

Optimal nonparametric change point analysis

Oscar Hernan Madrid Padilla

*Department of Statistics, University of California, Los Angeles
Los Angeles, California 90095-1554
e-mail: oscar.madrid@stat.ucla.edu*

Yi Yu

*Department of Statistics, University of Warwick
Coventry CV4 7AL, U.K.
e-mail: yi.yu.2@warwick.ac.uk*

Daren Wang

*Department of ACMS, University of Notre Dame
Notre Dame, IN 46556 USA
e-mail: ddwang24@nd.edu*

Alessandro Rinaldo

*Department of Statistics & Data Science, Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, U.S.A.
e-mail: arinaldo@cmu.edu*

Abstract: We study change point detection and localization for univariate data in fully nonparametric settings, in which at each time point, we acquire an independent and identically distributed sample from an unknown distribution that is piecewise constant. The magnitude of the distributional changes at the change points is quantified using the Kolmogorov–Smirnov distance. Our framework allows all the relevant parameters, namely the minimal spacing between two consecutive change points, the minimal magnitude of the changes in the Kolmogorov–Smirnov distance, and the number of sample points collected at each time point, to change with the length of the time series. We propose a novel change point detection algorithm based on the Kolmogorov–Smirnov statistic and show that it is nearly minimax rate optimal. Our theory demonstrates a phase transition in the space of model parameters. The phase transition separates parameter combinations for which consistent localization is possible from the ones for which this task is statistically infeasible. We provide extensive numerical experiments to support our theory.

MSC2020 subject classifications: Primary 62G05.

Keywords and phrases: Nonparametric, Kolmogorov–Smirnov statistic, CUSUM, minimax optimality, phase transition.

Received June 2020.

Contents

1	Introduction	1155
1.1	Problem formulation	1156
1.2	Summary of results	1158
2	Methodology	1159
3	Theory	1160
3.1	The consistency of the Kolmogorov–Smirnov detector algorithm	1160
3.2	Phase transition and minimax optimality	1163
3.3	Choice of tuning parameters	1164
4	Numerical experiments	1167
4.1	Simulations	1167
4.2	Real data analysis	1171
A	Comparisons	1173
B	Proof of Theorem 3.1	1175
C	Proofs of Section 3.2	1193
D	Proof of Theorem 3.2	1194
E	Sensitivity simulations	1196
	References	1197

1. Introduction

Change point analysis is a well-established topic in statistics that aims to detect and localize abrupt changes in the data generating distributions in time series data. Initiated during World War II (e.g. Wald, 1945), the field of change point analysis has since then produced a large body of literature as well as host of methods for statistical inference. These techniques are now widely used to address important real-life problems in a wide range of disciplines, including biology (e.g. Fan et al., 2015; Jewell et al., 2018), speech recognition (e.g. Fox et al., 2011; Chowdhury, Selouani and O’Shaughnessy, 2012), social sciences (e.g. Liu et al., 2013), climatology (e.g. Itoh and Kurths, 2010) and finance (e.g. Preuss, Puchstein and Dette, 2015; Russell and Rambaccussing, 2018).

The theoretical understanding of the statistical challenges associated with change point problems has also progressed considerably. The initial groundbreaking results of Yao (1988), Yao and Au (1989) and Yao and Davis (1986) that studied change point detection for a univariate piecewise-constant signal, have now been extended in several ways. For instance, Fryzlewicz (2014), Frick, Munk and Sieling (2014) and Kovács et al. (2020), among others, proposed computationally-efficient methods to deal with situations of potentially multiple mean change points (e.g. Wang, Yu and Rinaldo, 2020). More recently, Pein, Sieling and Munk (2017) constructed a method that can handle mean and variance changes simultaneously. Chan, Yau and Zhang (2014), Cho (2016), Cho and Fryzlewicz (2015) and Wang and Samworth (2018) studied high-dimensional mean change point detection problems. A different line of work, including efforts by Aue et al. (2009), Avanesov and Buzun (2016) and Wang, Yu and Rinaldo

(2021), has investigated scenarios where covariance matrices change. Cribben and Yu (2017), Liu et al. (2018) and Wang, Yu and Rinaldo (2018), among others, inspected dynamic network change point detection problems.

Most of the existing theoretical frameworks for statistical analysis of change point problems, however, rely heavily on strong modelling assumptions of parametric nature. Such approaches may be inadequate to capture the inherent complexity of modern, high-dimensional datasets. Indeed, the statistical literature on nonparametric change point analysis is surprisingly limited compared to its parametric counterpart. Among the nonparametric results, Carlstein (1988) considered the scenario where there is at most one change point. Harchaoui and Cappé (2007) utilized a kernel-based algorithm. Hawkins and Deng (2010) proposed a Mann–Whitney-type statistic to conduct online change point detection. Matteson and James (2014) established the consistency of change point estimators based on statistics originally introduced in Rizzo and Székely (2010). Zou et al. (2014) developed a nonparametric multiple change point detection method with consistency guarantees. Li et al. (2019) proposed two computationally-efficient kernel-based statistics for change point detection, which are inspired by the B -statistics. Padilla et al. (2018) proposed an algorithm for nonparametric change point detection based on the Kolmogorov–Smirnov statistic. Fearnhead and Rigaiil (2018) devised a mean change point detection method which is robust to outliers. Vanegas, Behr and Munk (2019) constructed a multiscale method for detecting changes in a fixed quantile of the distributions. Arlot, Celisse and Harchaoui (2019) considered a kernel version of the cumulative sum (CUSUM) statistic and an ℓ_0 -type optimization procedure which can be solved by dynamic programming. Arlot, Celisse and Harchaoui (2019) also derived an oracle inequality for their estimator, though without guarantees on change point localization.

In this paper we advance both the theory and methodology of nonparametric change point analysis by presenting a computationally-efficient procedure for univariate change point localization. The resulting method is proven to be consistent and in fact nearly minimax rate optimal for estimating the change points. Our analysis builds upon various recent contributions in the literature on parametric and high-dimensional change point analysis but allows for a fully nonparametric change point model. To the best of our knowledge, Zou et al. (2014) is one of very few examples yielding a procedure for nonparametric change point detection with provable guarantees on localization. A detailed comparison between our results and the ones of Zou et al. (2014) will be given in Appendix A, which also includes detailed comparisons of our work with Pein, Sieling and Munk (2017), Vanegas, Behr and Munk (2019) and Garreau and Arlot (2018).

1.1. Problem formulation

In this section we describe the change point model that we will consider. Our notation and settings are fairly standard, with one crucial difference from most of the contributions in the field: the changes in the underlying distribution at

the change points are not parametrically specified, but are instead quantified through a nonparametric measure of distance between distributions. This feature renders our methods and analysis applicable to a wide range of change point problems.

Condition 1.1 (Model). *Let $\{Y_{t,i}, t = 1, \dots, T, i = 1, \dots, n_t\} \subset \mathbb{R}$ be a collection of independent random variables such that $Y_{t,i} \sim F_t$, where F_1, \dots, F_T are cumulative distribution functions.*

Let $\{\eta_k\}_{k=0}^{K+1} \subset \{0, 1, \dots, T\}$ be a collection of unknown change points with $1 = \eta_0 < \eta_1 < \dots < \eta_K \leq T < \eta_{K+1} = T + 1$ such that $F_t \neq F_{t-1}$, if and only if $t \in \{\eta_1, \dots, \eta_K\}$.

The minimal spacing δ and the jump size κ are defined respectively as $\min_{k=1, \dots, K+1} \{\eta_k - \eta_{k-1}\} = \delta > 0$ and

$$\min_{k=1, \dots, K} \sup_{z \in \mathbb{R}} |F_{\eta_k}(z) - F_{\eta_{k-1}}(z)| = \min_{k=1, \dots, K} \kappa_k = \kappa > 0. \quad (1.1)$$

Furthermore, we set $n_{\min} = \min_{t=1, \dots, T} n_t$ and $n_{\max} = \max_{t=1, \dots, T} n_t$, and assume that $n_{\max} \asymp n_{\min} \asymp n$ for some n .

Throughout this paper, for any positive sequences a_T and b_T , we denote $a_T \gtrsim b_T$ if there exists $T_0 \in \mathbb{N}$ and an absolute constant $C > 0$ such that for any $T \geq T_0$, it holds $a_T \geq Cb_T$. We let $a_T \lesssim b_T$, if $b_T \gtrsim a_T$. We let $a_T \asymp b_T$, if $a_T \gtrsim b_T$ and $b_T \gtrsim a_T$.

We quantify the magnitude of the distributional changes between distribution functions at two consecutive change points using the Kolmogorov–Smirnov distance, a natural and widely-used metric for univariate probability distributions. Though stronger than the weak convergence, convergence in the Kolmogorov–Smirnov distance is weaker than many other forms of convergence. Examples of those stronger forms of convergence include the ones induced by the total variation distance and, provided that the distributions admit bounded Lebesgue densities, by the L_1 -Wasserstein distance. Nonetheless, reliance on the Kolmogorov–Smirnov distance offers significant advantages: (1) it allows for fully nonparametric change point analysis settings and procedures, that hold with minimal assumptions on the underlying distribution functions and (2) is amenable to statistical analysis. Although the Kolmogorov–Smirnov statistics is known to exhibit low power in goodness-of-fit testing problems (e.g. Duembgen and Wellner, 2014), this issue does not affect our results: as shown below in Lemma 3.2, our Kolmogorov–Smirnov-based change point localization procedure is nearly minimax rate optimal. Our numerical experiments further confirm that our procedure compares favourably with other nonparametric methods.

In Condition 1.1, we allow for multiple observations n_t to be collected at each time t . This generalizes the classical change point detection framework where $n_t = 1$ for all t (e.g. Zou et al., 2014). This flexibility is inspired by the recent interest in anomaly detection problems where multiple observations can be measured at a fixed time (e.g. Chan et al., 2014; Reinhart, Athey and Biegalski, 2014; Padilla et al., 2018). Our results remain valid even if $n_t = 1$ for all t .

The nonparametric change point model defined above in Condition 1.1 is specified by a few key parameters: the minimal spacing between two consecutive change points δ , the minimal jump size in terms of the Kolmogorov–Smirnov distance κ , and the number n_t of data points acquired at each time t . We adopt a high-dimensional framework in which all these quantities are allowed to change as functions of the total number of time points T .

We consider the change point localization problem of establishing consistent change point estimators $\{\hat{\eta}_k\}_{k=1}^{\hat{K}}$ of the true change points. These are measurable functions of data and return an increasing sequence of random time points $\hat{\eta}_1 < \dots < \hat{\eta}_{\hat{K}}$, such that, as $T \rightarrow \infty$, the following event holds with probability tending to 1: $\hat{K} = K$ and $\max_{k=1, \dots, K} |\hat{\eta}_k - \eta_k| \leq \epsilon$, where $\epsilon = \epsilon(T)$ is such that $\lim_{T \rightarrow \infty} \epsilon/\delta = 0$.

Throughout the rest of the paper, we refer to the quantity ϵ/δ as the localization rate. Our goal is to obtain change point estimators that, under the weakest possible conditions, are guaranteed to yield the smallest possible ϵ .

1.2. Summary of results

We will show that under Condition 1.1, the hardness of the change point localization task is fully captured by the quantity

$$\kappa\delta^{1/2}n^{1/2}, \quad (1.2)$$

which can be regarded as a signal-to-noise ratio of sort. We list our contributions as follows.

- We demonstrate the existence of a phase transition for the localization task in terms of the signal-to-noise ratio $\kappa\delta^{1/2}n^{1/2}$. We show that in the low signal-to-noise ratio regime $\kappa\delta^{1/2}n^{1/2} \lesssim 1$, no algorithm is guaranteed to be consistent. We also show that if $\kappa\delta^{1/2}n^{1/2} \geq \zeta_T$, where ζ_T is any diverging sequence as $T \rightarrow \infty$, then a minimax lower bound of the localization error rate is of order $(n\kappa^2T)^{-1}$.
- We develop a novel detection procedure, based on the Kolmogorov–Smirnov distance, given in Algorithm 1. We show that under suitable conditions, our method is consistent and nearly minimax rate optimal, up to logarithmic factors, in terms of both the required signal-to-noise ratio. The specific assumption is given in Condition 3.1, and the localization error rate in Theorem 3.1. Interestingly, for the lower bounds on the signal-to-noise ratio and the localization error rate, our rates match those derived for the univariate mean change point localization problem under sub-Gaussian noise, e.g. Wang, Yu and Rinaldo (2020).
- We provide extensive comparisons of our algorithm and theoretical guarantees with several competing methods and results from the literature. See Appendix A and Section 4, respectively. In particular, our simulations indicate that our procedure performs well across a variety of scenarios, in terms of both estimating the number of change points and their locations.

We point out that, although in deriving the theoretical guarantees for our methodologies we follow techniques proposed in existing work, namely Venkatraman (1992) and Fryzlewicz (2014), our results deliver improvements in two aspects. Firstly, the extension to nonparametric settings, in which the magnitude of the distributional changes is measured by the Kolmogorov–Smirnov distance, requires novel and nontrivial arguments, especially to quantify the order of the stochastic fluctuations of the associated CUSUM statistics. Secondly, the arguments used in Fryzlewicz (2014) for the theoretical analysis of the performance of the WBS algorithm have to be sharpened in order to allow for all the model parameters to vary as the sample size diverges and in order to yield optimal localization rates.

2. Methodology

In this section, we detail our Kolmogorov–Smirnov detector procedure, which is based on the CUSUM Kolmogorov–Smirnov statistic defined next. Similar or related Kolmogorov–Smirnov statistics based procedures have been considered previously in Darkhovski (1994), Boukai and Zhou (1997) and Padilla et al. (2018), among others.

Definition 2.1 (The CUSUM Kolmogorov–Smirnov statistic). *For any integer triplet (s, t, e) , $1 \leq s < t < e \leq T$, define the CUSUM Kolmogorov–Smirnov statistic as*

$$D_{s,e}^t = \sup_{z \in \mathbb{R}} |D_{s,e}^t(z)|, \quad (2.1)$$

with

$$D_{s,e}^t(z) = \sqrt{\frac{n_{s:t}n_{(t+1):e}}{n_{s:e}}} \left\{ \widehat{F}_{s:t}(z) - \widehat{F}_{(t+1):e}(z) \right\},$$

where for any integer pair (s, e) with $s < e$ and $z \in \mathbb{R}$, we write

$$\widehat{F}_{s:e}(z) = \frac{1}{n_{s:e}} \sum_{t=s}^e \sum_{i=1}^{n_t} 1_{\{Y_{t,i} \leq z\}} \quad \text{and} \quad n_{s:e} = \sum_{t=s}^e n_t.$$

In Definition 2.1, $n_{s:e}$ is the total number of observations collected in the integer interval $[s, e]$, and $\widehat{F}_{s:e}$ is the empirical cumulative distribution function estimated using the data collected in $[s, e]$.

The proposed procedure applies the wild binary segmentation procedure (Fryzlewicz, 2014). The latter was originally developed for univariate mean change point detection problems based on the univariate CUSUM statistic. We, instead, use the CUSUM Kolmogorov–Smirnov statistic. To be specific, we draw a collection of random time intervals, and within each interval, we search the time point which maximizes the CUSUM Kolmogorov–Smirnov statistic. If the corresponding maximal value of such statistic exceeds an appropriate threshold, then that time point is added to the collection of estimated change points. The process is then repeated separately for each of the resulting two time sub-intervals and stops when all the CUSUM Kolmogorov–Smirnov statistics are below the

threshold, or when the resulting time interval is too narrow. See Algorithm 1 for details.

Algorithm 1 Kolmogorov–Smirnov Detector. $\text{KSD}((s, e), \{(\alpha_m, \beta_m)\}_{m=1}^M, \tau)$

INPUT: Sample $\{Y_{t,i}\}_{t=s,i=1}^{e,n_t} \subset \mathbb{R}$, collection of intervals $\{(\alpha_m, \beta_m)\}_{m=1}^M$ and tuning parameter $\tau > 0$.

for $m = 1, \dots, M$ **do**
 $(s_m, e_m) \leftarrow (s, e) \cap (\alpha_m, \beta_m)$
 if $e_m - s_m > 2$ **then**
 $a_m \leftarrow \max_{t=s_m+1, \dots, e_m-1} D_{s_m, e_m}^t$
 $b_m \leftarrow \arg \max_{t=s_m+1, \dots, e_m-1} D_{s_m, e_m}^t$
 else
 $a_m \leftarrow -1$
 end if
end for
 $m^* \leftarrow \arg \max_{m=1, \dots, M} a_m$
if $a_{m^*} > \tau$ **then**
 Add b_{m^*} to the set of estimated change points
 $\text{KSD}((s, b_{m^*}), \{(\alpha_m, \beta_m)\}_{m=1}^M, \tau)$ ▷ Recursively call the function KSD
 $\text{KSD}((b_{m^*} + 1, e), \{(\alpha_m, \beta_m)\}_{m=1}^M, \tau)$
end if

OUTPUT: The set of estimated change points.

In Algorithm 1, the main input is the threshold τ , a tuning parameter controlling the number of returned change points, with larger values of τ producing smaller numbers of estimated change points. Our theory in the next section will shed some lights on how to choose τ .

For any integer triplet (s, t, e) , $1 \leq s < t < e \leq T$, the computational cost of calculating $D_{s,e}^t$ is of order $O\{(e-s)n_{s:e} \log(n_{s:e})\}$. This can be seen using a naïve calculation based on the merge sort algorithm (e.g. Knuth, 1998), and the fact that the supremum in (2.1) only needs to be taken over $z \in \{Y_{u,i} : s \leq u \leq e, i = 1, \dots, n_u\}$. Algorithm 1 therefore has the worst case running time of order $O\{MTn_{1:T} \log(n_{1:T})\}$ where M is the number of randomly drawn time intervals.

3. Theory

3.1. The consistency of the Kolmogorov–Smirnov detector algorithm

We first state a condition involving the minimal spacing, the minimal jump size and the total number of time points T , which overall amount to a minimal signal-to-noise ratio condition.

Condition 3.1. *There exists an absolute constant $C_{\text{SNR}} > 0$ such that*

$$\kappa \delta^{1/2} n^{1/2} > C_{\text{SNR}} \phi_T \log^{1/2}(n_{1:T}),$$

where ϕ_T is any diverging sequence as T grows unbounded.

The scaling exhibited in Condition 3.1 covers nearly all combinations of model parameters for which the localization task is feasible: in Section 3.2 we show that no estimator of the change points is guaranteed to be consistent when the model parameters violate Condition 3.1, up to a poly-logarithmic term in T .

We then show that, under Condition 3.1 and with appropriate choices of the input parameters, the Kolmogorov–Smirnov detector procedure will yield, with high probability, the correct number of change points and a vanishing localization rate. In fact, as shown below in Section 3.2, the resulting localization rates are nearly minimax rate optimal (achieving the minimax risk, e.g. Section 2.1 in Tsybakov, 2009).

Theorem 3.1 (Guarantees for the Kolmogorov–Smirnov detector). *Assume the inputs of Algorithm 1 are as follows. (i) The sequence $\{Y_{t,i}\}_{i=1,t=1}^{n_t,T}$ satisfies Conditions 1.1 and 3.1. (ii) The collection of intervals $\{(\alpha_m, \beta_m)\}_{m=1}^M \subset \{1, \dots, T\}$, with endpoints drawn independently and uniformly from $\{1, \dots, T\}$, satisfy*

$$\max_{m=1, \dots, M} (\beta_m - \alpha_m) \leq C_M \delta, \tag{3.1}$$

almost surely, for an absolute constant $C_M > 0$. (iii) The tuning parameter τ satisfies

$$c_{\tau,1} \log^{1/2}(n_{1:T}) \leq \tau \leq c_{\tau,2} \kappa \delta^{1/2} n^{1/2}, \tag{3.2}$$

where $c_{\tau,1}, c_{\tau,2} > 0$ are absolute constants.

Let $\{\hat{\eta}_k\}_{k=1}^{\hat{K}}$ be the corresponding output of Algorithm 1. Then, there exists a constant $C_\epsilon > 0$ such that

$$\begin{aligned} & \mathbb{P} \left\{ \hat{K} = K \quad \text{and} \quad \epsilon_k \leq C_\epsilon \kappa_k^{-2} \log(n_{1:T}) n^{-1}, \forall k = 1, \dots, K \right\} \\ & \geq 1 - \frac{24 \log(n_{1:T})}{T^3 n_{1:T}} - \frac{48T}{n_{1:T} \log(n_{1:T}) \delta} - \exp \left\{ \log \left(\frac{T}{\delta} \right) - \frac{M \delta^2}{16T^2} \right\}, \end{aligned} \tag{3.3}$$

where $\epsilon_k = |\hat{\eta}_k - \eta_k|$ for $k = 1, \dots, K$.

If $M \gtrsim \log(T/\delta) T^2 \delta^{-2}$, then the probability in (3.3) approaches to 1 as $T \rightarrow \infty$, which shows that Algorithm 1 is consistent.

Based on Condition 3.1, the range of tuning parameters τ defined in (3.2) is not empty, and the upper bound on the localization error rate satisfies

$$\frac{\epsilon}{\delta} = \max_{k=1, \dots, K} \frac{\epsilon_k}{\delta} \lesssim \phi_T^{-2} \rightarrow 0, \quad T \rightarrow \infty,$$

where the inequality follows from Condition 3.1. It is important to point out that Theorem 3.1 yields a family of localization rates that depend on how n , κ and δ scale with T . The slow rate ϕ_T^{-2} exhibited in the last display represents the worst case scenario corresponding to the weakest possible signal-to-noise ratio afforded in Condition 3.1. In fact, for most combinations of the model parameters, the rate can be much faster. For instance, when κ is bounded away from zero, the resulting rate is of order $O \{ \log(n_{1:T}) n^{-1} T^{-1} \phi_T \}$, if $\delta \asymp T$. Still assuming a non-vanishing κ , and provided that n increases with T , our Kolmogorov–Smirnov

detector estimator remains consistent even if δ is as small as $\log(n_{1:T})$, with a localization rate of order $O(n^{-1})$. Finally, if the change points are evenly spaced with $\delta \asymp T^{1/2}$, then the number of change points satisfies $K \asymp T^{1/2}$. In this case the Kolmogorov–Smirnov detector procedure will output a consistent estimator of the change points even with κ tending to 0 as T increases, as long as the rate of decay for κ is slower than $\phi_T \log^{1/2}(n_{1:T})n^{-1/2}T^{-1/4}$.

Remark 3.1 (Comments on (3.1)). *The condition (3.1) trivially holds when the number of change points K remains bounded as $T \rightarrow \infty$, and we sample the intervals $\{(\alpha_m, \beta_m)\}_{m=1}^M \subset \{1, \dots, T\}$ uniformly in $\{1, \dots, T\}$. However, if $K \rightarrow \infty$ as T grows, then (3.1) is somewhat unsatisfactory, as it assumes some knowledge about the rate of growth of the minimal spacing δ . This may not be available in practice, even though in many cases an educated guess on the minimal spacing is not too unreasonable to assume. Though (3.1) does not appear among the assumptions of Theorem 3.2 in Fryzlewicz (2014) about the performance of the wild binary segmentation algorithm, it is implicitly assumed there. More generally, some knowledge of δ , in one form or another, is routinely assumed in order to demonstrate consistency or optimality of wild binary segmentation-like algorithms, see, e.g. Wang and Samworth (2018), Wang, Yu and Rinaldo (2020), Baranowski, Chen and Fryzlewicz (2019), Anastasiou and Fryzlewicz (2019) and Eichinger and Kirch (2018). For more discussions on this, see Wang, Yu and Rinaldo (2020).*

In the proofs, we randomly draw intervals (a_m, b_m) and let $(\alpha_m, \beta_m) = (a_m, b_m)$ if $b_m - a_m \leq C_M \delta$, otherwise discard (a_m, b_m) . This procedure continues until we have M such intervals. In practice, as demonstrated in Section 4, there are always only finite number of change points and we therefore do not need to impose (3.1).

We remark that condition (3.1) is only needed to guarantee that the localization rate of the Kolmogorov–Smirnov detector procedure given in (3.2) is nearly minimax rate optimal, see the next section, but not to ensure consistency. In fact, (3.1) is not at all necessary for consistency. For instance, assuming that

$$\kappa(\delta n)^{1/2}(\delta/T) \gtrsim \phi_T \log^{1/2}(n_{1:T}),$$

a slightly stronger signal-to-noise ratio setting than those allowed in Condition 3.1, a simple adaptation of the proof of Theorem 3.1 shows that, even without (3.1), the Kolmogorov–Smirnov detector procedure will be consistent with a vanishing localization rate $\epsilon/\delta \lesssim \log(n_{1:T})\kappa^{-2}n^{-1}T^2\delta^{-3}$.

Theorem 3.1 delivers a strict improvement upon the guarantees claimed in other nonparametric change point detection papers, since it guarantees localization rates that are local. Indeed, according to Theorem 3.1, each change point η_k possesses its own localization rate, depending on κ_k , the magnitude of the corresponding distributional change. In particular, change points at which the distributional change is more significant can be estimated more accurately.

The detailed proof of Theorem 3.1 can be found in Appendix B. Here we provide a brief summary of the technical arguments used there. The first key

step is to obtain finite sample concentration inequalities to control the deviation between the CUSUM Kolmogorov–Smirnov statistics $\{D_{s,e}^t : 1 \leq s < t \leq e \leq T\}$ and their population versions. The rest of the proofs are conducted on the so-called good events, where with probabilities tending to 1 as T grows unbounded, the fluctuations remain within an appropriate range. Next, we show that the population versions of the CUSUM Kolmogorov–Smirnov statistics are maximized at the true change points and, crucially, decay rapidly away from them. Consequently, on the good events, our algorithm will correctly detect or reject the existence of change points and localize the change points accurately.

3.2. Phase transition and minimax optimality

In this subsection, we prove that Algorithm 1 is optimal, in the sense of guaranteeing nearly minimax optimal localization rates across almost all models for which the localization task is possible. Towards that end, recall that in Theorem 3.1 we have shown that Algorithm 1 provides consistent change point estimators under the assumption that

$$\kappa(\delta n)^{1/2} \gtrsim \phi_T \log^{1/2}(n_{1:T}). \tag{3.4}$$

In Lemma 3.1 below, we will show that if

$$\kappa(\delta n)^{1/2} \lesssim 1, \tag{3.5}$$

then no algorithm is guaranteed to be consistent uniformly over all possible change point problems. In light of (3.4), (3.5), Theorem 3.1 and Lemma 3.1 below, we have found a phase transition over the space of model parameters that separates scalings of the model parameters for which the localization task is impossible, from those for which Algorithm 1 is consistent. The separation between these two regions is sharp, saving for the term $\phi_T \log^{1/2}(n_{1:T})$.

Lemma 3.1. *Let $\{Y_{t,i}\}_{i=1,t=1}^{n,T}$ be a time series satisfying Condition 1.1 with one and only one change point. Let $P_{\kappa,n,\delta}^T$ denote the corresponding joint distribution. Set $\mathcal{P} = \left\{ P_{\kappa,n,\delta}^T : \delta = \min \left\{ \lfloor 2^{-1} \kappa^{-2} n^{-1} \rfloor, \lfloor T/3 \rfloor \right\} \right\}$ and, for each $P \in \mathcal{P}$, let $\eta(P)$ be the corresponding change point. Then,*

$$\inf_{\hat{\eta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P (|\hat{\eta} - \eta(P)|) \geq T/6,$$

where the infimum is over all possible estimators of the change point locations.

In our next result, we derive a minimax lower bound on the localization task, which applies to nearly all combinations of model parameters outside the impossibility region found in Lemma 3.1.

Lemma 3.2. *Let $\{Y_{t,i}\}_{i=1,t=1}^{n,T}$ be a time series satisfying Condition 1.1 with one and only one change point. Let $P_{\kappa,n,\delta}^T$ denote the corresponding joint distribution. Consider the class of distributions*

$$\mathcal{Q} = \left\{ P_{\kappa,n,\delta}^T : \delta \leq T/2, \kappa < 1/2, \kappa(\delta n)^{1/2} \geq \zeta_T \right\},$$

where $\{\zeta_T\}$ is any sequence such that $\lim_{T \rightarrow \infty} \zeta_T = \infty$. Then, for all T large enough, it holds that

$$\inf_{\hat{\eta}} \sup_{P \in \mathcal{Q}} \mathbb{E}_P(|\hat{\eta} - \eta(P)|) \geq \max \left\{ 1, (2e^2 n \kappa^2)^{-1} \right\},$$

where the infimum is over all possible estimators of the change point locations.

Note that the condition $\delta \leq T/2$ automatically holds due to the definition of δ . The above lower bound matches, saving for a poly-logarithmic factor, the localization rate we have established in Theorem 3.1, thus showing that Algorithm 1 is nearly minimax rate optimal.

3.3. Choice of tuning parameters

Algorithm 1 calls for three tuning parameters: the upper bound of random interval widths C_M , the number of random intervals M and the threshold τ . In this subsection, we discuss the practical guidances on choosing the tuning parameters.

The constant C_M is involved in (3.1) and purely for theoretical purposes. Since we allow the number of change points to diverge, to prompt the nearly minimax optimality, the constant C_M is required in Theorem 3.1. In practice, there are only finite change points in any given data sets, hence (3.1) automatically holds and this is not a tuning parameter in use.

For the number of random intervals M , as stated in Theorem 3.1, we need to choose M to satisfy that $M \gtrsim \log(T/\delta)T^2\delta^{-2}$. However, in practice, δ is likely unknown but it holds that $\delta \asymp T$, which leads to $M \gtrsim 1$. In all the numerical experiments in Section 4, we let $M = 120$, which works well in practice.

Arguably, the most important tuning parameter in Algorithm 1 is τ , whose value determines whether a candidate time point should be deemed a change point. If we let τ decrease from ∞ to 0, then the procedure produces more and more change points. In particular, if all the other inputs, namely $\{Y_{t,i}\}$ and $\{(\alpha_m, \beta_m)\}$, are kept fixed, then it holds that $\mathcal{B}(\tau_1) \subseteq \mathcal{B}(\tau_2)$, for $\tau_1 \geq \tau_2$, where $\mathcal{B}(\tau)$ is the collection of estimated change points returned by Algorithm 1 when a value of τ for the threshold parameter is used. We take advantage of such nesting in order to design a data-driven method for picking τ .

To proceed, we now introduce Algorithm 2. Algorithm 2 is a generic procedure that can be used for merging two collections of estimated change points \mathcal{B}_1 and \mathcal{B}_2 . Algorithm 2 deletes from $\mathcal{B}_1 \cup \mathcal{B}_2$ potential false positives by checking their validity one by one based on the CUSUM Kolmogorov–Smirnov statistics. However, Algorithm 2 does not scan for potential false positives in the set $\mathcal{B}_1 \cap \mathcal{B}_2$.

The criterion deployed in Algorithm 2 is based on the following check:

$$\sum_{t=\hat{\eta}_k+1}^{\eta} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \hat{z}\}} - \hat{F}_{(\hat{\eta}_k+1):\eta}(\hat{z}) \right\}^2 + \sum_{t=\hat{\eta}_k+1}^{\hat{\eta}_{k+1}} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \hat{z}\}} - \hat{F}_{(\eta+1):\hat{\eta}_{k+1}}(\hat{z}) \right\}^2$$

Algorithm 2 Model comparison

INPUT: Sample $\{Y_{t,i}\}_{t=1,i=1}^{T,n_t} \subset \mathbb{R}$, candidate models $\mathcal{B}_1, \mathcal{B}_2$.
 $\mathcal{C} \leftarrow (\mathcal{B}_2 \setminus \mathcal{B}_1) \cup (\mathcal{B}_1 \setminus \mathcal{B}_2)$
 $n_c \leftarrow |\mathcal{C}|$
 $\widehat{\mathcal{B}} \leftarrow \mathcal{B}_1 \cap \mathcal{B}_2$
for $i = 1, \dots, n_c$ **do**
 $\eta \leftarrow \eta_i \in \mathcal{C}$
 if $\eta \in \mathcal{B}_2 \setminus \mathcal{B}_1$ **then**
 Set k to be the integer satisfying $\eta \in (\widehat{\eta}_k, \widehat{\eta}_{k+1})$, where $\{\widehat{\eta}_k, \widehat{\eta}_{k+1}\} \subset \mathcal{B}_1$
 else
 Set k to be the integer satisfying $\eta \in (\widehat{\eta}_k, \widehat{\eta}_{k+1})$, where $\{\widehat{\eta}_k, \widehat{\eta}_{k+1}\} \subset \mathcal{B}_2$
 end if
 $\widehat{z} \leftarrow \min \arg \max_{z \in \{Y_{t,i}\}_{t=1,i=1}^{T,n_t}} \left| D_{\widehat{\eta}_k, \widehat{\eta}_{k+1}}^\eta(z) \right|$
 if (3.6) holds **then**
 $\widehat{\mathcal{B}} \leftarrow \widehat{\mathcal{B}} \cup \{\eta\}$
 end if
end for
OUTPUT: $\widehat{\mathcal{B}}$

$$+ \lambda < \sum_{t=\widehat{\eta}_k+1}^{\widehat{\eta}_{k+1}} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \widehat{z}\}} - \widehat{F}_{(\widehat{\eta}_k+1):\widehat{\eta}_{k+1}}(\widehat{z}) \right\}^2, \tag{3.6}$$

where $\lambda > 0$ is a specified tuning parameter to be discussed in Theorem 3.2.

In order to have a practical scheme for selecting τ , we now propose Algorithm 3. Algorithm 3 is a fully data-driven change point detection procedure with automated tuning parameter selection. Algorithm 3 requires two independent subsamples: $\{Y_{t,i}\}$ and $\{W_{t,i}\}$. In practice, one can perform sample splitting. If $n_t \geq 2$, for all t , then one can partition the data at every time point. If all or some n_t 's equal 1, then for $t \in \{j : n_j = 1\}$, we randomly assign the associated observation to $\{Y_{t,i}\}$ or $\{W_{t,i}\}$ with equal probability. In fact, there is no need to ensure that both subsamples have exactly the same number of sample points n_t for all t . Our theoretical guarantees in Theorem 3.2 still hold as long as the number of observations have the same scaling at each time point in the two samples $\{Y_{t,i}\}$ and $\{W_{t,i}\}$. When calling (3.6) in Algorithm 3, the empirical distribution functions are constructed based on $\{Y_{t,i}\}$. To present Algorithm 3, we slightly abuse the notation: in order to emphasize that Algorithm 1 is conducted on the sample $\{W_{t,i}\}$, we include $\{W_{t,i}\}$ as a formal input to the Kolmogorov–Smirnov detector algorithm. Since the CUSUM Kolmogorov–Smirnov statistics are based on $\{Y_{t,i}\}$, we use the notation $D_{\widehat{\eta}_k, \widehat{\eta}_{k+1}}^\eta(z, \{Y_{t,i}\})$ instead.

As for the implementation of Algorithm 3, we arrange all the candidate sets in increasing order of their corresponding τ_j values. This ensures a decreasing nesting of the candidate change points. We begin with the set corresponding to the smallest τ_j and compare consecutive sets. However, unlike Algorithm 2, we pick a single element from the difference sets, and decide to move on or to terminate the procedure. Theorem 3.2 provides suitable conditions guaranteeing that this procedure results in a consistent estimator of the change points.

Algorithm 3 Kolmogorov–Smirnov detector with tuning parameter selection

INPUT: Sample $\{Y_{t,i}\}_{t=1,i=1}^{T,n_t} \subset \mathbb{R}$, collection of intervals $\{(\alpha_m, \beta_m)\}_{m=1}^M$ and tuning parameter $\{\tau_j\}_{j=1}^J$.

for $j = 1, \dots, J$ **do**

$\mathcal{B}_j \leftarrow \text{KSD}((0, T), \{(\alpha_m, \beta_m)\}_{m=1}^M, \tau_j, \{W_{t,i}\})$ ▷ Algorithm 1

$\mathcal{O} \leftarrow \mathcal{B}_j$

end for

for $j = 1, \dots, J - 1$ **do**

if $\mathcal{B}_{j+1} \neq \mathcal{O}$ **then**

$\eta \leftarrow \min\{x : x \in \mathcal{O} \setminus \mathcal{B}_{j+1}\}$

Set k to be the integer satisfying $\eta \in (\hat{\eta}_k, \hat{\eta}_{k+1})$, where $\{\hat{\eta}_k, \hat{\eta}_{k+1}\} \subset \mathcal{B}_{j+1}$

$\hat{z} \leftarrow \min \arg \max_{z \in \{Y_{t,i}\}_{t=1,i=1}^{T,n_t}} |D_{\hat{\eta}_k, \hat{\eta}_{k+1}}^\eta(z, \{Y_{t,i}\})|$

if (3.6) holds **then**

$\mathcal{O} \leftarrow \mathcal{B}_{j+1}$

else

Terminate the algorithm

end if

else

$\mathcal{O} \leftarrow \mathcal{B}_{j+1}$

end if

end for

OUTPUT: \mathcal{O}

Theorem 3.2. *Suppose that the following holds. (i) The sequences $\{Y_{t,i}\}_{i=1,t=1}^{n_t,T}$, $\{W_{t,i}\}_{i=1,t=1}^{n_t,T}$ are independent and satisfy Conditions 1.1 and 3.1. (ii) The collection of intervals $\{(\alpha_m, \beta_m)\}_{m=1}^M \subset \{1, \dots, T\}$, whose endpoints are drawn independently and uniformly from $\{1, \dots, T\}$, satisfy $\max_{m=1, \dots, M} (\beta_m - \alpha_m) \leq C_M \delta$, almost surely, for an absolute constant $C_M > 1$. (iii) The tuning parameters $\{\tau_j\}_{j=1}^J$ satisfy*

$$\tau_J > \dots > c_{\tau,2} \kappa \delta^{1/2} n^{1/2} > \dots > \tau_{j^*} > \dots > c_{\tau,1} \log^{1/2}(n_{1:T}) > \dots > \tau_1, \quad (3.7)$$

where $c_{\tau,1}, c_{\tau,2} > 0$ are absolute constants and for some $j^* \in \{2, \dots, J - 1\}$.

Let $\hat{\mathcal{B}} = \{\hat{\eta}_1, \dots, \hat{\eta}_{\hat{K}}\}$ be the output of Algorithm 3 with inputs satisfying the conditions above. If $\lambda = C \log(n_{1:T})$, with a large enough constant $C > 0$, then

$$\begin{aligned} & \mathbb{P} \left\{ \hat{K} = K \quad \text{and} \quad \epsilon_k \leq C \epsilon \kappa_k^{-2} \log(n_{1:T}) n^{-1}, \forall k = 1, \dots, K \right\} \\ & \geq 1 - \frac{48 \log(n_{1:T})}{T^3 n_{1:T}} - \frac{96T}{n_{1:T} \log(n_{1:T}) \delta} - \exp \left\{ \log \left(\frac{T}{\delta} \right) - \frac{M \delta^2}{16T^2} \right\}. \end{aligned}$$

The proof of Theorem 3.2 can be found in Appendix D. It implicitly assumes that the nested sets $\{\mathcal{B}_j\}$ in Algorithm 3 satisfy $|\mathcal{B}_j \setminus \mathcal{B}_{j+1}| = 1$, for $j = 1, \dots, J$. If this condition is not met, then the conclusion of Theorem 3.2 still holds provided that we replace the inequality condition in Algorithm 3, with the inequality $\lambda > \max_{m=1, \dots, M} \sup_{z \in \mathbb{R}} |D_{a_m, b_m}^\eta(z, \{Y_{t,i}\})|^2$, where $(a_m, b_m) = (\hat{\eta}_k, \hat{\eta}_{k+1}) \cap (\alpha_m, \beta_m)$ for $m = 1, \dots, M$.

A similar proposal for selecting the threshold tuning parameter for the wild binary segmentation algorithm can be found in Theorem 3.3 in Fryzlewicz

(2014). The proof of our Theorem 3.2 delivers a more careful and precise analysis.

Finally, Theorem 3.2 suggests to choose the parameter λ in Algorithm 3 as $\lambda = C \log(n_{1:T})$. In practice, we set $C = 2/3$ and find this choice performing reasonably well. We also include further simulations in Appendix E showing that Algorithm 3 is less sensitive to the parameter C than Algorithm 1 is to the parameter τ .

4. Numerical experiments

4.1. Simulations

In this section we present the results of various simulation experiments aimed to assess the performances of our method in a wide range of scenarios and in relation to other competing methods. The R (R Core Team, 2019) code used in our simulations is available at <https://github.com/hernanmp/NWBS>. We measure the performance of an estimator \hat{K} of the true number of change points K by the absolute error $|\hat{K} - K|$. In all our examples, we report the average absolute errors over 100 repetitions. Furthermore, denoting by $\mathcal{C} = \{\eta_1, \dots, \eta_K\}$ the set of true change points, the performance of an estimator $\hat{\mathcal{C}}$ of \mathcal{C} is measured by the one-sided Hausdorff distance $d(\hat{\mathcal{C}}|\mathcal{C}) = \max_{\eta \in \mathcal{C}} \min_{x \in \hat{\mathcal{C}}} |x - \eta|$. By convention we set it to be infinity when $\hat{\mathcal{C}}$ is an empty set. If $\hat{\mathcal{C}} = \{1, \dots, T\}$, then $d(\hat{\mathcal{C}}|\mathcal{C}) = 0$. Thus, $d(\hat{\mathcal{C}}|\mathcal{C})$ can be insensitive to overestimation. To overcome this, we also calculate $d(\mathcal{C}|\hat{\mathcal{C}})$. In all of our simulations, for a method that produces an estimator $\hat{\mathcal{C}}$, we report the median of both $d(\hat{\mathcal{C}}|\mathcal{C})$ and $d(\mathcal{C}|\hat{\mathcal{C}})$ over 100 repetitions.

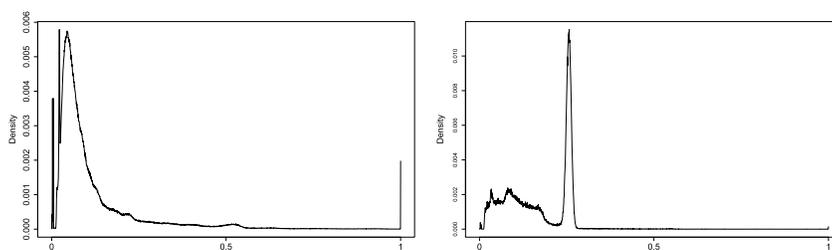


FIG 1. A plot showing the densities considered in Scenario 1 (Padilla et al., 2018).

We start by focusing on the case in which $n_t = 1$ for all $t = 1, \dots, T$. The methods that we benchmark against are the following: wild binary segmentation (Fryzlewicz, 2014; Baranowski and Fryzlewicz, 2019), Bai and Perron's method (Bai and Perron, 2003; Zeileis et al., 2002), pruned dynamic programming (Rigaill, 2010; Cleynen, Rigaill and Koskas, 2016), pruned exact linear time algorithm (Killick, Fearnhead and Eckley, 2012; Killick and Eckley, 2014), the simultaneous multiscale change point estimator of (Frick, Munk and Sieling,

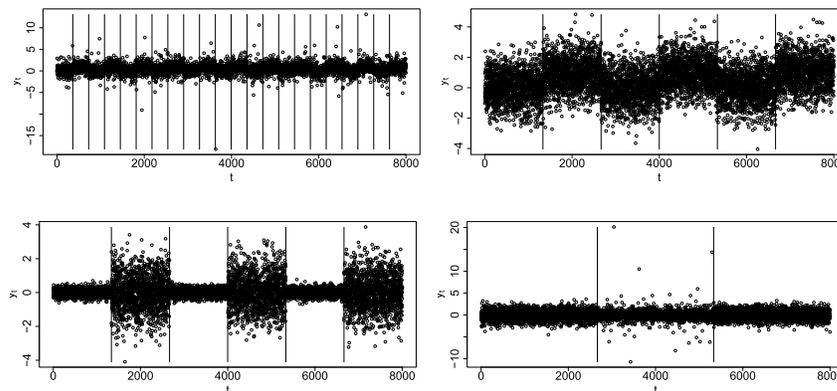


FIG 2. A plot showing realizations of different scenarios with $T = 8000$. From left to right and from top to bottom, the panels are from Scenarios 2, 3, 4 and 5, respectively.

2014; Pein et al., 2019), heterogeneous simultaneous multiscale change point estimators Pein, Sieling and Munk (2017); Pein et al. (2019), the nonparametric multiple change point detection of Zou et al. (2014) and Haynes, Fearnhead and Eckley (2017), the robust functional pruning optimal partitioning method of (Fearnhead and Rigaiil, 2018) and the kernel change point detection procedure of (Celisse et al., 2018; Arlot, Celisse and Harchaoui, 2019). For all the competing methods, we set the respective tuning parameters by default choices. For the kernel change point detection procedure, we use the function `KernSeg_MultiD` in the R (R Core Team, 2019) package `kernseg` (Marot et al., 2018). The function produces a sequence of candidate models, each of which is associated with a measure of fit. We choose the best model based on the elbow criterion from the sequence of measures of fit.

We apply Algorithm 3 with $\lambda = 2/3 \log(n_{1:T})$, a choice that is guided by Theorem 3.2 and that we find reasonable in practice (see Appendix E for sensitivity simulations comparing Algorithms 1 and 3). Moreover, we construct the samples $\{Y_{t,i}\}$ and $\{W_{t,i}\}$ by splitting the data into two time series having odd and even time indices. We set the number of random intervals as $M = 120$.

We explain the different generative models that are deployed in our simulations. For all scenarios, we consider $T \in \{1000, 4000, 8000\}$. Moreover, we consider the partition \mathcal{P} of $\{1, \dots, T\}$ induced by the change points $\eta_0 = 1 < \eta_1 < \dots < \eta_K < \eta_{K+1} = T + 1$, which are evenly spaced in $\{1, \dots, T\}$. The elements of \mathcal{P} are A_1, \dots, A_{K+1} , with $A_j = [\eta_{j-1}, \eta_j - 1]$. We consider the following scenarios.

Scenario 1. Let $K = 7$ for each instance of T . Define F_t to have probability density function as in the left panel of Figure 1 for $t \in A_j$ with odd j and as in the right panel of Figure 1 for $t \in A_j$ with even j .

Scenario 2. Let $K = \lfloor 2^{-1/2} T^{1/2} \log^{-1/2}(T) \rfloor$ and define $\theta \in \mathbb{R}^T$ as $\theta_t = 1$, $t \in A_j$ with odd j , and $\theta_t = 0$ otherwise. Let data be generated as $y_t =$

$\theta_t + 3^{-1/2}\varepsilon_t$, $t = 1, \dots, T$, where ε_t are independent and identically distributed as a t -distribution with 3 degrees of freedom.

Scenario 3. Replace the t -distribution in Scenario 2 with $\mathcal{N}(0, 1)$.

Scenario 4. Let $K = \lfloor 2^{-1/2}T^{1/2} \log^{-1/2}(T) \rfloor$ and define $\theta \in \mathbb{R}^T$ as $\theta_t = 1/5$, $t \in A_j$ with odd j , and $\theta_t = 1$ otherwise. Let data be generated as $y_t = \theta_t \varepsilon_t$, $t = 1, \dots, T$, where ε_t are independent and identically distributed as $\mathcal{N}(0, 1)$, $t = 1, \dots, T$.

Scenario 5. Let $K = 2$ and the data be generated as $y_t = \varepsilon_t$, $t \in A_j$ with odd j , and $y_t = 5^{-1/2}\xi_t$ otherwise, where ε_t are independent and identically distributed as $\mathcal{N}(0, 1)$ and ξ_t are independent and identically distributed as t -distribution with 2.5 degrees of freedom.

A visualization of different scenarios is given in Figure 2. We can see that these five scenarios capture a broad range of models that can allow us to assess the quality of different methods.

Based on the results in Tables 1 and 2, we can see that, generally, the best performance is attained by the Kolmogorov–Smirnov detector and kernel change point detection. In fact in some cases in terms of the localization error rate, the kernel change point detection outperforms the Kolmogorov–Smirnov detector, despite that to the best of our knowledge, the theoretical guarantees of the kernel change point detection’s localization error rate is not yet established. This is seen in Scenario 1, where, as the sample size grows, the Kolmogorov–Smirnov detector and robust functional pruning optimal partitioning provide the best estimates of K . This is not surprising as Scenario 1 presents a situation where the distributions are not members of usual parametric families. In Scenario 2, robust functional pruning optimal partitioning attains the best performance with heterogeneous simultaneous multiscale change point estimators, kernel change point detection and Kolmogorov–Smirnov detector as the closest competitors. This scenario poses a challenge for most methods due to the heavy tails nature of the t -distributions. In Scenario 3, Kolmogorov–Smirnov detector is outperformed by methods like wild binary segmentation, pruned dynamic programming, simultaneous multiscale change point estimators and Bai and Perron’s method. While Kolmogorov–Smirnov detector is still competitive in estimating the number of change points K , its localization errors are subpar. This should not come as a surprise, since all these methods are designed to work well in this particular mean change point model with sub-Gaussian errors. In Scenario 4, we see a clear advantage of using one of the nonparametric approaches, such as kernel change point detection, Kolmogorov–Smirnov detector and nonparametric multiple change point detection. Methods like wild binary segmentation, pruned exact linear time algorithm, Bai and Perron’s procedure, simultaneous multiscale change point estimators and heterogeneous simultaneous multiscale change point estimators perform poorly in this scenario. This is especially interesting for the heterogeneous simultaneous multiscale change point estimator, as such method has been proven to be effective in detecting changes in variance also when there are concurrent changes in mean. However, Scenario 4 only includes changes in variance with the mean remaining constant.

TABLE 1

KSD, Kolmogorov–Smirnov detector; *WBS*, wild binary segmentation; *PELT*, pruned exact linear time algorithm; *S3IB*, pruned dynamic programming algorithm; *NMCD*, nonparametric multiple change point detection; *SMUCE*, simultaneous multiscale change point estimators; *B&P*, Bai and Perron’s method; *HSMUCE*, heterogeneous simultaneous multiscale change point estimators; *RFPOP*, robust functional pruning optimal partitioning methods; *KCP*, kernel change point detection.

Scenario 1.

T	Metric	KSD	WBS	PELT	S3IB	NMCD	SMUCE	B&P	HSMUCE	RFPOP	KCP
1000	$ K - \hat{K} $	1.5	11.0	4.35	6.3	1.8	4.35	6.85	1.6	1.8	4.4
1000	$d(\hat{\mathcal{C}} \mathcal{C})$	32.0	63.0	64.5	∞	23.0	43.5	∞	75.5	13.0	637.0
1000	$d(\mathcal{C} \hat{\mathcal{C}})$	52.0	82.0	90.0	$-\infty$	36.0	93.5	$-\infty$	56.5	17.0	23.0
4000	$ K - \hat{K} $	0.2	0.5	21.4	6.9	2.2	52.0	4.9	8.3	0.24	1.4
4000	$d(\hat{\mathcal{C}} \mathcal{C})$	31.0	43.5	46.5	∞	13.5	448.0	1386.0	70.0	10.0	108.0
4000	$d(\mathcal{C} \hat{\mathcal{C}})$	31.0	364.5	448	$-\infty$	69.5	32.7	101.5	319.0	12.0	27.0
8000	$ K - \hat{K} $	0.0	0.9	41.2	5.35	3.25	62.7	3.7	17.6	0.24	0.9
8000	$d(\hat{\mathcal{C}} \mathcal{C})$	38.0	29.5	55.0	∞	10.0	70.0	1831	84.5	11.0	24.0
8000	$d(\mathcal{C} \hat{\mathcal{C}})$	38.0	818.5	955.0	$-\infty$	227.0	958.0	206.5	910.0	11.0	34.0

Scenario 2.

T	Metric	KSD	WBS	PELT	S3IB	NMCD	SMUCE	B&P	HSMUCE	RFPOP	KCP
1000	$ K - \hat{K} $	1.3	9.9	2.7	1.85	1.7	4.6	7.05	0.45	0.1	1.76
1000	$d(\hat{\mathcal{C}} \mathcal{C})$	11.0	8.5	9.0	15.5	7.0	10.5	738.5	23.0	6.0	12
1000	$d(\mathcal{C} \hat{\mathcal{C}})$	13.0	54.5	45.5	29.5	28.0	53.5	16.0	22.0	6.0	5.0
4000	$ K - \hat{K} $	0.0	15.4	10.05	14.7	3.35	14.8	11.95	0.0	0.1	0.6
4000	$d(\hat{\mathcal{C}} \mathcal{C})$	16.0	9.5	12.0	∞	9.0	35.0	1007	16.0	8.0	8.0
4000	$d(\mathcal{C} \hat{\mathcal{C}})$	16.0	176	163.0	$-\infty$	107.5	164	20.0	16.0	8.0	8.0
8000	$ K - \hat{K} $	1.3	8.4	18.45	20.8	4.75	28.5	18.4	0.1	0.2	0.0
8000	$d(\hat{\mathcal{C}} \mathcal{C})$	363.0	15.5	9.5	∞	13.5	40.0	2179	20.0	9.0	7.0
8000	$d(\mathcal{C} \hat{\mathcal{C}})$	18.0	254.5	2179	$-\infty$	129.5	257	19.5	20.0	10.0	7.0

Scenario 3.

T	Metric	KSD	WBS	PELT	S3IB	NMCD	SMUCE	B&P	HSMUCE	RFPOP	KCP
1000	$ K - \hat{K} $	0.8	0.0	0.0	0.1	0.8	0.0	0.0	0.2	0.0	2.3
1000	$d(\hat{\mathcal{C}} \mathcal{C})$	16.0	9.0	8.5	8.5	9.5	8.5	8.5	9.5	9.0	140.0
1000	$d(\mathcal{C} \hat{\mathcal{C}})$	19.0	9.5	8.5	0.9	10.5	8.5	8.5	9.5	9.0	7.0
4000	$ K - \hat{K} $	0.1	0.0	0.0	0.0	1.8	0.0	0.0	0.0	0.1	0.2
4000	$d(\hat{\mathcal{C}} \mathcal{C})$	22.0	8.0	9.0	8.0	11.5	8.0	8.0	9.5	8.0	11.0
4000	$d(\mathcal{C} \hat{\mathcal{C}})$	20.0	8.0	9.0	8.0	136.5	8.0	8.0	8.0	8.0	9.0
8000	$ K - \hat{K} $	0.2	0.0	0.0	0.1	1.7	0.0	0.0	0.2	0.0	0.2
8000	$d(\hat{\mathcal{C}} \mathcal{C})$	11.5	6.0	6.0	6.0	8.5	6.0	6.0	6.0	6.0	6.0
8000	$d(\mathcal{C} \hat{\mathcal{C}})$	11.5	6.0	6.0	6.0	353	6.0	6.0	6.5	6.0	6.0

Finally, Scenario 5 seems to be the most challenging one for all methods. In fact, Kolmogorov–Smirnov detector and kernel change point detection seem to be the only methods capable of estimating correctly the numbers of change points, with the kernel change point detection procedure yielding smaller localization rates.

In our second set of simulations we study the case where the number of data points collected at any time can be more than one. We consider the same 5 scenarios, same tuning parameter selection method and same performance

TABLE 2

KSD, Kolmogorov–Smirnov detector; *WBS*, wild binary segmentation; *PELT*, pruned exact linear time algorithm; *S3IB*, pruned dynamic programming algorithm; *NMCD*, nonparametric multiple change point detection; *SMUCE*, simultaneous multiscale change point estimators; *B&P*, Bai and Perron’s method; *HSMUCE*, heterogeneous simultaneous multiscale change point estimators; *RFPOP*, robust functional pruning optimal partitioning methods; *KCP*, kernel change point detection.

Scenario 4.

T	Metric	KSD	WBS	PELT	S3IB	NMCD	SMUCE	B&P	HSMUCE	RFPOP	KCP
1000	$ K - \hat{K} $	0.9	4.0	9.8	4.9	2.45	27.75	5.0	4.7	13.8	0.4
1000	$d(\hat{\mathcal{C}} \mathcal{C})$	36.0	∞	40.0	∞	4.0	24.5	∞	∞	82.0	5.0
1000	$d(\mathcal{C} \hat{\mathcal{C}})$	32.0	$-\infty$	153.5	$-\infty$	67.0	157	$-\infty$	$-\infty$	157.0	5.0
4000	$ K - \hat{K} $	0.0	3.8	36.3	5.0	2.7	71.1	5.0	4.5	44.8	0.1
4000	$d(\hat{\mathcal{C}} \mathcal{C})$	19.0	∞	106.5	∞	4.5	46.0	∞	∞	125.0	5.0
4000	$d(\mathcal{C} \hat{\mathcal{C}})$	19.0	$-\infty$	644.5	$-\infty$	66.0	651.5	$-\infty$	$-\infty$	647.0	5.0
8000	$ K - \hat{K} $	0.1	3.5	60.3	5.0	4.0	109.3	5.0	4.5	71.4	0.0
8000	$d(\hat{\mathcal{C}} \mathcal{C})$	23.0	6301.5	115.0	∞	2.5	47.5	∞	∞	135	6.0
8000	$d(\mathcal{C} \hat{\mathcal{C}})$	28.0	238	1300.5	$-\infty$	566.5	1316	$-\infty$	$-\infty$	1293	6.0

Scenario 5.

T	Metric	KSD	WBS	PELT	S3IB	NMCD	SMUCE	B&P	H-SMUCE	RFPOP	KCP
1000	$ K - \hat{K} $	0.4	3.6	7.2	5.0	1.5	26.95	5.0	4.8	1.96	0.1
1000	$d(\hat{\mathcal{C}} \mathcal{C})$	27.0	∞	70.0	∞	4.5	21.0	∞	∞	∞	9.0
1000	$d(\mathcal{C} \hat{\mathcal{C}})$	29.0	$-\infty$	147.0	$-\infty$	32	159.0	$-\infty$	$-\infty$	$-\infty$	8.0
4000	$ K - \hat{K} $	0.1	3.5	38.2	5.0	3.35	72.05	5.0	4.5	2.0	0.1
4000	$d(\hat{\mathcal{C}} \mathcal{C})$	24.0	∞	82.0	∞	3.0	39.5	∞	∞	∞	12.0
4000	$d(\mathcal{C} \hat{\mathcal{C}})$	25.0	$-\infty$	629.5	$-\infty$	275	640.5	$-\infty$	$-\infty$	$-\infty$	12.0
8000	$ K - \hat{K} $	0.0	4.2	63.9	5.0	4.4	114.0	5.0	4.6	1.84	0.1
8000	$d(\hat{\mathcal{C}} \mathcal{C})$	37.0	∞	107.5	∞	2.5	55.0	∞	∞	∞	12.0
8000	$d(\mathcal{C} \hat{\mathcal{C}})$	37.0	$-\infty$	1309	$-\infty$	552	1310.5	$-\infty$	$-\infty$	$-\infty$	12.0

metrics as in the first set of simulations. However, instead of setting $n_t = 1$, we let n_t to be fixed as 5, 15 and 30 or to be randomly distributed as Poisson(5), Poisson(15) and Poisson(5), for each t . We also set $T = 1000$. We only present the results of the Kolmogorov–Smirnov detector, because no other methods would automatically be able to handle this situation. We omit $d(\mathcal{C}|\hat{\mathcal{C}})$ as it does not provide additional information. The results on Table 3 show the effectiveness of the Kolmogorov–Smirnov detector for estimating the number of change points and their locations.

4.2. Real data analysis

We consider the array comparative genomic hybridization micro-array data set from Bleakley and Vert (2011). This consists of individuals with bladder tumours. The data set has been processed and can be obtained in the R (R Core Team, 2019) package *ecp* (Matteson and James, 2013). For the microarray corresponding to the first individual we consider change point localization using different methods.

TABLE 3
Performance evaluations for the KSD method in settings where n_t can be larger than 1.

Scenario 1.						
Metric	$n_t = 5$	$n_t = 15$	$n_t = 30$	$n_t \sim \text{Pois}(5)$	$n_t \sim \text{Pois}(15)$	$n_t \sim \text{Pois}(30)$
$ K - \hat{K} $	0.1	0.2	0.0	0.6	0.3	0.0
$d(\hat{\mathcal{C}} \mathcal{C})$	6.5	2.5	2.0	6.0	2.0	1.5

Scenario 2.						
Metric	$n_t = 5$	$n_t = 15$	$n_t = 30$	$n_t \sim \text{Pois}(5)$	$n_t \sim \text{Pois}(15)$	$n_t \sim \text{Pois}(30)$
$ K - \hat{K} $	0.1	0.0	0.0	0.4	0.0	0.0
$d(\hat{\mathcal{C}} \mathcal{C})$	3.0	1.0	0.0	3.0	1.0	0.0

Scenario 3.						
Metric	$n_t = 5$	$n_t = 15$	$n_t = 30$	$n_t \sim \text{Pois}(5)$	$n_t \sim \text{Pois}(15)$	$n_t \sim \text{Pois}(30)$
$ K - \hat{K} $	0.3	0.3	0.0	0.4	0.0	0.0
$d(\hat{\mathcal{C}} \mathcal{C})$	6.5	1.0	0.5	5.0	2.0	1.0

Scenario 4.						
Metric	$n_t = 5$	$n_t = 15$	$n_t = 30$	$n_t \sim \text{Pois}(5)$	$n_t \sim \text{Pois}(15)$	$n_t \sim \text{Pois}(30)$
$ K - \hat{K} $	0.2	0.0	0.0	0.0	0.0	0.0
$d(\hat{\mathcal{C}} \mathcal{C})$	6.0	2.0	0.0	5.0	1.0	1.0

Scenario 5.						
Metric	$n_t = 5$	$n_t = 15$	$n_t = 30$	$n_t \sim \text{Pois}(5)$	$n_t \sim \text{Pois}(15)$	$n_t \sim \text{Pois}(30)$
$ K - \hat{K} $	0.1	0.0	0.0	0.0	0.0	0.0
$d(\hat{\mathcal{C}} \mathcal{C})$	9.5	3.0	2.0	6.0	5.5	6.0

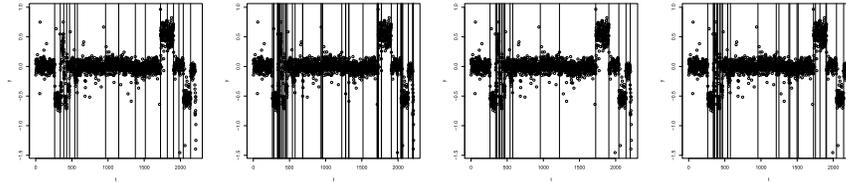


FIG 3. A plot showing individual 1 in the array comparative genomic hybridization data set, with estimated change points indicating by vertical lines. From left to right, the panels are based on the estimators from the Kolmogorov–Smirnov detector, wild binary segmentation, nonparametric multiple change point detection and robust functional pruning optimal partitioning methods, respectively.

Figure 3 shows that all the methods seem to recover the important change points in the time series associated with individual 1 in the data set. A potential advantage of the Kolmogorov–Smirnov detector method is that it seems less sensitive to potentially-spurious change points as opposed to wild binary segmentation, nonparametric multiple change point detection and robust functional pruning optimal partitioning.

Appendix A: Comparisons

We compare our rates with those in the univariate mean change point detection problem, which assumes sub-Gaussian data (e.g. Wang, Yu and Rinaldo, 2020). On one hand, this comparison inherits the main arguments when comparing parametric and nonparametric modelling methods in general. Especially with the general model assumption we impose on the underlying distribution functions, we enjoy risk-free from model mis-specification. On the other hand, seemingly surprisingly, we achieve the same rates of those in the univariate mean change point detection case, even though sub-Gaussianity is assumed thereof. In fact, this is expected. We are using the empirical distribution function in our CUSUM Kolmogorov–Smirnov statistic, which is essentially a weighted Bernoulli random variable at every $z \in R$. Due to the fact that Bernoulli random variables are sub-Gaussian, and that the empirical distribution functions are step functions with knots only at the sample points, we are indeed to expect the same rates.

Furthermore, the heterogeneous simultaneous multiscale change point estimator from Pein, Sieling and Munk (2017) can also be compared to the Kolmogorov–Smirnov detector. Assuming Gaussian errors, $\delta \asymp T$ and $K = O(1)$, Theorems 3.7 and 3.9 in Pein, Sieling and Munk (2017) proved that heterogeneous simultaneous multiscale change point estimator can consistently estimate the number of change points, and that $\epsilon \lesssim r(T)$, for any $r(T)$ sequence such that $r(T)/\log(T) \rightarrow \infty$. This is weaker than our upper bound that guarantees $\epsilon \lesssim \log(T)$. The Kolmogorov–Smirnov detector can handle changes in variance when the mean remains constant, a setting where it is unknown if the heterogeneous simultaneous multiscale change point estimator is consistent.

Another interesting contrast can be made between the Kolmogorov–Smirnov detector and the multiscale quantile segmentation method in Vanegas, Behr and Munk (2019). Both algorithms make no assumptions on the distributional form of the cumulative distribution functions. However, the multiscale quantile segmentation is designed to identify changes in one known quantile. This is not a requirement for the Kolmogorov–Smirnov detector which can detect any type of changes without previous knowledge of their nature. As for statistical guarantees, translating to our notation, provided that $\delta \gtrsim \log(T)$, multiscale quantile segmentation can consistently estimate the number of change points and have $\epsilon \lesssim \log(T)$. This matches our theoretical guarantees in Theorem 3.1.

We compare the theoretical properties of the Kolmogorov–Smirnov detector with the ones of the nonparametric multiple change point detection procedure in Zou et al. (2014). Both methods guarantee consistent change point localization of univariate sequences in fully nonparametric settings.

- We measure the magnitude κ of the distribution changes at the change points using the Kolmogorov–Smirnov distance, as in (1.1). In contrast, Zou et al. (2014) deploy a weighted Kullback–Leibler divergence, see Assumption A4 in their paper, which is stronger than the Kolmogorov–Smirnov distance, and therefore more discriminative. At the same time,

the authors require the cumulative distribution functions to be continuous, while our results hold with arbitrary distribution functions.

- We let κ change with T , thus allowing for the minimal jump size to decay to 0, and provide localization rates that depend explicitly on κ . In contrast, Zou et al. (2014) constraints the jump sizes to be bounded away from 0, and the resulting localization rate does not involve the jump sizes.
- Zou et al. (2014) impose stronger conditions on the rate at which the number of change point is permitted to diverge as a function of T . For example, assuming equally spaced change points, the assumptions in their Theorem 1 require, in our notation, that $K = o(T^{1/4})$. In contrast, we allow for K to grow as fast as $T/\log(T)$.
- If we let the number of change points be bounded in T , as assumed in Theorem 2 of Zou et al. (2014), then, again translating into our notation, their procedure yields $\epsilon \gtrsim \log^{2+c}(T)$, for $c > 0$, while we can achieve $\epsilon \lesssim \log(T)$.

Finally, we compare with the KCP method, studied in Arlot, Celisse and Harchaoui (2019) and Garreau and Arlot (2018). Specifically, the following discussions are based on translating Theorem 3.1 in Garreau and Arlot (2018) to our notation, and assuming $n_t = 1$, $t \in \{1, \dots, T\}$ in our setting.

- Both ours and theirs study a nonparametric change point localization problem, and both achieve parametric rates in the sense that the complexity of the class of the distribution functions are not shown in the results. In our case, this is because we adopt the Kolmogorov–Smirnov distance and transform general distributions to Bernoulli distributions. In Garreau and Arlot (2018), this is because they map the data to a reproducing kernel Hilbert space.
- The jump size κ are measured by different measurements in ours and theirs. In our paper, the jump size is taken to be the difference in terms of the Kolmogorov–Smirnov distance, which reflects the genuine differences between different distributions. In Garreau and Arlot (2018), it is a distance induced by the Hilbert space norm. More specifically, it is the distance between two different Bochner integrals. With this definition, despite the flexibility, as pointed out in Garreau and Arlot (2018), it is not guaranteed that the change points detected are the change points of the original distributions.
- In terms of the signal-to-noise ratio condition and the localization rates, there are two differences between ours and theirs. Firstly, we allow all model parameters to vary with the sample size n , while the variance parameter in Garreau and Arlot (2018) is assumed to be a constant. Secondly, in both the signal-to-noise ratio condition (see Eq. (3.3) in Garreau and Arlot, 2018) and the localization rate (see the final result in Theorem 3.1 in Garreau and Arlot, 2018), there is an additional factor K , the number of change points. Note that this factor does not appear in our results.

Appendix B: Proof of Theorem 3.1

Definition B.1. Denote the population version of the CUSUM Kolmogorov–Smirnov statistic as

$$\Delta_{s,e}^t = \sup_{z \in \mathbb{R}} |\Delta_{s,e}^t(z)|,$$

where

$$\Delta_{s,e}^t(z) = \left\{ \frac{n_{s:t} n_{(t+1):e}}{n_{s:e}} \right\}^{1/2} \{F_{s:t}(z) - F_{(t+1):e}(z)\}, \tag{B.1}$$

and

$$F_{s:e}(z) = \frac{1}{n_{s:e}} \sum_{t=s}^e n_t F_t(z).$$

Proof of Theorem 3.1. Let $\epsilon = C_\epsilon \kappa^{-2} \log(n_{1:T}) n_{\max}^7 n_{\min}^{-8}$. Since ϵ is the upper bound of the localisation error, by induction, it suffices to consider any interval $(s, e) \subset (1, T)$ that satisfies

$$\eta_{k-1} \leq s \leq \eta_k \leq \dots \leq \eta_{k+q} \leq e \leq \eta_{k+q+1}, \quad q \geq -1,$$

and

$$\max\{\min\{\eta_k - s, s - \eta_{k-1}\}, \min\{\eta_{k+q+1} - e, e - \eta_{k+q}\}\} \leq \epsilon,$$

where $q = -1$ indicates that there is no change point contained in (s, e) .

By Condition 3.1, there exists an absolute constant $c > 0$ such that

$$\epsilon \leq c \frac{n_{\max}^7}{n_{\min}^7} \delta \leq \delta/4.$$

It has to be the case that for any change point $\eta_k \in (0, T)$, either $|\eta_k - s| \leq \epsilon$ or $|\eta_k - s| \geq \delta - \epsilon \geq 3\delta/4$. This means that $\min\{|\eta_k - s|, |\eta_k - e|\} \leq \epsilon$ indicates that η_k is a detected change point in the previous induction step, even if $\eta_k \in (s, e)$. We refer to $\eta_k \in (s, e)$ an undetected change point if $\min\{|\eta_k - s|, |\eta_k - e|\} \geq 3\delta/4$.

In order to complete the induction step, it suffices to show that we (i) will not detect any new change point in (s, e) if all the change points in that interval have been previous detected, and (ii) will find a point $b \in (s, e)$ such that $|\eta_k - b| \leq \epsilon$ if there exists at least one undetected change point in (s, e) .

For $j = 1, 2$, define the events

$$\mathcal{A}_j(\gamma) = \left\{ \max_{1 \leq s < b < e \leq T} \sup_{z \in \mathbb{R}} \left| \sum_{k=s}^e w_k^{(j)} \sum_{i=1}^{n_k} \{1_{\{Y_{k,i} \leq z\}} - \mathbb{E}(1_{\{Y_{k,i} \leq z\}})\} \right| \leq \gamma \right\},$$

where

$$w_k^{(1)} = \begin{cases} \left(\frac{n_{(b+1):e}}{n_{s:b} n_{s:e}} \right)^{1/2}, & k = s, \dots, b, \\ - \left(\frac{n_{s:b}}{n_{(b+1):e} n_{s:e}} \right)^{1/2}, & k = b + 1, \dots, e, \end{cases} \quad \text{and } w_k^{(2)} = n_{s:e}^{-1/2}.$$

Define

$$\mathcal{S} = \bigcap_{k=1}^K \left\{ \alpha_s \in [\eta_k - 3\delta/4, \eta_k - \delta/2], \beta_s \in [\eta_k + \delta/2, \eta_k + 3\delta/4], \right. \\ \left. \text{for some } s = 1, \dots, M \right\}.$$

Set γ to be $C_\gamma \log^{1/2}(n_{1:T})$, with a sufficiently large constant $C_\gamma > 0$. The rest of the proof assumes the event $\mathcal{A}_1(\gamma) \cap \mathcal{A}_2(\gamma) \cap \mathcal{S}$, the probability of which can be lower bounded using Lemma B.3 and also Lemma 13 in Wang, Yu and Rinaldo (2020).

Step 1. In this step, we will show that we will consistently detect or reject the existence of undetected change points within (s, e) . Let a_m, b_m and m^* be defined as in Algorithm 1. Suppose there exists a change point $\eta_k \in (s, e)$ such that $\min\{\eta_k - s, e - \eta_k\} \geq 3\delta/4$. In the event \mathcal{S} , there exists an interval (α_m, β_m) selected such that $\alpha_m \in [\eta_k - 3\delta/4, \eta_k - \delta/2]$ and $\beta_m \in [\eta_k + \delta/2, \eta_k + 3\delta/4]$.

Following Algorithm 1, $(s_m, e_m) = (\alpha_m, \beta_m) \cap (s, e)$. We have that $\min\{\eta_k - s_m, e_m - \eta_k\} \geq (1/4)\delta$ and (s_m, e_m) contains at most one true change point.

It follows from Lemma B.4, with c_1 there chosen to be $1/4$, that

$$\max_{t=s_m+1, \dots, e_m-1} |\Delta_{s_m, e_m}^t| \geq \frac{\kappa \delta n_{\min}^{3/2}}{8(e_m - s_m)^{1/2} n_{\max}},$$

Therefore

$$a_m = \max_{t=s_m+1, \dots, e_m-1} D_{s_m, e_m}^t \geq \max_{t=s_m+1, \dots, e_m-1} \Delta_{s_m, e_m}^t - \gamma \\ \geq \frac{1}{8C_M^{1/2}} \kappa \delta^{1/2} \frac{n_{\min}^{3/2}}{n_{\max}} - \gamma.$$

Thus for any undetected change point $\eta_k \in (s, e)$, it holds that

$$a_{m^*} = \max_{m=1, \dots, M} a_m \geq \frac{1}{8C_M^{1/2}} \kappa \delta^{1/2} \frac{n_{\min}^{3/2}}{n_{\max}} - \gamma \geq c_{\tau, 2} \kappa \delta^{1/2} n^{1/2}, \quad (\text{B.2})$$

where the last inequality is from the choice of γ , the fact $n_{\min} \asymp n_{\max}$ and $c_{\tau, 2} > 0$ is achievable with a sufficiently large C_{SNR} in Condition 3.1. This means we accept the existence of undetected change points.

Recalling the notation $\{\epsilon_k\}_{k=1}^K$ we introduced in (3.3). Here with some abuse of notation, we let

$$\epsilon_k = C_\epsilon \kappa_k^{-2} \log(n_{1:T}) n^{-1}, \quad k = 1, \dots, K.$$

Suppose that there are no undetected change points within (s, e) , then for any (s_m, e_m) , one of the following situations must hold.

- (a) There is no change point within (s_m, e_m) ;

- (b) there exists only one change point $\eta_k \in (s_m, e_m)$ and $\min\{\eta_k - s_m, e_m - \eta_k\} \leq \epsilon_k$; or
(c) there exist two change points $\eta_k, \eta_{k+1} \in (s_m, e_m)$ and $\eta_k - s_m \leq \epsilon_k$, $e_m - \eta_{k+1} \leq \epsilon_{k+1}$.

Observe that if (a) holds, then we have

$$\max_{t=s_m+1, \dots, e_m-1} D_{s_m, e_m}^t \leq \max_{t=s_m+1, \dots, e_m-1} \Delta_{s_m, e_m}^t + \gamma = \gamma < \tau,$$

so no change points are detected.

Cases (b) and (c) are similar, and case (b) is simpler than (c), so we will only focus on case (c). It follows from Lemma B.2 that

$$\begin{aligned} \max_{t=s_m+1, \dots, e_m-1} \Delta_{s_m, e_m}^t &\leq n_{\max}^{1/2} (e_m - \eta_{k+1})^{1/2} \kappa_{k+1} + n_{\max}^{1/2} (\eta_k - s_m)^{1/2} \kappa_k \\ &\leq 2C_\epsilon^{1/2} \log^{1/2}(n_{1:T}), \end{aligned}$$

therefore

$$\begin{aligned} \max_{t=s_m+1, \dots, e_m-1} D_{s_m, e_m}^t &\leq \max_{t=s_m+1, \dots, e_m-1} \Delta_{s_m, e_m}^t + \gamma \\ &\leq 2C_\epsilon^{1/2} \log^{1/2}(n_{1:T}) + C_\gamma \log^{1/2}(n_{1:T}) < \tau. \end{aligned}$$

Under (3.2), we will always correctly reject the existence of undetected change points.

Step 2. Assume that there exists a change point $\eta_k \in (s, e)$ such that $\min\{\eta_k - s, e - \eta_k\} \geq 3\delta/4$. Let s_m, e_m and m^* be defined as in Algorithm 1. To complete the proof it suffices to show that, there exists a change point $\eta_k \in (s_{m^*}, e_{m^*})$ such that $\min\{\eta_k - s_{m^*}, \eta_k - e_{m^*}\} \geq \delta/4$ and $|b_{m^*} - \eta_k| \leq \epsilon$.

To this end, we are to ensure that the assumptions of Lemma B.9 are verified. Note that (B.26) follows from (B.2), (B.27) and (B.28) follow from the definitions of events $\mathcal{A}_1(\gamma)$ and $\mathcal{A}_2(\gamma)$, and (B.29) follows from Condition 3.1.

Thus, all the conditions in Lemma B.9 are met. Therefore, we conclude that there exists a change point η_k , satisfying

$$\min\{e_{m^*} - \eta_k, \eta_k - s_{m^*}\} > \delta/4 \tag{B.3}$$

and

$$|b_{m^*} - \eta_k| \leq C \frac{n_{\max}^9}{n_{\min}^{10}} \kappa^{-2} \gamma^2 \leq \epsilon,$$

where the last inequality holds from the choice of γ and Condition 3.1.

The proof is completed by noticing that (B.3) and $(s_{m^*}, e_{m^*}) \subset (s, e)$ imply that

$$\min\{e - \eta_k, \eta_k - s\} > \delta/4 > \epsilon.$$

As discussed in the argument before Step 1, this implies that η_k must be an undetected change point. \square

Below are a number of auxiliary lemmas. Lemma B.1 plays the role of Lemma 2.2 in Venkatraman (1992). Lemma B.3 controls the deviance between sample and population Kolmogorov–Smirnov statistics. Lemma B.4 is the density version of Lemma 2.4 in Venkatraman (1992). Lemma B.5 plays the role of Lemma 2.6 of Venkatraman (1992). Lemma B.6 is essentially Lemma 17 in Wang, Yu and Rinaldo (2020). Lemma B.8 is Lemma 19 in Wang, Yu and Rinaldo (2020).

Lemma B.1. *Under Condition 1.1, for any pair $(s, e) \subset (1, T)$ satisfying*

$$\eta_{k-1} \leq s \leq \eta_k \leq \dots \leq \eta_{k+q} \leq e \leq \eta_{k+q+1}, \quad q \geq 0,$$

we have the following.

(i) *Let*

$$b_1 \in \arg \max_{b=s+1, \dots, e-1} \Delta_{s,e}^b.$$

Then $b_1 \in \{\eta_1, \dots, \eta_K\}$.

(ii) *Let $z \in \arg \max_{x \in \mathbb{R}} |\Delta_{s,e}^{b_1}(x)|$. If $\Delta_{s,e}^t(z) \neq 0$ for some $t \in (s, e)$, then $|\Delta_{s,e}^t(z)|$ is either monotonic or decreases and then increases within each of the interval (s, η_k) , (η_k, η_{k+1}) , \dots , (η_{k+q}, e) .*

Proof. We prove by contradiction assuming that $b_1 \notin \{\eta_1, \dots, \eta_K\}$. Let z_0 be such that

$$z_0 \in \arg \max_{z \in \mathbb{R}} |\Delta_{s,e}^{b_1}(z)|.$$

Note that due to the fact for any cumulative distribution function $F : \mathbb{R} \rightarrow [0, 1]$, it holds that $F(-\infty) = 1 - F(\infty) = 0$, we have that $z_0 \in \mathbb{R}$ exists.

Therefore,

$$b_1 \in \arg \max_{b=s+1, \dots, e-1} |\Delta_{s,e}^b(z_0)|.$$

Next consider the time series $\{r_l(z_0)\}_{l=1}^{n_{s:e}}$ defined as

$$r_l(z_0) = \begin{cases} F_s(z_0) & l \in \{1, \dots, n_s\}, \\ F_{s+1}(z_0) & l \in \{n_s + 1, \dots, n_{s:(s+1)}\}, \\ \dots, & \\ F_e(z_0) & l \in \{n_{s:(e-1)} + 1, \dots, n_{s:e}\}, \end{cases}$$

and for $1 \leq l < n_{s:e}$ define

$$\tilde{r}_{1, n_{s:e}}^l(z_0) = \left(\frac{n_{s:e} - l}{n_{s:e} l} \right)^{1/2} \sum_{t=1}^l r_t(z_0) - \left\{ \frac{l}{n_{s:e}(n_{s:e} - l)} \right\}^{1/2} \sum_{t=l+1}^{n_{s:e}} r_t(z_0).$$

The set of change points of the time series $\{r_l(z_0)\}_{l=1}^{n_{s:e}}$ is

$$\{n_{s:\eta_k}, \dots, n_{s:\eta_{k+q}}\}.$$

Lemma 2.2 from Venkatraman (1992) applied to $\{r_l(z_0)\}_{l=1}^{n_{s:e}}$ leading to that

$$\begin{aligned} \Delta_{s,e}^{b_1} &= \Delta_{s,e}^{b_1}(z_0) = \tilde{r}_{1,n_{s:e}}^{n_{s:b_1}}(z_0) < \max_{j \in \{k, \dots, k+q\}} \tilde{r}_{1,n_{s:e}}^{n_{s:\eta_j}}(z_0) \\ &= \max_{j \in \{k, \dots, k+q\}} \Delta_{s,e}^{\eta_j}(z_0) \leq \max_{j \in \{k, \dots, k+q\}} \Delta_{s,e}^{\eta_j}, \end{aligned}$$

which is a contradiction.

As for (ii), it follows from applying Lemma 2.2 from Venkatraman (1992) to $\{r_l(z_0)\}_{l=1}^{n_{s:e}}$. □

Lemma B.2. *Under Condition 1.1, let $t \in (s, e)$. It holds that*

$$\Delta_{s,e}^t \leq 2n_{\max}^{1/2} \min\{(s - t + 1)^{1/2}, (e - t)^{1/2}\}. \tag{B.4}$$

If η_k is the only change point in (s, e) , then

$$\Delta_{s,e}^{\eta_k} \leq \kappa_k n_{\max}^{1/2} \min\{(s - \eta_k + 1)^{1/2}, (e - \eta_k)^{1/2}\}. \tag{B.5}$$

If $(s, e) \subset (1, T)$ contains two and only two change points η_k and η_{k+1} , then we have

$$\max_{t=s+1, \dots, e-1} \Delta_{s,e}^t \leq n_{\max}^{1/2} (e - \eta_{k+1})^{1/2} \kappa_{k+1} + n_{\max}^{1/2} (\eta_k - s)^{1/2} \kappa_k. \tag{B.6}$$

Proof. As for (B.4), it follows from that

$$\begin{aligned} \Delta_{s,e}^b &\leq \frac{2n_{s:b}^{1/2} n_{(b+1):e}^{1/2}}{n_{s:e}^{1/2}} \leq 2 \min\{n_{s:b}^{1/2}, n_{(b+1):e}^{1/2}\} \\ &\leq 2n_{\max}^{1/2} \min\{(s - b + 1)^{1/2}, (e - b)^{1/2}\}. \end{aligned}$$

As for (B.5), it is due to that

$$\begin{aligned} \Delta_{s,e}^{\eta_k} &= \frac{n_{s:\eta_k}^{1/2} n_{(\eta_k+1):e}^{1/2}}{n_{s:e}^{1/2}} \sup_{z \in \mathbb{R}} |F_{s:\eta_k}(z) - F_{(\eta_k+1):e}(z)| \\ &\leq \kappa_k n_{\max}^{1/2} \min\{(s - \eta_k + 1)^{1/2}, (e - \eta_k)^{1/2}\}. \end{aligned}$$

Equation (B.6) follows similarly. □

The following Lemma provides a concentration result for the sample CUSUM statistic around its population version. One natural way to proceed in its proof is to exploit Dvoretzky–Kiefer–Wolfowitz’s inequality. However, given the presence of multiple change points, an immediate application of Dvoretzky–Kiefer–Wolfowitz’s inequality produces an extra factor K in the upper bound in Lemma B.3. We instead obtain a better upper bound with a different proof technique.

Lemma B.3. *Under Condition 1.1, for any $1 \leq s < b < e \leq T$ and $z \in R$, define*

$$\Lambda_{s,e}^b(z) = D_{s,e}^b(z) - \Delta_{s,e}^b(z),$$

where $D_{s,e}^b(z)$ and $\Delta_{s,e}^b(z)$ are the sample and population versions of the Kolmogorov–Smirnov statistic defined in Definition 2.1 and (B.1), respectively. It holds that

$$\begin{aligned} & \text{pr} \left\{ \max_{1 \leq s < b < e \leq T} \sup_{z \in \mathbb{R}} |\Lambda_{s,e}^b(z)| > \left\{ \log \left(\frac{T^4}{12\delta} \right) + \log(n_{1:T}) \right\}^{1/2} + 6 \log^{1/2}(n_{1:T}) \right. \\ & \left. + \frac{48 \log(n_{1:T})}{n_{1:T}^{1/2}} \right\} \leq \frac{12 \log(n_{1:T})}{T^3 n_{1:T}} + \frac{24T}{n_{1:T} \log(n_{1:T}) \delta}. \end{aligned}$$

Moreover

$$\begin{aligned} & \text{pr} \left\{ \max_{1 \leq s < e \leq T} \sup_{z \in \mathbb{R}} \left| n_{s:e}^{-1/2} \sum_{t=s}^e \sum_{i=1}^{n_t} \{1_{Y_{t,i} \leq z} - \mathbb{E}(1_{Y_{t,i} \leq z})\} \right| \right. \\ & \left. > \left\{ \log \left(\frac{T^4}{12\delta} \right) + \log(n_{1:T}) \right\}^{1/2} + 6 \log^{1/2}(n_{1:T}) + \frac{48 \log(n_{1:T})}{n_{1:T}^{1/2}} \right\} \\ & \leq \frac{12 \log(n_{1:T})}{T^3 n_{1:T}} + \frac{24T}{n_{1:T} \log(n_{1:T}) \delta}. \end{aligned} \quad (\text{B.7})$$

Remark B.1. Lemma B.3 shows that as T diverges unbounded, it holds that

$$\max_{1 \leq s < b < e \leq T} \sup_{z \in \mathbb{R}} |\Lambda_{s,e}^b(z)| = O_p \left\{ \log^{1/2}(n_{1:T}) \right\}.$$

Proof of Lemma B.3. For any $1 \leq s < b < e \leq T$ and $z \in \mathbb{R}$, let

$$\begin{aligned} & \frac{n_{s:b}^{1/2} n_{(b+1):e}^{1/2}}{n_{s:e}^{1/2}} \left\{ \widehat{F}_{s:b}(z) - \widehat{F}_{(b+1):e}(z) \right\} \\ & = \sum_{k=s}^b \sum_{i=1}^{n_k} \frac{n_{(b+1):e}^{1/2}}{n_{s:b}^{1/2} n_{s:e}^{1/2}} 1_{\{Y_{k,i} \leq z\}} - \sum_{k=b+1}^e \sum_{i=1}^{n_k} \frac{n_{s:b}^{1/2}}{n_{(b+1):e}^{1/2} n_{s:e}^{1/2}} 1_{\{Y_{k,i} \leq z\}} \\ & = \sum_{k=s}^e w_k \sum_{i=1}^{n_k} 1_{\{Y_{k,i} \leq z\}}, \end{aligned}$$

where

$$w_k = \begin{cases} \left\{ \frac{n_{(b+1):e}}{n_{s:b} n_{s:e}} \right\}^{1/2}, & k = s, \dots, b; \\ - \left(\frac{n_{s:b}}{n_{(b+1):e} n_{s:e}} \right)^{1/2}, & k = b+1, \dots, e. \end{cases} \quad (\text{B.8})$$

Therefore, we have

$$\begin{aligned} & \left\{ \frac{n_{s:b} n_{(b+1):e}}{n_{s:e}} \right\}^{1/2} \left\{ F_{s:b}(z) - F_{(b+1):e}(z) \right\} = \sum_{k=s}^e w_k \sum_{i=1}^{n_k} \mathbb{E}(1_{\{Y_{k,i} \leq z\}}), \\ & D_{s,e}^b = \sup_{z \in \mathbb{R}} \left| \sum_{k=s}^e w_k \sum_{i=1}^{n_k} 1_{\{Y_{k,i} \leq z\}} \right| \quad \text{and} \quad \Delta_{s,e}^b = \sup_{z \in \mathbb{R}} \left| \sum_{k=s}^e w_k \sum_{i=1}^{n_k} \mathbb{E}(1_{\{Y_{k,i} \leq z\}}) \right|. \end{aligned}$$

Since

$$\begin{aligned} D_{s,e}^b &= \sup_{z \in R} \left| \sum_{k=s}^e w_k \sum_{i=1}^{n_k} \{ \mathbb{E}(1_{\{Y_{k,i} \leq z\}}) + 1_{\{Y_{k,i} \leq z\}} - \mathbb{E}(1_{\{Y_{k,i} \leq z\}}) \} \right| \\ &\leq \sup_{z \in \mathbb{R}} \left| \sum_{k=s}^e w_k \sum_{i=1}^{n_k} \mathbb{E}(1_{\{Y_{k,i} \leq z\}}) \right| + \sup_{z \in \mathbb{R}} \left| \sum_{k=s}^e w_k \sum_{i=1}^{n_k} \{ 1_{\{Y_{k,i} \leq z\}} - \mathbb{E}(1_{\{Y_{k,i} \leq z\}}) \} \right| \\ &= \Delta_{s,e}^b + \sup_{z \in \mathbb{R}} \left| \sum_{k=s}^e w_k \sum_{i=1}^{n_k} \{ 1_{\{Y_{k,i} \leq z\}} - \mathbb{E}(1_{\{Y_{k,i} \leq z\}}) \} \right|, \end{aligned}$$

we have

$$|D_{s,e}^b - \Delta_{s,e}^b| \leq \sup_{z \in \mathbb{R}} \left| \sum_{k=s}^e w_k \sum_{i=1}^{n_k} \{ 1_{\{Y_{k,i} \leq z\}} - \mathbb{E}(1_{\{Y_{k,i} \leq z\}}) \} \right|. \quad (\text{B.9})$$

Next for $z \in \mathbb{R}$ define

$$\Lambda_{s,e}^b(z) = \sum_{k=s}^e w_k \sum_{i=1}^{n_k} \{ 1_{\{Y_{k,i} \leq z\}} - \mathbb{E}(1_{\{Y_{k,i} \leq z\}}) \},$$

and let $\{s_1^k, \dots, s_{m-1}^k\} \subset \mathbb{R}$ satisfy

$$s_j^k = F_k^{-1}(j/m),$$

where m is a positive integer to be specified. Let $I_1^k = (-\infty, s_1^k]$, $I_j^k = (s_{j-1}^k, s_j^k]$, $j = 2, \dots, m-1$, and $I_m^k = (s_{m-1}^k, \infty)$. With this notation, for any $k \in \{1, \dots, n\}$, we get a partition of \mathbb{R} , namely $\mathcal{I}_k = \{I_1^k, \dots, I_m^k\}$. Let $\mathcal{I} = \cap_{k=1}^T \mathcal{I}_k = \{I_1, \dots, I_M\}$. Note that there are at most T/δ distinct \mathcal{I}_k 's, and therefore $M \leq Tm/\delta$.

Let also z_j be an interior point of I_j for all $j \in \{1, \dots, M\}$. Then

$$\sup_{z \in \mathbb{R}} |\Lambda_{s,e}^b(z)| \leq \max_{j=1, \dots, M} \left\{ |\Lambda_{s,e}^b(z_j)| + \sup_{z \in I_j} |\Lambda_{s,e}^b(z_j) - \Lambda_{s,e}^b(z)| \right\}. \quad (\text{B.10})$$

By Hoeffding's inequality and a union bound argument, we have for any $\varepsilon > 0$

$$\mathbb{P} \left\{ \max_{1 \leq s < b < e \leq T} \max_{j=1, \dots, M} |\Lambda_{s,e}^b(z_j)| > \varepsilon \right\} \leq \frac{2T^4 m}{\delta} \exp(-2\varepsilon^2), \quad (\text{B.11})$$

since

$$\sum_{k=s}^e \sum_{i=1}^{n_k} w_k^2 = 1.$$

On the other hand, for $j \in \{1, \dots, M\}$, let $z \in I_j$ and without loss of generality assume that $z_j < z$. Let

$$u_j = |\{(i, k) : k \in \{1, \dots, T\}, i \in \{1, \dots, n_k\} \text{ and } y_{k,i} \in I_j\}|.$$

Let $r(t)$ satisfy $\eta_{r(t)-1} + 1 \leq t \leq \eta_{r(t)}$, $t \in \{1, \dots, T\}$. For $I_j \in \mathcal{I}$, let $q(j, k)$ be such that $I_j \subset I_{q(j,k)}^{r(k)}$. Let

$$v_j = |\{(i, k) : k \in \{1, \dots, T\}, i \in \{1, \dots, n_k\} \text{ and } y_{k,i} \in I_{v(j,k)}^{r(k)}\}|.$$

It holds that $u_j \leq v_j$ and $E(v_j) = n_{1:T}/m$.

We have,

$$\begin{aligned} |\Lambda_{s,e}^b(z_j) - \Lambda_{s,e}^b(z)| &\leq \left| \sum_{k=s}^e \sum_{i=1}^{n_k} w_k \{1_{\{y_{k,i} \leq z_j\}} - 1_{\{y_{k,i} \leq z\}}\} \right| \\ &\quad + \left| \sum_{k=s}^e \sum_{i=1}^{n_k} w_k \{F_k(z_j) - F_k(z)\} \right| \\ &\leq \left| \sum_{k=s}^e \sum_{i=1}^{n_k} 1_{\{z_j < y_{k,i} \leq z\}} \right|^{1/2} + \left(\sum_{k=s}^e \sum_{i=1}^{n_k} |w_k| \right) \max_{k=s, \dots, e} |F_k(z) - F_k(z_j)| \\ &\leq \max_{1 \leq j \leq M} u_j^{1/2} + \frac{2}{m} \left\{ \frac{n_{(b+1):e} n_{s:b}}{n_{s:e}} \right\}^{1/2} \leq \max_{1 \leq j \leq M} u_j^{1/2} + \frac{2n_{1:T}^{1/2}}{m}. \end{aligned} \tag{B.12}$$

From the multiplicative Chernoff bound

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq j \leq M} u_j \geq \frac{3n_{1:T}}{2m} \right\} &\leq M\mathbb{P} \left\{ u_j \geq \frac{3n_{1:T}}{2m} \right\} \leq M\mathbb{P} \left\{ v_j \geq \frac{3n_{1:T}}{2m} \right\} \\ &< \frac{Tm}{\delta} \exp \left(-\frac{n_{1:T}}{12m} \right) \leq \exp \left\{ -\frac{n_{1:T}}{12m} + \log(T) + \log(m) - \log(\delta) \right\}. \end{aligned} \tag{B.13}$$

Combining (B.10), (B.11), (B.12) and (B.13), we have

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq s < b < e \leq T} \sup_{z \in \mathbb{R}} |\Lambda_{s,e}^b(z)| > \epsilon + \left(\frac{3n_{1:T}}{2m} \right)^{1/2} + \frac{2n_{1:T}^{1/2}}{m} \right\} \\ \leq \frac{2T^4 m}{\delta} \exp(-2\epsilon^2) + \exp \left\{ -\frac{n_{1:T}}{12m} + \log(T) + \log(m) - \log(\delta) \right\}. \end{aligned} \tag{B.14}$$

Choosing

$$\epsilon = \log^{1/2} \left(\frac{2T^4 m}{\delta} \right) \quad \text{and} \quad m = \frac{n_{1:T}}{24 \log(n_{1:T})},$$

Eq. (B.14) results in

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq s < b < e \leq T} \sup_{z \in \mathbb{R}} |\Lambda_{s,e}^b(z)| > \left\{ \log \left(\frac{T^4}{12\delta} \right) + \log(n_{1:T}) \right\}^{1/2} \right. \\ \left. + 6 \log^{1/2}(n_{1:T}) + \frac{48 \log(n_{1:T})}{n_{1:T}^{1/2}} \right\} \end{aligned}$$

$$\leq \frac{12 \log(n_{1:T})}{T^3 n_{1:T}} + \frac{24T}{n_{1:T} \log(n_{1:T}) \delta}.$$

As for the result (B.7), we only need to change (B.8) to $w_k = (n_{s:e})^{-1/2}$. \square

Lemma B.4. *Under Condition 1.1, let $1 \leq s < \eta_k < e \leq T$ be any interval satisfying*

$$\min\{\eta_k - s, e - \eta_k\} \geq c_1 \delta,$$

with $c_1 > 0$. Then we have that

$$\max_{t=s+1, \dots, e-1} \Delta_{s,e}^t \geq \frac{c_1 \kappa \delta n_{\min}}{2(e-s)^{1/2} n_{\max}^{1/2}} \geq \frac{c_1 \kappa \delta n_{\min}^{3/2}}{2(e-s)^{1/2} n_{\max}^{1/2}}.$$

Proof. Let

$$z_0 \in \arg \max_{z \in \mathbb{R}} |F_{\eta_k}(z) - F_{\eta_{k+1}}(z)|.$$

Without loss of generality, assume that $F_{\eta_k}(z_0) > F_{\eta_{k+1}}(z_0)$. For $s < t < e$, it holds that

$$\begin{aligned} \Delta_{s,e}^t(z_0) &= \left\{ \frac{n_{s:e} n_{s:t}}{n_{(t+1):e}} \right\}^{1/2} \left\{ \frac{1}{n_{s:t}} \sum_{l=s}^t n_l F_l(z_0) - \frac{1}{n_{s:e}} \sum_{l=s}^e n_l F_l(z_0) \right\} \\ &= \left\{ \frac{n_{s:e}}{n_{s:t} n_{(t+1):e}} \right\}^{1/2} \sum_{l=s}^t n_l \tilde{F}_l(z_0), \end{aligned}$$

where $\tilde{F}_l(z_0) = F_l(z_0) - (n_{s:e})^{-1} \sum_{l=s}^e n_l F_l(z_0)$.

Due to Condition 1.1, it holds that $\tilde{F}_{\eta_k}(z_0) > \kappa/2$. Therefore

$$\sum_{l=s}^{\eta_k} n_l \tilde{F}_l(z_0) \geq (c_1/2) \kappa n_{\min} \delta$$

and

$$\left\{ \frac{n_{s:e}}{n_{s:t} n_{(t+1):e}} \right\}^{1/2} \geq (e-s)^{-1/2} n_{\max}^{-1/2} \geq \frac{n_{\min}^{1/2}}{(e-s)^{1/2} n_{\max}}.$$

Then

$$\max_{t=s+1, \dots, e-1} \Delta_{s,e}^t \geq \frac{1}{2(e-s)^{1/2} n_{\max}^{1/2}} \geq \frac{c_1 \kappa \delta n_{\min}^{3/2}}{2(e-s)^{1/2} n_{\max}}. \quad \square$$

In the following lemma, the condition (B.16) follows from Lemma B.4, and (B.17) follows from Lemma B.3.

Lemma B.5. *Let $z_0 \in \mathbb{R}$, $(s, e) \subset (1, T)$. Suppose that there exists a true change point $\eta_k \in (s, e)$ such that*

$$\min\{\eta_k - s, e - \eta_k\} \geq c_1 \delta, \tag{B.15}$$

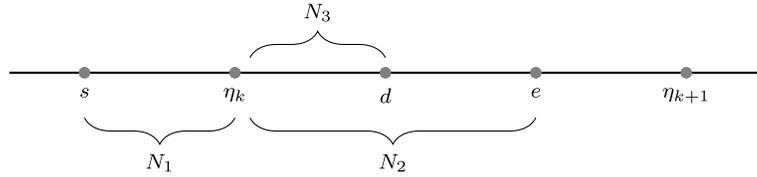


FIG 4. A graph showing Case (i) in the proof of Lemma B.5.

and

$$\Delta_{s,e}^{\eta_k}(z_0) \geq (c_1/2) \frac{n_{\min}^{3/2}}{n_{\max}} \frac{\kappa\delta}{(e-s)^{1/2}}, \tag{B.16}$$

where $c_1 > 0$ is a sufficiently small constant. In addition, assume that

$$\max_{s < t < e} |\Delta_{s,e}^t(z_0)| - \Delta_{s,e}^{\eta_k}(z_0) \leq 3 \log \left(\frac{T^4}{\delta} \right) + 3 \log(n_{1:T}) \leq \frac{\kappa\delta^4 n_{\min}^5}{(e-s)^{7/2} n_{\max}^{9/2}}. \tag{B.17}$$

Then there exists $d \in (s, e)$ satisfying

$$|d - \eta_k| \leq \frac{c_1 \delta n_{\min}^2}{32 n_{\max}^2}, \tag{B.18}$$

and

$$\Delta_{s,e}^{\eta_k}(z_0) - \Delta_{s,e}^d(z_0) > c |d - \eta_k| \delta \frac{n_{\min}^2}{n_{\max}^2} \Delta_{s,e}^{\eta_k}(z_0) (e-s)^{-2},$$

where $c > 0$ is a sufficiently small constant.

Proof. Let us assume without loss of generality that $d \geq \eta_k$. Following the argument of Lemma 2.6 in Venkatraman (1992), it suffices to consider two cases: (i) $\eta_{k+1} > e$ and (ii) $\eta_{k+1} \leq e$.

Case (i) $\eta_{k+1} > e$. It holds that

$$\Delta_{s,e}^{\eta_k}(z_0) = \left(\frac{N_1 N_2}{N_1 + N_2} \right)^{1/2} \{F_{\eta_k}(z_0) - F_{\eta_{k+1}}(z_0)\}$$

and

$$\Delta_{s,e}^d(z_0) = N_1 \left\{ \frac{N_2 - N_3}{(N_1 + N_3)(N_1 + N_2)} \right\}^{1/2} \{F_{\eta_k}(z_0) - F_{\eta_{k+1}}(z_0)\},$$

where $N_1 = n_{s:\eta_k}$, $N_2 = n_{(\eta_k+1):e}$ and $N_3 = n_{(\eta_k+1):d}$. Therefore, due to (B.15), we have

$$\begin{aligned} E_l &= \Delta_{s,e}^{\eta_k}(z_0) - \Delta_{s,e}^d(z_0) = \left[1 - \left\{ \frac{N_1(N_2 - N_3)}{N_2(N_1 + N_3)} \right\}^{1/2} \right] \Delta_{s,e}^{\eta_k}(z_0) \\ &= \frac{N_1 + N_2}{\{N_2(N_1 + N_3)\}^{1/2} [\{N_2(N_1 + N_3)\}^{1/2} + \{N_1(N_2 - N_3)\}^{1/2}]} N_3 \Delta_{s,e}^{\eta_k}(z_0) \end{aligned}$$

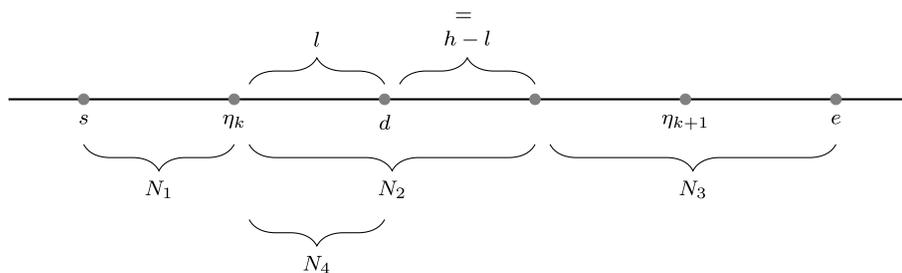


FIG 5. Illustrations of Case (ii) in the proof of Lemma B.5.

$$\geq c_1 \frac{n_{\min}^2}{n_{\max}^2} |d - \eta_k| \delta \Delta_{s,e}^{\eta_k}(z_0) (e - s)^{-2}. \tag{B.19}$$

Case (ii) $\eta_{k+1} \leq e$. Let $N_1 = n_{s:\eta_k}$, $N_2 = n_{(\eta_k+1):(\eta_k+h)}$ and $N_3 = n_{(\eta_k+h+1):e}$, where $h = c_1 \delta / 8$. Then,

$$\Delta_{s,e}^{\eta_k}(z_0) = a \left(\frac{N_1 + N_2 + N_3}{N_1(N_2 + N_3)} \right)^{1/2} \text{ and}$$

$$\Delta_{s,e}^{\eta_k+h}(z_0) = (a + N_2 \theta) \left(\frac{N_1 + N_2 + N_3}{N_3(N_1 + N_2)} \right)^{1/2}.$$

where

$$a = \sum_{l=s}^{\eta_k} n_l F_l(z_0) - c_0, \quad c_0 = \frac{1}{n_{s:e}} \sum_{l=s}^e n_l F_l(z_0)$$

and

$$\theta = \frac{a \{(N_1 + N_2)N_3\}^{1/2}}{N_2} \left\{ \frac{1}{\{N_1(N_2 + N_3)\}^{1/2}} - \frac{1}{(N_1 + N_2)N_3} + \frac{b}{a(N_1 + N_2 + N_3)^{1/2}} \right\},$$

with $b = \Delta_{s,e}^{\eta_k+h}(z_0) - \Delta_{s,e}^{\eta_k}(z_0)$.

Next, we set $l = d - \eta_k \leq h/2$ and $N_4 = n_{(\eta_k+1):d}$. Therefore, as in the proof of Lemma 2.6 in Venkatraman (1992), we have that

$$E_l = \Delta_{s,e}^{\eta_k}(z_0) - \Delta_{s,e}^{\eta_k+l}(z_0) = E_{1l}(1 + E_{2l}) + E_{3l}, \tag{B.20}$$

where

$$E_{1l} = \frac{aN_4(N_2 - N_4)(N_1 + N_2 + N_3)^{1/2}}{\{N_1(N_2 + N_3)(N_1 + N_4)(N_2 + N_3 - N_4)\}^{1/2}} \times \frac{1}{\{(N_1 + N_4)(N_2 + N_3 - N_4)\}^{1/2} + \{N_1(N_2 + N_3)\}^{1/2}},$$

$$E_{2l} = \frac{(N_3 - N_1)(N_3 - N_1 - N_4)}{\{(N_1 + N_4)(N_2 + N_3 - N_4)\}^{1/2} + \{(N_1 + N_2)N_3\}^{1/2}} \times \frac{1}{\{N_1(N_2 + N_3)\}^{1/2} + \{(N_1 + N_2)N_3\}^{1/2}},$$

and

$$E_{3l} = -\frac{bN_4}{N_2} \left\{ \frac{(N_1 + N_2)N_3}{(N_1 + N_4)(N_2 + N_3 - N_4)} \right\}^{1/2}.$$

Since $N_2 - N_4 \geq n_{\min}c_1\delta/16$, it holds that

$$E_{1l} \geq c_{1l}|d - \eta_k|\delta \frac{n_{\min}^2}{n_{\max}^2} \Delta_{s,e}^{\eta_k}(z_0)(e - s)^{-2}, \tag{B.21}$$

where $c_{1l} > 0$ is a sufficiently small constant depending on c_1 . As for E_{2l} , due to (B.18), we have

$$E_{2l} \geq -1/2. \tag{B.22}$$

As for E_{3l} , we have

$$\begin{aligned} E_{3l} &\geq -\left\{ 3 \log\left(\frac{T^4}{\delta}\right) + 3 \log(n_{1:T}) \right\} |d - \eta_k| \frac{n_{\min}^2}{n_{\max}^2} \frac{e - s}{c_1^2 \delta^2} \\ &\geq -c_{3l} \left\{ 3 \log\left(\frac{T^4}{\delta}\right) + 3 \log(n_{1:T}) \right\} |d - \eta_k| \Delta_{s,e}^{\eta_k}(z_0) \delta (e - s)^{-2} \frac{n_{\min}^2}{n_{\max}^2} \\ &\quad \times \frac{n_{\max}^{9/2}}{n_{\min}^5} (e - s)^{7/2} \frac{\log(n_{1:T})}{\kappa \delta^4} \\ &\geq -c_{1l}/2 |d - \eta_k| \delta \frac{n_{\min}^2}{n_{\max}^2} \Delta_{s,e}^{\eta_k}(z_0)(e - s)^{-2}, \end{aligned} \tag{B.23}$$

where the first inequality follows from (B.17), the second inequality from (B.16), and the last from (B.17).

Combining (B.20), (B.21), (B.22) and (B.23), we have

$$\Delta_{s,e}^{\eta_k}(z_0) - \Delta_{s,e}^d(z_0) \geq c|d - \eta_k|\delta \frac{n_{\min}^2}{n_{\max}^2} \Delta_{s,e}^{\eta_k}(z_0)(e - s)^{-2}, \tag{B.24}$$

where $c > 0$ is a sufficiently small constant.

In view of (B.19) and (B.24), we conclude the proof. □

Lemma B.6. *Suppose $(s, e) \subset (1, T)$ such that $e - s \leq C_M\delta$ and that*

$$\eta_{k-1} \leq s \leq \eta_k \leq \dots \leq \eta_{k+q} \leq e \leq \eta_{k+q+1}, \quad q \geq 0.$$

Denote

$$\kappa_{\max}^{s,e} = \max\{\kappa_p : p = k, \dots, k + q\}.$$

Then for any $p \in \{k - 1, \dots, k + q\}$, it holds that

$$\sup_{z \in \mathbb{R}} \left| \frac{1}{n_{s:e}} \sum_{t=s}^e n_t F_t(z) - F_{\eta_p}(z) \right| \leq (C_M + 1) \kappa_{\max}^{s,e}.$$

Proof. Since $e - s \leq C_M \delta$, the interval (s, e) contains at most $C_M + 1$ true change points. Note that

$$\begin{aligned}
 & \sup_{z \in \mathbb{R}} \left| \frac{1}{n_{s:e}} \sum_{t=s}^e n_t F_t(z) - F_{\eta_p}(z) \right| \\
 = & \sup_{z \in \mathbb{R}} \left\{ \frac{1}{n_{s:e}} \left| \sum_{t=s}^{\eta_k} n_t \{F_{\eta_{k-1}}(z) - F_{\eta_p}(z)\} + \sum_{t=\eta_{k+1}}^{\eta_{k+1}} n_t \{F_{\eta_k}(z) - F_{\eta_p}(z)\} + \dots \right. \right. \\
 & \left. \left. + \sum_{t=\eta_{k+q}+1}^e n_t \{F_{\eta_{k+q}}(z) - F_{\eta_p}(z)\} \right| \right\} \\
 \leq & \frac{|p-k| \sum_{t=s}^{\eta_k} n_t + |p-k-1| \sum_{t=\eta_{k+1}}^{\eta_{k+1}} n_t + \dots + |p-k-q-1| \sum_{t=\eta_{k+q}+1}^e n_t}{n_{s:e}} \\
 & \cdot \kappa_{\max}^{s,e} \\
 \leq & (C_M + 1) \kappa_{\max}^{s,e}. \quad \square
 \end{aligned}$$

For any $x = (x_i) \in \mathbb{R}^{n_{s:e}}$, define

$$\mathcal{P}_{s,e}^d(x) = \frac{1}{n_{s:e}} \sum_{i=1}^{n_{s:e}} x_i + \langle x, \psi_{s,e}^d \rangle \psi_{s,e}^d,$$

where $\langle \cdot, \cdot \rangle$ is the inner product in Euclidean space, and $\psi_{s,e}^d \in \mathbb{R}^{n_{s:e}}$ with

$$(\psi_{s,e}^d)_i = \begin{cases} \left\{ \frac{n_{(d+1):e}}{n_{s:e} n_{s:d}} \right\}^{1/2}, & i = 1, \dots, n_{s:d}, \\ - \left\{ \frac{n_{s:d}}{n_{s:e} n_{(d+1):e}} \right\}^{1/2}, & i = n_{s:d} + 1, \dots, n_{s:e}, \end{cases}$$

i.e. the i -th entry of $\mathcal{P}_{s,e}^d(x)$ satisfies

$$\mathcal{P}_{s,e}^d(x)_i = \begin{cases} \frac{1}{n_{s:d}} \sum_{j=1}^{n_{s:d}} x_j, & i = 1, \dots, n_{s:d}, \\ \frac{1}{n_{(d+1):e}} \sum_{j=n_{s:d}+1}^{n_{s:e}} x_j, & i = n_{s:d} + 1, \dots, n_{s:e}. \end{cases}$$

Lemma B.7. *Suppose Condition 1.1 holds and consider any interval $(s, e) \subset (1, T)$ satisfying that there exists a true change point $\eta_k \in (s, e)$. Let*

$$b \in \arg \max_{s < t < e} D_{s,e}^t \quad \text{and} \quad z_0 \in \arg \max_{z \in \mathbb{R}} |D_{s,e}^b(z)|.$$

Let

$$\mu_{s,e} = \underbrace{(F_s(z_0), \dots, F_s(z_0))}_{n_s}, \dots, \underbrace{(F_e(z_0), \dots, F_e(z_0))}_{n_e} \top \in \mathbb{R}^{n_{s:e}}$$

and

$$Y_{s,e} = \underbrace{(1_{\{Y_{s,1} \leq z_0\}}, \dots, 1_{\{Y_{s,n_s} \leq z_0\}})}_{n_s}, \dots, \underbrace{(1_{\{Y_{e,1} \leq z_0\}}, \dots, 1_{\{Y_{e,n_e} \leq z_0\}})}_{n_e} \top \in \mathbb{R}^{n_{s:e}}.$$

We have

$$\|Y_{s,e} - \mathcal{P}_{s,e}^b(Y_{s,e})\|^2 \leq \|Y_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(Y_{s,e})\|^2 \leq \|Y_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e})\|^2. \quad (\text{B.25})$$

Proof. Note that for any $d \in (s, e)$, we have

$$\begin{aligned} \|Y_{s,e} - \mathcal{P}_{s,e}^d(Y_{s,e})\|^2 &= n_{s:d}(Y_1 - Y_1^2) + n_{(d+1):e}(Y_2 - Y_2^2) \\ &= -\{D_{s,e}^t(z_0)\}^2 + \frac{\{\sum_{t=s}^e \sum_{i=1}^{n_t} 1_{\{Y_{t,i} \leq z_0\}}\}^2}{n_{s:e}} - \sum_{t=s}^e \sum_{i=1}^{n_t} 1_{\{Y_{t,i} \leq z_0\}}, \end{aligned}$$

where

$$Y_1 = \frac{1}{n_{s:d}} \sum_{t=s}^d \sum_{i=1}^{n_t} 1_{\{Y_{t,i} \leq z_0\}}, \quad \text{and} \quad Y_2 = \frac{1}{n_{(d+1):e}} \sum_{t=d+1}^e \sum_{i=1}^{n_t} 1_{\{Y_{t,i} \leq z_0\}}.$$

It follow from the definition of b , we have that

$$\|Y_{s,e} - \mathcal{P}_{s,e}^b(Y_{s,e})\|^2 \leq \|Y_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(Y_{s,e})\|^2.$$

The second inequality in (B.25) follows from the observation that the sum of the squares of errors is minimized by the sample mean. \square

Lemma B.8. *Let $(s, e) \subset (1, T)$ contains two or more change points such that*

$$\eta_{k-1} \leq s \leq \eta_k \leq \dots \leq \eta_{k+q} \leq e \leq \eta_{k+q+1}, \quad q \geq 1.$$

If $\eta_k - s \leq c_1 \delta$, for $c_1 > 0$, then

$$\Delta_{s,e}^{\eta_k} \leq \left(\frac{c_1 n_{\max}}{n_{\min}}\right)^{1/2} \Delta_{s,e}^{\eta_{k+1}} + 2n_{s:\eta_k}^{1/2} \kappa_k.$$

Proof. Consider the distribution sequence $\{G_t\}_{t=s}^e$ be such that

$$G_t = \begin{cases} F_{\eta_{k+1}}, & t = s + 1, \dots, \eta_k, \\ F_t, & t = \eta_k + 1, \dots, e. \end{cases}$$

For any $s < t < e$, define

$$\mathcal{G}_{s,e}^t = \sup_{z \in \mathbb{R}} |\mathcal{G}_{s,e}^t(z)|,$$

where

$$\mathcal{G}_{s,e}^t(z) = \left(\frac{n_{s:t}n_{(t+1):e}}{n_{s:e}}\right)^{1/2} \left\{ \frac{1}{n_{s:t}} \sum_{l=s}^t n_l G_l(z) - \frac{1}{n_{(t+1):e}} \sum_{l=t+1}^e n_l G_l(z) \right\}.$$

For any $t \geq \eta_k$ and $z \in \mathbb{R}$, it holds that

$$|\Delta_{s,e}^t(z) - \mathcal{G}_{s,e}^t(z)| = \left\{ \frac{n_{(t+1):e}}{n_{s:e}n_{s:t}} \right\}^{1/2} n_{s:\eta_k} |F_{\eta_{k+1}}(z) - F_{\eta_k}(z)| \leq n_{s:\eta_k}^{1/2} \kappa_k.$$

Thus we have

$$\begin{aligned} \Delta_{s,e}^{\eta_k} &= \sup_{z \in R} \left| \Delta_{s,e}^{\eta_k}(z) - \mathcal{G}_{s,e}^{\eta_k}(z) + \mathcal{G}_{s,e}^{\eta_k}(z) \right| \leq \sup_{z \in \mathbb{R}} \left| \Delta_{s,e}^{\eta_k}(z) - \mathcal{G}_{s,e}^{\eta_k}(z) \right| + \mathcal{G}_{s,e}^{\eta_k} \\ &\leq \mathcal{G}_{s,e}^{\eta_k} + n_{s:\eta_k}^{1/2} \kappa_k \leq \left\{ \frac{n_{s:\eta_k} n_{(\eta_{k+1}+1):e}}{n_{s:\eta_{k+1}} n_{(\eta_k+1):e}} \right\}^{1/2} \mathcal{G}_{s,e}^{\eta_{k+1}} + n_{s:\eta_k}^{1/2} \kappa_k \\ &\leq \left(\frac{c_1 n_{\max}}{n_{\min}} \right)^{1/2} \Delta_{s,e}^{\eta_{k+1}} + 2n_{s:\eta_k}^{1/2} \kappa_k. \quad \square \end{aligned}$$

Lemma B.9. Under Condition 1.1, let (s_0, e_0) be an interval with $e_0 - s_0 \leq C_M \delta$ and contain at least one change point η_k such that

$$\eta_{k-1} \leq s_0 \leq \eta_k \leq \dots \leq \eta_{k+q} \leq e_0 \leq \eta_{k+q+1}, \quad q \geq 0.$$

Suppose that there exists k' such that

$$\min\{\eta_{k'} - s_0, e_0 - \eta_{k'}\} \geq \delta/16.$$

Let

$$\kappa_{s,e}^{\max} = \max\{\kappa_p : \min\{\eta_p - s_0, e_0 - \eta_p\} \geq \delta/16\}.$$

Consider any generic $(s, e) \subset (s_0, e_0)$, satisfying

$$\min\{\eta_k - s_0, e_0 - \eta_k\} \geq \delta/16, \quad \eta_k \in (s, e).$$

Let $b \in \arg \max_{s < t < e} D_{s,e}^t$. For some $c_1 > 0$ and $\gamma > 0$, suppose that

$$D_{s,e}^b \geq c_1 \kappa_{s,e}^{\max} \delta^{1/2} \frac{n_{\min}^{3/2}}{n_{\max}}, \quad (\text{B.26})$$

$$\max_{s < t < e} \sup_{z \in \mathbb{R}} \left| \Lambda_{s,e}^t(z) \right| \leq \gamma, \quad (\text{B.27})$$

and

$$\max_{1 \leq s < e \leq T} \sup_{z \in \mathbb{R}} \left| n_{s:e}^{-1/2} \sum_{t=s}^e \sum_{i=1}^{n_t} \{1_{\{Y_{t,i} \leq z\}} - F_t(z)\} \right| \leq \gamma. \quad (\text{B.28})$$

If there exists a sufficiently small $0 < c_2 < c_1/2$ such that

$$\gamma \leq c_2 \kappa_{s,e}^{\max} \delta^{1/2} \frac{n_{\min}^{3/2}}{n_{\max}}, \quad (\text{B.29})$$

then there exists a change point $\eta_k \in (s, e)$ such that

$$\min\{e - \eta_k, \eta_k - s\} \geq \delta/4 \quad \text{and} \quad |\eta_k - b| \leq C \frac{n_{\max}^9}{n_{\min}^{10}} \kappa^{-2} \gamma^2,$$

where $C > 0$ is a sufficiently large constant.

Proof. Without loss of generality, assume that $\Delta_{s,e}^b > 0$ and that $\Delta_{s,e}^t$ is locally decreasing at b . Observe that there has to be a change point $\eta_k \in (s, b)$, or otherwise $\Delta_{s,e}^b > 0$ implies that $\Delta_{s,e}^t$ is decreasing, as a consequence of Lemma B.1. Thus, if $s \leq \eta_k \leq b \leq e$, then

$$\Delta_{s,e}^{\eta_k} \geq \Delta_{s,e}^b \geq D_{s,e}^b - \gamma \geq (c_1 - c_2)\kappa_{s,e}^{\max} \delta^{1/2} \frac{n_{\min}^{3/2}}{n_{\max}} \geq (c_1/2)\kappa_{s,e}^{\max} \delta^{1/2} \frac{n_{\min}^{3/2}}{n_{\max}}, \tag{B.30}$$

where the second inequality follows from (B.27), and the second inequality follows from (B.26) and (B.29). Observe that $e - s \leq e_0 - s_0 \leq C_M \delta$ and that (s, e) has to contain at least one change point or otherwise $\max_{s < t < e} \Delta_{s,e}^t = 0$, which contradicts (B.30).

Step 1. In this step, we are to show that

$$\min\{\eta_k - s, e - \eta_k\} \geq \min\{1, c_1^2\} \delta / 16. \tag{B.31}$$

Suppose that η_k is the only change point in (s, e) . Then (B.31) must hold or otherwise it follows from (B.5) in Lemma B.2, we have

$$\Delta_{s,e}^{\eta_k} \leq \kappa_k n_{\max}^{1/2} \frac{c_1 \delta^{1/2}}{4},$$

which contradicts (B.30).

Suppose (s, e) contains at least two change points. Then $\eta_k - s < \min\{1, c_1^2\} \delta / 16$ implies that η_k is the most left change point in (s, e) . Therefore it follows from Lemma B.8 that

$$\begin{aligned} \Delta_{s,e}^{\eta_k} &\leq \frac{c_1}{4} \left(\frac{n_{\max}}{n_{\min}}\right)^{1/2} \Delta_{s,e}^{\eta_{k+1}} + 2n_{s:\eta_k}^{1/2} \kappa_k \\ &\leq \frac{c_1}{4} \left(\frac{n_{\max}}{n_{\min}}\right)^{1/2} \max_{s < t < e} \Delta_{s,e}^t + \frac{\delta^{1/2}}{4} c_1 n_{\max}^{1/2} \kappa_k \\ &\leq \frac{c_1}{4} \left(\frac{n_{\max}}{n_{\min}}\right)^{1/2} \max_{s < t < e} D_{s,e}^t + \frac{c_1}{4} \left(\frac{n_{\max}}{n_{\min}}\right)^{1/2} \gamma + \frac{\delta^{1/2}}{4} c_1 n_{\max}^{1/2} \kappa_k \\ &\leq \max_{s < t < e} D_{s,e}^t - \gamma, \end{aligned}$$

which contradicts with (B.30).

Step 2. It follows from Lemma B.5 that there exists $d \in (\eta_k, \eta_k + c_1 \delta n_{\min}^2 n_{\max}^{-2} / 32)$ and that

$$\Delta_{s,e}^{\eta_k} - \Delta_{s,e}^d \geq 2\gamma. \tag{B.32}$$

We claim that $b \in (\eta_k, d) \subset (\eta_k, \eta_k + c_1 \delta n_{\min}^2 n_{\max}^{-2} / 16)$. By contradiction, suppose that $b \geq d$. Then

$$\Delta_{s,e}^b \leq \Delta_{s,e}^d < \Delta_{s,e}^{\eta_k} - 2\gamma \leq \max_{s < t < e} \Delta_{s,e}^t - 2\gamma \leq \max_{s < t < e} D_{s,e}^t - \gamma = D_{s,e}^b - \gamma, \tag{B.33}$$

where the first inequality follows from Lemma B.1, the second follows from (B.32), and the fourth follows from (B.27). Note that (B.33) is a contradiction with (B.30), therefore we have $b \in (\eta_k, \eta_k + c_1 \delta n_{\min}^2 n_{\max}^{-2} / 32)$.

Step 3. It follows from (B.25) in Lemma B.7 that

$$\|Y_{s,e} - \mathcal{P}_{s,e}^b(Y_{s,e})\|^2 \leq \|Y_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(Y_{s,e})\|^2 \leq \|Y_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e})\|^2,$$

with the notation defined in Lemma B.7. By contradiction, we assume that

$$\eta_k + C \frac{n_{\max}^9}{n_{\min}^{10}} \kappa^{-2} \gamma^2 < b, \quad (\text{B.34})$$

where $C > 0$ is a sufficiently large constant. We are to show that this leads to the bound that

$$\|Y_{s,e} - \mathcal{P}_{s,e}^b(Y_{s,e})\|^2 > \|Y_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e})\|^2, \quad (\text{B.35})$$

which is a contradiction.

We have $\min\{\eta_k - s, e - \eta_k\} \geq \min\{1, c_1^2\} \delta / 16$ and $|b - \eta_k| \leq c_1 \delta n_{\min}^2 n_{\max}^{-2} / 32$. For properly chose c_1 , we have

$$\min\{e - b, b - s\} \geq \min\{1, c_1^2\} \delta / 32.$$

It holds that

$$\begin{aligned} & \|Y_{s,e} - \mathcal{P}_{s,e}^b(Y_{s,e})\|^2 - \|Y_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e})\|^2 \\ &= \|\mu_{s,e} - \mathcal{P}_{s,e}^b(\mu_{s,e})\|^2 - \|\mu_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e})\|^2 + \\ & \quad 2\langle Y_{s,e} - \mu_{s,e}, \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e}) - \mathcal{P}_{s,e}^b(Y_{s,e}) \rangle. \end{aligned}$$

Therefore if we can show that

$$2\langle Y_{s,e} - \mu_{s,e}, \mathcal{P}_{s,e}^b(Y_{s,e}) - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e}) \rangle < \|\mu_{s,e} - \mathcal{P}_{s,e}^b(\mu_{s,e})\|^2 - \|\mu_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e})\|^2, \quad (\text{B.36})$$

then (B.35) holds.

As for the right-hand side of (B.36), we have

$$\begin{aligned} & \|\mu_{s,e} - \mathcal{P}_{s,e}^b(\mu_{s,e})\|^2 - \|\mu_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e})\|^2 = (\Delta_{s,e}^{\eta_k}(z_0))^2 - (\Delta_{s,e}^b(z_0))^2 \\ & \geq \{|\Delta_{s,e}^{\eta_k}(z_0) - \Delta_{s,e}^b(z_0)|\} |\Delta_{s,e}^{\eta_k}(z_0)|. \end{aligned} \quad (\text{B.37})$$

We are then to utilize the result of Lemma B.5. Note that z_0 there can be any $z_0 \in R$ satisfying conditions thereof. Equation (B.16) holds due to the fact that here we have

$$\begin{aligned} |\Delta_{s,e}^{\eta_k}(z_0)| & \geq |\Delta_{s,e}^b(z_0)| \geq |D_{s,e}^b(z_0)| - \gamma \geq c_1 \kappa_{s,e}^{\max} \delta^{1/2} \frac{n_{\min}^{3/2}}{n_{\max}} - c_2 \kappa_{s,e}^{\max} \delta^{1/2} \frac{n_{\min}^{3/2}}{n_{\max}} \\ & \geq c_1 / 2 \kappa_{s,e}^{\max} \delta^{1/2} \frac{n_{\min}^{3/2}}{n_{\max}}, \end{aligned} \quad (\text{B.38})$$

where the first inequality follows from the fact that η_k is a true change point, the second inequality from (B.27), the third inequality follows from (B.26) and

(B.29), and the final inequality follows from the condition that $0 < c_2 < c_1/2$. Towards this end, it follows from Lemma B.5 that

$$\Delta_{s,e}^{\eta_k}(z_0) - \Delta_{s,e}^b(z_0) \geq c|b - \eta_k| \delta \frac{n_{\min}^2}{n_{\max}^2} \Delta_{s,e}^{\eta_k}(z_0) (e-s)^{-2}. \quad (\text{B.39})$$

Combining (B.37), (B.38) and (B.39), we have

$$\|\mu_{s,e} - \mathcal{P}_{s,e}^b(\mu_{s,e})\|^2 - \|\mu_{s,e} - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e})\|^2 \geq \frac{cc_1^2}{4} \delta^2 \frac{n_{\min}^5}{n_{\max}^4} \kappa^2 (e-s)^{-2} |b - \eta_k|. \quad (\text{B.40})$$

The left-hand side of (B.36) can be decomposed as follows.

$$\begin{aligned} & 2\langle Y_{s,e} - \mu_{s,e}, \mathcal{P}_{s,e}^b(Y_{s,e}) - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e}) \rangle \\ &= 2\langle Y_{s,e} - \mu_{s,e}, \mathcal{P}_{s,e}^b(Y_{s,e}) - \mathcal{P}_{s,e}^b(\mu_{s,e}) \rangle + 2\langle Y_{s,e} - \mu_{s,e}, \mathcal{P}_{s,e}^b(\mu_{s,e}) - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e}) \rangle \\ &= (I) + 2 \left(\sum_{i=1}^{n_{s:\eta_k}} + \sum_{i=n_{s:\eta_k}+1}^{n_{s:b}} + \sum_{i=n_{s:b}+1}^{n_{s:e}} \right) (Y_{s,e} - \mu_{s,e})_i (\mathcal{P}_{s,e}^b(\mu_{s,e}) - \mathcal{P}_{s,e}^{\eta_k}(\mu_{s,e}))_i \\ &= (I) + (II.1) + (II.2) + (II.3). \end{aligned} \quad (\text{B.41})$$

Term (I). It holds that

$$(I) = \frac{2}{n_{s:b}} \left\{ \sum_{j=1}^{n_{s:b}} (Y_{s,e} - \mu_{s,e})_j \right\}^2 + \frac{2}{n_{(b+1):e}} \left\{ \sum_{j=n_{s:b}+1}^{n_{s:e}} (Y_{s,e} - \mu_{s,e})_j \right\}^2 \leq 2\gamma^2, \quad (\text{B.42})$$

where the inequality follows from the definition of the CUSUM statistics and (B.27).

Term (II). It holds that

$$(II.1) = 2n_{s:\eta_k}^{1/2} \left\{ n_{s:\eta_k}^{-1/2} \sum_{i=1}^{n_{s:\eta_k}} (Y_{s,e} - \mu_{s,e})_i \right\} \left\{ \frac{1}{n_{s:b}} \sum_{i=1}^{n_{s:b}} (\mu_{s,e})_i - \frac{1}{n_{s:\eta_k}} \sum_{i=1}^{\eta_k} (\mu_{s,e})_i \right\}.$$

In addition, it holds that

$$\begin{aligned} & \left| \frac{1}{n_{s:b}} \sum_{i=1}^{n_{s:b}} (\mu_{s,e})_i - \frac{1}{n_{s:\eta_k}} \sum_{i=1}^{\eta_k} (\mu_{s,e})_i \right| \\ &= \frac{n_{(\eta_k+1):b}}{n_{s:b}} \left| -\frac{1}{n_{s:\eta_k}} \sum_{i=1}^{n_{s:\eta_k}} (\mu_{s,e})_i + F_{\eta_k+1}(z_0) \right| \\ &\leq \frac{n_{(\eta_k+1):b}}{n_{s:b}} (C_M + 1) \kappa_{s,e}^{\max}, \end{aligned}$$

where the inequality follows from Lemma B.6. Combining with (B.28), it leads to that

$$(II.1) \leq 2n_{s:\eta_k}^{1/2} \gamma \frac{n_{(\eta_k+1):b}}{n_{s:b}} (C_M + 1) \kappa_{s,e}^{\max}$$

$$\leq 2 \frac{n_{\max}^{3/2}}{n_{\min}} \frac{4}{\min\{1, c_1^2\}} \delta^{-1/2} \gamma |b - \eta_k| (C_M + 1) \kappa_{s,e}^{\max}. \quad (\text{B.43})$$

As for the term (II.2), it holds that

$$(II.2) \leq 2n_{\max}^{1/2} |b - \eta_k|^{1/2} \gamma (2C_M + 3) \kappa_{s,e}^{\max}. \quad (\text{B.44})$$

As for the term (II.3), it holds that

$$(II.3) \leq 2 \frac{n_{\max}^{3/2}}{n_{\min}} \frac{4}{\min\{1, c_1^2\}} \delta^{-1/2} \gamma |b - \eta_k| (C_M + 1) \kappa_{s,e}^{\max}. \quad (\text{B.45})$$

Therefore, combining (B.40), (B.41), (B.42), (B.43), (B.44) and (B.44), we have that (B.36) holds if

$$\delta^2 \frac{n_{\min}^5}{n_{\max}^4} \kappa^2 (e-s)^{-2} |b - \eta_k| \gtrsim \max \left\{ \gamma^2, \frac{n_{\max}^{3/2}}{n_{\min}} \delta^{-1/2} \gamma |b - \eta_k| \kappa, n_{\max}^{1/2} |b - \eta_k|^{1/2} \gamma \kappa \right\}.$$

The second inequality holds due to Condition 3.1, the third inequality holds due to (B.34) and the first inequality is a consequence of the third inequality and Condition 3.1. \square

Appendix C: Proofs of Section 3.2

Proof of Lemma 3.1. Let P_0 denote the joint distribution of the independent random variables $\{Y_{t,i}\}_{i=1,t=1}^{n,T}$ such that $Y_{1,1}, \dots, Y_{\delta,n}$ are independent and identically distributed as δ_0 and $Y_{\delta+1,1}, \dots, Y_{T,n}$ are independent and identically distributed as δ_1 , where $\delta_c, c \in \mathbb{R}$, is the Dirac distribution having point mass at point c .

Let P_1 denote the joint distribution of the independent random variables $\{Z_{t,i}\}_{i=1,t=1}^{n,T}$ such that $Z_{1,1}, \dots, Z_{T-\delta,n}$ are independent and identically distributed as δ_1 and $Z_{T-\delta+1,1}, \dots, Z_{T,n}$ are independent and identically distributed as δ_0 .

It holds that $\eta(P_0) = \delta$ and $\eta(P_1) = T - \delta$. Since $\delta \leq T/3$, it holds that

$$\inf_{\hat{\eta}} \sup_{P \in \mathcal{P}} E_P (|\hat{\eta} - \eta|) \geq (T/3) \{1 - d_{\text{TV}}(P_0, P_1)\} \geq (T/3) \{1 - 2\delta n\} \geq \frac{1 - 1/2}{3} T,$$

where $d_{\text{TV}}(\cdot, \cdot)$ is the total variation distance. In the last display, the first inequality follows from Le Cam's lemma (see, e.g. Yu, 1997), and the second inequality follows from Eq.(1.2) in Steerneman (1983). \square

Proof of Lemma 3.2. Let P_0 denote the joint distribution of the independent random variables $\{Y_{t,i}\}_{i=1,t=1}^{n,T}$ such that $Y_{1,1}, \dots, Y_{\delta,n}$ are independent and identically distributed as F and $Y_{\delta+1,1}, \dots, Y_{T,n}$ are independent and identically distributed as G .

Let P_1 be the joint distribution of the independent random variables $\{Z_{t,i}\}_{i=1,t=1}^{n,T}$ such that $Z_{1,1}, \dots, Z_{\delta+\xi,n}$ are independent and identically distributed as F and $Z_{\delta+\xi+1,1}, \dots, Z_{T,n}$ are independent and identically distributed as G , where ξ is a positive integer no larger than $n - 1 - \delta$,

$$F(x) = \begin{cases} 0, & x \leq 0, \\ x, & 0 < x \leq 1, \\ 1, & x \geq 1, \end{cases}$$

and

$$G(x) = \begin{cases} 0, & x \leq 0, \\ (1 - 2\kappa)x, & 0 < x \leq 1/2, \\ (1/2 - \kappa) + (1 + 2\kappa)(x - 1/2), & 1/2 < x \leq 1, \\ 1, & x \geq 1. \end{cases}$$

It holds that

$$\sup_{z \in \mathbb{R}} |F(z) - G(z)| = \kappa,$$

$\eta(P_0) = \delta$ and $\eta(P_1) = \delta + \xi$. By Le Cam’s Lemma (e.g. Yu, 1997) and Lemma 2.6 in Tsybakov (2009), it holds that

$$\inf_{\hat{\eta}} \sup_{P \in \mathcal{Q}} E_P(|\hat{\eta} - \eta|) \geq \xi \{1 - d_{TV}(P_0, P_1)\} \geq \frac{\xi}{2} \exp(-\text{KL}(P_0, P_1)), \quad (\text{C.1})$$

where $\text{KL}(\cdot, \cdot)$ denotes the Kullback–Leibler divergence.

Since

$$\text{KL}(P_0, P_1) = \sum_{i \in \{\delta+1, \dots, \delta+\xi\}} \text{KL}(P_{0i}, P_{1i}) = \frac{n\xi}{2} \log(1 - 4\kappa^2) \leq 2n\xi\kappa^2,$$

we have

$$\inf_{\hat{\eta}} \sup_{P \in \mathcal{Q}} \mathbb{E}_P(|\hat{\eta} - \eta|) \geq \frac{\xi}{2} \exp(-2n\xi\kappa^2).$$

Set $\xi = \min\{\lceil \frac{1}{n\kappa^2} \rceil, T - 1 - \delta\}$. By the assumption on ζ_T , for all T large enough we must have that $\xi = \lceil \frac{1}{n\kappa^2} \rceil$. Thus, for all T large enough, using (C.1),

$$\inf_{\hat{\eta}} \sup_{P \in \mathcal{Q}} E_P(|\hat{\eta} - \eta|) \geq \max \left\{ 1, \frac{1}{2} \left\lceil \frac{1}{n\kappa^2} \right\rceil e^{-2} \right\}. \quad \square$$

Appendix D: Proof of Theorem 3.2

Proof. It follows from Theorem 3.1 and the proof thereof that applying Algorithm 1 to $\{W_{t,i}\}$ and the τ sequence defined in (3.7), with probability at least

$$1 - \frac{24 \log(n_{1:T})}{T^3 n_{1:T}} - \frac{48T}{n_{1:T} \log(n_{1:T}) \delta} - \exp \left\{ \log \left(\frac{T}{\delta} \right) - \frac{M\delta^2}{16T^2} \right\},$$

the event \mathcal{A} , which is defined as follows holds.

A1 if $\tau > c_{\tau,2}\kappa\delta^{1/2}n_{\min}^{3/2}n_{\max}^{-1}$, then the corresponding change point estimators satisfying $\widehat{K} < K$, but for any $\hat{\eta}$ in the estimator set, there exists $k \in \{1, \dots, K\}$ such that

$$|\hat{\eta} - \eta_k| \leq C_\epsilon \kappa_k^{-2} \log(n_{1:T}) n_{\max}^9 n_{\min}^{-10};$$

A2 if $c_{\tau,2}\kappa\delta^{1/2}n_{\min}^{3/2}n_{\max}^{-1} \geq \tau \geq c_{\tau,1} \log^{1/2}(n_{1:T})$, then the corresponding change point estimators satisfying $\widehat{K} = K$, and for any $\hat{\eta}$ in the estimator set, there exists $k \in \{1, \dots, K\}$ such that

$$|\hat{\eta} - \eta_k| \leq C_\epsilon \kappa_k^{-2} \log(n_{1:T}) n_{\max}^9 n_{\min}^{-10};$$

A3 if $\tau < c_{\tau,1} \log^{1/2}(n_{1:T})$, then the corresponding change point estimators satisfying $\widehat{K} > K$, and for any true change point η_k , there exists $\hat{\eta}$ in the estimators such that

$$|\hat{\eta} - \eta_k| \leq C_\epsilon \kappa_k^{-2} \log(n_{1:T}) n_{\max}^9 n_{\min}^{-10}.$$

The rest of the proof is conducted conditionally on the event \mathcal{A} .

Different τ_j 's may return the same collections of the change point estimators. For simplicity, in the rest of the proof, we assume that distinct candidate τ_j 's in (3.7) return distinct and nested \mathcal{B}_j with $|\mathcal{B}_j| = K_j$.

Step 1. Let $\hat{\eta}_0 = 0$ and $\hat{\eta}_{K+1} = T$. In this step, it suffices to show that for any $k \in \{0, \dots, K-1\}$, it holds that with large probability

$$\begin{aligned} & \sum_{l=k}^{k+1} \sum_{t=\hat{\eta}_l+1}^{\hat{\eta}_{l+1}} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \hat{z}\}} - \widehat{F}_{(\hat{\eta}_l+1):\hat{\eta}_{l+1}}^Y(\hat{z}) \right\}^2 + \lambda \\ & < \sum_{t=\hat{\eta}_k+1}^{\hat{\eta}_{k+2}} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \hat{z}\}} - \widehat{F}_{(\hat{\eta}_k+1):\hat{\eta}_{k+2}}^Y(\hat{z}) \right\}^2. \end{aligned} \tag{D.1}$$

Without loss of generality, we consider the case when $k = 0$.

With probability at least

$$1 - \frac{24 \log(n_{1:T})}{T^3 n_{1:T}} - \frac{48T}{n_{1:T} \log(n_{1:T}) \delta} - \exp \left\{ \log \left(\frac{T}{\delta} \right) - \frac{M\delta^2}{16T^2} \right\},$$

it holds that

$$\begin{aligned} & \sum_{t=1}^{\hat{\eta}_2} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \hat{z}\}} - \widehat{F}_{1:\hat{\eta}_2}(\hat{z}) \right\}^2 - \sum_{l=0}^1 \sum_{t=\hat{\eta}_l+1}^{\hat{\eta}_{l+1}} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \hat{z}\}} - \widehat{F}_{(\hat{\eta}_l+1):\hat{\eta}_{l+1}}(\hat{z}) \right\}^2 \\ & = \left\{ D_{1,\hat{\eta}_2}^{\hat{\eta}_1}(\{Y_{t,i}\}) \right\}^2 \geq c_{\tau,2}^2 \kappa^2 \delta \frac{n_{\min}^3}{n_{\max}^2}, \end{aligned} \tag{D.2}$$

where the last inequality follows from the proof of Theorem 3.1.

Therefore, for $\lambda = C \log(n_{1:T})$, (D.1) holds due to Condition 3.1.

Step 2. In this step, we are to show with large probability, Algorithm 3 will not over select. For simplicity, assume $\mathcal{B}_2 = \{\hat{\eta}_1\}$ and $\mathcal{B}_1 = \{\hat{\eta}, \hat{\eta}_1\}$ with $0 < \hat{\eta} < \hat{\eta}_1$. Let \hat{z} be the one defined in Algorithm 3 using the triplet $\{0, \hat{\eta}, \hat{\eta}_1\}$.

Since

$$\begin{aligned} & \sum_{t=1}^{\hat{\eta}_1} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \hat{z}\}} - \widehat{F}_{1:\hat{\eta}_1}^Y(\hat{z}) \right\}^2 - \sum_{t=1}^{\hat{\eta}} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \hat{z}\}} - \widehat{F}_{1:\hat{\eta}}^Y(\hat{z}) \right\}^2 \\ & - \sum_{t=\hat{\eta}+1}^{\hat{\eta}_1} \sum_{i=1}^{n_t} \left\{ 1_{\{Y_{t,i} \leq \hat{z}\}} - \widehat{F}_{(\hat{\eta}+1):\hat{\eta}_1}^Y(\hat{z}) \right\}^2 = \left\{ D_{0,\hat{\eta}_1}^{\hat{\eta}}(\{Y_{t,i}\}) \right\}^2 \leq c_{\tau,1}^2 \log(n_{1:T}) \end{aligned}$$

holds with probability at least

$$1 - \frac{24 \log(n_{1:T})}{T^3 n_{1:T}} - \frac{48T}{n_{1:T} \log(n_{1:T}) \delta} - \exp \left\{ \log \left(\frac{T}{\delta} \right) - \frac{S\delta^2}{16T^2} \right\}.$$

Therefore, for $\lambda = C \log(n_{1:T})$, (D.1) holds.

Combining both steps above and the fact that these two steps are conducted in the event \mathcal{A} , we have that

$$\begin{aligned} & \mathbb{P} \left\{ \widehat{K} = K \quad \text{and} \quad \epsilon_k \leq C_\epsilon \kappa_k^{-2} \log(n_{1:T}) n_{\max}^9 n_{\min}^{-10}, \forall k = 1, \dots, K \right\} \\ & \geq 1 - \frac{48 \log(n_{1:T})}{T^3 n_{1:T}} - \frac{96T}{n_{1:T} \log(n_{1:T}) \delta} - \exp \left\{ \log \left(\frac{T}{\delta} \right) - \frac{M\delta^2}{16T^2} \right\}. \quad \square \end{aligned}$$

Appendix E: Sensitivity simulations

We now explore the sensitivity of the tuning parameter C ($\lambda = C \log(n_{1:T})$) in Algorithm 3 and the tuning parameter τ in Algorithm 1. We choose $C, \tau \in \{0.1, 0.2, 0.3, \dots, 2.9, 3.0\}$ and for each algorithm calculate the corresponding number of change point for data generated under Scenarios 2 and 3 with $T = 1000$ and $n_t = 1$ for all t , in Section 4. Specifically, for each scenario we generate 50 data sets and report the median number of estimated change points for Algorithms 1 and 3 based on different choices of their tuning parameters.

The results in Figure 6 clearly show that Algorithm 3 is less sensitive to C than Algorithm 1 is to τ .

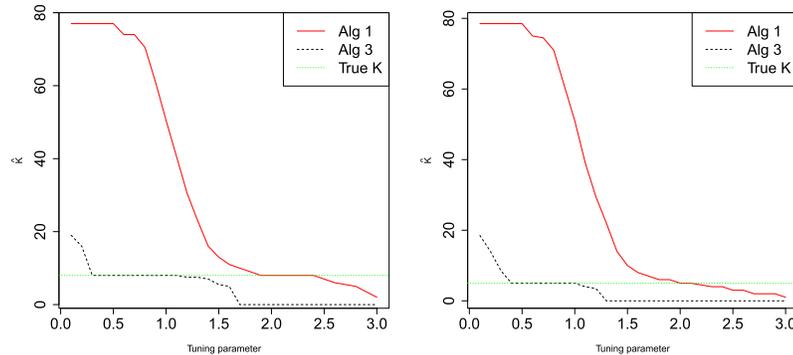


FIG 6. Sensitivity of the tuning parameters in Algorithms 1 and 3. The left panel shows the median of the estimated number of change points by Algorithms 1 and 3 under 50 Monte Carlo simulations based on Scenario 2 in Section 4. The right panel shows plot the corresponding for Scenario 3 in Section 4.

References

- ANASTASIOU, A. and FRYZLEWICZ, P. (2019). Detecting multiple generalized change-points by isolating single ones. *arXiv preprint 1901.10852*.
- ARLOT, S., CELISSE, A. and HARCHAOU, Z. (2019). A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research* **20** 1–56. [MR4048973](#)
- AUE, A., HÖMANN, S., HORVÁTH, L. and REIMHERR, M. (2009). Break detection in the covariance structure of multivariate nonlinear time series models. *The Annals of Statistics* **37** 4046–4087. [MR2572452](#)
- AVANESOV, V. and BUZUN, N. (2016). Change-point detection in high-dimensional covariance structure. *arXiv preprint 1610.03783*. [MR3861282](#)
- BAI, J. and PERRON, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics* **18** 1–22.
- BARANOWSKI, R., CHEN, Y. and FRYZLEWICZ, P. (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81** 649–672. [MR3961502](#)
- BARANOWSKI, R. and FRYZLEWICZ, P. (2019). wbs: Wild Binary Segmentation for Multiple Change-Point Detection R package version 1.4.
- BLEAKLEY, K. and VERT, J.-P. (2011). The group fused lasso for multiple change-point detection. *arXiv preprint 1106.4199*.
- BOUKAI, B. and ZHOU, H. (1997). Nonparametric estimation in a two change-point model. *Journal of Nonparametric Statistics* **8** 275–292. [MR1487005](#)
- CARLSTEIN, E. (1988). Nonparametric change-point estimation. *The Annals of Statistics* **16** 188–197. [MR0924865](#)
- CELISSE, A., MAROT, G., PIERRE-JEAN, M. and RIGAILL, G. (2018). New efficient algorithms for multiple change-point detection with reproducing kernels.

- Computational Statistics & Data Analysis* **128** 200–220. [MR3850633](#)
- CHAN, N. H., YAU, C. Y. and ZHANG, R.-M. (2014). Group LASSO for structural break time series. *Journal of the American Statistical Association* **109** 590–599. [MR3223735](#)
- CHAN, K.-S., LI, J., EICHINGER, W. and BAI, E.-W. (2014). A distribution-free test for anomalous gamma-ray spectra. *Radiation Measurements* **63** 18–25.
- CHO, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics* **10** 2000–2038. [MR3522667](#)
- CHO, H. and FRYZLEWICZ, P. (2015). Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77** 475–507. [MR3310536](#)
- CHOWDHURY, M. F. R., SELOUANI, S.-A. and O’SHAUGHNESSY, D. (2012). Bayesian on-line spectral change point detection: a soft computing approach for on-line ASR. *International Journal of Speech Technology* **15** 5–23.
- CLEYNEN, A., RIGAILL, G. and KOSKAS, M. (2016). Segmentor3IsBack: A Fast Segmentation Algorithm R package version 2.0.
- CRIBBEN, I. and YU, Y. (2017). Estimating whole-brain dynamics by using spectral clustering. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **66** 607–627. [MR3632344](#)
- DARKHOVSKI, B. S. (1994). Nonparametric methods in change-point problems: A general approach and some concrete algorithms. *Lecture Notes-Monograph Series* 99–107. [MR1477917](#)
- DUENBGEN, L. and WELLNER, J. A. (2014). Confidence bands for distribution functions: A new look at the law of the iterated logarithm. *arXiv preprint 1402.2918*.
- EICHINGER, B. and KIRCH, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **24** 526–564. [MR3706768](#)
- FAN, Z., DROR, R. O., MILDORF, T. J., PIANA, S. and SHAW, D. E. (2015). Identifying localized changes in large systems: Change-point detection for biomolecular simulations. *Proceedings of the National Academy of Sciences* **112** 7454–7459.
- FEARNHEAD, P. and RIGAILL, G. (2018). Change-point detection in the presence of outliers. *Journal of the American Statistical Association* 1–15. [MR3941246](#)
- FOX, E. B., SUDDERTH, E. B., JORDAN, M. I. and WILLSKY, A. S. (2011). A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics* **5** 1020–1056. [MR2840185](#)
- FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76** 495–580. [MR3210728](#)
- FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics* **42** 2243–2281. [MR3269979](#)
- GARREAU, D. and ARLOT, S. (2018). Consistent change-point detection with kernels. *Electronic Journal of Statistics* **12** 4440–4486. [MR3892345](#)
- HARCHAOUI, Z. and CAPPÉ, O. (2007). Retrospective multiple change-point es-

- timation with kernels. In *2007 IEEE/SP 14th Workshop on Statistical Signal Processing* 768–772. IEEE.
- HAWKINS, D. M. and DENG, Q. (2010). A nonparametric change-point control chart. *Journal of Quality Technology* **42** 165–173.
- HAYNES, K., FEARNHEAD, P. and ECKLEY, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing* **27** 1293–1305. [MR3647098](#)
- ITOH, N. and KURTHS, J. (2010). Change-point detection of climate time series by nonparametric method. In *Proceedings of the world congress on engineering and computer science* **1** 445–448. Citeseer.
- JEWELL, S., HOCKING, T. D., FEARNHEAD, P. and WITTEN, D. (2018). Fast nonconvex deconvolution of calcium imaging data. *arXiv preprint* [1802.07380](#). [MR4164053](#)
- KILLICK, R. and ECKLEY, I. A. (2014). changepoint: An R Package for Change-point Analysis. *Journal of Statistical Software* **58** 1–19.
- 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** 1590–1598. [MR3036418](#)
- KNUTH, D. (1998). Section 5.2. 4: Sorting by merging. *The Art of Computer Programming* **3** 158–168. [MR3077154](#)
- KOVÁCS, S., LI, H., BÜHLMANN, P. and MUNK, A. (2020). Seeded Binary Segmentation: A general methodology for fast and optimal change point detection. *arXiv preprint* [2002.06633](#).
- LI, S., XIE, Y., DAI, H. and SONG, L. (2019). Scan B-statistic for kernel change-point detection. *Sequential Analysis* **38** 503–544. [MR4057156](#)
- LIU, S., YAMADA, M., COLLIER, N. and SUGIYAMA, M. (2013). Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks* **43** 72–83.
- LIU, F., CHOI, D., XIE, L. and ROEDER, K. (2018). Global spectral clustering in dynamic networks. *Proceedings of the National Academy of Sciences* **115** 927–932. [MR3763702](#)
- MAROT, G., RIGAILL, G., PIERRE-JEAN, M. and BRUNIN, M. (2018). package KernSeg.
- MATTESON, D. S. and JAMES, N. A. (2013). ecp: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data Technical Report, Cornell University. [MR3180567](#)
- MATTESON, D. S. and JAMES, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association* **109** 334–345. [MR3180567](#)
- PADILLA, O. H. M., ATHEY, A., REINHART, A. and SCOTT, J. G. (2018). Sequential nonparametric tests for a change in distribution: an application to detecting radiological anomalies. *Journal of the American Statistical Association* 1–15. [MR3963159](#)
- PEIN, F., SIELING, H. and MUNK, A. (2017). Heterogeneous change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 1207–1227. [MR3689315](#)

- PEIN, F., HOTZ, T., SIELING, H. and ASPELMEIER, T. (2019). stepR: Multi-scale change-point inference R package version 2.0-4.
- PREUSS, P., PUCHSTEIN, R. and DETTE, H. (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association* **110** 654–668. [MR3367255](#)
- REINHART, A., ATHEY, A. and BIEGALSKI, S. (2014). Spatially-aware temporal anomaly mapping of gamma spectra. *IEEE Transactions on Nuclear Science* **61** 1284–1289.
- RIGAILL, G. (2010). Pruned dynamic programming for optimal multiple change-point detection. *arXiv preprint 1004.0887* **17**.
- RIZZO, M. L. and SZÉKELY, G. J. (2010). Disco analysis: A nonparametric extension of analysis of variance. *The Annals of Applied Statistics* **4** 1034–1055. [MR2758432](#)
- RUSSELL, B. and RAMBACCUSSING, D. (2018). Breaks and the statistical process of inflation: the case of estimating the ‘modern’ long-run Phillips curve. *Empirical Economics* 1–21.
- R CORE TEAM (2019). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.
- STEERNEMAN, T. (1983). On the total variation and Hellinger distance between signed measures; an application to product measures. *Proceedings of the American Mathematical Society* **88** 684–688. [MR0702299](#)
- TSYBAKOV, A. (2009). *Introduction to Nonparametric Estimation*. Springer. [MR2724359](#)
- VANEGAS, L. J., BEHR, M. and MUNK, A. (2019). Multiscale quantile regression. *arXiv preprint 1902.09321*.
- VENKATRAMAN, E. S. (1992). Consistency results in multiple change-point problems, PhD thesis, Stanford University. [MR2687536](#)
- WALD, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics* **16** 117–186. [MR0013275](#)
- WANG, T. and SAMWORTH, R. J. (2018). High-dimensional changepoint estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 57–83. [MR3744712](#)
- WANG, D., YU, Y. and RINALDO, A. (2018). Optimal change point detection and localization in sparse dynamic networks. *arXiv preprint 1809.09602*, *Annals of Statistics*, to appear. [MR4206675](#)
- WANG, D., YU, Y. and RINALDO, A. (2020). Univariate mean change point detection: Penalization, cusum and optimality. *Electronic Journal of Statistics* **14** 1917–1961. [MR4091859](#)
- WANG, D., YU, Y. and RINALDO, A. (2021). Optimal Covariance Change Point Detection in High Dimension. *Bernoulli* **27** 554–575. [MR4177380](#)
- YAO, Y. C. (1988). Estimating the number of change-points via Schwarz’ criterion. *Statistics & Probability Letters* **6** 181–189. [MR0919373](#)
- YAO, Y.-C. and AU, S.-T. (1989). Least-squares estimation of a stop function. *Sankhyā: The Indian Journal of Statistics, Series A* 370–381. [MR1175613](#)
- YAO, Y. C. and DAVIS, R. A. (1986). The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent

- normal variates. *Sankhyā: The Indian Journal of Statistics, Series A* 339–353. [MR0905446](#)
- YU, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer. [MR1462963](#)
- ZEILEIS, A., LEISCH, F., HORNIK, K. and KLEIBER, C. (2002). strucchange: An R Package for Testing for Structural Change in Linear Regression Models. *Journal of Statistical Software* **7** 1–38.
- ZOU, C., YIN, G., FENG, L. and WANG, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems. *The Annals of Statistics* **42** 970–1002. [MR3210993](#)