|\\
\leq&2\kappa_k|\widehat{\nu}_{k+1}|_1\Big|\sum_{t\in\mathcal{S}_{2i-1}}(X_tX_t^\top-\Sigma)v_k-\sum_{t\in(s_k+\{2j-1\}S,\eta_k]}(X_tX_t^\top-\Sigma)v_k\Big|_\infty\\
&+2\kappa_k|\widehat{\nu}_k|_1\Big|\sum_{t\in(\eta_k,s_k+2jS]}(X_tX_t^\top-\Sigma)v_k\Big|_\infty+2(S-h_1)\kappa_k|\widehat{\nu}_{k+1}|_2\Lambda_{\max}+2h_2\kappa_k|\widehat{\nu}_k|_2\Lambda_{\max}\\
&+2\big|\widehat{\nu}_k\big|_1^2\Big|\sum_{t\in\mathcal{S}_{2i-1}}(X_tX_t^\top-\Sigma)-\sum_{t\in\mathcal{S}_{2i}}(X_tX_t^\top-\Sigma)\Big|_\infty+2|\widehat{\nu}_k|_1\Big|\sum_{t\in\mathcal{S}_{2i-1}}\epsilon_tX_t-\sum_{t\in\mathcal{S}_{2i}}\epsilon_tX_t\Big|_\infty\\
&+2\big|\widehat{\nu}_{k+1}\big|_1^2\Big|\sum_{t\in\mathcal{S}_{2i-1}}(X_tX_t^\top-\Sigma)-\sum_{t\in\mathcal{S}_{2i}}(X_tX_t^\top-\Sigma)\Big|_\infty+2|\widehat{\nu}_{k+1}|_1\Big|\sum_{t\in\mathcal{S}_{2i-1}}\epsilon_tX_t-\sum_{t\in\mathcal{S}_{2i}}\epsilon_tX_t\Big|_\infty.
\end{aligned}
$$

By Lemma 5 and Lemma 6, it follows with probability at least $1-c(n^{-3}+S^{-5})$ that

$$
|D_j-D_j^*|\leq C_3\big(\kappa_k^2\mathfrak{s}^{1/2}\alpha_n^{-1/2}+\kappa_k^2\mathfrak{s}\alpha_n^{-1}+\kappa_k\mathfrak{s}^{1/2}\alpha_n^{-1/2}\big)\bigg(\sqrt{\log(p)}+\frac{\log^{1/\gamma}(p)}{\sqrt{S}}\bigg)+C_4\sqrt{S}\kappa_k^2\alpha_n^{-1/2}.
$$

Combining all the above cases, it follows with probability at least $1 - c(n^{-2} + RS^{-5})$ that

$$
\begin{aligned}
&\frac{1}{R\kappa_k^2}\sum_{i=1}^{R}(D_i - D_i^*)^2 = \frac{1}{R\kappa_k^2}\sum_{i\neq j}(D_i - D_i^*)^2 + \frac{1}{R\kappa_k^2}(D_j - D_j^*)^2 \\
\leq & C_5\big(\kappa_k \mathfrak{s}^{1/2}\alpha_n^{-1/2} + \kappa_k \mathfrak{s}\alpha_n^{-1} + \mathfrak{s}^{1/2}\alpha_n^{-1/2}\big)^2\left(\sqrt{\log(p)} + \frac{\log^{1/\gamma}(p)}{\sqrt{S}}\right)^2 + C_6\frac{S\kappa_k^2\alpha_n^{-1}}{R}.
\end{aligned} \tag{116}
$$

Combining (115) and (116), we have that with probability at least $1 - cn^{-3}$

$$
\left|\frac{\widehat{\kappa}_k^2}{\kappa_k^2}\widehat{\sigma}_\infty^2 - \check{\sigma}_\infty^2\right| \leq C_7\frac{\mathfrak{s}}{\alpha_n}\left(\sqrt{\log(p)} + \frac{\log^{1/\gamma}(p)}{\sqrt{S}}\right)^2.
$$

Furthermore, applying Lemma 7, Lemma 23, Slutsky's theorem, we have that as $n \to \infty$,

$$
\left|\widehat{\sigma}_\infty^2 - \sigma_\infty^2\right| \xrightarrow{P.} 0.
$$

□

Proof of Proposition 9. This proof uses the results in the **Preliminary.** step of the proof of Theorem 3, which are omitted here for brevity.

$$
\begin{aligned}
&\widehat{\varpi}_k - \varpi_k = \frac{1}{n\widehat{\kappa}_k^2}\sum_{t=1}^{n}\widehat{\Psi}_k^\top X_t X_t^\top \widehat{\Psi}_k - v_k^\top \Sigma v_k \\
=&\frac{\kappa_k^2}{\widehat{\kappa}_k^2}\frac{1}{n\kappa_k^2}\sum_{t=1}^{n}\{\widehat{\Psi}_k^\top X_t X_t^\top \widehat{\Psi}_k - \Psi_k^\top X_t X_t^\top \Psi_k\} + \left(\frac{\kappa_k^2}{\widehat{\kappa}_k^2} - 1\right)\frac{1}{n}\sum_{t=1}^{n} v_k^\top X_t X_t^\top v_k \\
&+ \frac{1}{n}\sum_{t=1}^{n}\{v_k^\top X_t X_t^\top v_k - v_k^\top \Sigma v_k\} \\
=&I + II + III.
\end{aligned}
$$

Step 1. We consider the term I. It follows that

$$
\begin{aligned}
&\frac{1}{n\kappa_k^2}\sum_{t=1}^{n}\{\widehat{\Psi}_k^\top X_t X_t^\top \widehat{\Psi}_k - \Psi_k^\top X_t X_t^\top \Psi_k\} \\
=&\frac{1}{n\kappa_k^2}\sum_{t=1}^{n}(\widehat{\beta}_k - \widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_k} + \beta^*_{\mathcal{I}_{k-1}})^\top (X_t X_t^\top - \Sigma)(\widehat{\beta}_k - \widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_k} + \beta^*_{\mathcal{I}_{k-1}}) \\
&- \frac{1}{n\kappa_k^2}\sum_{t=1}^{n}(\beta^*_{\eta_k} - \beta^*_{\eta_{k-1}} - \beta^*_{\mathcal{I}_k} + \beta^*_{\mathcal{I}_{k-1}})^\top (X_t X_t^\top - \Sigma)(\beta^*_{\eta_k} - \beta^*_{\eta_{k-1}} - \beta^*_{\mathcal{I}_k} + \beta^*_{\mathcal{I}_{k-1}}) \\
&+ \frac{2}{n\kappa_k^2}\sum_{t=1}^{n}(\widehat{\beta}_k - \widehat{\beta}_{k-1} - \beta^*_{\eta_k} + \beta^*_{\eta_{k-1}})^\top (X_t X_t^\top - \Sigma)(\beta^*_{\mathcal{I}_k} - \beta^*_{\mathcal{I}_{k-1}})
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{\kappa_k^2}(\widehat{\beta}_k-\widehat{\beta}_{k-1}-\beta^*_{\mathcal{I}_k}+\beta^*_{\mathcal{I}_{k-1}})^\top\Sigma(\widehat{\beta}_k-\widehat{\beta}_{k-1}-\beta^*_{\mathcal{I}_k}+\beta^*_{\mathcal{I}_{k-1}})\\
&-\frac{1}{\kappa_k^2}(\beta^*_{\eta_k}-\beta^*_{\eta_{k-1}}-\beta^*_{\mathcal{I}_k}+\beta^*_{\mathcal{I}_{k-1}})^\top\Sigma(\beta^*_{\eta_k}-\beta^*_{\eta_{k-1}}-\beta^*_{\mathcal{I}_k}+\beta^*_{\mathcal{I}_{k-1}})\\
&+\frac{2}{\kappa_k^2}(\widehat{\beta}_k-\widehat{\beta}_{k-1}-\beta^*_{\eta_k}+\beta^*_{\eta_{k-1}})^\top\Sigma(\beta^*_{\mathcal{I}_k}-\beta^*_{\mathcal{I}_{k-1}})\\
=&I_1-I_2+2I_3+I_4-I_5+2I_6.
\end{aligned}
$$

Term I_1. It holds that

$$
|I_1|\le\frac{2}{n\kappa_k^2}\sum_{t=1}^n(\widehat{\beta}_k-\beta^*_{\mathcal{I}_k})^\top(X_tX_t^\top-\Sigma)(\widehat{\beta}_k-\beta^*_{\mathcal{I}_k})+\frac{2}{n\kappa_k^2}\sum_{t=1}^n(\widehat{\beta}_{k-1}-\beta^*_{\mathcal{I}_{k-1}})^\top(X_tX_t^\top-\Sigma)(\widehat{\beta}_{k-1}-\beta^*_{\mathcal{I}_{k-1}}),
$$

where the inequality follows from the Cauchy–Schwarz inequality and the symmetry of $X_tX_t^\top-\Sigma$. It follows from Theorem 33**b**, (39) and (41) that with probability at least $1-n^{-5}$,

$$
|I_1|\le C_1\sqrt{\frac{\mathfrak{s}\log(pn)}{\Delta}}\frac{\mathfrak{s}\log(pn)}{\Delta\kappa_k^2}\le C_2\alpha_n^{-3/2}.
$$

Term I_2. It holds that

$$
|I_2|\le\frac{2}{n\kappa_k^2}\sum_{t=1}^n(\beta^*_{\eta_k}-\beta^*_{\mathcal{I}_k})^\top(X_tX_t^\top-\Sigma)(\beta^*_{\eta_k}-\beta^*_{\mathcal{I}_k})+\frac{2}{n\kappa_k^2}\sum_{t=1}^n(\beta_{\eta^*_{k-1}}-\beta^*_{\mathcal{I}_{k-1}})^\top(X_tX_t^\top-\Sigma)(\beta^*_{\eta_{k-1}}-\beta^*_{\mathcal{I}_{k-1}}).
$$

Let

$$
z_t=\frac{1}{|\beta^*_{\mathcal{I}_{k-1}}-\beta^*_{\eta_{k-1}}|_2^2}(\beta^*_{\mathcal{I}_{k-1}}-\beta^*_{\eta_{k-1}})^\top\big(X_tX_t^\top-\Sigma\big)(\beta^*_{\mathcal{I}_{k-1}}-\beta^*_{\eta_{k-1}}).
$$

Note that

$$
\begin{aligned}
&\big\|z_t-z_{t,\{0\}}\big\|_2\\
\le&\left\|\frac{1}{|\beta^*_{\mathcal{I}_{k-1}}-\beta^*_{\eta_{k-1}}|_2^2}(\beta^*_{\mathcal{I}_{k-1}}-\beta^*_{\eta_{k-1}})^\top\Big\{X_t(X_t-X_{t,\{0\}})^\top+(X_t-X_{t,\{0\}})X_{t,\{0\}}^\top\Big\}(\beta^*_{\mathcal{I}_{k-1}}-\beta^*_{\eta_{k-1}})\right\|_2\\
\le&2\sup_{|v|_2=1}\big\|v^\top X_1\big\|_4\delta^X_{t,4}.
\end{aligned}
$$

By Assumption 2

$$
\sup_{m\ge0}\exp(cm^{\gamma_1(X)})\sum_{s=m}^\infty\big\|z_s-z_{s,\{0\}}\big\|_2\le 2D_X\sup_{|v|_2=1}\big\|v^\top X_1\big\|_4,
$$

and by Lemma 26, there exists C_1 depending on C_X such that

$$
\mathbb{P}(|z_t|\ge\tau)\le 2\exp(-(\tau/C_1)^{\gamma_2/2}).
$$

Then, by Theorem 4, we have with probability at least $1-n^{-5}$ that

$$
\left|\frac{1}{n}\sum_{t=1}^n z_t\right|\le C_2\Big\{\sqrt{\frac{\log(n)}{n}}+\frac{\log^{1/\gamma}(n)}{n}\Big\}.
$$

Consequently,

$$
\begin{aligned}
|I_2| \leq & C_3\kappa_k^{-2}\big(|\beta^*_{\mathcal{I}_k} - \beta^*_{\eta_k}|_2^2 + |\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\eta_{k-1}}|_2^2\big)\left\{\sqrt{\frac{\log(n)}{n}} + \frac{\log^{1/\gamma}(n)}{n}\right\} \\
\leq & C_4\alpha_n^{-2}\left\{\sqrt{\frac{\log(n)}{n}} + \frac{\log^{1/\gamma}(n)}{n}\right\} \leq C_4\alpha_n^{-2},
\end{aligned}
$$

where the second inequality follows from (43) and (44).

Term I_3. It holds that

$$
|I_3| \leq \frac{1}{n\kappa_k^2}\big(|\widehat{\beta}_k - \beta^*_{\eta_k}|_1 + |\widehat{\beta}_{k-1} - \beta^*_{\eta_{k-1}}|_1\big)\left|\sum_{t=1}^n (X_tX_t^\top - \Sigma)(\beta^*_{\mathcal{I}_k} - \beta^*_{\mathcal{I}_{k-1}})\right|_\infty.
$$

For any $j \in \{1, \dots, p\}$, let

$$
z_{tj} = \frac{1}{|\beta^*_{\mathcal{I}_k} - \beta^*_{\mathcal{I}_{k-1}}|_2}\big(X_{tj}X_t^\top - \Sigma_{j\cdot}\big)(\beta^*_{\mathcal{I}_{k-1}} - \beta^*_{\mathcal{I}_{k-1}}).
$$

Similarly, we have for any $j \in \{1, \dots, p\}$ that

$$
\big\|z_{tj} - z_{tj,\{0\}}\big\|_2 \leq 2\sup_{|v|_2=1}\big\|v^\top X_1\big\|_4\delta^X_{t,4}.
$$

Under assumption 2, we can verify that

$$
\sup_{m\geq 0}\exp(cm^{\gamma_1(X)})\sum_{s=m}^\infty \big\|z_{sj} - z_{sj,\{0\}}\big\|_2 \leq 2D_X\sup_{|v|_2=1}\big\|v^\top X_1\big\|_4,
$$

and

$$
\mathbb{P}(|z_{tj}| \geq \tau) \leq 2\exp(-(\tau/C_5)^{\gamma_2/2}).
$$

Then, by Theorem 4 and the union bound, we have with probability at least $1 - n^{-5}$ that

$$
\max_{1\leq j\leq p}\left|\frac{1}{n}\sum_{t=1}^n z_{tj}\right| \leq C_6\left\{\sqrt{\frac{\log(pn)}{n}} + \frac{\log^{1/\gamma}(pn)}{n}\right\}.
$$

Consequently,

$$
\begin{aligned}
|I_3| \leq & C_7\kappa_k^{-2}\big(|\widehat{\beta}_k - \beta^*_{\eta_k}|_1 + |\widehat{\beta}_{k-1} - \beta^*_{\eta_{k-1}}|_1\big)|\beta^*_{\mathcal{I}_k} - \beta^*_{\mathcal{I}_{k-1}}|_2\left\{\sqrt{\frac{\log(pn)}{n}} + \frac{\log^{1/\gamma}(pn)}{n}\right\} \\
\leq & C_8\alpha_n^{-1/2}\left\{\sqrt{\frac{\mathfrak{s}\log(pn)}{n}} + \frac{\mathfrak{s}^{1/2}\log^{1/\gamma}(pn)}{n}\right\} \\
\leq & C_8\alpha_n^{-1/2}\left\{\sqrt{\frac{\mathfrak{s}\log(pn)}{\Delta}} + \frac{\mathfrak{s}^{1/2}\log^{1/\gamma}(pn)}{\Delta}\right\} \\
\leq & C_9\alpha_n^{-1},
\end{aligned}
$$

where the second inequality follows from Lemma 6, (44) and (45), and the last inequality follows from (24).

Terms I_4, I_5 and I_6. By (39), it holds that

$$|I_4| \leq 2\kappa_k^{-2}\Lambda_{\max}(\Sigma)\big(|\widehat{\beta}_k - \beta^*_{\mathcal{I}_k}|_2^2 + |\widehat{\beta}_{k-1} - \beta^*_{\mathcal{I}_{k-1}}|_2^2\big) \leq C_{10}\frac{\mathfrak{s}\log(pn)}{\Delta\kappa_k^2} \leq C_{11}\alpha_n^{-1}.$$

By (43) and (44)

$$|I_5| \leq 2\kappa_k^{-2}\Lambda_{\max}(\Sigma)\big(|\beta^*_{\eta_k} - \beta^*_{\mathcal{I}_k}|_2^2 + |\beta^*_{\eta_{k-1}} - \beta^*_{\mathcal{I}_{k-1}}|_2^2\big) \leq C_{12}\alpha_n^{-2}.$$

By Lemma 6, (44) and (45)

$$|I_6| \leq \kappa_k^{-2}\Lambda_{\max}(\Sigma)\big(|\widehat{\beta}_k - \beta^*_{\eta_k}|_2 + |\widehat{\beta}_{k-1} - \beta^*_{\eta_{k-1}}|_2\big)|\beta^*_{\mathcal{I}_k} - \beta^*_{\mathcal{I}_{k-1}}|_2 \leq C_{13}\alpha_n^{-1/2}.$$

Combining the upper bounds of $|I_i|$ for $i = 1, \ldots, 6$, we have that

$$\frac{1}{n\kappa_k^2}\sum_{t=1}^{n}\big\{\widehat{\Psi}_k^\top X_t X_t^\top \widehat{\Psi}_k - \Psi_k^\top X_t X_t^\top \Psi_k\big\} = o_p(1).$$

Since $\frac{\kappa_k}{\widehat{\kappa}_k} - 1 = o_p(1)$ due to Lemma 7, by Slutsky's theorem, we have that

$$I = o_p(1).$$

Step 2. We consider the terms II and III. Using the similar arguments as for the term I_2, we have that with probability at least $1 - n^{-5}$ that

$$\left|\frac{1}{n}\sum_{t=1}^{n} v_k^\top (X_t X_t^\top - \Sigma) v_k\right| \leq C_2\left\{\sqrt{\frac{\log(n)}{n}} + \frac{\log^{1/\gamma}(n)}{n}\right\}.$$

It follows that

$$II = o_p(1),$$

since $\frac{\kappa_k}{\widehat{\kappa}_k} - 1 = o_p(1)$ due to Lemma 7. Moreover,

$$III = o_p(1).$$

Therefore, we have that as $n \to \infty$

$$\widehat{\varpi}_k \xrightarrow{P.} \varpi_k.$$

$\square$

Proof of Theorem 10. Note that when $M = \infty$, (32) can be equivalently written as

$$\widehat{u}^{(b)} = \arg\min_{r\in\mathbb{R}}\big\{\widehat{\varpi}_k|r| + \widehat{\sigma}_\infty(k)\mathbb{W}^{(b)}(r)\big\},$$

Denote $\widehat{u} = \widehat{u}^{(b)}$ and $z_i = z_i^{(b)}$ for simplicity.

Step 1. We show $\widehat{u}$ is uniformly tight. A similar argument was used in the proof of Theorem 3. Without loss of generality, assume that $\widehat{u} > 0$. If $\widehat{u} \leq 1$, there is nothing to show. Suppose $\widehat{u} \geq 1$. By Lemma 27 with $\nu = 1$, uniformly for all $r \geq 1$ and $n \in \mathbb{Z}_+$,

$$\frac{1}{\sqrt{nr}\log(r)} \sum_{i=1}^{nr} z_i = O_p(1).$$

Since $\widehat{\sigma}_\infty(k) - \sigma_\infty(k) = o_p(1)$ and $\widehat{\varpi}_k - \varpi_k = o_p(1)$, it follows with probability approaching to 1,

$$0 < \varpi_k/2 \leq \widehat{\varpi}_k \text{ and } \widehat{\sigma}_\infty(k) \in [0, \infty).$$

Since $\widehat{u}$ is the minimizer,

$$\widehat{u}\widehat{\varpi}_k + \widehat{\sigma}_\infty(k) \frac{1}{\sqrt{n}} \sum_{i=1}^{n\widehat{u}} z_i \leq 0.$$

Therefore,

$$\widehat{u}\varpi_k/2 \leq \widehat{u}\widehat{\varpi}_k \leq -\widehat{\sigma}_\infty(k) \frac{1}{\sqrt{n}} \sum_{i=1}^{n\widehat{u}} z_i = O_p(\sqrt{\widehat{u}}\log(\widehat{u})).$$

This implies that $\widehat{u} = O_p(1)$.

Step 2. We show that for any fixed $M > 0$,

$$\widehat{\varpi}_k|r| + \widehat{\sigma}_\infty(k)\mathbb{W}^{(b)}(r) \xrightarrow{\mathcal{D}} \varpi_k|r| + \sigma_\infty(k)\mathbb{W}(r) \text{ for } |r| \leq M.$$

Uniformly for all $r \leq M$, the functional CLT for i.i.d. random variables leads to

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{nr} z_i \xrightarrow{\mathcal{D}} \mathbb{B}_1(r).$$

Since $\widehat{\sigma}_\infty(k) - \sigma_\infty(k) = o_p(1)$ and $\widehat{\varpi}_k - \varpi_k = o_p(1)$, the Argmax (or Argmin) continuous mapping theorem and Slutsky's theorem lead to

$$\widehat{u} \xrightarrow{\mathcal{D}} \arg\min_r \{\varpi_k|r| + \sigma_\infty(k)\mathbb{W}(r)\}.$$

Finally, Lemma 7 and Slutsky's theorem directly give the desired result. $\square$

E Additional Technical Lemmas

Lemma 24 (Burkholder's inequality). *Let $q > 1$, $q' = \min\{q, 2\}$ and $K_q = \max\{(q-1)^{-1}, \sqrt{q-1}\}$. Let $\{X_t\}_{t=1}^n$ be a martingale difference sequence with $\|X_i\|_q < \infty$. That is, there exists a martingale $\{S_k\}_{k=0}^n$ such that $X_k = S_k - S_{k-1}$. Then*

$$\|S_n\|_q^{q'} \leq K_q^{q'} \sum_{t=1}^n \|X_t\|_q^{q'}.$$

Proof. See Rio (2009) for $q > 2$ and Burkholder (1988) for $1 < q \leq 2$. □

Lemma 25. *Suppose $\{Z_t\}_{t\in\mathbb{Z}}$ is a (possibly) nonstationary process in the form of (10). Then*

$$\|\mathcal{P}_{t-i}Z_t\|_q = \|\mathbb{E}(Z_t|\mathcal{F}_{t-i}) - \mathbb{E}(Z_t|\mathcal{F}_{t-i-1})\|_q \leq \|Z_t - Z_{t,\{t-i\}}\|_q.$$

Proof. Note that

$$\mathbb{E}(Z_t|\mathcal{F}_{t-i-1}) = \mathbb{E}(Z_{t,\{t-i\}}|\mathcal{F}_{t-i-1}) = \mathbb{E}(Z_{t,\{t-i\}}|\mathcal{F}_{t-i}).$$

Therefore

$$\begin{aligned}\|\mathcal{P}_{t-i}Z_t\|_q =& \|\mathbb{E}(Z_t|\mathcal{F}_{t-i}) - \mathbb{E}(Z_t|\mathcal{F}_{t-i-1})\|_q \\ =& \|\mathbb{E}(Z_t|\mathcal{F}_{t-i}) - \mathbb{E}(Z_{t,\{t-i\}}|\mathcal{F}_{t-i-1})\|_q \\ =& \|\mathbb{E}(Z_t - Z_{t,\{t-i\}}|\mathcal{F}_{t-i})\|_q \\ \leq& \|Z_t - Z_{t,\{t-i\}}\|_q\end{aligned}$$

where the last inequality follows from Jensen's inequality. □

Lemma 26. *Suppose Z_1 and Z_2 are two random variables such that*

$$\mathbb{P}(|Z_i| \geq \tau) \leq 2\exp(-(\tau/C_i)^{\gamma_2}). \tag{117}$$

Then there exists a constant C depending on only C_1 and C_2 such that

$$\mathbb{P}(|Z_1Z_2 - \mathbb{E}(Z_1Z_2)| \geq \tau) \leq 2\exp(-(\tau/C)^{\gamma_2/2}).$$

Proof. This is a well-known properties for sub-Weibull random variables. See Wong et al. (2020) for a detailed introduction of subweibull random variables. □

E.1 A maximal inequality for weighted partial sums under dependence

Consider a stationary process $\{Z_t\}_{t\in\mathbb{Z}} \subset \mathbb{R}$ with $\mathbb{E}[Z_t] = 0$ and the functional dependence measure $\delta_{s,q}$ defined similarly in (9). Let $S_n = \sum_{k=1}^n Z_k$ and $S_n^* = \max_{1\leq i\leq n}|S_i|$. The following lemma is a maximal inequality for weighted partial sums.

Lemma 27. *Assume $\sum_{s=1}^\infty \delta_{s,q} < \infty$ for some $q > 2$. Let $\nu > 0$ be given. For any $0 < a < 1$, it holds that*

$$\mathbb{P}\bigg(\bigg|\sum_{i=1}^r Z_i\bigg| \leq \frac{C}{a}\sqrt{r}\Big\{\log(r\nu) + 1\Big\} \text{ for all } r \geq 1/\nu\bigg) \geq 1 - a^2,$$

where $C > 0$ is an absolute constant.

Proof. Let $s \in \mathbb{Z}_+$ and $\mathcal{T}_s = [2^s/\nu, 2^{s+1}/\nu]$. By Lemma 29, for all $x > 0$,

$$\mathbb{P}\bigg(\sup_{r\in\mathcal{T}_s}\frac{\big|\sum_{i=1}^r Z_i\big|}{\sqrt{r}} \geq x\bigg) \leq C_1x^{-2}.$$

Therefore by a union bound, for any $0 < a < 1$,

$$\mathbb{P}\bigg(\exists s \in \mathbb{Z}_+ : \sup_{r\in\mathcal{T}_s}\frac{\big|\sum_{i=1}^r Z_i\big|}{\sqrt{r}} \geq \frac{\sqrt{C_1}\pi}{\sqrt{6}a}(s+1)\bigg) \leq \sum_{s=0}^\infty \frac{6a^2}{\pi^2(s+1)^2} = a^2. \tag{118}$$

For any $r \in \mathcal{T}_s$, $s \le \log(r\nu)/\log(2)$, and therefore

$$\mathbb{P}\left(\exists s \in \mathbb{Z}_+ : \sup_{r \in \mathcal{T}_s} \frac{\left|\sum_{i=1}^r Z_i\right|}{\sqrt{r}} \ge \frac{\sqrt{C_1}\pi}{\sqrt{6}a}\left\{\frac{\log(r\nu)}{\log(2)} + 1\right\}\right) \le a^2. \tag{119}$$

Equation (119) directly gives

$$\mathbb{P}\left(\sup_{r \in \mathcal{T}_s} \frac{\left|\sum_{i=1}^r Z_i\right|}{\sqrt{r}} \le \frac{C}{a}\left\{\log(r\nu) + 1\right\} \text{ for all } s \le \log(r\nu)/\log(2)\right) \ge 1 - a^2.$$

□

The proof of Lemma 27 relies on the following two lemmas. The next Lemma is a slight modification of the Rosenthal-type inequality given in Theorem 1 of Liu et al. (2013).

Lemma 28. *It holds for $q > 2$ that*

$$\begin{aligned}
\|S_N^*\|_p \le & N^{1/2}\left\{\frac{87p}{\log(p)}\sum_{j=1}^N \delta_{j,2} + 3(p-1)^{1/2}\sum_{j=1}^\infty \delta_{j,p} + \frac{29p}{\log(p)}\|Z_1\|_2\right\} \\
& + N^{1/2}\left\{\frac{87p(p-1)^{1/2}}{\log(p)}\sum_{j=1}^N \delta_{j,p} + \frac{29p}{\log(p)}\|Z_1\|_p\right\}.
\end{aligned}$$

Proof. By Theorem 1 of Liu et al. (2013),

$$\begin{aligned}
\|S_N^*\|_p \le & N^{1/2}\left\{\frac{87p}{\log(p)}\sum_{j=1}^N \delta_{j,2} + 3(p-1)^{1/2}\sum_{j=1}^\infty \delta_{j,p} + \frac{29p}{\log(p)}\|Z_1\|_2\right\} \\
& + N^{1/p}\left\{\frac{87p(p-1)^{1/2}}{\log(p)}\sum_{j=1}^N j^{1/2-1/p}\delta_{j,p} + \frac{29p}{\log(p)}\|Z_1\|_p\right\} \\
\le & N^{1/2}\left\{\frac{87p}{\log(p)}\sum_{j=1}^N \delta_{j,2} + 3(p-1)^{1/2}\sum_{j=1}^\infty \delta_{j,p} + \frac{29p}{\log(p)}\|Z_1\|_2\right\} \\
& + N^{1/p}\left\{\frac{87p(p-1)^{1/2}}{\log(p)}N^{1/2-1/p}\sum_{j=1}^N \delta_{j,p} + \frac{29p}{\log(p)}\|Z_1\|_p\right\} \\
\le & N^{1/2}\left\{\frac{87p}{\log(p)}\sum_{j=1}^N \delta_{j,2} + 3(p-1)^{1/2}\sum_{j=1}^\infty \delta_{j,p} + \frac{29p}{\log(p)}\|Z_1\|_2\right\} \\
& + N^{1/2}\left\{\frac{87p(p-1)^{1/2}}{\log(p)}\sum_{j=1}^N \delta_{j,p} + \frac{29p}{\log(p)}\|Z_1\|_p\right\}.
\end{aligned}$$

□

The key to proving Lemma 27 is applying the uniform bound on a series of maximal inequalities for the weighted partial sums of dependent random variables within a progressive block. The following lemma gives such maximal inequality.

Lemma 29. *Assume $\sum_{s=1}^{\infty} \delta_{s,q} < \infty$ for some $q > 2$. Then, it follows that for any $d \in \mathbb{Z}_+$, $\alpha > 0$ and $x > 0$,*

$$\mathbb{P}\bigg(\max_{k\in[d,(\alpha+1)d]} \frac{|\sum_{i=1}^{k} Z_i|}{\sqrt{k}} \geq x \bigg) \leq Cx^{-2},$$

where $C > 0$ is an absolute constant .

Proof. Let $N = (\alpha + 1)d$ and recall that

$$S_N^* = \max_{k=1,\dots,N} \bigg| \sum_{i=1}^{k} Z_i \bigg|.$$

Since $\sum_{s=1}^{\infty} \delta_{s,2} < \infty$, it follows from Lemma 28 that

$$\|S_N^*\|_2 \leq C_1 N^{1/2}.$$

Therefore it holds that

$$\mathbb{P}\bigg(\bigg| \frac{S_N^*}{\sqrt{N}} \bigg| \geq x \bigg) = \mathbb{P}\bigg(\bigg| \frac{S_N^*}{\sqrt{N}} \bigg|^2 \geq x^2 \bigg) \leq C_1^2 x^{-2}.$$

Observe that

$$\frac{|S_N^*|}{\sqrt{N}} = \max_{k=1,\dots,N} \frac{|\sum_{i=1}^{k} Z_i|}{\sqrt{N}} \geq \max_{k\in[d,(\alpha+1)d]} \frac{|\sum_{i=1}^{k} Z_i|}{\sqrt{N}} \geq \max_{k\in[d,(\alpha+1)d]} \frac{|\sum_{i=1}^{k} Z_i|}{\sqrt{(\alpha+1)k}}.$$

Therefore

$$\mathbb{P}\bigg(\max_{k\in[d,(\alpha+1)d]} \frac{|\sum_{i=1}^{k} Z_i|}{\sqrt{k}} \geq \sqrt{(\alpha+1)}x \bigg) \leq \mathbb{P}\bigg(\bigg| \frac{S_N^*}{\sqrt{N}} \bigg| \geq x \bigg) \leq C_1^2 x^{-2},$$

which gives

$$\mathbb{P}\bigg(\max_{k\in[d,(\alpha+1)d]} \frac{|\sum_{i=1}^{k} Z_i|}{\sqrt{k}} \geq x \bigg) \leq C_2 x^{-2}.$$

□

E.2 Pointwise deviation bounds under dependence

Lemma 30. *Let $1/\gamma = 1/\gamma_1 + 2/\gamma_2 > 1$. Under Assumption 2 and Assumption 3, it holds for any integer $n \geq 3$ that*

$$\mathbb{P}\Bigg(\bigg| \frac{1}{n} \sum_{t=1}^{n} X_t \epsilon_t \bigg|_\infty \geq C \bigg\{ \sqrt{\frac{\log(pn)}{n}} + \frac{\log^{1/\gamma}(pn)}{n} \bigg\} \Bigg) \leq n^{-5}$$

for some sufficiently large constant C depending on C_ϵ, D_ϵ, C_X and D_X. Thus when $n \geq C_\varsigma(\mathfrak{s}\log(pn))^{2/\gamma-1}$,

$$\mathbb{P}\Bigg(\bigg| \frac{1}{n} \sum_{t=1}^{n} X_t \epsilon_t \bigg|_\infty \geq 2C \sqrt{\frac{\log(pn)}{n}} \Bigg) \leq n^{-5}$$

Proof. For any $j \in \{1, \dots, p\}$, let $Z_t = X_{tj}\epsilon_t$. Note that $\{Z_t\}_{t\in\mathbb{Z}}$ is a stationary process. Observe that by Lemma 26, there exists C_1 depending on C_X and C_ϵ such that

$$\mathbb{P}(|Z_t| \geq \tau) \leq 2\exp(-(\tau/C_1)^{\gamma_2/2}).$$

For any $t \in \mathbb{Z}$, we have that

$$\begin{aligned}
\delta^Z_{s,2} =& \|Z_t - Z_{t,\{t-s\}}\|_2 = \|X_{tj}\epsilon_t - X_{tj,\{t-s\}}\epsilon_{t,\{t-s\}}\|_2 \\
\leq& \|X_{tj}(\epsilon_t - \epsilon_{t,\{t-s\}})\|_2 + \|(X_{tj} - X_{tj,\{t-s\}})\epsilon_{t,\{t-s\}}\|_2 \\
\leq& \|X_{tj}\|_4\|\epsilon_t - \epsilon_{t,\{t-s\}}\|_4 + \|X_{tj} - X_{tj,\{t-s\}}\|_4\|\epsilon_{t,\{t-s\}}\|_4 \\
\leq& \delta^\epsilon_{s,4}\max_{1\leq j\leq p}\|X_{0j}\|_4 + \delta^X_{s,4}\|\epsilon_0\|_4 = \delta^\epsilon_{s,4}\|X_{0\cdot}\|_4 + \delta^X_{s,4}\|\epsilon_0\|_4.
\end{aligned}$$

As a result, it holds that

$$\Delta^Z_{m,2} = \sum_{s=m}^{\infty} \delta^Z_{s,2} \leq \|X_{0\cdot}\|_4\Delta^\epsilon_{m,4} + \|\epsilon_0\|_4\Delta^X_{m,4}.$$

Consequently,

$$\begin{aligned}
\sup_{m\geq 0}\exp(cm^{\gamma_1})\Delta^Z_{m,2} \leq& \|X_{0\cdot}\|_4\sup_{m\geq 0}\exp(cm^{\gamma_1})\Delta^\epsilon_{m,4} + \|\epsilon_0\|_4\sup_{m\geq 0}\exp(cm^{\gamma_1})\Delta^X_{m,4} \\
\leq& 2\max\{\|\epsilon_0\|_4 D_X, \|X_{0\cdot}\|_4 D_\epsilon\} < \infty.
\end{aligned}$$

By Theorem 4, it holds that for any $\tau \geq 1$ and interger $n \geq 3$,

$$\mathbb{P}\left(\left|\sum_{t=1}^n X_{tj}\epsilon_t\right| \geq \tau\right) = \mathbb{P}\left(\left|\sum_{t=1}^n Z_t\right| \geq \tau\right) \leq n\exp\left(-c_1\tau^\gamma\right) + 2\exp\left(-\frac{c_2\tau^2}{n}\right).$$

By the union bound argument, it follows that for any $\tau \geq 1$,

$$\mathbb{P}\left(\left|\sum_{t=1}^n X_t\epsilon_t\right|_\infty \geq \tau\right) \leq np\exp\left(-c_1\tau^\gamma\right) + 2p\exp\left(-\frac{c_2\tau^2}{n}\right).$$

Therefore, with sufficiently large constant C_2,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^n X_t\epsilon_t\right|_\infty \geq C_2\left\{\sqrt{\frac{\log(pn)}{n}} + \frac{\log^{1/\gamma}(pn)}{n}\right\}\right) \leq n^{-5}.$$

$\square$

Lemma 31. *Suppose Assumption 2 holds and that $\{\alpha_t\}_{t=1}^n \in \mathbb{R}^p$ is a collection of deterministic vector such that*

$$\max_{1\leq t\leq n}|\alpha_t|_2 \leq 1 \quad \textit{and that} \quad \sum_{t=1}^n \alpha_t = 0.$$

Then it holds for any integer $n \geq 3$ that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^n X_t X_t^\top\alpha_t\right|_\infty \geq C\left\{\sqrt{\frac{\log(pn)}{n}} + \frac{\log^{1/\gamma}(np)}{n}\right\}\right) \leq n^{-5},$$

for some sufficiently large constant C depending on C_X and D_X. Thus when $n \geq C_\zeta(\mathfrak{s}\log(pn))^{2/\gamma-1}$,

$$\mathbb{P}\bigg(\bigg|\frac{1}{n}\sum_{t=1}^{n} X_t X_t^\top \alpha_t\bigg|_\infty \geq 2C\sqrt{\frac{\log(pn)}{n}}\bigg) \leq n^{-5}.$$

Proof. Since $\sum_{t=1}^{n}\alpha_t = 0$, we have

$$\mathbb{E}\bigg(\sum_{t=1}^{n} X_t X_t^\top \alpha_t\bigg) = \sum_{t=1}^{n}\Sigma\alpha_t = 0.$$

For any $j \in \{1,\dots,p\}$, let $Z_t = X_{tj}X_t^\top\alpha_t - \mathbb{E}(X_{tj}X_t^\top\alpha_t)$. By Lemma 26, there exists C_1 depending on C_X such that

$$\mathbb{P}(|Z_t| \geq \tau) \leq 2\exp(-(\tau/C_1)^{\gamma_2/2}).$$

For any $t \in \mathbb{Z}$, we have that

$$\begin{aligned}
\delta^Z_{s,2} =&\|Z_t - Z_{t,\{t-s\}}\|_2 = \|X_{tj}X_t^\top\alpha_t - X_{tj,\{t-s\}}X_{t,\{t-s\}}^\top\alpha_t\|_2\\
\leq&\|(X_{tj} - X_{tj,\{t-s\}})X_t^\top\alpha_t\|_2 + \|X_{tj,\{t-s\}}(X_t^\top\alpha_t - X_{tj,\{t-s\}}^\top\alpha_t)\|_2\\
\leq&\|X_{tj} - X_{tj,\{t-s\}}\|_4\|X_t^\top\alpha_t\|_4 + \|X_{tj,\{t-s\}}\|_4\|X_t^\top\alpha_t - X_{tj,\{t-s\}}^\top\alpha_t\|_4\\
\leq&2\delta^X_{s,4}\sup_{|v|_2=1}\|v^\top X_0\|_4.
\end{aligned}$$

As a result, it holds that

$$\Delta^Z_{m,2} = \sum_{s=m}^{\infty}\delta^Z_{s,2} \leq 2\Delta^X_{s,4}\sup_{|v|_2=1}\|v^\top X_0\|_4.$$

Consequently,

$$\sup_{m\geq 0}\exp(cm^{\gamma_1})\Delta^Z_{m,2} \leq 2D_X\sup_{|v|_2=1}\|v^\top X_0\|_4 < \infty.$$

By Theorem 4, it holds that for any $\tau \geq 1$ and integer $n \geq 3$,

$$\mathbb{P}\bigg(\bigg|\sum_{t=1}^{n} X_{tj}X_t^\top\alpha_t\bigg| \geq \tau\bigg) = \mathbb{P}\bigg(\bigg|\sum_{t=1}^{n} Z_i\bigg| \geq \tau\bigg) \leq n\exp\Big(-c_1\tau^\gamma\Big) + 2\exp\Big(-\frac{c_2\tau^2}{n}\Big),$$

where $\mathbb{E}(\sum_{t=1}^{n} Z_t) = 0$ is used in the equality. By the union bound argument, it follows that for any $\tau \geq 1$ and integer $n \geq 3$,

$$\mathbb{P}\bigg(\bigg|\sum_{t=1}^{n} X_t X_t^\top\alpha_t\bigg|_\infty \geq \tau\bigg) \leq np\exp\Big(-c_1\tau^\gamma\Big) + 2p\exp\Big(-\frac{c_2\tau^2}{n}\Big).$$

Therefore with a sufficiently large constant C_2,

$$\mathbb{P}\bigg(\bigg|\frac{1}{n}\sum_{t=1}^{n} X_t X_t^\top\alpha_t\bigg|_\infty \geq C_2\bigg\{\sqrt{\frac{\log(pn)}{n}} + \frac{\log^{1/\gamma}(np)}{n}\bigg\}\bigg) \leq n^{-5}.$$

$\square$

Throughout the rest of this subsection, denote

$$\mathbb{B}_0(a) = \{v \in \mathbb{R}^p : |v|_0 \le a\}, \quad \mathbb{B}_1(a) = \{v \in \mathbb{R}^p : |v|_1 \le a\} \quad \text{and} \quad \mathbb{B}_2(a) = \{v \in \mathbb{R}^p : |v|_2 \le a\}.$$

Theorem 32 (REC Version I)**.** *Denote $\widehat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n} X_t X_t^\top$. Under Assumption 2, it holds that for $n \ge 3$,*

$$\mathbb{P}\biggl(v^\top \widehat{\Sigma} v \ge \frac{\Lambda_{\min}(\Sigma)}{2}|v|_2^2 - C\frac{\log(pn)}{n^\gamma}|v|_1^2 \quad \textit{for all } v \in \mathbb{R}^p\biggr) \ge 1 - 2n^{-5},$$

where $C > 0$ is an absolute constant.

Proof. By Lemma 35 for any integer $s \ge 1$, with $\mathcal{K}(s) = \mathbb{B}_0(s) \cap \mathbb{B}_2(1)$, there exist sufficiently large constants C_1, C_2 such that

$$\mathbb{P}\biggl\{\sup_{v \in \mathcal{K}(2s)} \Bigl|v^\top(\widehat{\Sigma} - \Sigma)v\Bigr| \ge \frac{\tau}{n}\biggr\} \le \exp\Bigl(-c_1\tau^\gamma + \log(n) + C_1 s\log(p)\Bigr) + 2\exp\Bigl(-\frac{c_2\tau^2}{n} + C_2 s\log(p)\Bigr).$$

By Lemma 12 in the supplementary material of Loh and Wainwright (2012), it follows that for any $\tau \ge 1$,

$$\begin{aligned}&\mathbb{P}\biggl(\Bigl|v^\top(\widehat{\Sigma} - \Sigma)v\Bigr| \le \frac{C_3\tau}{n}\bigl\{|v|_2^2 + \frac{1}{s}|v|_1^2\bigr\} \quad \text{for all } v \in \mathbb{R}^p\biggr)\\ \ge &1 - \exp\Bigl(-c_1\tau^\gamma + \log(n) + C_1 s\log(p)\Bigr) - 2\exp\Bigl(-\frac{c_2\tau^2}{n} + C_2 s\log(p)\Bigr).\end{aligned}$$

For s to be chosen later, set

$$\tau = \tau_s = C_4\biggl\{\sqrt{ns\log(pn)} + (s\log(pn))^{1/\gamma}\biggr\},$$

for some sufficiently large constant C_4. Since $\tau_s \ge 1$, with probability at least $1 - 2n^{-5}$, it follows for all $v \in \mathbb{R}^p$ that

$$\Bigl|v^\top(\widehat{\Sigma} - \Sigma)v\Bigr| \le C_5\biggl\{\sqrt{\frac{s\log(pn)}{n}} + \frac{(s\log(pn))^{1/\gamma}}{n}\biggr\}|v|_2^2 + C_5\biggl\{\sqrt{\frac{\log(pn)}{sn}} + \frac{(s\log(pn))^{1/\gamma}}{sn}\biggr\}|v|_1^2.$$

Set

$$s = \min\left\{\left\lfloor\frac{\Lambda_{\min}^2(\Sigma)n}{16C_5^2\log(pn)}\right\rfloor, \left\lfloor\frac{\Lambda_{\min}^\gamma(\Sigma)n^\gamma}{4^\gamma C_5^\gamma\log(pn)}\right\rfloor\right\}.$$

Since $\gamma < 1$, $s \asymp (\frac{n^\gamma}{\log(pn)})$. With probability at least $1 - 2n^{-5}$, it holds for all $v \in \mathbb{R}^p$ that

$$\Bigl|v^\top(\widehat{\Sigma} - \Sigma)v\Bigr| \le \frac{\Lambda_{\min}(\Sigma)}{2}|v|_2^2 + C_6\biggl(\frac{\log(pn)}{n^{(1+\gamma)/2}} + \frac{\log(pn)}{n^\gamma}\biggr)|v|_1^2 \le \frac{\Lambda_{\min}(\Sigma)}{2}|v|_2^2 + C_7\frac{\log(pn)}{n^\gamma}|v|_1^2.$$

The desired result follows from the assumption that $\gamma < 1$ and the observation that $v^\top\Sigma v \ge \Lambda_{\min}|v|_2^2$ for all $v \in \mathbb{R}^p$. $\square$

Theorem 33 (REC Version II)**.** *Let $\alpha > 1$. Denote*

$$\mathcal{C} = \{v \in \mathbb{R}^p : |v_{U^c}|_1 \le \alpha |v_U|_1 \quad \textit{for any } U \subset [1, \dots, p] \textit{ such that } |U| = \mathfrak{s}\}.$$

Suppose Assumption 2 holds.
a. *There exists an absolute constant $C > 0$ such that for any integer $n \ge 3$,*

$$\mathbb{P}\bigg\{ \sup_{v \in \mathcal{C}} \Big| v^\top (\widehat{\Sigma} - \Sigma) v \Big| \ge C |v|_2^2 \bigg(\sqrt{\frac{\mathfrak{s} \log(np)}{n}} + \frac{\{\mathfrak{s} \log(np)\}^{1/\gamma}}{n} \bigg) \bigg\} \le n^{-5}.$$

b. *Suppose in addition that*

$$n \ge \zeta = C_\zeta \{\mathfrak{s} \log(pn)\}^{2/\gamma - 1},$$

for sufficiently large $C_\zeta > 0$. Then there exists an absolute constant $C > 0$ such that

$$\mathbb{P}\bigg\{ \sup_{v \in \mathcal{C}} \Big| v^\top (\widehat{\Sigma} - \Sigma) v \Big| \ge C |v|_2^2 \sqrt{\frac{\mathfrak{s} \log(pn)}{n}} \bigg\} \le n^{-5}.$$

Consequently, it holds that

$$\mathbb{P}\bigg\{ v^\top \widehat{\Sigma} v \ge \frac{\Lambda_{\min}(\Sigma)}{2} |v|_2^2 \quad \textit{for all } v \in \mathcal{C} \bigg\} \ge 1 - n^{-5}. \tag{120}$$

Proof. For any set $A \subset \mathbb{R}^p$, let $\mathrm{Cl}(A)$ denote the closure of A and $\mathrm{conv}(A)$ denote the convex hull of A. Let

$$\mathcal{K}(\mathfrak{s}) = \mathbb{B}_0(\mathfrak{s}) \cap \mathbb{B}_2(1).$$

By Lemma 35 for any $\tau \ge 1$, there exist sufficiently large constants C_1, C_2 such that

$$\mathbb{P}\bigg\{ \sup_{v \in \mathcal{K}(2\mathfrak{s})} \Big| v^\top (\widehat{\Sigma} - \Sigma) v \Big| \ge \frac{\tau}{n} \bigg\} \le \exp\Big(- c_1 \tau^\gamma + \log(n) + C_1 \mathfrak{s} \log(p) \Big) + 2 \exp\Big(- \frac{c_2 \tau^2}{n} + C_2 \mathfrak{s} \log(p) \Big).$$

By Lemma F.1 in the supplementary of Basu and Michailidis (2015)

$$\mathcal{C} \cap \mathbb{B}_2(1) \subset (\alpha + 2)\, \mathrm{Cl}\{\mathrm{conv}(\mathcal{K}(\mathfrak{s}))\},$$

it follows that

$$\sup_{v \in \mathcal{C} \cap \mathbb{B}_2(1)} \Big| v^\top (\widehat{\Sigma} - \Sigma) v \Big| \le \sup_{v \in (\alpha+2)\, \mathrm{Cl}\{\mathrm{conv}(\mathcal{K}(\mathfrak{s}))\}} \Big| v^\top (\widehat{\Sigma} - \Sigma) v \Big|.$$

By Lemma F.3 in the supplementary of Basu and Michailidis (2015)

$$\sup_{v \in \mathrm{Cl}\{\mathrm{conv}(\mathcal{K}(\mathfrak{s}))\}} |v^\top D v| \le 3 \sup_{v \in \mathcal{K}(2\mathfrak{s})} |v^\top D v|.$$

Therefore

$$\sup_{v \in (\alpha+2)\, \mathrm{Cl}\{\mathrm{conv}(\mathcal{K}(\mathfrak{s}))\}} \Big| v^\top (\widehat{\Sigma} - \Sigma) v \Big| \le 3(\alpha + 2)^2 \sup_{v \in \mathcal{K}(2\mathfrak{s})} |v^\top D v|.$$

So

$$\frac{1}{3(\alpha + 2)^2} \sup_{v \in \mathcal{C} \cap \mathbb{B}_2(1)} \Big| v^\top (\widehat{\Sigma} - \Sigma) v \Big| \le \sup_{v \in \mathcal{K}(2\mathfrak{s})} |v^\top D v|.$$

It follows that

$$\mathbb{P}\Big\{\sup_{v\in\mathcal{C}\cap\mathbb{B}_2(1)}\Big|v^\top(\widehat{\Sigma}-\Sigma)v\Big|\geq C_1\frac{\tau}{n}\Big\}\leq \exp\Big(-c_1\tau^\gamma+\log(n)+C_1\mathfrak{s}\log(p)\Big)+2\exp\Big(-\frac{c_2\tau^2}{n}+C_2\mathfrak{s}\log(p)\Big).$$

where $C_1=3(\alpha+2)^2$. For some sufficiently large C_2, let

$$\frac{\tau}{n}=C_2\Big\{\sqrt{\frac{\mathfrak{s}\log(pn)}{n}}+\frac{\{\mathfrak{s}\log(np)\}^{1/\gamma}}{n}\Big\},$$

Therefore

$$\mathbb{P}\Big\{\sup_{v\in\mathcal{C}\cap\mathbb{B}_2(1)}\Big|v^\top(\widehat{\Sigma}-\Sigma)v\Big|\geq 2C_2\Big(\sqrt{\frac{\mathfrak{s}\log(pn)}{n}}+\frac{\{\mathfrak{s}\log(np)\}^{1/\gamma}}{n}\Big)\Big\}\leq 2\exp\Big(-5\mathfrak{s}\log(pn)\Big)\leq n^{-5}.$$

In addition, if $n\geq C_\zeta\{\mathfrak{s}\log(pn)\}^{2/\gamma-1}$, then

$$\frac{\tau}{n}=C_2\Big\{\sqrt{\frac{\mathfrak{s}\log(pn)}{n}}+\frac{\{\mathfrak{s}\log(np)\}^{1/\gamma}}{n}\Big\}\leq 2C_2\sqrt{\frac{\mathfrak{s}\log(pn)}{n}},$$

and therefore

$$\mathbb{P}\Big\{\sup_{v\in\mathcal{C}\cap\mathbb{B}_2(1)}\Big|v^\top(\widehat{\Sigma}-\Sigma)v\Big|\geq 2C_2\sqrt{\frac{\mathfrak{s}\log(pn)}{n}}\Big\}\leq 2\exp\Big(-5\mathfrak{s}\log(pn)\Big)\leq n^{-5}. \tag{121}$$

Finally, Equation (120) is a direct consequence of Equation (121) and the fact that

$$n\geq C_\zeta\{\mathfrak{s}\log(pn)\}^{2/\gamma-1}\geq C_\zeta\mathfrak{s}\log(pn).$$

□

Definition 1. *Let V be a subset of $\mathbb{R}^p$. The ϵ-net $\mathcal{N}_\epsilon$ of V is a subset of V such that for any $v\in V$, there exists a vector $w\in\mathcal{N}_\epsilon$ such that*

$$|v-w|_2\leq\epsilon.$$

Lemma 34. *Let $\frac{1}{\gamma}=\frac{1}{\gamma_1}+\frac{2}{\gamma_2}$. Under Assumption 2, for any $\tau\geq 1$ and any non-random $v\in\mathbb{R}^p$, it holds that*

$$\mathbb{P}\Big\{\big|v^\top(\widehat{\Sigma}-\Sigma)v\big|\geq\frac{\tau}{n}\Big\}\leq n\exp\Big(-c_1\tau^\gamma\Big)+2\exp\Big(-\frac{c_2\tau^2}{n}\Big).$$

Proof. Note that

$$v^\top(\widehat{\Sigma}-\Sigma)v=\frac{1}{n}\sum_{t=1}^n(v^\top X_t)^2-v^\top\Sigma v.$$

Let $Z_t=(v^\top X_t)^2-v^\top\Sigma v$. Then $\mathbb{E}(Z_t)=0$. By Lemma 26, there exists C_1 depending on C_X such that

$$\mathbb{P}(|Z_t|\geq\tau)\leq 2\exp(-(\tau/C_1)^{\gamma_2/2}).$$

For any $t \in \mathbb{Z}$, we have that

$$\begin{aligned}\|Z_t - Z_{t,\{t-s\}}\|_2 =& \|(v^\top X_t)^2 - (v^\top X_{t,\{t-s\}})^2\|_2 \\ \leq& \big(\|v^\top X_t\|_4 + \|v^\top X_{t,\{t-s\}}\|_4\big)\|v^\top X_t - v^\top X_{t,\{t-s\}}\|_4 \\ \leq& 2\delta^X_{s,4} \sup_{|v|_2=1} \|v^\top X_0\|_4.\end{aligned}$$

The rest of the proof is similar to the proof of Lemma 31 and is omitted for brevity. $\square$

Lemma 35. *For any integer $s \in \mathbb{Z}_+$, denote $\mathcal{K}(s) = B_0(s) \cap B_2(1)$. Under Assumption 2, for any $\tau \geq 1$ and any $s \in \mathbb{Z}_+$, there exist sufficiently large constants C_1, C_2 such that*

$$\mathbb{P}\bigg\{ \sup_{v \in \mathcal{K}(s)} \bigg| v^\top (\widehat{\Sigma} - \Sigma) v \bigg| \geq \frac{\tau}{n} \bigg\} \leq \exp\Big(-c_1\tau^\gamma + \log(n) + C_1 s \log(p)\Big) + 2\exp\Big(-\frac{c_2\tau^2}{n} + C_2 s\log(p)\Big),$$

Proof. Denote $D = \widehat{\Sigma} - \Sigma$.

Step 1. For any vector v, let $\text{supp}(v)$ be the support of v. Let $U \subset [1, \ldots, p]$ be such that $|U| = s$. Denote

$$S_U = \{v \in \mathbb{R}^p : |v|_2 \leq 1, \text{supp}(v) \subset U\}.$$

Let $\mathcal{N}_{1/10}$ be the 1/10-net of S_U. Then by standard covering lemma, $|\mathcal{N}_{1/10}| \leq 21^s$. For every $v \in S_U$, there exists some $u_i \in \mathcal{N}_{1/10}$ such that $|v - u_i|_2 \leq 1/10$. Denote

$$\varphi = \sup_{v \in S_U} |v^\top D v|.$$

Then for any $v \in S_U$, we have

$$v^\top D v = (v - u_i + u_i)^\top D(v - u_i + u_i) = u_i^\top D u_i + 2u_i^\top D(v - u_i) + (v - u_i)^\top D(v - u_i).$$

and therefore

$$|v^\top D v| \leq |u_i^\top D u_i| + 2|u_i^\top D(v - u_i)| + |(v - u_i)^\top D(v - u_i)|.$$

Note that

$$|u_i^\top D u_i| \leq \sup_{u \in \mathcal{N}_{1/10}} |u^\top D u|.$$

Since $10(v - u_i) \in S_U$, it follows that

$$|(v - u_i)^\top D(v - u_i)| \leq \varphi/100.$$

In addition, denote $\Delta v = 10(v - u_i)$. Then $\Delta v \in S_U$.

$$\begin{aligned}2\bigg|u_i^\top D\Delta v\bigg| =& \bigg|(\Delta v + u_i)^\top D(\Delta v + u_i) - u_i^\top D u_i - \Delta v^\top D\Delta v\bigg| \\ =& \bigg|4(\frac{\Delta v + u_i}{2})^\top D(\frac{\Delta v + u_i}{2})\bigg| + \bigg|u_i^\top D u_i\bigg| + \bigg|\Delta v^\top D\Delta v\bigg| \\ \leq& 4\varphi + \sup_{u \in \mathcal{N}_{1/10}} |u^\top D u| + \varphi,\end{aligned}$$

where $\frac{\Delta v + u_i}{2} \in S_U$ because $|\Delta v + u_i|_2 \leq 2$. So

$$2|u_i^\top D(v - u_i)| = \frac{1}{10}\bigg|u_i^\top D\Delta v\bigg| \leq \frac{4}{10}\varphi + \frac{\sup_{u\in\mathcal{N}_{1/10}} |u^\top Du|}{10} + \frac{\varphi}{10}.$$

Putting these calculations together, it holds that for any $v \in S_U$,

$$\begin{aligned}|v^\top Dv| \leq& |u_i^\top Du_i| + 2|u_i^\top D(v - u_i)| + |(v - u_i)^\top D(v - u_i)| \\ \leq& \sup_{u\in\mathcal{N}_{1/10}} |u^\top Du| + \frac{4}{10}\varphi + \frac{\sup_{u\in\mathcal{N}_{1/10}} |u^\top Du|}{10} + \frac{\varphi}{10} + \varphi/100.\end{aligned}$$

Taking the sup over all $v \in S_U$,

$$\varphi = \sup_{v\in S_U} |v^\top Dv| \leq \sup_{u\in\mathcal{N}_{1/10}} |u^\top Du| + \frac{4}{10}\varphi + \frac{\sup_{u\in\mathcal{N}_{1/10}} |u^\top Du|}{10} + \frac{\varphi}{10} + \varphi/100.$$

Rearranging this gives

$$\varphi(1 - \frac{4}{10} - \frac{1}{10} - \frac{1}{100}) \leq \frac{11}{10} \sup_{u\in\mathcal{N}_{1/10}} |u^\top Du|,$$

which leads to

$$\sup_{v\in S_U} |v^\top Dv| = \varphi \leq 3 \sup_{u\in\mathcal{N}_{1/10}} |u^\top Du|.$$

Step 2. It suffices to consider

$$\sup_{u\in\mathcal{N}_{1/10}} |u^\top Du|.$$

Note that by union bound,

$$\begin{aligned}\mathbb{P}\bigg(\sup_{u\in\mathcal{N}_{1/10}} |u^\top Du| \geq \tau\bigg) \leq& |\mathcal{N}_{1/10}| \sup_{u\in\mathcal{N}_{1/10}} \mathbb{P}(|u^\top Du| \geq \tau) \\ \leq& n21^s \exp\Big(-c_1\tau^\gamma\Big) + 2 \cdot 21^s \exp\Big(-\frac{c_2\tau^2}{n}\Big),\end{aligned}$$

where the second inequality follows from Lemma 34. It follows from **Step 1** that

$$\mathbb{P}\bigg(\sup_{v\in S_U} |v^\top Dv| \geq 3\tau\bigg) \leq n21^s \exp\Big(-c_1\tau^\gamma\Big) + 2 \cdot 21^s \exp\Big(-\frac{c_2\tau^2}{n}\Big),$$

Step 3. Note that

$$\mathcal{K}(s) = \bigcup_{|U|=s} S_U.$$

There are at most $\binom{p}{s} \leq p^s$ choices of U. So taking another union bound, we have

$$\mathbb{P}\bigg(\sup_{v\in\mathcal{K}(s)} |u^\top Du| \geq 3\tau\bigg) \leq np^s21^s \exp\Big(-c_1\tau^\gamma\Big) + 2p^s21^s \exp\Big(-\frac{c_2\tau^2}{n}\Big).$$

This immediately leads to the desired result. $\square$

E.3 Uniform deviation bounds under dependence

Without loss of generality, assume that $p \geq n^{\alpha}$ for some $\alpha > 0$.

Lemma 36. *Suppose Assumption 2 and Assumption 3 hold. Then*

$$\mathbb{P}\left(\left|\sum_{t=1}^{r} X_t \epsilon_t\right|_{\infty} \geq C_2\left(\sqrt{r\log(p)} + \log^{1/\gamma}(p)\right) \textit{ for all } 3 \leq r \leq n\right) \leq n^{-4}$$

Suppose in addition $u \in \mathbb{R}^p$ is a deterministic vector such that $|u|_2 = 1$. Then it holds that

$$\mathbb{P}\left(\left|\sum_{t=1}^{r} u^{\top} X_t X_{tj} - u^{\top}\Sigma_{\cdot j}\right| \geq C_1\left(\sqrt{r\log(np)} + \log^{1/\gamma}(np)\right) \textit{ for all } 3 \leq r \leq n\right) \leq n^{-4}.$$

Proof. For the first probability bound, by Lemma 30, for any $r \in \{3, 4, \ldots, n\}$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{r} X_i \epsilon_i\right|_{\infty} \geq C_2\left(\left(\sqrt{r\log(p)} + \log^{1/\gamma}(p)\right)\right)\right) \leq n^{-5}$$

So by a union bound,

$$\mathbb{P}\left(\left|\sum_{i=1}^{r} X_i \epsilon_i\right|_{\infty} \geq C_2\left(\sqrt{r\log(p)} + \log^{1/\gamma}(p)\right) \text{ for all } 3 \leq r \leq n\right) \leq n^{-4}$$

For the second probability bound, for fixed $j \in [1, \ldots, p]$, let

$$W_t = u^{\top} X_t X_{tj} - u^{\top}\Sigma_{\cdot j},$$

where $\Sigma_{\cdot j}$ denote the j-th column of Σ. For any $t \in \mathbb{Z}$, we have that

$$\begin{aligned}
|W_t - W_{t,\{t-s\}}|_2 =& |u^{\top} X_t X_{tj} - u^{\top}\Sigma_{\dot{j}} - \left\{u^{\top} X_{t,\{t-s\}} X_{tj,\{t-s\}} - u^{\top}\Sigma_{\dot{j}}\right\}|_2 \\
\leq& |u^{\top} X_t X_{tj} - u^{\top} X_{t,\{t-s\}} X_{tj}|_2 - |u^{\top} X_{t,\{t-s\}} X_{tj} - u^{\top} X_{t,\{t-s\}} X_{tj,\{t-s\}}|_2 \\
\leq& |X_{tj}|_4 |u^{\top} X_t - u^{\top} X_{t,\{t-s\}}|_4 + |X_{tj} - X_{tj,\{t-s\}}|_4 |u^{\top} X_{t,\{t-s\}}|_4 \\
\leq& \Delta^X_{0,4}\delta^X_{s,4} + \delta^X_{s,4}\Delta^X_{0,4}.
\end{aligned}$$

As a result, it holds that

$$\delta^W_{s,2} := \sup_{t\in\mathbb{Z}} |W_t - W_{t,\{t-s\}}|_2 \leq \Delta^X_{0,4}\delta^X_{s,4} + \delta^X_{s,4}\Delta^X_{0,4}$$

and that

$$\Delta^W_{m,2} = \sum_{s=m}^{\infty} \delta^W_{s,2} \leq \Delta^X_{0,4}\sum_{s=m}^{\infty}\delta^X_{s,4} + \Delta^X_{0,4}\sum_{s=m}^{\infty}\delta^X_{s,4} = 2\Delta^X_{0,4}\Delta^X_{m,4}.$$

Consequently,

$$\sup_{m\geq 0} \exp(cm^{\gamma_1})\Delta^Z_{m,2} \leq 2\Delta^X_{0,4} \sup_{m\geq 0} \exp(cm^{\gamma_1})\Delta^X_{m,4} \leq 2D^2_X.$$

Let $\frac{1}{\gamma} = \frac{1}{\gamma_1} + \frac{2}{\gamma_2}$. By Theorem 4, it holds that for any $\tau \geq 1$ and $r \in \{3, 4, \ldots, n\}$,

$$\mathbb{P}\Big(\Big|\sum_{t=1}^{r} u^\top X_t X_{tj} - u^\top \Sigma_{\cdot j}\Big| \geq \tau\Big) = \mathbb{P}\Big(\Big|\sum_{t=1}^{r} Z_t\Big| \geq \tau\Big) \leq r \exp\Big(-c_1 \tau^\gamma\Big) + 2\exp\Big(-\frac{c_2 \tau^2}{r}\Big).$$

By union bound, it follows that for any $\tau \geq 1$,

$$\mathbb{P}\Big(\Big|\sum_{t=1}^{r} u^\top X_t X_{tj} - u^\top \Sigma_{\cdot j}\Big| \geq \tau\Big) \leq rp \exp\Big(-c_1 \tau^\gamma\Big) + 2p\exp\Big(-\frac{c_2 \tau^2}{r}\Big).$$

Therefore, with sufficiently large constant C_2,

$$\mathbb{P}\Big(\Big|\sum_{t=1}^{r} u^\top X_t X_{tj} - u^\top \Sigma_{\cdot j}\Big|_\infty \geq C_2\Big\{\sqrt{r\log(pn)} + \log^{1/\gamma}(pn)\Big\}\Big) \leq n^{-5}.$$

By another union bound,

$$\mathbb{P}\Big(\Big|\sum_{t=1}^{r} u^\top X_t X_{tj} - u^\top \Sigma_{\cdot j}\Big| \geq C_1\Big(\sqrt{r\log(np)} + \log^{1/\gamma}(np)\Big) \text{ for all } 3 \leq r \leq n\Big) \leq n^{-4}.$$

□

Lemma 37. *Suppose Assumption 2 holds. Let $\alpha > 1$ and*

$$\mathcal{C} = \{v \in \mathbb{R}^p : |v_{U^c}|_1 \leq \alpha |v_U|_1 \quad \text{for any } U \subset [1, \ldots, p] \text{ such that } |U| = \mathfrak{s}\}.$$

Then there exists an absolute constant $C > 0$ such that

$$\mathbb{P}\Big\{\Big|v^\top\Big\{\sum_{i=1}^{r} X_i X_i^\top - \Sigma\Big\}v\Big| \geq C|v|_2^2\Big(\sqrt{r\mathfrak{s}\log(np)} + \{\mathfrak{s}\log(np)\}^{1/\gamma}\Big)\ \forall r \in \{3, \ldots, n\}\ \forall v \in \mathcal{C}\Big\} \leq n^{-4}.$$

Proof. By Theorem 33,

$$\mathbb{P}\Big\{\sup_{v\in\mathcal{C}}\Big|v^\top\Big\{\sum_{i=1}^{r} X_i X_i^\top - \Sigma\Big\}v\Big| \geq C|v|_2^2\Big(\sqrt{r\mathfrak{s}\log(np)} + \{\mathfrak{s}\log(np)\}^{1/\gamma}\Big)\ \forall v \in \mathcal{C}\Big\} \leq n^{-5}.$$

The desired result follows from a union bound. □

F Proof of Theorem 4

F.1 Introduction of nonstationary processes

Let $\{Z_t\}_{t=\mathbb{Z}}$ be an $\mathbb{R}$-valued (possibly) nonstationary process, such that

$$Z_t = g_t(\mathcal{F}_t), \tag{122}$$

where $\{g_t(\cdot)\}_{t\in\mathbb{Z}}$ are time-dependent $\mathbb{R}$-valued measurable functions and $\mathcal{F}_t = \{\epsilon_s\}_{s\leq t}$ for any $t \in \mathbb{Z}$ and $\mathcal{F}_{-\infty} = \{\emptyset, \Omega\}$. Let $s_1, s_2 \in \mathbb{Z}$, such that $s_2 \leq s_1 \leq t$. Let ϵ_t^* be an independent copy of ϵ_t.

For notational convenience, let $\{\epsilon_{s_1+1},\ldots,\epsilon_t\}=\emptyset$ when $s_1+1>t$. Define $Z_{t,\{s_1,s_2\}}=g_t(\mathcal{F}_{t,\{s_1,s_2\}})$ with $\mathcal{F}_{t,\{s_1,s_2\}}=\{\ldots,\epsilon_{s_2-1},\epsilon^*_{s_2},\ldots,\epsilon^*_{s_1},\epsilon_{s_1+1},\ldots,\epsilon_t\}$. Write $Z_{t,\{s,s\}}=Z_{t,\{s\}}$. Under the general nonstationary representation (122), the functional dependence measure and its tail cumulative version are defined respectively as

$$\delta_{s,q}=\sup_{t\in\mathbb{Z}}\|Z_t-Z_{t,\{t-s\}}\|_q \text{ and } \Delta_{m,q}=\sum_{s=m}^{\infty}\delta_{s,q}.$$

Note that in the special case when $g_t(\cdot)=g(\cdot)$, for all $t=1,\ldots,n$, the representation (122) reduces to that of a stationary time series

$$Z_t=g(\mathcal{F}_t),$$

and the functional dependence measure and related conditions can be defined accordingly.

F.2 Auxiliary lemmas

The first lemma can be used to relate the Laplace transform of a partial sum to the products of the Laplace transforms of each individual summand.

Lemma 38. *Let $\{Z_i\}_{i\in\mathbb{Z}}$ be an $\mathbb{R}$-valued possibly nonstationary process of form (122). Assume there exists a positive M such that $|Z_i|\le M$ for any $i\in\mathbb{Z}$. Then for any $a>0$, we have*

$$\begin{aligned}\left|\mathbb{E}\Big[\exp\Big(a\sum_{i=1}^n Z_i\Big)\Big]-\prod_{i=1}^n\mathbb{E}[\exp(aZ_i)]\right| &\le a\exp(anM)\sum_{i=2}^n\|Z_i-Z_{i,\{i-1,-\infty\}}\|_2\\ &\le a\exp(anM)(n-1)\Delta_{1,2}.\end{aligned}$$

Proof. For the product of the Laplace transforms of each individual random variable, we have the telescoping decomposition as

$$\begin{aligned}\prod_{i=1}^n\mathbb{E}[\exp(aZ_i)]=&\,\mathbb{E}\Big[\exp\Big(a\sum_{i=1}^{n-1}Z_i\Big)\Big]\mathbb{E}[\exp(aZ_n)]\\ &+\mathbb{E}\Big[\exp\Big(a\sum_{i=1}^{n-2}Z_i\Big)\Big]\prod_{j=n-1}^{n}\mathbb{E}[\exp(aZ_j)]-\mathbb{E}\Big[\exp\Big(a\sum_{i=1}^{n-1}Z_i\Big)\Big]\mathbb{E}[\exp(aZ_n)]\\ &+\mathbb{E}\Big[\exp\Big(a\sum_{i=1}^{n-3}Z_i\Big)\Big]\prod_{j=n-2}^{n}\mathbb{E}[\exp(aZ_j)]-\mathbb{E}\Big[\exp\Big(a\sum_{i=1}^{n-2}Z_i\Big)\Big]\prod_{j=n-1}^{n}\mathbb{E}[\exp(aZ_n)]\\ &+\ldots\\ &+\prod_{j=1}^{n}\mathbb{E}[\exp(aZ_j)]-\mathbb{E}\Big[\exp\Big(a\sum_{i=1}^{2}Z_i\Big)\Big]\prod_{j=3}^{n}\mathbb{E}[\exp(aZ_n)].\end{aligned}$$

For notational simplicity, given a real value sequence $\{b_i\}_{i=1}^n$, we write $\prod_{i=n+1}^n b_i=1$. Then, it satisfies that

$$\mathbb{E}\Big[\exp\Big(a\sum_{i=1}^n Z_i\Big)\Big]-\prod_{i=1}^n\mathbb{E}[\exp(aZ_i)]$$

$$= \sum_{s=2}^{n} \left\{ \left(\mathbb{E}\Big[\exp\Big(a \sum_{i=1}^{s} Z_i \Big) \Big] - \mathbb{E}\Big[\exp\Big(a \sum_{i=1}^{s-1} Z_i \Big) \Big] \mathbb{E}[\exp(aZ_s)] \right) \prod_{j=s+1}^{n} \mathbb{E}[\exp(aZ_j)] \right\}. \tag{123}$$

Using coupling, we have that $Z_{s,\{s-1,-\infty\}}$ and $Z_{s'}$ are independent for any $s' \le s-1$, and $Z_{s,\{s-1,-\infty\}}$ and Z_s have the same distribution. We have that

$$\begin{aligned}
&\left| \mathbb{E}\Big[\exp\Big(a \sum_{i=1}^{s} Z_i \Big) \Big] - \mathbb{E}\Big[\exp\Big(a \sum_{i=1}^{s-1} Z_i \Big) \Big] \mathbb{E}[\exp(aZ_s)] \right| \prod_{j=s+1}^{n} \mathbb{E}[\exp(aZ_j)] \\
=& \left| \mathbb{E}\Big[\exp\Big(a \sum_{i=1}^{s-1} Z_i \Big) \Big(\exp(aZ_s) - \exp(aZ_{s,\{s-1,-\infty\}}) \Big) \Big] \right| \prod_{j=s+1}^{n} \mathbb{E}[\exp(aZ_j)] \\
\le& a \exp(anM) \mathbb{E}\big| Z_s - Z_{s,\{s-1,-\infty\}} \big| \le a \exp(anM) \big\| Z_s - Z_{s,\{s-1,-\infty\}} \big\|_2,
\end{aligned} \tag{124}$$

where the first inequality is due to the mean value theorem and the fact that Z_i are bounded, and the second inequality follows from Hölder's inequality. Combining (123) and (124), we have

$$\begin{aligned}
\left| \mathbb{E}\Big[\exp\Big(a \sum_{i=1}^{n} Z_i \Big) \Big] - \prod_{i=1}^{n} \mathbb{E}[\exp(aZ_i)] \right| \le& a \exp(anM) \sum_{i=2}^{n} \big\| Z_i - Z_{i,\{i-1,-\infty\}} \big\|_2 \\
\le& a(n-1) \exp(anM) \sum_{j=1}^{\infty} \delta_{j,2},
\end{aligned}$$

where the second inequality follows from the definition of $\delta_{s,q}$. □

The following lemma can be used to relate the bound of the log-Laplace transform of sum of random variables to that of each individual random variable. This lemma is the Lemma 3 in Merlevède et al. (2011), we reproduce it for completeness.

Lemma 39. *Let $Z_1, Z_2, \dots$ be a sequence of $\mathbb{R}$-valued random variables. Assume that there exist positive constants $\sigma_1, \sigma_2, \dots$ and $c_1, c_2, \dots$ such that, for any positive integer i and any $t \in [0, 1/c_i)$,*

$$\log \mathbb{E}[\exp(tZ_i)] \le (\sigma_i t)^2 / (1 - c_i t).$$

Then, for any positive n and any t in $[0, 1/(c_1 + c_2 + \cdots + c_n))$,

$$\log \mathbb{E}[\exp(t(Z_1 + Z_2 + \cdots + Z_n))] \le (\sigma t)^2 / (1 - Ct), \tag{125}$$

where $\sigma = \sigma_1 + \sigma_2 + \cdots + \sigma_n$ and $C = c_1 + c_2 + \cdots + c_n$.

Proof. For $i \ge 1$, denote the partial sums $S_i = \sum_{j=1}^{i} Z_j$. The proof is by induction. For $n = 1$, we have $\sigma = \sigma_1$ and $C = c_1$, and (125) holds obviously.

Assuming (125) holds for $n = k$, i.e. for any $t \in [0, 1/(c_1 + c_2 + \cdots + c_k))$, it satisfies that

$$\log \mathbb{E}[\exp(tS_k)] \le \Big(t \sum_{j=1}^{k} \sigma_j \Big)^2 \Big/ \Big(1 - t \sum_{j=1}^{k} c_j \Big).$$

For $n = k + 1$, by Hölder's inequality we have for any $u \in (0, 1)$ that

$$\log \mathbb{E}[\exp(t(S_k + Z_{k+1}))] \le u \log \mathbb{E}[\exp(u^{-1} t S_k)] + (1 - u) \log \mathbb{E}[\exp((1-u)^{-1} t Z_{k+1})]. \tag{126}$$

Choose

$$u = \Big(\sum_{j=1}^{k}\sigma_j \Big/ \sum_{j=1}^{k+1}\sigma_j\Big)\Big(1 - t\sum_{j=1}^{k+1} c_j\Big) + t\sum_{j=1}^{k} c_j,$$

and thus

$$1 - u = \Big(\sigma_{k+1} \Big/ \sum_{j=1}^{k+1}\sigma_j\Big)\Big(1 - t\sum_{j=1}^{k+1} c_j\Big) + tc_{k+1}.$$

Since $1/(c_1 + c_2 + \cdots + c_{k+1})$ is less than $1/(c_1 + c_2 + \cdots + c_k)$ and $1/c_{k+1}$, we have for any $t \in [0, 1/(c_1 + c_2 + \cdots + c_{k+1}))$ that

$$(126) \le \frac{t^2\big(\sum_{j=1}^{k}\sigma_j\big)^2}{u - t\sum_{j=1}^{k} c_j} + \frac{t^2\sigma_{k+1}^2}{(1-u) - tc_{k+1}} = \frac{\big(t\sum_{j=1}^{k+1}\sigma_j\big)^2}{1 - t\sum_{j=1}^{k+1} c_j},$$

which completes the proof. □

The following lemma is the Burkholder's moment inequality for martingale.

Lemma 40 (Theorem 2.1 in Rio (2009))**.** *Let $M_n = \sum_{i=1}^{n}\xi_i$, where $\{\xi_i\}_{i\in\mathbb{Z}}$ are martingale differences. Then for any $q \ge 2$,*

$$\|M_n\|_q \le \sqrt{q-1}\Big(\sum_{i=1}^{n}\|\xi_i\|_q^2\Big)^{1/2}.$$

For a possibly nonstationary process with dependence measure decays exponentially, the next lemma gives an upper bound on its long-run variance.

Lemma 41. *Let $\{X_i\}_{i\in\mathbb{Z}}$ be a sequence of $\mathbb{R}$-valued random variables in the form of (122). Assume $\sup_{i\in\mathbb{Z}}\|X_i\|_2 < \infty$ and there exist some absolute constants $c, C_{\mathrm{FDM}}, \gamma_1 > 0$ such that*

$$\sup_{m\ge 0}\exp(cm^{\gamma_1})\Delta_{m,2} \le C_{\mathrm{FDM}}.$$

Then, the long-run variance of $\{X_t\}_{t\in\mathbb{Z}}$ satisfies that

$$\lim_{n\to\infty}\mathrm{Var}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\Big) \le \sum_{l=-\infty}^{\infty}\sup_{t\in\mathbb{Z}}\big|\mathrm{Cov}(X_t, X_{t+l})\big| = C_{\mathrm{LRV}} < \infty. \tag{127}$$

Proof. Define the projection operator $\mathcal{P}_j\cdot = \mathbb{E}[\cdot|\mathcal{F}_j] - \mathbb{E}[\cdot|\mathcal{F}_{j-1}]$ with $j \in \mathbb{Z}$. A random variable X_i is decomposed as

$$X_i - \mathbb{E}[X_i] = \sum_{k=0}^{\infty}\big(\mathbb{E}[X_i|\mathcal{F}_{i-k}] - \mathbb{E}[X_i|\mathcal{F}_{i-k-1}]\big) = \sum_{k=0}^{\infty}\mathcal{P}_{i-k}X_i.$$

It holds for any $l \geq 0$ that

$$\begin{aligned}
&\sup_{t\in\mathbb{Z}} \big| \operatorname{Cov}(X_t, X_{t+l})\big| = \sup_{t\in\mathbb{Z}} \bigg|\mathbb{E}\Big[\Big(\sum_{k=0}^{\infty} \mathcal{P}_{-k}X_t\Big)\Big(\sum_{k=0}^{\infty} \mathcal{P}_{l-k}X_{t+l}\Big)\Big]\bigg| \\
=&\sup_{t\in\mathbb{Z}} \bigg|\sum_{k=0}^{\infty} \mathbb{E}\big[(\mathcal{P}_{-k}X_t)(\mathcal{P}_{-k}X_{t+l})\big]\bigg| \leq \sum_{k=0}^{\infty} \sup_{t\in\mathbb{Z}} \Big|\mathbb{E}\big[(\mathcal{P}_{-k}X_t)(\mathcal{P}_{-k}X_{t+l})\big]\Big| \\
\leq&\sum_{k=0}^{\infty} \sup_{t\in\mathbb{Z}} \|\mathcal{P}_{-k}X_t\|_2 \|\mathcal{P}_{-k}X_{t+l}\|_2 \leq \sum_{k=0}^{\infty} \delta_{k,2}\delta_{k+l,2} \leq \sqrt{\sum_{k=0}^{\infty} \delta_{k,2}^2}\sqrt{\sum_{k=0}^{\infty} \delta_{k+l,2}^2} \\
\leq&\Big(\sum_{k=0}^{\infty} \delta_{k,2}\Big)\Big(\sum_{k=l}^{\infty} \delta_{k,2}\Big) = \Delta_{0,2}\Delta_{l,2},
\end{aligned}$$

where the first inequality follows the triangle inequality and the second and fourth inequalities follow Hölder's inequality. The second equality also follows the orthogonality of $\mathcal{P}_j\cdot$, i.e. for $i < j$

$$\mathbb{E}[(\mathcal{P}_iX_r)(\mathcal{P}_jX_s)] = \mathbb{E}[\mathbb{E}[(\mathcal{P}_iX_r)(\mathcal{P}_jX_s)|\mathcal{F}_i]] = \mathbb{E}[(\mathcal{P}_iX_r)\mathbb{E}[X_s - X_s|\mathcal{F}_i]] = 0,$$

and the orthogonality also holds for $i > j$ by symmetry. The third inequality is due to the fact that

$$\begin{aligned}
\|\mathcal{P}_jX_i\|_2 =&\|\mathbb{E}[X_i|\mathcal{F}_j] - \mathbb{E}[X_i|\mathcal{F}_{j-1}]\|_2 = \|\mathbb{E}[X_i|\mathcal{F}_j] - \mathbb{E}[X_{i,\{j\}}|\mathcal{F}_{j-1}]\|_2 \\
=&\|\mathbb{E}[X_i - X_{i,\{j\}}|\mathcal{F}_j]\|_2 \leq \|X_i - X_{i,\{j\}}\|_2 \leq \delta_{i-j,2},
\end{aligned}$$

where the second and the third equality follows the definition of the coupled random variables $X_{i,\{j\}}$, the first inequality follows Jensen's inequality, and the second inequality follows the definition of the functional dependence measure. By the same arguments, we have for any $l < 0$ that

$$\sup_{t\in\mathbb{Z}} \big| \operatorname{Cov}(X_t, X_{t+l})\big| \leq \Delta_{0,2}\Delta_{-l,2}.$$

Therefore, we have that

$$C_{\mathrm{LRV}} = \sum_{l=-\infty}^{\infty} \sup_{t\in\mathbb{Z}} \big| \operatorname{Cov}(X_t, X_{t+l})\big| \leq 2\Delta_{0,2} \sum_{l=0}^{\infty} \Delta_{l,2} \leq 2C_{\mathrm{FDM}}^2 \sum_{l=0}^{\infty} \exp(-cl^{\gamma_1}). \tag{128}$$

To bound (128), we compare the series $\{a_m = \exp(-cm^{\gamma_1})\}_{m=1}^{\infty}$ and $\{b_m = (m+1)^{-\nu}\}_{m=1}^{\infty}$ with $\nu > 1$. Since by the properties of the Riemann zeta function, it holds that $\sum_{m=1}^{\infty} b_m < \infty$.

We embed the series $\{a_m\}_{m=1}^{\infty}$ and $\{b_m\}_{m=1}^{\infty}$ into continuous time processes by defining $a_x = a_{\lceil x\rceil}$ and $b_x = b_{\lceil x\rceil}$ for $x \in [1,\infty)$. Since for any $x \in [1,\infty)$, $a_x, b_x > 0$ and we define a function $g(x)$ on $x \in [1,\infty)$ as

$$g(x) = \log\left(\frac{a_x}{b_x}\right) = -cx^{\gamma_1} + \nu\log(x+1).$$

We have the derivative as

$$g'(x) = -c\gamma_1 x^{\gamma_1-1} + \frac{\nu}{x+1}.$$

Since for any absolute constants $c, \gamma_1 > 0$ and $\nu > 1$, there exists a finite $x^* \geq 1$ such that for any $x \geq x^*$, it holds that

$$\exp(-cx^{\gamma_1}) \leq (x+1)^{-\nu} \;\text{ and }\; g'(x) < 0.$$

Letting $m^* = \lceil x^* \rceil$, we have that

$$
\begin{aligned}
C_{\mathrm{LRV}} \leq & 2C_{\mathrm{FDM}}^2 \sum_{m=0}^{\infty} \exp(-cm^{\gamma_1}) = 2C_{\mathrm{FDM}}^2 \sum_{m=0}^{m^*} \exp(-cm^{\gamma_1}) + 2C_{\mathrm{FDM}}^2 \sum_{m=m^*+1}^{\infty} \exp(-cm^{\gamma_1}) \\
\leq & 2C_{\mathrm{FDM}}^2 \sum_{m=0}^{m^*} \exp(-cm^{\gamma_1}) + 2C_{\mathrm{FDM}}^2 \sum_{m=m^*+1}^{\infty} (m+1)^{-\nu} < \infty.
\end{aligned}
$$

□

We compare the following two conditions in the next lemma. (i) There exist some absolute constants $c', \nu > 0$ such that

$$
c' \sup_{m \geq 0} (m+1)^{\nu} \Delta_{m,2} \leq 1. \tag{129}
$$

(ii) There exist some absolute constants $c, C_{\mathrm{FDM}}, \gamma_1 > 0$ such that

$$
\sup_{m \geq 0} \exp(cm^{\gamma_1}) \Delta_{m,2} \leq C_{\mathrm{FDM}}. \tag{130}
$$

Lemma 42. *Let $\{X_t\}_{t \in \mathbb{Z}}$ be a sequence of $\mathbb{R}$-valued random variables in the form of (122). If there exist some absolute constants $c, \gamma_1 > 0$ such that (130) holds, then (129) holds for any $c', \nu > 0$.*

Proof. If (130) holds, then we have for any $m \geq 0$ that

$$
\Delta_{m,2} = \sum_{k=m}^{\infty} \delta_{k,2} \leq C_{\mathrm{FDM}} \exp(-cm^{\gamma_1}),
$$

and

$$
\delta_{m,2} = \sum_{k=m}^{\infty} \delta_{k,2} - \sum_{k=m+1}^{\infty} \delta_{k,2} \leq C_{\mathrm{FDM}} \exp(-cm^{\gamma_1}).
$$

This further leads to

$$
\sup_{m \geq 0} (m+1)^{\nu} \sum_{k=m}^{\infty} \delta_{k,2} \leq C_{\mathrm{FDM}} \sup_{m \geq 0} (m+1)^{\nu} \sum_{k=m}^{\infty} \exp(-ck^{\gamma_1}) \leq C_{\mathrm{FDM}} \sum_{k=0}^{\infty} (k+1)^{\nu} \exp(-ck^{\gamma_1}). \tag{131}
$$

Define the following function on $x \in [1, \infty)$ as

$$
g(x) = \log \left(\frac{\exp(-cx^{\gamma_1})}{(x+1)^{-\nu'}} \right) = -cx^{\gamma_1} + \nu' \log(x+1),
$$

where ν' is some absolute constant such that $\nu' - \nu > 1$. We have the derivative as

$$
g'(x) = -c\gamma_1 x^{\gamma_1 - 1} + \frac{\nu'}{x+1}.
$$

Since for any absolute constants $c, \gamma_1 > 0$ and $\nu' > \nu + 1$, there exists a finite $x^* \geq 1$ such that for any $x \geq x^*$, it holds that

$$
g'(x) < 0 \ \text{ and } \ \exp(-cx^{\gamma_1}) \leq (x+1)^{-\nu'}.
$$

Letting $k^* = \lceil x^* \rceil$, we have that

$$
\sup_{m \geq 0} (m+1)^{\nu} \sum_{k=m}^{\infty} \delta_{k,2} \leq C_{\mathrm{FDM}} \sum_{k=0}^{k^*} (k+1)^{\nu} \exp(-ck^{\gamma_1}) + C_{\mathrm{FDM}} \sum_{k=k^*+1}^{\infty} (k+1)^{-(\nu' - \nu)} < \infty.
$$

Thus, there exists such c'. □

F.3 Bernstein-type inequality for nonstationary processes

Theorem 43. *Let $\{X_i\}_{i\in\mathbb{Z}}$ be an $\mathbb{R}$-valued possibly nonstationary process of the form (122). Assume that $\mathbb{E}[X_i] = 0$ for any $i \in \mathbb{Z}$ and that there exist absolute constants $\gamma_1, \gamma_2, c, C_{\mathrm{FDM}} > 0$ such that*

$$\sup_{m\geq 0} \exp(cm^{\gamma_1})\Delta_{m,2} \leq C_{\mathrm{FDM}}, \tag{132}$$

for any $x > 0$

$$\sup_{i\in\mathbb{Z}} \mathbb{P}(|X_i| > x) \leq \exp(1 - x^{\gamma_2}), \tag{133}$$

and $1/\gamma = 1/\gamma_1 + 1/\gamma_2 > 1$. Then, we have for any $n \geq 3$ and $x > 0$,

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{i=1}^n X_i\Big| \geq x\Big) \leq & n \exp\Big(-\frac{x^\gamma}{C_1}\Big) + \exp\Big(-\frac{x^2}{C_2(1+nC_{\mathrm{LRV}})}\Big) \\ & + \exp\Big(-\frac{x^2}{C_3 n}\exp\Big(\frac{x^{\gamma(1-\gamma)}}{C_4(\log x)^\gamma}\Big)\Big),\end{aligned}$$

where $C_1, C_2, C_3, C_4 > 0$ are absolute constants and C_{LRV} is defined in (127).

Remark 2. For completeness, we give the proof of Theorem 43. The proof closely follows that of Theorem 1 in Merlevède et al. (2011). There are only two differences. (1) We replace Lemma 2 therein with Lemma 38, which is based on the functional dependence measure. (2) Lemma 5 in Dedecker and Prieur (2004) gives a coupling property of the τ-mixing coefficient, i.e. letting $\mathcal{M}$ be a sigma algebra and X be an integrable real-valued random variable, there exists a random variable $\tilde{X}$ independent of $\mathcal{M}$ and distributed as X such that $\mathbb{E}|X - \tilde{X}|$ equals the τ-mixing coefficient associated with $\mathcal{M}$. Instead, for random variables $\{X_t\}_{t\in\mathbb{Z}}$ of the form (122), we use coupling method (see e.g. Berkes et al., 2009b) to directly construct $X^{(l)}_{s,\{k,-\infty\}}$ with $s > k$, which is independent of $\mathcal{F}_k$ and distributed as X_s.

Before proving Theorem 43, we introduce the truncation argument.

Truncation. For $\mathbb{R}$-valued random variables $\{X_i\}_{i\in\mathbb{Z}}$, define the truncated and centered version with parameter $M > 0$ as

$$\overline{X}_M(i) = \psi_M(X_i) - \mathbb{E}[\psi_M(X_i)],$$

where $\psi_M(x) = \mathrm{sign}(x)(|x| \wedge M)$ for any $x \in \mathbb{R}$. The choice of M will be specified later. We list below some facts about $\overline{X}_M(i)$.

- Due to the triangle inequality and the boundness, we have that

$$|\overline{X}_M(i)| \leq 2M.$$

- Due to the triangle inequality, we have that

$$\mathbb{E}[|X_i - \overline{X}_M(i)|] \leq 2\mathbb{E}[|X_i - \psi_M(X_i)|].$$

- $|X_i - \psi_M(X_i)| = (|X_i| - M)\mathbb{1}\{|X_i| > M\} \leq |X_i|\mathbb{1}\{|X_i| > M\}.$

Proof of Theorem 43. Without loss of generality, we let $C_{\mathrm{FDM}} = 1$ and the assumption (132) becomes

$$\sup_{m\geq 0} \exp(cm^{\gamma_1})\Delta_{m,2} \leq 1, \tag{134}$$

To see this, if $C_{\mathrm{FDM}} < 1$, (132) implies (134). If $C_{\mathrm{FDM}} > 1$, we can consider the normalized process $\{C_{\mathrm{FDM}}^{-1} X_i\}_{i\in\mathbb{Z}}$. In this case, by the assumption (133) we have for any $x > 0$ that

$$\sup_{i\in\mathbb{Z}} \mathbb{P}(C_{\mathrm{FDM}}^{-1}|X_i| > x) \leq \exp(1 - C_{\mathrm{FDM}} x^{\gamma_2}) \leq \exp(1 - x^{\gamma_2}),$$

We define function $H(x) = \exp(1 - x^{\gamma_2})$ on $x > 0$. Our goal is to obtain a Bernstein-type inequality for the partial sum $\sum_{i=1}^n X_i$ for any $t > 0$ and any $n \geq 3$. Using the truncation argument, we decompose the partial sum as

$$\sum_{i=1}^n X_i = \sum_{i=1}^n (X_i - \overline{X}_M(i)) + \sum_{i=1}^n \overline{X}_M(i) = \sum_{i=1}^n (X_i - \overline{X}_M(i)) + \overline{S}_M. \tag{135}$$

The first term on the right-hand side of (135) can be handled using tail probability of the marginal distributions. To be specific, for any $x > 0$, it satisfies that

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{i=1}^n (X_i - \overline{X}_M(i))\Big| > x\Big) &\leq \frac{1}{x}\sum_{i=1}^n \mathbb{E}\big[\big|X_i - \overline{X}_M(i)\big|\big] \leq \frac{2}{x}\sum_{i=1}^n \mathbb{E}[|X_i - \psi_M(X_i)|] \\ &\leq \frac{2}{x}\sum_{i=1}^n \mathbb{E}[|X_i| \mathbb{1}\{|X_i| \geq M\}] = \frac{2}{x}\sum_{i=1}^n \int_M^\infty \mathbb{P}(|X_i| \geq t)dt \leq \frac{2n}{x}\int_M^\infty H(t)dt,\end{aligned}$$

where the first inequality follows from the triangle and Markov's inequalities, the second, the third and the fourth inequalities follow from the definitions of $\overline{X}_M(i)$, $\psi_M(X_i)$ and $H(t)$ respectively. Note that $\log(H(t)) = 1 - t^{\gamma_2}$. It follows from that the function $t \mapsto t^2 H(t)$ is nonincreasing for $t \geq (2/\gamma_2)^{1/\gamma_2}$. Hence, assuming

$$M \geq (2/\gamma_2)^{1/\gamma_2}, \tag{136}$$

we have that

$$\int_M^\infty H(t)dt \leq M^2 H(M) \int_M^\infty \frac{dt}{t^2} = MH(M).$$

Therefore, with any truncation parameter $M \geq (2/\gamma_2)^{1/\gamma_2}$, we have for any $x > 0$ that

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n (X_i - \overline{X}_M(i))\Big| > x\Big) \leq 2nx^{-1} MH(M). \tag{137}$$

Bounding the tail probability of $\overline{S}_M$ is the key for obtaining the Bernstein-type inequality of $\sum_{i=1}^n X_i$. Letting $u > 0$, we consider the following three cases.

Case 1. Suppose that $u \geq n^{\gamma_1/\gamma}$. Letting $M = u/n$, we have that

$$\big|\overline{S}_M\big| \leq 2nM = 2u.$$

We require $M = u/n \geq (2/\gamma_2)^{1/\gamma_2}$. Since $u/n \geq u^{1-\gamma/\gamma_1} = u^{\gamma/\gamma_2}$, this requirement holds if $u \geq (2/\gamma_2)^{1/\gamma}$. We have by (135) and (137) that for any $u \geq (n^{\gamma_1} \vee 2/\gamma_2)^{1/\gamma}$ that

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n X_i\Big| > 3u\Big) \leq \mathbb{P}\Big(\Big|\sum_{i=1}^n (X_i - \overline{X}_M(i))\Big| > u\Big) \leq 2H(M) = 2\exp(1 - (u/n)^{\gamma_2}) \leq 2e\exp(-u^\gamma),$$

where the last inequality follows from that $(u/n)^{\gamma_2} \geq u^{\gamma}$. While, if $n^{\gamma_1} < 2/\gamma_2$, for $n^{\gamma_1/\gamma} \leq u < (2/\gamma_2)^{1/\gamma}$, it holds trivially that

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n} X_i\Big| > 3u\Big) \leq 1 \leq e\exp(-u^{\gamma}/C),$$

where C is some positive constant such that $C \geq 2/\gamma_2$. Therefore, we have for any $u \geq n^{\gamma_1/\gamma}$ that

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n} X_i\Big| > 3u\Big) \leq \exp(-u^{\gamma}/C_1),$$

for some sufficient large $C_1 > 0$.

Case 2. Suppose that $2^{(1-\gamma_1)/\gamma_2}(4\zeta)^{\gamma_1/\gamma} \leq u < n^{\gamma_1/\gamma}$, where $\zeta = \mu \vee (2/\gamma_2)^{1/\gamma_1}$ and $\mu = (2(2 \vee 4c_0^{-1})/(1-\gamma))^{2/(1-\gamma)}$. Let $r \in [1, n/2]$ be some real value to be chosen later. Let

$$A = \Big\lfloor \frac{n}{2r} \Big\rfloor, \quad k = \Big\lfloor \frac{n}{2A} \Big\rfloor \quad \text{and} \quad M = H^{-1}(\exp(-A^{\gamma_1})).$$

Denoted $I_j = \{1 + (j-1)A, \ldots, jA\}$ for $j = 1, \ldots, 2k$, and also $I_{2k+1} = \{1 + 2kA, \ldots, n\}$. Denote $\overline{S}_M(I_j) = \sum_{i \in I_j} \overline{X}_M(i)$ for any $j = 1, \ldots, 2k+1$. We have that

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} (X_i - \overline{X}_M(i)) + \sum_{j=1}^{k} \overline{S}_M(I_{2j-1}) + \sum_{j=1}^{k} \overline{S}_M(I_{2j}) + \overline{S}_M(I_{2k+1}). \tag{138}$$

We now use the coupling method introduced by Berkes et al. (2009b) to construct independent block sums for odd blocks and even blocks respectively. To be specific, for any $j = 3, 4, \ldots, 2k$, let $\{\epsilon_t^{(j)}\}_{t \in \mathbb{Z}}$ be an i.i.d. sequence such that $\epsilon_t^{(j)}$ is an independent copy of ϵ_t and $\epsilon_t, \epsilon_t^{(3)}, \epsilon_t^{(4)}, \ldots, \epsilon_t^{(2k)}$ are mutually independent. This is always possible by enlarging the original probability space. For $s > h$, let

$$X_{s,\{h,-\infty\}}^{(j)} = g_s(\mathcal{F}_{s,\{h,-\infty\}}^{(j)}), \tag{139}$$

where $\mathcal{F}_{s,\{h,-\infty\}}^{(l)} = \sigma(\ldots, \epsilon_{h-1}^{(l)}, \epsilon_h^{(l)}, \epsilon_{h+1}, \ldots, \epsilon_s)$.

We only consider the odd blocks, since the same argument applies for the even blocks. Using the notation in (139), for any $j = 2, \ldots, k$ and $i \in I_{2j-1}$, let

$$\overline{X}_M^{(2j-1)}(i, \{(2j-3)A, -\infty\}) = \psi_M(X_{i,\{(2j-3)A,-\infty\}}^{(2j-1)}) - \mathbb{E}[\psi_M(X_{i,\{(2j-3)A,-\infty\}}^{(2j-1)})],$$

and denote

$$\overline{S}_M^*(I_{2j-1}) = \sum_{i = I_{2j-1}} \overline{X}_M^{(2j-1)}(i, \{(2j-3)A, -\infty\})$$

By the above construction, we have $\big\{\overline{S}_M(I_1), \overline{S}_M^*(I_3), \ldots, \overline{S}_M^*(I_{2k-1})\big\}$ are mutually independent, and for any $j = 2, \ldots, k$, $\overline{S}_M^*(I_{2j-1})$ and $\overline{S}_M(I_{2j-1})$ have the same distribution. Moreover, we have for any $j = 2, \ldots, k$ that

$$\mathbb{E}\big|\overline{S}_M(I_{2j-1}) - \overline{S}_M^*(I_{2j-1})\big| \leq \sum_{i \in I_{2j-1}} \big|\overline{X}_M^{(2j-1)}(i) - \overline{X}_M^{(2j-1)}(i, \{(2j-3)A, -\infty\})\big|$$

$$
\begin{aligned}
&\leq \sum_{i\in I_{2j-1}} \big\|\overline{X}_M^{(2j-1)}(i) - \overline{X}_M^{(2j-1)}(i,\{(2j-3)A,-\infty\})\big\|_2 \leq 2\sum_{i\in I_{2j-1}} \big\|X_i - X^{(2j-1)}_{i,\{(2j-3)A,-\infty\}}\big\|_2 \\
&\leq 2\sum_{i\in I_{2j-1}}\sum_{s=i-(2j-3)A}^{\infty} \delta_{s,2} \leq 2A\sum_{s=A+1}^{\infty}\delta_{s,2} \leq 2A\exp(-cA^{\gamma_1}).
\end{aligned}
$$

where the first inequality follows from the triangle inequality, the second inequality follows from Hölder's inequality, the third inequality follows from the second fact about the truncaton, the fourth inequality follows from the definition of $\delta_{s,2}$, and by writing the difference as telescoping sums, and the last two inequalities follow from (132). The same argument can be applied to the even blocks, with block sums denoted as

$$
\overline{S}_M^*(I_{2j}) = \sum_{i=I_{2j}} \overline{X}_M^{(2j)}(i,\{(2j-2)A,-\infty\}) \text{ for } j=2,\ldots,k.
$$

Denote also $\overline{S}_M^*(I_1) = \overline{S}_M(I_1)$ and $\overline{S}_M^*(I_2) = \overline{S}_M(I_2)$. Then, (138) can be further decomposed as

$$
\begin{aligned}
\sum_{i=1}^n X_i =& \sum_{i=1}^n \{X_i - \overline{X}_M(i)\} + \sum_{j=1}^k \overline{S}_M^*(I_{2j-1}) + \sum_{j=1}^k \overline{S}_M^*(I_{2j}) + \overline{S}_M(I_{2k+1}) \\
&+ \sum_{j=2}^k \big\{\overline{S}_M(I_{2j-1}) - \overline{S}_M^*(I_{2j-1})\big\} + \sum_{j=2}^k \big\{\overline{S}_M(I_{2j}) - \overline{S}_M^*(I_{2j})\big\}.
\end{aligned}
\tag{140}
$$

Since $\big|\overline{S}_M(I_{2k+1})\big| \leq 2AM$, for any $u \geq 2AM$, it holds that

$$
\begin{aligned}
\mathbb{P}\Big(\Big|\sum_{i=1}^n X_i\Big| > 6u\Big) \leq& \mathbb{P}\Big(\Big|\sum_{i=1}^n (X_i - \overline{X}_M(i))\Big| > u\Big) + \mathbb{P}\Big(\Big|\sum_{j=1}^k \overline{S}_M^*(I_{2j-1})\Big| > u\Big) \\
&+ \mathbb{P}\Big(\Big|\sum_{j=1}^k \overline{S}_M^*(I_{2j})\Big| > u\Big) + 4(k-1)Au^{-1}\exp(-cA^{\gamma_1}) \\
=& I + II + III + 4(k-1)Au^{-1}\exp(-cA^{\gamma_1}),
\end{aligned}
\tag{141}
$$

where the first inequality follows from Markov's inequality We bound I using (137). Since $H^{-1}(y) = (\log(e/y))^{1/\gamma_2}$, we have $M = (1+A^{\gamma_1})^{1/\gamma_2} \geq (A^{\gamma_1})^{1/\gamma_2}$. For any $A \geq (2/\gamma_2)^{1/\gamma_1}$, which satisfies the requirement (136), we have for any $u > 0$ that

$$
I \leq 2nu^{-1}M\exp(-A^{\gamma_1}).
$$

Since $\big\{\overline{S}_M^*(I_{2j-1})\big\}_{j=1}^k$ are mutually independent, by Chebyshev's inequality, we have for any $t > 0$ that

$$
II \leq 2e^{-ut}\prod_{i=1}^k \mathbb{E}\Big[\exp\big(t\overline{S}_M(I_{2j-1})\big)\Big].
$$

Applying Proposition 45 for each Laplace transform of block sum, if $A \geq \mu$, we have for any $t < \nu A^{-\gamma_1(1-\gamma)/\gamma}$ that

$$
\sum_{i=1}^k \log\mathbb{E}\Big[\exp\big(t\overline{S}_M(I_{2j-1})\big)\Big] \leq \frac{kAV(A)t^2}{1-\nu^{-1}A^{\gamma_1(1-\gamma)/\gamma}t}.
$$

Then, if $A \geq \zeta$, we have for any $u \geq 0$ that

$$II \leq 2\exp\Big(-ut+\frac{kAV(A)t^2}{1-\nu^{-1}A^{\gamma_1(1-\gamma)/\gamma}t}\Big)=2\exp\Big(-\frac{u^2}{4kAV(A)+2\nu^{-1}A^{\gamma_1(1-\gamma)/\gamma}u}\Big),$$

where the equality follows by letting $t=u\big(2kAV(A)+\nu^{-1}A^{\gamma_1(1-\gamma)/\gamma}u\big)^{-1}$. Using the same argument, we obtain the same bound for III.

Therefore, if $A \geq \zeta$, we have for any $u \geq 2AM$ that

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{i=1}^n X_i\Big|>6u\Big)=&4(k-1)Au^{-1}\exp(-cA^{\gamma_1})+2nu^{-1}M\exp(-A^{\gamma_1})\\&+4\exp\Big(-\frac{u^2}{4kAV(A)+2\nu^{-1}A^{\gamma_1(1-\gamma)/\gamma}u}\Big).\end{aligned}\tag{142}$$

We now choose $r=2^{(1-\gamma_1)/(\gamma_1+\gamma_2)}nu^{-\gamma/\gamma_1}$. By the definition of A, we have that

$$2^{(\gamma_1-1)/(\gamma_1+\gamma_2)}u^{\gamma/\gamma_1}-2\leq 2A\leq 2^{(\gamma_1-1)/(\gamma_1+\gamma_2)}u^{\gamma/\gamma_1}.$$

Since $M=(1+A^{\gamma_1})^{1/\gamma_2}\leq 2^{1/\gamma_2}A^{\gamma_1/\gamma_2}$, the choice of r implies that

$$2AM\leq 2^{1+1/\gamma_2}A^{\gamma_1/\gamma}\leq u.$$

Moreover, by the definition of ζ and μ, we have that $\zeta\geq\mu\geq 16$ and $u^{\gamma/\gamma_1}\geq 4\zeta\geq 64$. Since $u\geq 2^{(1-\gamma_1)/\gamma_2}(4\zeta)^{\gamma_1/\gamma}$, it holds that

$$A\geq 2^{(\gamma_1-1)/(\gamma_1+\gamma_2)}u^{\gamma/\gamma_1}/2-1\geq(4\zeta)/2-1\geq(4\zeta)/4\geq\zeta.$$

Therefore, we have that (142) holds for any $2^{(1-\gamma_1)/\gamma_2}(4\zeta)^{\gamma_1/\gamma}\leq u<n^{\gamma_1/\gamma}$. We list the following facts.

- By the definition of k and the fact that $4\zeta\geq 2^6$, we have that
$$4(k-1)Au^{-1}\leq 2nu^{-1}\leq 2^{1+(\gamma_1-1)/\gamma_2}n(4\zeta)^{-\gamma_1/\gamma}\leq 2^{-(5\gamma_1+1)/\gamma_2-5}n.$$
- Since $A\geq 2\zeta-1\geq 31$, we have that $2nu^{-1}M\leq n/A\leq n/31$.
- Since $2A\leq 2^{(\gamma_1-1)/(\gamma_1+\gamma_2)}u^{\gamma/\gamma_1}$, $0<\gamma<1$ and $\gamma_2>0$, we have that
$$A^{\gamma_1(1-\gamma)/\gamma}u\leq 2^{-(1+\gamma_2)(1-\gamma)\gamma_2}u^{2-\gamma}\leq u^{2-\gamma}.$$

We have that

$$(142)\leq\frac{n}{2^{5+(1+5\gamma_1)/\gamma_2}}\exp\Big(-c\frac{u^\gamma}{2^{\gamma_1}}\Big)+\frac{n}{31}\exp\Big(-\frac{u^\gamma}{2^{\gamma_1}}\Big)+4\exp\Big(-\frac{u^2}{2nV(A)+2\nu^{-1}u^{2-\gamma}}\Big).$$

Recall that $V(A)=(\nu_1C_{\mathrm{LRV}}+\nu_2\exp\big(-\nu_3A^{\gamma_1(1-\gamma)}(\log A)^{-\gamma}\big)$, where $\nu_1,\nu_2,\nu_3>0$ are constants depending only on c, γ_1 and γ_2. Therefore, we have for any $2^{(1-\gamma_1)/\gamma_2}(4\zeta)^{\gamma_1/\gamma}\leq u<n^{\gamma_1/\gamma}$ that

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{i=1}^n X_i\Big|>6u\Big)=&n\exp\Big(-\frac{u^\gamma}{C_1}\Big)+\exp\Big(-\frac{u^2}{C_2(1+nC_{\mathrm{LRV}})}\Big)\\&+\exp\Big(-\frac{u^2}{C_3n}\exp\Big(\frac{x^{\gamma(1-\gamma)}}{C_4(\log A)^\gamma}\Big)\Big),\end{aligned}$$

where $C_1, C_2, C_3, C_4 > 0$ are absolute constants depending only on γ_1, γ_2 and c.

Case 3. Suppose that $u < 6 \times 2^{(1-\gamma_1)/\gamma_2}(4\zeta)^{\gamma_1/\gamma}$. It holds trivially that

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n} X_i\Big| > u\Big) \leq 1 \leq e \exp\Big(-\frac{u^\gamma}{6^\gamma 2^{(1-\gamma_1)\gamma/\gamma_2}(4\zeta)^{\gamma_1}}\Big),$$

which is smaller than $n\exp(-u^\gamma/C_1)$, if $n \geq 3$ and $C_1 \geq 6^\gamma 2^{(1-\gamma_1)\gamma/\gamma_2}(4\zeta)^{\gamma_1}$. Combining the three cases completes the proof. □

Theorem 43 relies on the upper bound on log-Laplace transform of a block sum of bounded random variables, where the block size A is relatively smaller than the sample size n. In the following, we describe the idea to obtain such bound. The results are summarized in Propositions 44 and 45.

Let A be a positive integer and $\{X_i\}_{i=1}^{A}$ be a sequence of random variables of the form (122) and with mean zero. Our goal is to upper bound the log-Laplace transform of partial sums i.e. $\log \mathbb{E}\big[\exp\big(t\sum_{i=1}^{A} X_i\big)\big]$, for any small $t > 0$. To this end, we introduce the construction of the Cantor-like set $K_A^{(l)} = \bigcup_{k=1}^{2^l} I_{l,k} \subset \{1, \ldots, A\}$, where $\{I_{l,k}\}_{k=1}^{2^l}$ are integer intervals with same cardinality n_l, and all neighboring integer intervals are separated by some integer intervals. Heuristically, this construction reduces the dependence by creating gaps.

Construction of $K_A^{(l)}$. The construction of $K_A^{(l)}$ of $\{1, \ldots, A\}$ involves l recursive steps, where l satisfies $A2^{-l} \geq (2 \vee 4c_0^{-1})$ is a positive integer. Let $c_0 \in (0, (2^{1/\gamma - 1} - 1)(2(2^{1/\gamma} - 1))^{-1}]$ be some constant. We will see that such requirements of l and c_0 justify the following construction of $K_A^{(l)}$. We do not need the specific values of l and c_0 until we give them respectively in (178) and (148).

Step C1. Let $n_0 = A$. Define integer d_0 as

$$d_0 = \begin{cases} \max\{m \in 2\mathbb{N}, m \leq c_0 n_0\} & \text{if } n_0 \text{ is even,} \\ \max\{m \in 2\mathbb{N}+1, m \leq c_0 n_0\} & \text{if } n_0 \text{ is odd.} \end{cases}$$

Note that $n_0 - d_0$ is even. Define $n_1 = (n_0 - d_0)/2$ and two integer intervals of cardinality n_1 separated by a gap of d_0 integers as

$$I_{1,1} = \{1, \ldots, n_1\} \text{ and } I_{1,2} = \{n_1 + d_0 + 1, \ldots, n_0\}.$$

We discard the gap of d_0 integers for the next step.

Step C2. Define integer d_1 as

$$d_1 = \begin{cases} \max\{m \in 2\mathbb{N}, m \leq c_0 n_0 2^{-(l \wedge 1/\gamma)}\} & \text{if } n_1 \text{ is even,} \\ \max\{m \in 2\mathbb{N}+1, m \leq c_0 n_0 2^{-(l \wedge 1/\gamma)}\} & \text{if } n_1 \text{ is odd.} \end{cases}$$

Note that $n_1 - d_1$ is even, and define $n_2 = (n_1 - d_1)/2$. Based on $I_{1,1}$, define two integer intervals of cardinality n_2 separated by a gap of d_1 integers as

$$I_{2,1} = \{1, \ldots, n_2\} \text{ and } I_{2,2} = \{n_2 + d_1 + 1, \ldots, n_1\}.$$

Based on $I_{1,2}$, define similarly two integer intervals of cardinality n_2 separated by a gap of d_1 integers as

$$I_{2,3} = (n_1 + d_0) + I_{2,1} \text{ and } I_{2,4} = (n_1 + d_0) + I_{2,2}.$$

We discard the two gaps of d_1 integers for the next step.

Step Ck. After Step C$(k-1)$ with $2 \leq k \leq l$, we obtain integer intervals $\{I_{k-1,j}\}_{j=1}^{2^{k-1}}$. Denote for $1 \leq j \leq 2^{k-1}$, $I_{k-1,j} = \{a_{k-1,j}, \ldots, b_{k-1,j}\}$. Define integer d_{k-1} as

$$d_{k-1} = \begin{cases} \max\{m \in 2\mathbb{N}, m \leq c_0 n_0 2^{-(l \wedge (k-1)/\gamma)}\} & \text{if } n_{k-1} \text{ is even,} \\ \max\{m \in 2\mathbb{N}+1, m \leq c_0 n_0 2^{-(l \wedge (k-1)/\gamma)}\} & \text{if } n_{k-1} \text{ is odd.} \end{cases} \tag{143}$$

Note that $n_{k-1} - d_{k-1}$ is even, and define $n_k = (n_{k-1} - d_{k-1})/2$. For each $j = 1, \ldots, 2^{k-1}$, based on $I_{k-1,j}$, we define two integer intervals of cardinality n_k separated by a gap of d_{k-1} integers as

$$I_{k,2j-1} = \{a_{k,2j-1}, \ldots, b_{k,2j-1}\} \text{ and } I_{k,2j} = \{a_{k,2j}, \ldots, b_{k,2j}\},$$

where $a_{k,2j-1} = a_{k-1,j}$, $b_{k,2j-1} = a_{k,2j-1} + n_k - 1$, $a_{k,2j} = b_{k,2j} - n_k + 1$ and $b_{k,2j} = b_{k-1,j}$. Moreover, we have $a_{k,2j} - b_{k,2j-1} - 1 = d_{k-1}$. We discard the 2^{k-1} gaps of d_{k-1} integers.

Finally, after Step Cl, we obtain $\{I_{l,1}, \ldots, I_{l,2^l}\}$ and $K_A^{(l)} = \bigcup_{i=1}^{2^l} I_{l,i}$. Moreover, for any $k = 0, 1, \ldots, l$ and $j = 1, \ldots, 2^k$, we also define

$$K_{A,k,j}^{(l)} = \bigcup_{i=(j-1)2^{l-k}+1}^{j2^{l-k}} I_{l,i}, \tag{144}$$

and we have

$$K_A^{(l)} = \bigcup_{j=1}^{2^k} K_{A,k,j}^{(l)}. \tag{145}$$

Moreover, it follows that

$$\begin{aligned} A - \mathrm{Card}(K_A^{(l)}) &= \sum_{j=0}^{l-1} 2^j d_j \leq A c_0 \Big(\sum_{j=0}^{\lfloor \gamma l \rfloor} 2^{j(1-1/\gamma)} + \sum_{j=\lfloor \gamma l \rfloor + 1}^{l-1} 2^{j-l} \Big) \\ &\leq A c_0 \Big(\frac{2 - 2^{1-1/\gamma} - 2^{\lfloor \gamma l \rfloor - l}}{1 - 2^{1-1/\gamma}} \Big) \leq A c_0 \Big(\frac{2 - 2^{1-1/\gamma}}{1 - 2^{1-1/\gamma}} \Big). \end{aligned}$$

To ensure $\mathrm{Card}(K_A^{(l)}) \geq A/2$, we choose some c_0 such that

$$0 < c_0 \leq \frac{2^{1/\gamma - 1} - 1}{2(2^{1/\gamma} - 1)}. \tag{146}$$

Also, by definition, $n_k = 2^{-k}\big(n_0 - \lfloor c_0 n_0 \rfloor - \sum_{i=1}^{k-1} 2^i \lfloor c_0 n_0 2^{-(l \wedge i/\gamma)} \rfloor\big)$. For any k such that $(k-1)/\gamma \leq l$, we need $n_k \geq 1$ to ensure the construction of $K_A^{(l)}$. Since

$$\begin{aligned} n_k \geq & 2^{-k} \Big(n_0 - c_0 n_0 - c_0 n_0 \sum_{i=1}^{k-1} 2^{i(1-1/\gamma)} \Big) \\ = & 2^{-k} \Big(n_0 - c_0 n_0 - c_0 n_0 \frac{2^{1-1/\gamma}(1 - 2^{(k-1)(1-1/\gamma)})}{1 - 2^{1-1/\gamma}} \Big) \\ \geq & 2^{-k} \Big(n_0 - c_0 n_0 - c_0 n_0 \frac{2^{1-1/\gamma}}{1 - 2^{1-1/\gamma}} \Big) = 2^{-k} A \Big(1 - c_0 \frac{1}{1 - 2^{1-1/\gamma}} \Big). \end{aligned} \tag{147}$$

We choose

$$c_0 = \frac{2^{1/\gamma-1}-1}{2(2^{1/\gamma}-1)}. \tag{148}$$

Plugging (148) into (147), we have for any $\gamma \in (0,1)$ and any k such that $(k-1)/\gamma \le l$ that

$$n_k \ge 4c_0^{-1}\Big(1 - c_0\frac{1}{1-2^{1-1/\gamma}}\Big) = 4\Big(c_0^{-1} - \frac{1}{1-2^{1-1/\gamma}}\Big) = 12 + \frac{8}{2^{1/\gamma}-2} \ge 12,$$

where the first inequality follows from the definition of l, i.e. $A2^{-l} \ge (2 \vee 4c_0^{-1})$ and $k \le l$. The second inequality follows from the fact $2^{1/\gamma} > 2$ for any $\gamma \in (0,1)$. Moreover, since we require l such that $A2^{-l} \ge (2 \vee 4c_0^{-1})$, we have that $n_l \ge A2^{-l-1} \ge 1$.

We list the following facts regarding $K_A^{(l)}$.

- $\mathrm{Card}(I_{l,j}) = n_l$ for any $j = 1, \dots, 2^l$, and $\mathrm{Card}(K_A^{(l)}) = 2^l n_l$.
- $A/2 \le \mathrm{Card}(K_A^{(l)}) \le A$.

Recall c_0 and c given respectively in (146) and (132), we define the following positive constants depending only on c, γ_1 and γ_2.

$$c_1 = \min\{c^{1/\gamma_1}c_0/4, 2^{-1/\gamma}\},\ c_2 = 2^{-(2+\gamma_1/\gamma+1/\gamma_2)}c_1^{\gamma_1},\ c_3 = 2^{-(1+1/\gamma_2)}, \tag{149}$$

and

$$\kappa = \min\{c_2, c_3\}. \tag{150}$$

Note that by definition, we have $c_1, c_2, c_3, \kappa \in (0,1)$.

The next proposition gives upper bounds on the log-Laplace transform for sum of random variables on a Cantor-like set.

Proposition 44. *Let A be a positive integer and $\{X_i\}_{i=1}^A$ be a sequence of random variables of the form (122), with mean zero. Assume that (132) and (133) hold. Let $M = H^{-1}(\exp(-A^{\gamma_1}))$ and l be a positive integer such that $A2^{-l} \ge (1 \vee 2c_0^{-1})$. Then, we have the following results.*

(i) For any $t \le 2^{-(1+1/\gamma_2)}A^{-\gamma_1/\gamma}$, it holds that

$$\log \mathbb{E}\Big[\exp\Big(t\sum_{j\in K_A^{(l)}} \overline{X}_M(j)\Big)\Big] \le C_{\mathrm{LRV}}t^2A,$$

where C_{LRV} is defined in (127).

(ii) For any $2^{-(1+1/\gamma_2)}A^{-\gamma_1/\gamma} \le t \le \kappa\big(A^{\gamma-1} \wedge (2^l/A)\big)^{\gamma_1/\gamma}$, it holds that

$$\log \mathbb{E}\Big[\exp\Big(t\sum_{j\in K_A^{(l)}} \overline{X}_M(j)\Big)\Big] \le (tAl + 6A^{\gamma})\exp(-(2^{-l}c_1A)^{\gamma_1}/2) + C_{\mathrm{LRV}}t^2A.$$

Decomposition of partial sums. After obtaining a Cantor-like set from an integer interval $\{1, \dots, A\}$, there are $A - \mathrm{Card}(K_A^l)$ indices left over. These indices can be reassembled and reused to construct Canto-like set. We perform such reassembling and construction recursively until the number of remaining indices is small enough, and the final remainder can be dealt easily. Recall l is a positive integer, to be chosen later, which satisfies $A2^{-l} \geq (2 \vee 4c_0^{-1})$. For any $j = 0, 1, \dots,$ we will define A_j as the number of indices to start with at Step Dj of the decomposition. Since A_j decreases as j increases, the decomposition stops at Step D$m(A)$, where

$$m(A) = \min\{m \in \mathbb{N}, A_m \leq A2^{-l}\}. \tag{151}$$

Let the initial truncation parameter $T_0 = M = H^{-1}(\exp(-A_0^{\gamma_1}))$ with $A_0 = A$. We need to refine the truncation parameter at each step, define for $1 \leq j \leq m(A) - 1$ that

$$T_j = H^{-1}(\exp(-A_j^{\gamma_1})). \tag{152}$$

Also, for $1 \leq j \leq m(A) - 1$, A_j may varies. We choose l_j as

$$l_j = \min\{k \in \mathbb{N}, A_j 2^{-k} \leq A_0 2^{-l}\}.$$

Step D0 Let $A_0 = A$, $T_0 = M$ and $\overline{X}_{T_0}^{(0)}(k) = \overline{X}_M(k)$ for any $k = 1, \dots, A_0$. Let $l_0 = l$. Let $K_{A_0}^{(l_0)}$ be the Cantor-like set defined from $\{1, \dots, A_0\}$.

Step D1 Let $A_1 = A_0 - \mathrm{Card}(K_{A_0}^{(l_0)})$ and $\{i_1, \dots, i_{A_1}\} = \{1, \dots, A\} \backslash K_{A_0}^{(l_0)}$. Define for any $k = 1, \dots, A_1$, $\overline{X}_{T_1}^{(1)}(k) = \overline{X}_{T_1}(i_k)$. Let $K_{A_1}^{(l_1)}$ be the Cantor-like set defined from $\{i_1, \dots, i_{A_1}\}$.

Step Dj For $1 \leq j \leq m(A) - 1$, let $A_j = A_{j-1} - \mathrm{Card}(K_{A_{j-1}}^{(l_{j-1})})$ and $\{i_1, \dots, i_{A_j}\} = \{1, \dots, A\} \backslash \left(\bigcup_{k=0}^{j-1} K_{A_k}^{(l_k)}\right)$. Define for any $k = 1, \dots, A_j$, $\overline{X}_{T_j}^{(j)}(k) = \overline{X}_{T_j}(i_k)$. Let $K_{A_j}^{(l_j)}$ be the Cantor-like set defined from $\{i_1, \dots, i_{A_j}\}$.

Step D$m(A)$ Let $A_{m(A)} = A_{m(A)-1} - \mathrm{Card}(K_{A_{m(A)-1}}^{(l_{m(A)-1})})$ and $\{i_1, \dots, i_{A_{m(A)}}\} = \{1, \dots, A\} \backslash \left(\bigcup_{k=0}^{m(A)-1} K_{A_k}^{(l_k)}\right)$. Define for any $k = 1, \dots, A_{m(A)}$, $\overline{X}_{T_{m(A)-1}}^{(m(A))}(k) = \overline{X}_{T_{m(A)-1}}(i_k)$.

Regarding $m(A)$, we have the following facts.

- For $0 \leq j \leq m(A)$, $A_j \leq A2^{-j}$, which follows from the fact of $K_A^{(l)}$.
- $m(A) \geq 1$, since $l \geq 1$ and $A_0 > A2^{-l}$. Moreover, $m(A) \leq l$, which follows from the definition of $m(A)$.

Therefore, we have the following decomposition.

$$\sum_{i=1}^{A} \overline{X}_M(i) = \sum_{i=1}^{A_0} \overline{X}_{T_0}^{(0)}(i) = \sum_{i=K_{A_0}^{(l_0)}} \overline{X}_{T_0}^{(0)}(i) + \sum_{i=1}^{A_1} \overline{X}_{T_0}^{(1)}(i)$$

$$= \sum_{i=K_{A_0}^{(l_0)}} \overline{X}_{T_0}^{(0)}(i) + \sum_{i=1}^{A_1} \left\{\overline{X}_{T_0}^{(1)}(i) - \overline{X}_{T_1}^{(1)}(i)\right\} + \sum_{i=1}^{A_1} \overline{X}_{T_1}^{(1)}(i)$$

$$= \sum_{i=K_{A_0}^{(l_0)}} \overline{X}_{T_0}^{(0)}(i) + \sum_{i=1}^{A_1} \left\{\overline{X}_{T_0}^{(1)}(i) - \overline{X}_{T_1}^{(1)}(i)\right\} + \sum_{i=K_{A_1}^{(l_1)}} \overline{X}_{T_1}^{(1)}(i) + \sum_{i=1}^{A_2} \overline{X}_{T_1}^{(2)}(i)$$

$$
\begin{aligned}
&= \sum_{i=K_{A_0}^{(l_0)}} \overline{X}_{T_0}^{(0)}(i) + \sum_{i=1}^{A_1} \{\overline{X}_{T_0}^{(1)}(i) - \overline{X}_{T_1}^{(1)}(i)\} + \sum_{i=K_{A_1}^{(l_1)}} \overline{X}_{T_1}^{(1)}(i) + \sum_{i=1}^{A_2} \{\overline{X}_{T_1}^{(2)}(i) - \overline{X}_{T_2}^{(2)}(i)\} + \sum_{i=1}^{A_2} \overline{X}_{T_2}^{(2)}(i) \\
&= \sum_{j=0}^{1} \sum_{i=K_{A_j}^{(l_j)}} \overline{X}_{T_j}^{(j)}(i) + \sum_{j=1}^{2} \sum_{i=1}^{A_j} \{\overline{X}_{T_{j-1}}^{(j)}(i) - \overline{X}_{T_j}^{(j)}(i)\} + \sum_{i=1}^{A_2} \overline{X}_{T_2}^{(2)}(i) \\
&= \dots \\
&= \sum_{j=0}^{m(A)-2} \sum_{i \in K_{A_j}^{(l_j)}} \overline{X}_{T_j}^{(j)}(i) + \sum_{j=1}^{m(A)-1} \sum_{i=1}^{A_j} \{\overline{X}_{T_{j-1}}^{(j)}(i) - \overline{X}_{T_j}^{(j)}(i)\} + \sum_{i=1}^{A_{m(A)-1}} \overline{X}_{T_{m(A)-1}}^{(m(A)-1)}(i) \\
&= \sum_{j=0}^{m(A)-1} \sum_{i \in K_{A_j}^{(l_j)}} \overline{X}_{T_j}^{(j)}(i) + \sum_{j=1}^{m(A)-1} \sum_{i=1}^{A_j} \{\overline{X}_{T_{j-1}}^{(j)}(i) - \overline{X}_{T_j}^{(j)}(i)\} + \sum_{i=1}^{A_{m(A)}} \overline{X}_{T_{m(A)-1}}^{(m(A))}(i) \\
&= I + II + III. \qquad (153)
\end{aligned}
$$

The next proposition gives an upper bound on the log-Laplace transform of (153). Recall κ given in (150), we define the following positive constants depending only on c, γ_1 and γ_2.

$$
\mu = (2(2 \vee 4c_0^{-1})/(1-\gamma))^{2/(1-\gamma)}, \quad c_4 = 2^{1+1/\gamma_2} 3^{\gamma_1/\gamma_2} c_0^{-\gamma_1/\gamma_2} \qquad (154)
$$

and

$$
\nu = (3c_4 + \kappa^{-1})^{-1}(1 - 2^{-\gamma_1(1-\gamma)/\gamma}). \qquad (155)
$$

Note that by definition, we have $\nu \in (0, 1)$ and $c_4 > 1$.

Proposition 45. *Let A be a positive integer and $\{X_i\}_{i=1}^A$ be a sequence of random variables of the form (122), with mean zero. Assume (132) and (133) hold. Let $M = H^{-1}(\exp(-A^{\gamma_1}))$. Then, if $A > \mu$, we have for any $t \in [0, \nu A^{-\gamma_1(1-\gamma)/\gamma})$ that*

$$
\log \mathbb{E}\Big[\exp\Big(t \sum_{i=1}^{A} \overline{X}_M(i)\Big)\Big] \le \frac{AV(A)t^2}{1 - \nu^{-1} A^{\gamma_1(1-\gamma)/\gamma} t},
$$

where $V(A) = (\nu_1 C_{\mathrm{LRV}} + \nu_2 \exp\big(-\nu_3 A^{\gamma_1(1-\gamma)}(\log A)^{-\gamma}\big)$ with constants $\nu_1, \nu_2, \nu_3 > 0$ depending only on c, γ_1 and γ_2, and C_{LRV} is defined in (127).

F.4 Proofs of propositions

Proof of Proposition 44. Recall the definitions (144) and (145), we have for $k = 2, \dots, l$ and $j = 1, \dots, 2^k$ that

$$
K_A^{(l)} = \bigcup_{i=1}^{2} K_{A,1,i}^{(l)} \text{ and } K_{A,k-1,j}^{(l)} = \bigcup_{i=2j-1}^{2j} K_{A,k,i}^{(l)}.
$$

Each block at previous step is split into two sub-blocks at the current step. The key of this proof is to upper bound the telescoping distance between the log-Laplace transform of the block and the sum of the log-Laplace transforms of two sub-blocks at each step. Our proof contains l steps. Let $k_l = \max\{k \in \mathbb{N}, k/\gamma < l\}$. Since $\gamma < 1$, $k_l < \gamma l \le l-1$, we need at least $k_l + 2$ steps. At each step $k = 1, \dots, k_l + 1$, we apply Lemma 38 and adapt the truncation parameter to a smaller value. While at steps $k_l + 2, \dots, l$, the truncation parameter no longer needs to be adapted. Let $M_0 = M$. Define the truncation parameter at Step k for any $k = 1, \dots, k_l + 1$ as

$$M_k = H^{-1}(\exp(-A^{\gamma_1} 2^{-\gamma_1 (l \wedge k/\gamma)})),$$

and the truncation parameter maintains M_{k_l+1} after Step $k_l + 1$. Since $H^{-1}(y) = (\log(e/y))^{1/\gamma_2}$ for any $y \le e$, we have for any $x \ge 1$,

$$H^{-1}(\exp(-x^{\gamma_1})) = (1 + x^{\gamma_1})^{1/\gamma_2} \le 2^{1/\gamma_2} x^{\gamma_1/\gamma_2}. \tag{156}$$

Note that for any $k = 0, \dots, k_l$, $A 2^{-k/\gamma} \ge A 2^{-l} \ge 1$, we have that

$$M_k \le 2^{1/\gamma_2} (A 2^{-(l \wedge k/\gamma)})^{\gamma_1/\gamma_2} \text{ for any } k = 0, \dots, k_l + 1. \tag{157}$$

Step 1. Recall that for $i = 1, 2$, $\mathrm{Card}(K_{A,1,i}^{(l)}) \le A/2$ and for any $j = 1, \dots, A$, $|\overline{X}_{M_0}(j)| \le 2M_0$. Applying Lemma 38, we have for any $t > 0$ that

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\exp\Big(t \sum_{j \in K_A^{(l)}} \overline{X}_{M_0}(j)\Big)\Big] - \prod_{i=1}^{2} \mathbb{E}\Big[\exp\Big(t \sum_{j \in K_{A,1,i}^{(l)}} \overline{X}_{M_0}(j)\Big)\Big]\Big| \\
\le& t \exp(2tAM_0) \Big\| \sum_{j \in K_{A,1,2}} \{\overline{X}_{M_0}(j) - \overline{X}_{M_0, \{n_1, -\infty\}}(j)\} \Big\|_2 \\
\le& t \exp(2tAM_0) \sum_{j \in K_{A,1,2}^{(l)}} \sum_{s=j-n_1}^{\infty} \delta_{s,2} \le \frac{tA}{2} \exp(2tAM_0) \exp(-c d_0^{\gamma_1}),
\end{aligned}$$

where n_1 is the maximal integer in $K_{A,1,1}^{(l)}$, d_0 is the gap between $K_{A,1,1}^{(l)}$ and $K_{A,1,2}^{(l)}$, the second inequality follows from the triangle inequality and the definition of $\delta_{j,2}$, and the third inequality follows from the facts that $\mathrm{Card}(K_{A,1,2}^{(l)}) = A/2$ and $j - n_1 \ge d_0$ for any $j \in K_{A,1,2}^{(l)}$ and the assumption (132). Since $\overline{X}_{M_0}(j)$ are centered and by Jensen's inequality, we have that the Laplace transforms are all greater than 1. By the inequality

$$|\log(x) - \log(y)| \le |x - y| \text{ for } x, y \ge 1, \tag{158}$$

we have that

$$\begin{aligned}
&\Big|\log \mathbb{E}\Big[\exp\Big(t \sum_{j \in K_A^{(l)}} \overline{X}_{M_0}(j)\Big)\Big] - \sum_{i=1}^{2} \log \mathbb{E}\Big[\exp\Big(t \sum_{j \in K_{A,1,i}^{(l)}} \overline{X}_{M_0}(j)\Big)\Big]\Big| \\
\le& \frac{tA}{2} \exp(2tAM_0) \exp(-c d_0^{\gamma_1}).
\end{aligned} \tag{159}$$

We adapt the truncation parameter to M_1. By the inequality

$$|e^{tx} - e^{ty}| \le |t||x - y|(e^{|tx|} \vee e^{|ty|}), \tag{160}$$

we have for $i = 1, 2$ and any $t > 0$ that

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[\exp\Big(t\sum_{j\in K^{(l)}_{A,1,i}}\overline{X}_{M_0}(j)\Big)\Big]-\mathbb{E}\Big[\exp\Big(t\sum_{j\in K^{(l)}_{A,1,i}}\overline{X}_{M_1}(j)\Big)\Big]\Big|\\
\le& te^{tAM_0}\sum_{j\in K^{(l)}_{A,1,i}}\mathbb{E}[|\overline{X}_{M_0}(j)-\overline{X}_{M_1}(j)|]\le 2te^{tAM_0}\sum_{j\in K^{(l)}_{A,1,i}}\mathbb{E}[|\psi_{M_0}(X_j)-\psi_{M_1}(X_j)|]\\
\le& 2tM_0e^{tAM_0}\sum_{j\in K^{(l)}_{A,1,i}}\mathbb{P}(|X_j|\ge M_1)\le 2tAM_0e^{tAM_0}H(M_1)\\
=& 2tAM_0e^{tAM_0}\exp(-A^{\gamma_1}2^{-\gamma_1/\gamma})\le 2e^{2tAM_0}\exp(-A^{\gamma_1}2^{-\gamma_1/\gamma}),
\end{aligned}
$$

where the first inequality follows from that

$$
\Big|t\sum_{j\in K^{(l)}_{A,1,i}}\overline{X}_{M_1}(j)\Big|\le\Big|t\sum_{j\in K^{(l)}_{A,1,i}}\overline{X}_{M_0}(j)\Big|\le tAM_0,
$$

and the triangle inequality, the second and the third inequalities follow from the facts about $\overline{X}_M(j)$, the fourth inequality follows from the assumption (133), the equality follows from the definition of M_1, and the last inequality follows from that $1 + x \le e^x$. By the inequality (158), we have for $i = 1, 2$ and any $t > 0$ that

$$
\begin{aligned}
&\Big|\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K^{(l)}_{A,1,i}}\overline{X}_{M_0}(j)\Big)\Big]-\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K^{(l)}_{A,1,i}}\overline{X}_{M_1}(j)\Big)\Big]\Big|\\
\le& 2e^{2tAM_0}\exp(-A^{\gamma_1}2^{-\gamma_1/\gamma}),
\end{aligned}\tag{161}
$$

Combining (159) and (161), we have that

$$
\begin{aligned}
&\Big|\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K^{(l)}_{A}}\overline{X}_{M_0}(j)\Big)\Big]-\sum_{i=1}^{2}\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K^{(l)}_{A,1,i}}\overline{X}_{M_1}(j)\Big)\Big]\Big|\\
\le&\frac{tA}{2}\exp(2tAM_0)\exp(-cd_0^{\gamma_1})+4\exp(2tAM_0)\exp(-A^{\gamma_1}2^{-\gamma_1/\gamma}).
\end{aligned}\tag{162}
$$

Step k. For any $k = 2, \dots, k_l + 1$, we use the same arguments as in Step 1. Note that $\mathrm{Card}(K^{(l)}_{A,k,i}) \le 2^{-k}A$, for any $i = 1, \dots, 2^k$, we have that

$$
\begin{aligned}
&\Big|\sum_{i=1}^{2^{k-1}}\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K^{(l)}_{A,k-1,i}}\overline{X}_{M_{k-1}}(j)\Big)\Big]-\sum_{i=1}^{2^k}\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K^{(l)}_{A,k,i}}\overline{X}_{M_k}(j)\Big)\Big]\Big|\\
\le& 2^{k-1}\frac{tA}{2^k}\exp(t2M_{k-1}A/2^{k-1})\exp(-cd_{k-1}^{\gamma_1})+2^{k+1}\exp(t2M_{k-1}A/2^{k-1})\exp(-A^{\gamma_1}2^{-\gamma_1(l\wedge k/\gamma)})\\
=&\Big[\frac{tA}{2}\exp(-cd_{k-1}^{\gamma_1})+2^{k+1}\exp(-A^{\gamma_1}2^{-\gamma_1(l\wedge k/\gamma)})\Big]\exp(2^{2-k}tM_{k-1}A).
\end{aligned}
$$

After k_l+1 steps, we upper bound all telescoping distances to the sum of log-Laplace transforms for $K_{A,k_l+1,i}$, $i=1,\ldots,2^{k_l+1}$. Combining them together using the triangle inequality, we have that

$$\begin{aligned}
&\Big|\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K_A^{(l)}}\overline{X}_{M_0}(j)\Big)\Big]-\sum_{i=1}^{2^{k_l+1}}\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K_{A,k_l+1,i}^{(l)}}\overline{X}_{M_{k_l+1}}(j)\Big)\Big]\Big|\\
\le&\frac{tA}{2}\sum_{k=1}^{k_l+1}\big\{\exp(-cd_{k-1}^{\gamma_1})\exp(2^{2-k}tM_{k-1}A)\big\}+\sum_{k=1}^{k_l}\big\{2^{k+1}\exp(-A^{\gamma_1}2^{-k\gamma_1/\gamma})\exp(2^{2-k}tM_{k-1}A)\big\}\\
&+2^{k_l+1}\exp(-A^{\gamma_1}2^{-\gamma_1 l})\exp(2^{1-k_l}tM_{k_l}A).
\end{aligned}\tag{163}$$

Now we consider the steps $k=k_l+2,\ldots,l$. These steps can be simplified as one step, since at these steps, by the definitions of d_{k-1} and M_{k-1}, both the size of gaps and the truncation parameter remain unchanged. We consider this extra step. For any $i=1,\ldots,2^{k_l+1}$, $\sum_{j\in K_{A,k_l+1,i}^{(l)}}\overline{X}_{M_{k_l}+1}$ is a sum of 2^{l-k_l-1} sub-blocks, each of size n_l and bounded by $2M_{k_l+1}n_l$. In additional, the sub-blocks are equidistant with size d_{k_l+1} between two neighbouring sub-blocks. By Lemma 38, we have that

$$\begin{aligned}
&\Big|\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K_{A,k_l+1,i}^{(l)}}\overline{X}_{M_{k_l+1}}(j)\Big)\Big]-\sum_{s=(i-1)2^{l-k_l-1}+1}^{i2^{l-k_l-1}}\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in I_{l,s}}\overline{X}_{M_{k_l+1}}(j)\Big)\Big]\Big|\\
\le&tn_l2^{l-k_l-1}\exp(2^{l-k_l}tM_{k_l+1}n_l).
\end{aligned}\tag{164}$$

Combining (163) and (164), we have that

$$\begin{aligned}
&\Big|\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in K_A^{(l)}}\overline{X}_{M_0}(j)\Big)\Big]-\sum_{i=1}^{2^l}\log\mathbb{E}\Big[\exp\Big(t\sum_{j\in I_{l,i}}\overline{X}_{M_{k_l+1}}(j)\Big)\Big]\Big|\\
\le&\frac{tA}{2}\sum_{k=1}^{k_l+1}\big\{\exp(-cd_{k-1}^{\gamma_1})\exp(2^{2-k}tM_{k-1}A)\big\}+\sum_{k=1}^{k_l}\big\{2^{k+1}\exp(-A^{\gamma_1}2^{-k\gamma_1/\gamma})\exp(2^{2-k}tM_{k-1}A)\big\}\\
&+2^{k_l+1}\exp(-A^{\gamma_1}2^{-\gamma_1 l})\exp(2^{1-k_l}tM_{k_l}A)+tn_l2^l\exp(-cd_{k_l+1}^{\gamma_1})\exp(2^{l-k_l}tM_{k_l+1}n_l).
\end{aligned}\tag{165}$$

Due to the definition of d_j given in (143), we have for any $j=0,\ldots,l-1$ that

$$d_j+1\ge\lfloor c_0A2^{-(l\wedge j/\gamma)}\rfloor,$$

where for $x\in\mathbb{R}$, $\lfloor x\rfloor$ is the largest integer no greater than x. Moreover, since $c_0A2^{-(l\wedge j/\gamma)}\ge 2$ by assumption, we have

$$d_j\ge(d_j+1)/2\ge c_0A2^{-(l\wedge j/\gamma)-2}.$$

Recall also that $\gamma l<k_l+1\le l-1$ and $2^ln_l\le A$. Thus, we have that

$$\begin{aligned}
(165)\le&\frac{tA}{2}\sum_{k=1}^{k_l+1}\big\{\exp(-c(c_02^{-2}A2^{-(k-1)/\gamma})^{\gamma_1}+2^{2-k}tM_{k-1}A)\big\}\\
&+\sum_{k=1}^{k_l}\big\{2^{k+1}\exp(-A^{\gamma_1}2^{-k\gamma_1/\gamma}+2^{2-k}tM_{k-1}A)\big\}+2^{k_l+1}\exp(-A^{\gamma_1}2^{-\gamma_1 l}+2^{1-k_l}tM_{k_l}A)
\end{aligned}$$

$$
\begin{aligned}
&+ tA\exp(-c(c_0 2^{-2}A2^{-l})^{\gamma_1})\exp(2^{-k_l}tM_{k_l+1}A)\\
=&\frac{tA}{2}\sum_{k=0}^{k_l}\{\exp(-(c_1A2^{-k/\gamma})^{\gamma_1}+2^{1-k}tM_kA)\}\\
&+\sum_{k=0}^{k_l-1}\{2^{k+2}\exp(-(c_1A2^{-k/\gamma})^{\gamma_1}+2^{1-k}tM_kA)\}\\
&+2^{k_l+1}\exp(-(A2^{-l})^{\gamma_1}+2^{1-k_l}tM_{k_l}A)+tA\exp(-(c_1A2^{-l})^{\gamma_1}+2^{-k_l}tM_{k_l+1}A),
\end{aligned}
$$

where the equality is due to the definition $c_1=\min\{c^{1/\gamma_1}c_0/4, 2^{-1/\gamma}\}$.

By (157), $1/\gamma = 1/\gamma_1 + 1/\gamma_2$ with $\gamma<1$ and $\gamma l < k_l+1 \le l-1$, we have the following facts

- $2^{1-k}M_kA \le 2^{1+1/\gamma_2}(2^{-k}A)^{\gamma_1/\gamma}$ for any $k=0,\dots,k_l$,
- $M_{k_l+1}\le 2^{1/\gamma_2}(A2^{-l})^{\gamma_1/\gamma_2}\le 2^{1/\gamma_2}(A2^{-\gamma l})^{\gamma_1/\gamma_2}$,
- $2^{-k_l}M_{k_l+1}A = 2AM_{k_l+1}2^{-(k_l+1)}\le 2^{1+1/\gamma_2}A^{\gamma_1/\gamma}2^{-\gamma l\gamma_1/\gamma_2-\gamma l}\le 2^{1+1/\gamma_2}A^{\gamma_1/\gamma}2^{-\gamma_1 l}$,
- and

$$
\begin{aligned}
&2^{1-k_l}M_{k_l}A \le 2^{1+1/\gamma_2}A^{\gamma_1/\gamma}2^{-k_l}(2^{-k_l/\gamma})^{\gamma_1/\gamma_2}\le 2^{1+1/\gamma_2+\gamma_1/\gamma}(A2^{-k_l-1})^{\gamma_1/\gamma}\\
\le&2^{2+(1+\gamma_1)/\gamma_2}(A2^{-\gamma l})^{\gamma_1/\gamma}\le 2^{2+(1+\gamma_1)/\gamma_2}A^{\gamma_1/\gamma}2^{-\gamma_1 l}.
\end{aligned}
$$

Recall that $c_2 = 2^{-(2+\gamma_1/\gamma+1/\gamma_2)}c_1^{\gamma_1}$, and for $t\le c_2A^{\gamma_1(\gamma-1)/\gamma}$, we have that

$$
\begin{aligned}
(165)\le&\frac{tA}{2}\sum_{k=0}^{k_l}\{\exp(-(c_1A2^{-k/\gamma})^{\gamma_1}+2^{-(1+\gamma_1/\gamma)}(c_12^{-k/\gamma}A)^{\gamma_1})\}\\
&+\sum_{k=0}^{k_l-1}\{2^{k+2}\exp(-(c_1A2^{-k/\gamma})^{\gamma_1}+2^{-(1+\gamma_1/\gamma)}(c_12^{-k/\gamma}A)^{\gamma_1})\}\\
&+2^{k_l+1}\exp(-(c_1A2^{-l})^{\gamma_1}+2^{-1}(c_12^{-l}A)^{\gamma_1})+tA\exp(-(c_1A2^{-l})^{\gamma_1}+2^{-(1+\gamma_1/\gamma)}(c_12^{-l}A)^{\gamma_1})\\
\le&\frac{tA}{2}\sum_{k=0}^{k_l}\exp(-(c_1A2^{-k/\gamma})^{\gamma_1}/2)+\sum_{k=0}^{k_l-1}\{2^{k+2}\exp(-(c_1A2^{-k/\gamma})^{\gamma_1}/2)\}\\
&+(2^{k_l+1}+tA)\exp(-(c_1A2^{-l})^{\gamma_1}/2)\\
\le&\frac{tA}{2}(k_l+1)\exp(-(c_1A2^{-k_l/\gamma})^{\gamma_1}/2)+(2^{k_l+2}-2)\exp(-(c_1A2^{-(k_l-1)/\gamma})^{\gamma_1}/2)\\
&+(2^{k_l+1}+tA)\exp(-(c_1A2^{-l})^{\gamma_1}/2)\\
\le&(tAl+6A^\gamma)\exp(-2^{-1}(2^{-l}c_1A)^{\gamma_1}),
\end{aligned}
$$

where the last inequality follows from $2^{k_l}\le 2^{l\gamma}\le A^\gamma$, $k_l+1\le l-1$ and $l>2$.

We upper bound the log-Laplace transform of $\sum_{j\in I_{l,i}}\overline{X}_{M_{k_l+1}}(j)$ in (165). The sum is bounded by $2M_{k_l+1}n_l \le n_l 2^{1+1/\gamma_2}(A2^{-l})^{\gamma_1/\gamma_2}\le 2^{1+1/\gamma_2}(A2^{-l})^{\gamma_1/\gamma}$, which is due to (157) and $n_l\le A2^{-l}$. Note that the function $x\mapsto g(x)=x^{-2}(e^x-x-1)$ is increasing on $\mathbb{R}$ and $g(1)<1$. For any centered random variable $U\in\mathbb{R}$ such that $|U|\le M$, and any $t>0$, we have that

$$
\mathbb{E}[\exp(tU)]\le 1+t^2g(tM)\mathbb{E}[U^2]\le\exp\big(t^2g(tM)\mathbb{E}[U^2]\big). \tag{166}
$$

Let $t \le 2^{-(1+1/\gamma_2)}(A2^{-l})^{-\gamma_1/\gamma} = c_3(A2^{-l})^{-\gamma_1/\gamma}$, we have $t\sum_{j\in I_{l,i}} \overline{X}_{M_{k_l+1}}(j)$ is bounded by 1. Thus, by (166) it holds that

$$\log \mathbb{E}\Big[\exp\Big(t\sum_{j\in I_{l,i}} \overline{X}_{M_{k_l+1}}(j)\Big)\Big] \le t^2 g(1)\mathbb{E}\Big[\Big(\sum_{j\in I_{l,i}} \overline{X}_{M_{k_l+1}}(j)\Big)^2\Big] \le C_{\mathrm{LRV}} t^2 n_l, \tag{167}$$

where the second inequality follows from Lemma 41 and the fact that the functional dependence measure of the truncated process is bounded by that of the original process, i.e. for any $M > 0$

$$\big\|\overline{X}_M(s) - \overline{X}^*_M(s)\big\|_2 = \|\psi_M(X_s) - \psi_M(X^*_s)\|_2 \le \delta_{s,2},$$

where the inequality follows from that $\psi_M(x)$ is Lipschitz continuous.

Recall (157) that $\big|\sum_{j\in K_A^{(l)}} \overline{X}_{M_0}(j)\big| \le 2M_0A \le 2^{1+1/\gamma_2}A^{\gamma_1/\gamma}$. If $t \le 2^{-(1+1/\gamma_2)}A^{-\gamma_1/\gamma}$, then $t\sum_{j\in K_A^{(l)}} \overline{X}_{M_0}(j)$ is bounded by 1. By (166), we have for any $t \le 2^{-(1+1/\gamma_2)}A^{-\gamma_1/\gamma}$

$$\log \mathbb{E}\Big[\exp\Big(t\sum_{j\in K_A^{(l)}} \overline{X}_{M_0}(j)\Big)\Big] \le C_{\mathrm{LRV}} t^2 A.$$

Recall the definition of κ given in (150). For any $2^{-(1+1/\gamma_2)}A^{-\gamma_1/\gamma} \le t \le \kappa\big(A^{\gamma_1(\gamma-1)/\gamma} \vee (A2^{-l})^{-\gamma_1/\gamma}\big)$, combining (165) and (167) leads to

$$\begin{aligned}\log \mathbb{E}\Big[\exp\Big(t\sum_{j\in K_A^{(l)}} \overline{X}_{M_0}(j)\Big)\Big] &\le (tAl + 6A^{\gamma})\exp(-(2^{-l}c_1A)^{\gamma_1}/2) + C_{\mathrm{LRV}} t^2 n_l 2^l\\ &\le (tAl + 6A^{\gamma})\exp(-(2^{-l}c_1A)^{\gamma_1}/2) + C_{\mathrm{LRV}} t^2 A.\end{aligned}$$

where the first inequality follows from the triangle inequality and that the log-Laplace transform of a centered random is nonnegative. □

Proof of Proposition 45. The proof is essentially bounding all the three terms in (153).

Log-Laplace transform of the term I in (153). For any $j = 0, \dots, m(A) - 1$, by definition we have that

$$2^{-l_j}A_j = 2^{1-l_j}A_j/2 > 2^{-l}A/2 \ge (1 \vee 2c_0^{-1}). \tag{168}$$

By Proposition 44, for any $j = 0, \dots, m(A) - 1$, we have for any $t \le 2^{-(1+1/\gamma_2)}A_j^{-\gamma_1/\gamma}$ that

$$\log \mathbb{E}\Big[\exp\Big(t\sum_{i\in K_{A_j}^{(l_j)}} \overline{X}^{(j)}_{T_j}(i)\Big)\Big] \le C_{\mathrm{LRV}} t^2 A_j. \tag{169}$$

Moreover, for any $2^{-(1+1/\gamma_2)}A_j^{-\gamma_1/\gamma} \le t \le \kappa(A_j^{\gamma-1} \wedge (2^{l_j}/A_j))^{\gamma_1/\gamma}$, it holds that

$$\begin{aligned}&\log \mathbb{E}\Big[\exp\Big(t\sum_{i\in K_{A_j}^{(l_j)}} \overline{X}^{(j)}_{T_j}(i)\Big)\Big] \le (tA_jl_j + 4A_j^{\gamma})\exp(-(2^{-l_j}c_1A_j)^{\gamma_1}/2) + C_{\mathrm{LRV}} t^2 A_j\\ \le& t^2\Big((C_{\mathrm{LRV}}A_j)^{1/2} + \big(l_i^{1/2}A_j^{1/2}2^{1/2+1/(2\gamma_2)}A_j^{\gamma_1/(2\gamma)} + 2A_j^{\gamma/2}2^{1+1/\gamma_2}A_j^{\gamma_1/\gamma}\big)\exp(-(2^{-l_j}c_1A_j)^{\gamma_1}/4)\Big)^2.\end{aligned} \tag{170}$$

Note that, by construction, we have the following facts.

- For each $j = 0, \ldots, m(A) - 1$, it holds that $l_j \leq l \leq A$.
- For any $j = 0, \ldots, m(A)$, it holds that $A_j \leq A2^{-j}$.
- For any $j = 0, \ldots, m(A) - 1$, it holds that $2^{-l-1}A < 2^{-l_j}A_j \leq 2^{-l}A$.

We have that

$$
\begin{aligned}
(170) \leq & t^2 \bigg[\bigg(\frac{C_{\mathrm{LRV}} A}{2^j} \bigg)^{1/2} + \bigg(\frac{A^{1+\frac{\gamma_1}{2\gamma}}}{(2^j)^{\frac{1}{2}+\frac{\gamma_1}{2\gamma}}} 2^{\frac{1}{2}+\frac{1}{2\gamma_2}} + 2^{2+\frac{1}{\gamma_2}} \frac{A^{\frac{\gamma}{2}+\frac{\gamma_1}{\gamma}}}{(2^j)^{\frac{\gamma}{2}+\frac{\gamma_1}{\gamma}}} \bigg) \exp\bigg(- \frac{c_1^{\gamma_1}}{2^{2+\gamma_1}} \Big(\frac{A}{2^l} \Big)^{\gamma_1} \bigg) \bigg]^2 \\
\leq & t^2 \bigg[\bigg(\frac{C_{\mathrm{LRV}} A}{2^j} \bigg)^{1/2} + \bigg(\frac{A^{1+\frac{\gamma_1}{\gamma}}}{(2^j)^{\frac{\gamma}{2}+\frac{\gamma_1}{2\gamma}}} 2^{3+\frac{1}{\gamma_2}} \bigg) \exp\bigg(- \frac{c_1^{\gamma_1}}{2^{2+\gamma_1}} \Big(\frac{A}{2^l} \Big)^{\gamma_1} \bigg) \bigg]^2 = t^2 \sigma_{I,j}^2.
\end{aligned}
$$

Combining (169) and (170), we have for any $t \leq \kappa(A_j^{\gamma-1} \wedge (2^{l_j}/A_j))^{\gamma_1/\gamma}$, it holds that

$$
\log \mathbb{E}\Big[\exp\Big(t \sum_{i \in K_{A_j}^{(l_j)}} \overline{X}_{T_j}^{(j)}(i) \Big) \Big] \leq t^2 \sigma_{I,i}^2. \tag{171}
$$

Log-Laplace transform of the term II in (153). Before proceeding, we state the following fact. For any $j = 0, \ldots, m(A) - 1$, it holds that

$$
A_{j+1} \geq c_0 A_j / 3. \tag{172}
$$

Recall in the construction of $K_A^{(l)}$ that $A - \mathrm{Card}(K_A^{(l)}) = \sum_{i=0}^{l-1} 2^i d_j \geq \lfloor c_0 A \rfloor - 1$ by the definition of d_i. We thus have that $A_{j+1} \geq \lfloor c_0 A_j \rfloor - 1$. Since by (168) $c_0 A_j \geq 2$, we have that $\lfloor c_0 A_j \rfloor - 1 \geq (\lfloor c_0 A_j \rfloor + 1)/3 \geq c_0 A_j / 3$, which leads to (172).

By the property of the truncation function, for any $j = 1, \ldots, m(A) - 1$, we have that

$$
\begin{aligned}
& \Big| \sum_{i=1}^{A_j} \big(\overline{X}_{T_{j-1}}^{(j)}(i) - \overline{X}_{T_j}^{(j)}(i) \big) \Big| \leq \sum_{i=1}^{A_j} \big\{ |\psi_{T_{j-1}}(X_i) - \psi_{T_j}(X_i)| + \mathbb{E}[|\psi_{T_{j-1}}(X_i) - \psi_{T_j}(X_i)|] \big\} \\
\leq & T_{j-1} \sum_{i=1}^{A_j} \big(\mathbb{1}\{|X_i| > T_j\} + \mathbb{P}(|X_i| > T_j) \big).
\end{aligned} \tag{173}
$$

Then, (173) leads to

$$
\begin{aligned}
& \mathbb{E}\Big[\Big| \sum_{i=1}^{A_j} \big(\overline{X}_{T_{j-1}}^{(j)}(i) - \overline{X}_{T_j}^{(j)}(i) \big) \Big|^2 \Big] \leq (2T_{j-1})^2 \mathbb{E}\Big[\Big(\sum_{i=1}^{A_j} \frac{\mathbb{1}\{|X_i| > T_j\} + \mathbb{P}(|X_i| > T_j)}{2} \Big)^2 \Big] \\
\leq & (2T_{j-1})^2 A_j \sum_{i=1}^{A_j} \mathbb{E}\Big[\Big(\frac{\mathbb{1}\{|X_i| > T_j\} + \mathbb{P}(|X_i| > T_j)}{2} \Big)^2 \Big] \leq (2A_j T_{j-1})^2 H(T_j).
\end{aligned} \tag{174}
$$

By (152) and (172), (173) and (174) can be further bounded respectively as

$$
\Big| \sum_{i=1}^{A_j} \big(\overline{X}_{T_{j-1}}^{(j)}(i) - \overline{X}_{T_j}^{(j)}(i) \big) \Big| \leq 2A_j T_{j-1} \leq 2A_j 2^{1/\gamma_2} A_{j-1}^{\gamma_1/\gamma_2} \leq (3/c_0)^{\gamma_1/\gamma_2} 2^{1+1/\gamma_2} A_j^{\gamma_1/\gamma} = c_4 A_j^{\gamma_1/\gamma}, \tag{175}
$$

and

$$\mathbb{E}\Big[\Big|\sum_{i=1}^{A_j}\big(\overline{X}^{(j)}_{T_{j-1}}(i)-\overline{X}^{(j)}_{T_j}(i)\big)\Big|^2\Big]\le (c_4A_j^{\gamma_1/\gamma})^2\exp(-A_j^{\gamma_1}),$$

where $c_4=2^{1+1/\gamma_2}3^{\gamma_1/\gamma_2}c_0^{-\gamma_1/\gamma_2}$ defined in (154). Since $A_j\le A2^{-j}$ and for any

$$t\le (2c_4)^{-1}(2^j/A)^{\gamma_1(1-\gamma)/\gamma},$$

we have

$$c_4tA_j^{\gamma_1/\gamma}\le 2^{-1}A_j^{\gamma_1-\gamma_1/\gamma}A_j^{\gamma_1/\gamma}=A_j^{\gamma_1}/2.$$

Then by (166), we have for any $t\le (2c_4)^{-1}(2^j/A)^{\gamma_1(1-\gamma)/\gamma}$ that

$$\begin{aligned}\log\mathbb{E}\Big[\exp\Big(t\sum_{i=1}^{A_j}\big(\overline{X}^{(j)}_{T_{j-1}}(i)-\overline{X}^{(j)}_{T_j}(i)\big)\Big)\Big]&\le t^2c_4^2\Big(\frac{A}{2^j}\Big)^{\frac{2\gamma_1}{\gamma}}\exp\Big(-\frac{A_j^{\gamma_1}}{2}\Big)\\ \le& t^2\Big(c_4\Big(\frac{A}{2^j}\Big)^{\frac{\gamma_1}{\gamma}}\exp\Big(-\frac{1}{4}\Big(\frac{A}{2^l}\Big)^{\gamma_1}\Big)\Big)^2=t^2\sigma_{II,i}^2,\end{aligned}\tag{176}$$

where the last inequality follows from that $A_j\ge A_{m(A)-1}>A2^{-l}$ by the definition of $m(A)$ given in (151).

Log-Laplace transform of the term III in (153). We have that

$$\Big|\sum_{i=1}^{A_{m(A)}}\overline{X}^{(m(A))}_{T_{m(A)-1}}(i)\Big|\le 2T_{m(A)-1}A_{m(A)}\le c_4A_{m(A)}^{\gamma_1/\gamma}\le c_4(A2^{-l})^{\gamma_1/\gamma},$$

where the second inequality follows from the same arguments in (175), and the third inequality follows from that $A_{m(A)}\le A2^{-l}$ by the definition of $m(A)$. Then by (166), for any $t\le c_4^{-1}(A2^{-l})^{-\gamma_1/\gamma}$, we have that

$$\log\mathbb{E}\Big[\exp\Big(t\sum_{i=1}^{A_{m(A)}}\overline{X}^{(m(A))}_{T_{m(A)-1}}(i)\Big)\Big]\le t^2C_{\mathrm{LRV}}A2^{-l}\le t^2\Big[\Big(\frac{C_{\mathrm{LRV}}A}{2^l}\Big)^{1/2}\Big]^2=t^2\sigma_{III}^2.\tag{177}$$

Let

$$C=\frac{1}{\kappa}\sum_{j=0}^{m(A)-1}\Big(\Big(\frac{A}{2^j}\Big)^{1-\gamma}\vee\frac{A}{2^l}\Big)^{\gamma_1/\gamma}+2c_4\sum_{j=1}^{m(A)-1}\Big(\frac{A}{2^j}\Big)^{\gamma_1(1-\gamma)/\gamma}+c_4\Big(\frac{A}{2^l}\Big)^{\gamma_1/\gamma},$$

and

$$\begin{aligned}\sigma=&\sum_{j=0}^{m(A)-1}\sigma_{I,j}+\sum_{j=1}^{m(A)-1}\sigma_{II,j}+\sigma_{III}\\ =&\sum_{j=0}^{m(A)-1}\Big\{\Big(\frac{C_{\mathrm{LRV}}A}{2^j}\Big)^{1/2}+\Big(\frac{A^{1+\frac{\gamma_1}{\gamma}}}{(2^j)^{\frac{\gamma}{2}+\frac{\gamma_1}{2\gamma}}}2^{3+\frac{1}{\gamma_2}}\Big)\exp\Big(-\frac{c_1^{\gamma_1}}{2^{2+\gamma_1}}\Big(\frac{A}{2^l}\Big)^{\gamma_1}\Big)\Big\}\\ &+\sum_{j=1}^{m(A)-1}\Big\{c_4\Big(\frac{A}{2^j}\Big)^{\gamma_1/\gamma}\exp\Big(-\frac{1}{4}\Big(\frac{A}{2^l}\Big)^{\gamma_1}\Big)\Big\}+\Big(\frac{C_{\mathrm{LRV}}A}{2^l}\Big)^{1/2}.\end{aligned}$$

Note that $m(A) \le l$ and $l \le \log A / \log 2$. We select l as

$$l = \min\{k \in \mathbb{N}, 2^k \ge A^{\gamma} (\log A)^{\gamma/\gamma_1}\}. \tag{178}$$

This selection satisfy our initial requirement for l, i.e. $A2^{-l} \ge (2 \vee 4c_0^{-1})$, if

$$(2 \vee 4c_0^{-1})(\log A)^{\gamma/\gamma_1} \le A^{1-\gamma}. \tag{179}$$

We now show that (179) holds. For $A \ge 3$, since $\gamma_1/\gamma > 1$, we have that

$$(2 \vee 4c_0^{-1})(\log A)^{\gamma/\gamma_1} \le (2 \vee 4c_0^{-1}) \log A \le 2(1-\gamma)^{-1}(2 \vee 4c_0^{-1}) A^{(1-\gamma)/2},$$

where the last inequality follows from the inequality $\delta \log u \le u^{\delta}$ for any $\delta, u > 0$. Recall the definition $\mu = \big(2(2 \vee 4c_0^{-1})/(1-\gamma)\big)^{\frac{2}{1-\gamma}}$ given in (154). This implies that if $A > \mu$, then (179) holds and the selection of l is valid. Moreover, recall $\nu = \big(3c_4 + \kappa^{-1}\big)^{-1}(1 - 2^{-\gamma_1(1-\gamma)/\gamma})$ given in (155). Note that

$$\sum_{j=0}^{m(A)-1} \left(\frac{A}{2^j}\right)^{\gamma_1(1-\gamma)/\gamma} \le \frac{A^{\gamma_1(1-\gamma)/\gamma}}{1 - 2^{-\gamma_1(1-\gamma)/\gamma}},$$

$$\sum_{j=0}^{m(A)-1} \left(\frac{A}{2^l}\right)^{\gamma_1/\gamma} \le l\left(\frac{A}{2^l}\right)^{\gamma_1/\gamma} \le \left(\frac{A}{2^l}(\log A)^{\gamma/\gamma_1}\right)^{\gamma_1/\gamma} \le (A^{\gamma_1(1-\gamma)/\gamma}),$$

$$\sum_{j=1}^{m(A)-1} \left(\frac{A}{2^j}\right)^{\gamma_1(1-\gamma)/\gamma} \le \sum_{j=0}^{m(A)-1} \left(\frac{A}{2^j}\right)^{\gamma_1(1-\gamma)/\gamma},$$

and

$$\left(\frac{A}{2^l}\right)^{\gamma_1/\gamma} \le l\left(\frac{A}{2^l}\right)^{\gamma_1/\gamma} \le (A^{\gamma_1(1-\gamma)/\gamma}).$$

Therefore, it holds that

$$C \le \left(3c_4 + \frac{1}{\kappa}\right)\frac{A^{\gamma_1(1-\gamma)/\gamma}}{1 - 2^{-\gamma_1(1-\gamma)/\gamma}} = \nu^{-1} A^{\gamma_1(1-\gamma)/\gamma}.$$

Since $\gamma_1/\gamma > 1$, it holds that

$$\sum_{j=0}^{m(A)-1} 2^{-j/2} \le \frac{1}{1 - 2^{-1/2}}, \quad \sum_{j=0}^{m(A)-1} 2^{-j\left(\frac{\gamma}{2} + \frac{\gamma_1}{2\gamma}\right)} \le \frac{1}{1 - 2^{-\left(\frac{\gamma}{2} + \frac{\gamma_1}{2\gamma}\right)}} \le \frac{1}{1 - 2^{-1/2}},$$

and

$$\sum_{j=1}^{m(A)-1} 2^{-j\frac{\gamma_1}{\gamma}} \le \frac{2^{-\frac{\gamma_1}{\gamma}}}{1 - 2^{-\frac{\gamma_1}{\gamma}}} \le 1.$$

Therefore, it holds that

$$\sigma \le 4(C_{\mathrm{LRV}} A)^{1/2} + 28 \times 2^{\frac{1}{\gamma_2}} A^{1+\frac{\gamma_1}{\gamma}} \exp\left(-\frac{c_1^{\gamma_1}}{2^{2+\gamma_1}}\left(\frac{A}{2^l}\right)^{\gamma_1}\right) + c_4 \exp\left(-\frac{1}{4}\left(\frac{A}{2^l}\right)^{\gamma_1}\right).$$

Since by our choice of l, $A2^{-l} \geq 2^{-1}A^{1-\gamma}(\log A)^{-\gamma/\gamma_1}$, there exist constants $\nu_1, \nu_2, \nu_3 > 0$ depending only on c, γ_1 and γ_2 such that

$$\sigma^2 \leq A(\nu_1 C_{\text{LRV}} + \nu_2 \exp\big(-\nu_3 A^{\gamma_1(1-\gamma)}(\log A)^{-\gamma}\big) = AV(A).$$

Combining (153), (171), (176) and (177), and by Lemma 39, we have for any $t \in [0, \nu A^{-\gamma_1(1-\gamma)/\gamma})$ that

$$\log \mathbb{E}\Big[\exp\Big(t\sum_{i=1}^{A} \overline{X}_M(i)\Big)\Big] \leq \frac{AV(A)t^2}{1-\nu^{-1}A^{\gamma_1(1-\gamma)/\gamma}t}.$$

$\square$