Supplementary document for "Two-stage data segmentation permitting multiscale change points, heavy tails and dependence"

Haeran Cho and Claudia Kirch

September 6, 2021

We provide an additional real data example in Section A, the proofs of the theoretical results in Sections B–D, and discuss a CUSUM-based candidate generating procedure in Section E. The discussion on the computational complexity of the localised pruning is provided in Section F, and Section G provides a complete description of the simulation results. Finally, Section H gives the pseudo-codes of `LocAlg` and `PrunAlg` introduced in Section 3 of the main text.

A Real data analysis: Kepler light curve data

Kepler light curve dataset contains regularly measured luminosity of stars. The transit of an orbiting planet results in periodically recurring segments of reduced luminosity, which can be used for detecting exoplanets via the transit method (Sartoretti and Schneider, 1999). Regarding segments of dimmed luminosity as collective anomalies, Fisch et al. (2018) apply their anomaly detection methodology to the light curve data obtained from Kepler-1132 (available in the R package `anomaly` (Fisch et al., 2018)), which is known to host at least one orbiting planet (Rein, 2018). In their paper, the data is pre-processed into equally sized bins aggregating the luminosity from different orbits using the known periodicity (62.89 days) of the orbiting planet. This amplifies the signal and transforms the irregularly sampled time series data into a regular one. From the aggregated data, they detect a short interval of collective anomalies over $[649, 660]$ (at the scale of bins).

We apply the proposed localised pruning to both raw Kepler-1132 data and its binned and aggregated version. For the former, we ignore the presence of missing observations, which yields $n = 51405$, and set the penalty at $\xi_n = \log^{1.1}(n)$ to account for possible outliers and heavy tails. For the binned and aggregated data of length $N = 3078$, we choose the penalty $\xi_N = \log^{1.01}(N)$ on the basis of its Gaussian-like tail behaviour.

Top panel of Figure 2 plots the raw data and the estimated change points. MoLP (with $\alpha = 0.2$ and $\eta = 0.4$) detects 14 estimators in total, while CuLP (with $C_\zeta = 0.5$) returns 16

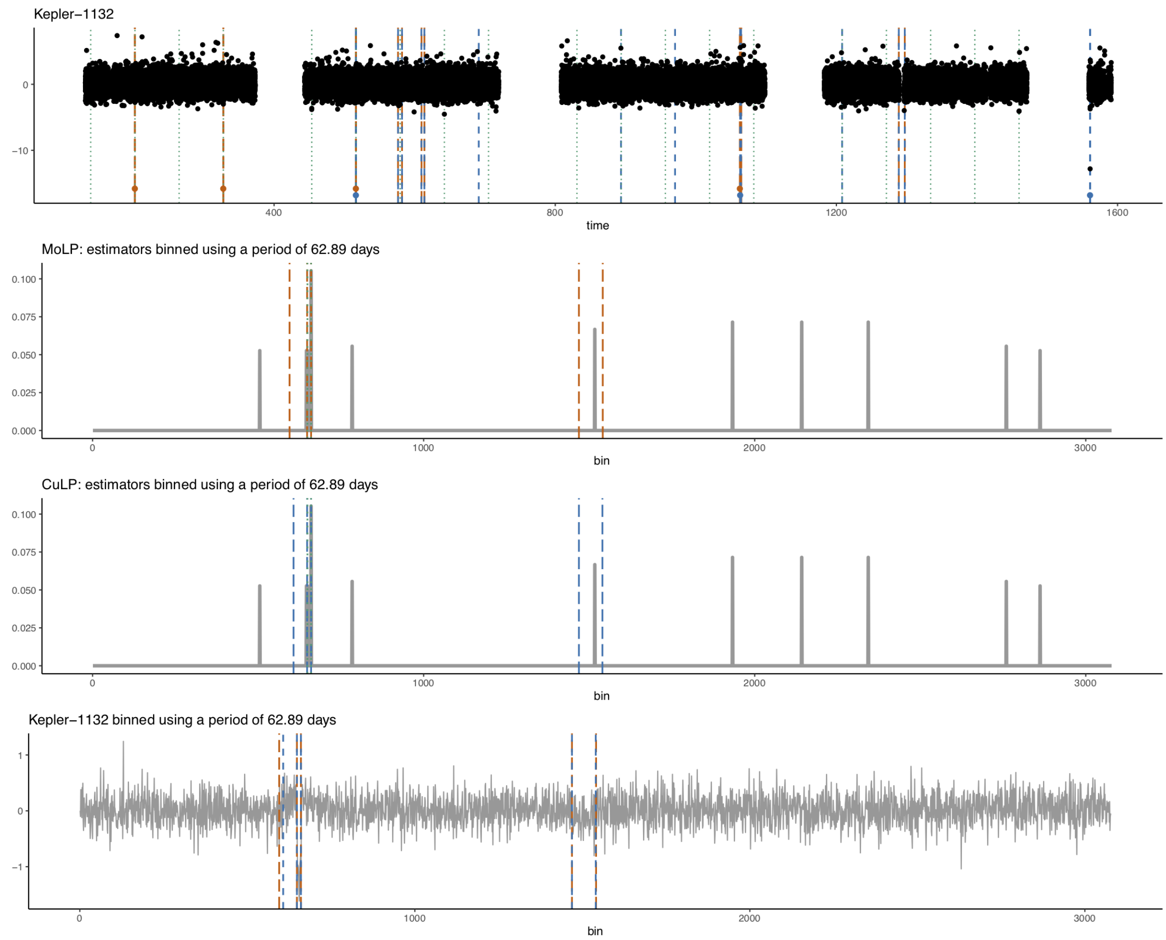

Figure 2: Top: Luminosity of Kepler-1132 measured every half an hour (approximately) with change point estimators (vertical lines; longdashed: MoLP, dashed: CuLP) and the beginnings of the anomalous intervals detected by Fisch et al. (2018) periodically repeated every 62.89 days (vertical dotted lines). Where two estimators returned by the same method lie too close to each other to be distinguished, a filled circle is added. Second, third: Change point estimators from the top panel binned using the periodicity of 62.89. Bottom: Kepler-1132 data binned and aggregated using the periodicity of 62.89 days. In the second, third and bottom panels, change point estimators from the aggregated data are also given as vertical lines (longdashed: MoLP, dashed: CuLP, dotted: Fisch et al. (2018)).

estimators, out of which there are 10 overlapping estimators in the sense that either they are identical or very close to one another. In analysing the raw data, we do not use the known periodicity to accumulate the information obtained from different orbits, nor do we utilise the knowledge that the changes are of epidemic nature as in Fisch et al. (2018). Nonetheless, both MoLP and CuLP identify the anomalous interval detected by Fisch et al. (2018) at some orbits (see the second and the third panels of Figure 2). Additionally detected change points may be attributed to the missingness in the data which is not accounted for by our methodology, particularly the pair in the vicinity of 1290 in the observation time scale. From the binned and aggregated data, both MoLP and CuLP detect 5 estimators including 648 and 660 (bottom panel of Figure 2), correctly identifying the anomalous segment reported in Fisch et al. (2018). In summary, our methodology is able to detect the periodic reduction in luminosity of Kepler-1132 without aggregating the signal using the extra information of periodicity which, in the problem of detecting exoplanets, may not be readily available.

B Proof of the result in Section 2

B.1 Proof of Proposition 1

We first prove assertion (a): By Hoeffding's inequality (see Theorem 2.6.3 of Vershynin (2018)), we have

$$\mathsf{P}\left(\max_{0\le s<e\le n}\frac{1}{\sqrt{e-s}}\left|\sum_{t=s+1}^{e}\varepsilon_t\right|\ge\omega_n\right)\le n(n+1)\exp\left(-c_\varepsilon\omega_n^2\right)$$

where c_ε is an absolute constant depending on the distribution of ε_t (via its Orlicz norm). Consequently, $\omega_n=\sqrt{3\log(n)/c_\varepsilon}$ fulfils Assumption 1. Next, define $S_\ell^\pm=\pm\sum_{t=1}^{\ell}\varepsilon_t$ and note that $\{\exp(S_\ell^\pm)\}$ is a non-negative sub-martingale. Then following the proof of Lemma 5 of Wang and Samworth (2018b), by Doob's martingale inequality and Proposition 2.5.2 (v) of Vershynin (2018), we get for some constant $c'_\varepsilon>0$ only depending on the distribution of ε_t (via its Orlicz norm) and any $0<l_n<u_n<\infty$,

$$\begin{aligned}
&\mathsf{P}\left(\max_{l_n\le\ell\le u_n}\frac{\sqrt{l_n}}{\ell}S_\ell^\pm\ge\omega_n^{(1)}\right)\le\sum_{i=\lfloor\log_2 l_n\rfloor}^{\lceil\log_2 u_n\rceil}\mathsf{P}\left(\max_{2^{i-1}\le\ell<2^i}S_\ell^\pm\ge\frac{2^{i-1}\omega_n^{(1)}}{\sqrt{l_n}}\right)\\
&=\sum_{i=\lfloor\log_2 l_n\rfloor}^{\lceil\log_2 u_n\rceil}\inf_{\lambda>0}\mathsf{P}\left(\max_{2^{i-1}\le\ell<2^i}e^{\lambda S_\ell^\pm}\ge\exp\left(\frac{\lambda 2^{i-1}\omega_n^{(1)}}{\sqrt{l_n}}\right)\right)\\
&\le\sum_{i=\lfloor\log_2 l_n\rfloor}^{\lceil\log_2 u_n\rceil}\inf_{\lambda>0}\exp\left(2^i c'_\varepsilon\lambda^2-\frac{\lambda 2^{i-1}\omega_n^{(1)}}{\sqrt{l_n}}\right)=\sum_{i=\lfloor\log_2 l_n\rfloor}^{\lceil\log_2 u_n\rceil}\exp\left(-\frac{2^{i+1}\,(\omega_n^{(1)})^2}{32\,c'_\varepsilon l_n}\right)\\
&\le\sum_{j=0}^{\lceil\log_2 u_n\rceil-\lfloor\log_2 l_n\rfloor}\exp\left(-\frac{2^j(\omega_n^{(1)})^2}{32\,c'_\varepsilon}\right)\le\sum_{j\ge0}(\max(q_n,\nu_n))^{-2^{j+1}}\le\frac{2}{(\max(q_n,\nu_n))^2}
\end{aligned}$$

for $\omega_n^{(1)}=\sqrt{64c'_\varepsilon\log(\max(q_n,\nu_n))}$ provided that $\max(q_n,\nu_n)\ge2$. Consequently,

$$\mathsf{P}\left(\max_{l_n\le\ell\le u_n}\frac{\sqrt{l_n}}{\ell}\left|\sum_{t=1}^{\ell}\varepsilon_t\right|\ge\omega_n^{(1)}\right)\le\frac{4}{(\max(q_n,\nu_n))^2}.$$

By sub-additivity and the i.i.d. assumption of the errors (distributional equality) and since $\nu_n\to\infty$ at an arbitrary rate, the above choice for $\omega_n^{(1)}$ fulfils Assumption 1. Furthermore, by

Lévy's reflection principle (see e.g. Theorem 3.1.11 of Giné and Nickl (2016)) and Hoeffding's inequality, it holds for some constant c''_ε and any $u_n > 1$,

$$\mathsf{P}\left(\max_{1\le \ell\le u_n} \frac{1}{\sqrt{u_n}} S_\ell^{\pm} \ge \omega_n^{(2)}\right) \le 2\,\mathsf{P}\left(\frac{1}{\sqrt{u_n}} S_{u_n}^{\pm} \ge \omega_n^{(2)}\right) \le 4\,\exp\left(-c''_\varepsilon(\omega_n^{(2)})^2\right).$$

Therefore, with $\omega_n^{(2)} = \sqrt{2\,\log(\max(q_n,\nu_n))/c''_\varepsilon}$, we have

$$\mathsf{P}\left(\max_{1\le \ell\le u_n} \frac{1}{\sqrt{u_n}}\left|\sum_{j=1}^{\ell}\varepsilon_t\right| \ge \omega_n^{(2)}\right) \le \frac{8}{(\max(q_n,\nu_n))^2},$$

such that the proof can be completed as for $\omega_n^{(1)}$, concluding the proof of (a).
The assertion in (c.i) follows directly from the invariance principle in addition to (a) ($\omega_n \asymp \sqrt{\log(n)}$ derived for the increments of the Wiener process). To prove (c.ii), let

$$M_n(j) = \max_{d_j^{-2}a_n \le \ell \le \theta_j-\theta_{j-1}} \frac{\sqrt{d_j^{-2}a_n}}{\ell}\left|\sum_{t=\theta_j-\ell+1}^{\theta_j}\varepsilon_t\right|.$$

Then, by Theorem B.3 in Kirch (2006), it holds uniformly in j that

$$\mathsf{E}|M_n(j)|^\gamma = O(1)\left(\frac{1}{(d_j^{-2}a_n)^{\gamma/2}}\sum_{i=1}^{d_j^{-2}a_n} i^{\gamma/2-1} + (d_j^{-2}a_n)^{\gamma/2}\sum_{i>d_j^{-2}a_n}\frac{1}{i^{\gamma/2+1}}\right) = O(1),$$

where $O(1)$ does not depend on j. From this and Markov's inequality, we yield

$$\mathsf{P}\left(\max_{1\le j\le q_n} M_n(j) \ge \omega_n^{(1)}\right) \le \frac{q_n\,\max_{1\le j\le q_n}\mathsf{E}\left(|M_n(j)|^\gamma\right)}{(\omega_n^{(1)})^\gamma} = O(1)\left(\frac{q_n^{1/\gamma}}{\omega_n^{(1)}}\right)^\gamma$$

such that the claim follows with $\omega_n^{(1)} \asymp q_n^{1/\gamma}\,\nu_n$. The assertion for $\omega_n^{(2)}$ follows analogously.
The assertion in (b) for ω_n follows directly from Theorem 1.1 of Mikosch and Račkauskas (2010). The assertion for $\omega_n^{(1)}$ and $\omega_n^{(2)}$ follows analogously as in the proof of (c.ii): The moments $\mathsf{E}(\varepsilon_t^{\beta'})$ and $\mathsf{E}(\varepsilon_t^{\beta})$ for all $\beta < \beta' < \alpha$ exist and independent and centred sequences fulfil the moment condition in (c.ii), see e.g., Theorem 3.7.8 of Stout (1974). By (c.ii), it gives $\omega_n^{(1)} \asymp q_n^{1/\beta'}\,\nu_n$ so that the given choice $\omega_n^{(1)} \asymp \max(q_n^{1/\beta},\nu_n)$ is also valid.

C Proof of the result in Section 3

In this section, we provide the proofs of Theorems 1–2 which establish the consistency of the localised pruning algorithm combining `LocAlg` and `PrunAlg`. They are based on Propositions C.1–C.3, whose proofs can be found in Section C.4. Throughout, we assume that Assumptions 1 and 4 (a) (and Assumption 5 for Theorem 2) hold. In addition, we work under the following non-asymptotic bound:

$$\max\left(\frac{\omega_n^{(1)}}{\sqrt{\nu_n\,\rho_n}}, \frac{\omega_n^{(2)}}{\nu_n\sqrt{\rho_n}}, \frac{Q_n\omega_n^2}{n}, \frac{\xi_n}{D_n}, \frac{\rho_n\nu_n}{\xi_n}, \frac{\omega_n^2}{\xi_n}, \frac{1}{\nu_n}\right) \le \frac{1}{M} \tag{C.1}$$

for some $M > 0$, which holds for all $n \ge n(M)$ for some large enough $n(M)$. This replaces the asymptotic conditions in Assumptions 2, 3 and 4. Here, we regard ρ_n as the precision originally attained by a candidate generating mechanism. If $\max(\omega_n^{(1)}, \omega_n^{(2)})^2 = O(\rho_n)$ as in Assumption 4 (a), (C.1) is fulfilled by $\nu_n \to \infty$ arbitrarily slowly as stated in the theorem. If not, the assertions still hold for any ν_n fulfilling the above. Also, when $\rho_n = O(\omega_n^2)$ is not met, the assertions continue to hold but with a penalty parameter greater than the acceptable precision, which is reflected in (C.1). In the proofs of Propositions C.1–C.3, we state the precise requirement on the ratios in the LHS of (C.1) each instance they appear; while this allows to make a tighter bound on each term with which non-asymptotic results are readily derived, we omit such a detailed analysis here and simply state that the assertion in (C.1) holds for n large enough.

We write $\mathrm{SC}(\mathcal{A}) = \mathrm{SC}(\mathcal{A}|\mathcal{C}, \widehat{\Theta}, s, e)$ where there is no confusion since, for given s and e, the difference between $\mathrm{SC}(\mathcal{A}|\mathcal{C}, \widehat{\Theta}, s, e)$ and $\mathrm{SC}(\mathcal{A}'|\mathcal{C}, \widehat{\Theta}, s, e)$ does not depend on candidates outside (s, e) for any $\mathcal{A}, \mathcal{A}' \subset \mathcal{C} \cap (s, e)$. For a change point currently under consideration, say $\theta_\circ$, we write its neighbouring change points as $\theta_\pm$ (i.e. $\Theta \cap (\theta_-, \theta_+) = \{\theta_\circ\}$) allowing for $\theta_- = 0$ and $\theta_+ = n$, and denote the associated jump sizes by $d_\circ$ and $d_\pm$, respectively.

For any given interval $(s, e]$, Proposition C.1 establishes the sure detectability of any change point in $\Theta^{(s,e)}$ as defined in (6), as well as the *un*detectability of any change point not belonging to $\bar{\Theta}^{(s,e)}$ as defined in (7).

Proposition C.1. For any $0 \le s < e \le n$ (with $\Theta \cap (s, e) \ne \emptyset$) and $\theta_\circ \in \Theta \cap (s, e)$, let $\mathcal{A} \subset \mathcal{D} = \mathcal{K} \cap (s, e)$ denote a set of candidate estimators where $k_\pm \in \mathcal{A} \cup \{s, e\}$ satisfy $\theta_\circ \in (k_-, k_+)$ as well as $\mathcal{A} \cap (k_-, k_+) = \emptyset$. Then, there exist universal constants $c^*, C^* \in (0, \infty)$ with $c^* < C^*$, with which the following statements hold on $\mathcal{M}_n$ for n large enough: Let

$$\max\left\{d_+^2(k_+ - \theta_+) \cdot \mathbb{I}_{k_+ \ge \theta_+}, d_-^2(\theta_- - k_-) \cdot \mathbb{I}_{k_- \le \theta_-}\right\} \le C^*\xi_n.$$

(a) If $d_\circ^2 \min(\theta_\circ - k_-, k_+ - \theta_\circ) \ge C^*\xi_n$, we have $\mathrm{SC}(\mathcal{A}) > \mathrm{SC}(\mathcal{A} \cup \{k'_\circ\})$ for all $k'_\circ \in \mathcal{V}'_\circ$.

(b) Suppose $\theta_- < k_-$ and $d_\circ^2(\theta_\circ - k_-) < c^*\xi_n$. Then, if either $\theta_+ > k_+$ or $|k - \theta_\circ| < (\theta_+ - k)$, we have $\mathrm{SC}(\mathcal{A}) < \mathrm{SC}(\mathcal{A} \cup \{k\})$.

(c) Suppose $k_+ < \theta_+$ and $d_\circ^2(k_+ - \theta_\circ) < c^*\xi_n$. Then, if either $k_- > \theta_-$ or $|k - \theta_\circ| < (k - \theta_-)$, we have $\mathrm{SC}(\mathcal{A}) < \mathrm{SC}(\mathcal{A} \cup \{k\})$.

Throughout, for any $k_\pm, k_\circ \in \mathcal{K} \cup \{0, n\}$ with $k_- < k_\circ < k_+$, we refer to $k_\circ$ as detecting $\theta_\circ \in \Theta \cap (k_-, k_+]$ if $\theta_\circ = \arg\min_{\theta \in \Theta \cap (k_-, k_+]} |k_\circ - \theta|$, i.e. its nearest change point within $(k_-, k_+]$ is $\theta_\circ$, even though there may be some $\theta_j \notin (k_-, k_+]$ closer to $k_\circ$ than $\theta_\circ$.
Proposition C.2 states that when a given set $\mathcal{A}$ already contains an acceptable candidate for a change point in a local environment, SC increases if another candidate detecting the same change point is added to $\mathcal{A}$, as well as that adding spurious candidates increases SC.

Proposition C.2. For any $0 \le s < e \le n$ and some $k_\circ \in \mathcal{D} = \mathcal{K} \cap (s, e)$, let $\mathcal{A} \subset \mathcal{D} \setminus \{k_\circ\}$ with $k_\pm \in \mathcal{A} \cup \{s, e\}$ chosen such that $k_- < k_\circ < k_+$ and $(k_-, k_+) \cap \mathcal{A} = \emptyset$. Further, we suppose that $k_\pm$ satisfy

(a) $\Theta \cap (k_-, k_+] = \emptyset$, or

(b) if $\Theta \cap (k_-, k_+] \neq \emptyset$, then for any $\theta_j \in \Theta \cap (k_-, k_+]$, we have $d_j^2 \min(\theta_j - k_-, k_+ - \theta_j) \le C^* \xi_n$. Additionally, for $\theta_\circ \in \Theta \cap (k_-, k_+]$ detected by $k_\circ$, either

  (i) at least one of $k_\pm$ is acceptable, i.e. $d_\circ^2 \min(\theta_\circ - k_-, k_+ - \theta_\circ) \le \rho_n \nu_n$, or

  (ii) $d_\circ^2 |k_\circ - \theta_\circ| > \widetilde{C}\, \xi_n$ for $\widetilde{C} > \max(C^*, \bar{C}(C^*)^2)$ with $\bar{C}$ as defined in Lemma C.4.

Then, adding $k_\circ$ to $\mathcal{A}$ yields an increase of SC, i.e. for n large enough,

$$\mathrm{SC}(\mathcal{A}) < \mathrm{SC}(\mathcal{A} \cup \{k_\circ\}) \quad \text{on} \quad \mathcal{M}_n.$$

The next proposition asserts that a set containing an unacceptable candidate yields larger SC than the one replacing it with a strictly valid estimator, when the corresponding change point is detectable in the interval of consideration.

Proposition C.3. For any $0 \le s < e \le n$ (with $\bar{\Theta}^{(s,e)} \neq \emptyset$) and $\theta_\circ \in \bar{\Theta}^{(s,e)}$, let $\mathcal{A} \subset \mathcal{D} = \mathcal{K} \cap (s, e)$ be any candidate subset with $k_\pm \in \mathcal{A} \cup \{s, e\}$ satisfying $\theta_\circ \in (k_-, k_+)$, $\mathcal{A} \cap (k_-, k_+) = \emptyset$, $d_\circ^2 |k_\pm - \theta_\circ| \ge c^* \xi_n$, as well as

$$\max\left\{d_+^2 (k_+ - \theta_+) \cdot \mathbb{I}_{k_+ \ge \theta_+},\, d_-^2 (\theta_- - k_-) \cdot \mathbb{I}_{k_- \le \theta_-}\right\} \le C^* \xi_n.$$

Denote by $k_\circ^* \in \mathcal{V}_\circ^*$ a strictly valid estimator for $\theta_\circ$, and by $k_\circ$ an estimator detecting $\theta_\circ$ within $(k_-, k_+]$ which satisfies $d_\circ^2 |k_\circ - \theta_\circ| \le \widetilde{C} \xi_n$ with $\widetilde{C}$ as in Proposition C.2, while being unacceptable for $\theta_\circ$. Then, adding $k_\circ^*$ to $\mathcal{A}$ yields a greater reduction in the RSS than adding $k_\circ$, i.e. for n large enough,

$$\mathrm{SC}(\mathcal{A} \cup \{k_\circ\}) > \mathrm{SC}(\mathcal{A} \cup \{k_\circ^*\}) \quad \text{on} \quad \mathcal{M}_n.$$

C.1 Proof of Theorem 1

On $\mathcal{M}_n$, the following arguments hold uniformly in $0 \le s < e \le n$ and the corresponding $\mathcal{D} = \mathcal{K} \cap (s, e)$ for n large enough. First, we note that

(D1) any set $\mathcal{A} \subset \mathcal{D}$ fulfilling (C1) contains at least one estimator satisfying $d_j^2 \min_{k \in \mathcal{A}} |k - \theta_j| \le C^* \xi_n$ for all $\theta_j \in \Theta^{(s,e)}$.

We prove (D1) by contradiction. Suppose that for some $\theta_\circ \in \Theta^{(s,e)}$, the set $\mathcal{A}$ does not contain any candidate within its $(C^* d_\circ^{-2} \xi_n)$-environment. To such $\mathcal{A}$, we can add, if necessary, strictly valid candidates until the resultant set contains one strictly valid candidate for each $\theta_j \in \Theta \cap (s, e) \setminus \{\theta_\circ\}$. Then, the conditions of Proposition C.1 (a) are met, and adding any $k'_\circ \in \mathcal{V}'_\circ$ to such a set results in a decrease of SC.
Also, we can always find a subset of $\mathcal{D}$ that fulfils (C1), since

(D2) any $\mathcal{A} \subset \mathcal{D}$ containing exactly one acceptable estimator for all $\bar{\Theta}^{(s,e)}$ with $|\mathcal{A}| = |\bar{\Theta}^{(s,e)}|$ satisfies (C1).

To see this, adding candidates detecting $\theta_j \in \bar{\Theta}^{(s,e)}$ to $\mathcal{A}$ incurs monotonic increase of SC by Proposition C.2 since in each step, either (a) or (b.i) therein is fulfilled for any candidates $k_\circ \in \mathcal{D} \setminus \mathcal{A}$ (since $\rho_n \nu_n < C^* \xi_n$ under (C.1) for n large enough). Similarly, when adding those detecting $\theta_j \in \Theta \cap (s, e) \setminus \bar{\Theta}^{(s,e)}$ to $\mathcal{A}$, Proposition C.1 (b)–(c) applies.
Denoting by $\mathcal{F}_{[m]}$ the collection of the subsets of $\mathcal{D}$ of cardinality m that fulfil (C1). By (D1), we have $|\mathcal{F}_{[m]}| = 0$ for $m < |\Theta^{(s,e)}|$. Also, defining $m^* = \min\{1 \le m \le |\mathcal{D}| : |\mathcal{F}_{[m]}| \ne \emptyset\}$, we have $m^* \le |\bar{\Theta}^{(s,e)}| \le |\Theta^{(s,e)}| + 2 \le m^* + 2$ by (D2). Suppose now that there exists $\mathcal{A} \in \bigcup_{m^* \le m \le m^*+2} \mathcal{F}_{[m]}$ for which

(a) $|\mathcal{A} \cap \mathcal{V}'_j| \ne 1$ for $\theta_j \in \Theta^{(s,e)}$, or

(b) $|\mathcal{A} \cap \mathcal{V}'_j| > 1$ for $\theta_j \in \bar{\Theta}^{(s,e)} \setminus \Theta^{(s,e)}$, or

(c) $\mathcal{A} \setminus \bigcup_{j:\, \theta_j \in \bar{\Theta}^{(s,e)}} \mathcal{V}'_j \ne \emptyset$.

We show that such a set $\mathcal{A}$ cannot be returned by (C2). To this end, we apply the following operations to $\mathcal{A}$. Because the set changes after each operation, we denote the active set by $\mathcal{A}'$ in the following which is initially set as $\mathcal{A}' = \mathcal{A}$.

Step 1: If $\mathcal{A}'$ contains any estimator of $\Theta \cap (s, e) \setminus \bar{\Theta}^{(s,e)}$, iteratively remove such estimators from $\mathcal{A}'$ one at a time which, by Proposition C.1 (b)–(c) and (D1), strictly reduces the SC monotonically. Also remove any estimator $k_\circ \in \mathcal{A}'$ one at a time which is too far from its nearest change point, say $\theta_\circ$, in the sense that $d_\circ^2 |k_\circ - \theta_\circ| > \widetilde{C} \xi_n$; this strictly reduces the SC by Proposition C.2 (a), (b.ii) and (D1).

Step 2: If $\mathcal{A}' \cap \mathcal{V}'_\circ = \emptyset$ for some $\theta_\circ \in \Theta^{(s,e)}$, by (D1), we have at least one $k_\circ \in \mathcal{A}'$ satisfying $d_\circ^2 |k_\circ - \theta_\circ| \le C^* \xi_n$. Let $k_\circ$ be the closest estimator of $\theta_\circ$ in $\mathcal{A}'$ and identify $k_\pm \in \mathcal{A} \cup \{s, e\}$ such that $(k_-, k_+) \cap \mathcal{A} = \{k_\circ\}$. When $d_\circ^2 \min(\theta_\circ - k_-, k_+ - \theta_\circ) < c^* \xi_n$, we can remove one of $k_\pm$ closer to $\theta_\circ$ while decreasing the SC. To see this, suppose without loss of generality (otherwise consider the time series in reverse) that this is k_+. Then, $k_+ > \theta_\circ$ since $k_\circ$ is the estimator closest to $\theta_\circ$ in $\mathcal{A}'$. Denote by $\tilde{k}_\circ = k_+$ and define $\tilde{k}_\pm$ analogously as $k_\pm$ with regards to $\tilde{k}_\circ$ (such that $\tilde{k}_- = k_\circ$), and let $\tilde{d}_\circ$ denote the jump size associated with a change point $\tilde{\theta}_\circ$. Then, one of the followings applies.

- Conditions of Proposition C.1 (b) are met by $\tilde{k}_\pm$ if $k_\circ = \tilde{k}_- \le \theta_\circ = \tilde{\theta}_\circ < k_+ = \tilde{k}_\circ$.
- Conditions of Proposition C.2 (a) are met by $\tilde{k}_\pm$ if $\theta_\circ < \tilde{k}_- < \tilde{k}_\circ < \tilde{k}_+ \le \theta_+$.
- Conditions of Proposition C.2 (b.ii) hold for $\tilde{k}_\pm$ and $\tilde{\theta}_\circ$ if $\theta_\circ < \tilde{k}_- < \tilde{k}_\circ < \theta_+ = \tilde{\theta}_\circ < \tilde{k}_+$, since in this case, $\tilde{d}_\circ^2(\tilde{\theta}_\circ - \tilde{k}_\circ) = \tilde{d}_\circ^2(\tilde{\theta}_\circ - \theta_\circ)\{1 - (\tilde{k}_\circ - \theta_\circ)/(\tilde{\theta}_\circ - \theta_\circ)\} \ge D_n - c^*\xi_n > \widetilde{C}\xi_n$ for n large enough.

In all cases, removing $\tilde{k}_\circ = k_+$ results in a decrease of SC. Iteratively repeat the removal and re-defining of $k_\circ$ and $k_\pm$ until $d_\circ^2 \min(\theta_\circ - k_-, k_+ - \theta_\circ) \ge c^*\xi_n$. Then, the resultant $\mathcal{A}'$ and $k_\circ$ are such that $\mathcal{A}' \setminus \{k_\circ\}$ meets the conditions of Proposition C.3 for $\theta_\circ$. Therefore, replacing $k_\circ$ with any of $k_\circ^* \in \mathcal{V}_\circ^*$ yields a reduction in the SC. Repeat the above until $|\mathcal{A}' \cap \mathcal{V}_j'| = 1$ for all $\theta_j \in \Theta^{(s,e)}$, which strictly decreases $\text{SC}(\mathcal{A}')$ monotonically.

Step 3: If $\mathcal{A}' \cap \mathcal{V}_j' = \emptyset$ for some $\theta_j \in \bar{\Theta}^{(s,e)} \setminus \Theta^{(s,e)}$ yet $\mathcal{A}'$ contains an estimator of θ_j, we take the same steps as in Step 2 for all such θ_j so that $|\mathcal{A}' \cap \mathcal{V}_j'| = 1$, which strictly decreases $\text{SC}(\mathcal{A}')$ monotonically.

Step 4: If there exists $\theta_j \in \bar{\Theta}^{(s,e)}$ for which there are more than one estimator in $\mathcal{A}'$, through Steps 2–3, we have $\mathcal{A}' \cap \mathcal{V}_j' \ne \emptyset$. Remove the duplicate estimators one at a time until all θ_j with $\mathcal{A}' \cap \mathcal{V}_j' \ne \emptyset$ have exactly one acceptable estimator in $\mathcal{A}'$ which, by Proposition C.1 (b)–(c) or by Proposition C.2 (a) and (b.i), results in a strictly monotonic reduction of SC.

After Steps 1–4, we have $\mathcal{A}'$ that satisfies $\mathcal{A}' \setminus \bigcup_{j:\,\theta_j \in \bar{\Theta}^{(s,e)}} \mathcal{V}_j' = \emptyset$, with $|\mathcal{A}' \cap \mathcal{V}_j'| = 1$ for $\theta_j \in \Theta^{(s,e)}$ and $|\mathcal{A}' \cap \mathcal{V}_j'| \le 1$ for $\theta_j \in \bar{\Theta}^{(s,e)} \setminus \Theta^{(s,e)}$, as well as $\text{SC}(\mathcal{A}') < \text{SC}(\mathcal{A})$ because under (a)–(c), at least one of Steps 1–4 above has to take place. Further, if necessary, by adding strictly valid candidates to $\mathcal{A}'$ for all those $\theta_j \in \bar{\Theta}^{(s,e)} \setminus \Theta^{(s,e)}$ with $|\mathcal{A}' \cap \mathcal{V}_j'| = 0$, we yield $\mathcal{A}'' \supset \mathcal{A}'$ fulfilling (C1) by (D2) and of cardinality $|\bar{\Theta}^{(s,e)}|$, i.e. $\mathcal{A}'' \in \bigcup_{m^* \le m \le m^*+2} \mathcal{F}_{[m]}$. Since $\mathcal{A}' \subset_R \mathcal{A}''$ with $\subset_R$ defined below (C2) and $\text{SC}(\mathcal{A}') < \text{SC}(\mathcal{A})$, this shows that $\mathcal{A}$ with candidates belonging to either of (a)–(c) cannot be returned in (C2). In conclusion, $\widehat{\Theta}^{(s,e)}$ obtained from (C2) satisfies the assertion of the theorem.

C.2 Proof of Theorem 2

Under (C.1), we make the following observations: For all $j = 1, \ldots, q_n$,

(a) $d_j^2|\widehat{\theta}_j - \theta_j| \le \rho_n\nu_n < c^*\xi_n$ for any $\widehat{\theta}_j \in \mathcal{V}_j'$, and

(b) $d_j^2 \min(\theta_j - \theta_{j-1}, \theta_{j+1} - \theta_j) \ge D_n > 2\max(C^*\xi_n, \rho_n\nu_n)$

for n large enough.

In iteratively applying Steps 1–4 of `LocAlg`, Theorem 1 guarantees that $\widehat{\Theta}$ contains only acceptable estimators of $\theta_j \in \Theta$. Also, each change point can belong to $\Theta^{(s,e)}$ defined by the interval of consideration $(s, e] = (k_L, k_R]$ at most once: When $\theta_j \in \Theta^{(s,e)}$ for the first time, it gets detected by some $\widehat{\theta}_j \in \mathcal{V}'_j$ by Theorem 1. Then, in the following iterations, either $\theta_j \notin (s, e)$, or some $k \in (\mathcal{C} \cup \widehat{\Theta}) \cap [\min(\theta_j, \widehat{\theta}_j), \max(\theta_j, \widehat{\theta}_j)]$ defines the endpoints of the local environment by Step 2. In the latter case, θ_j cannot be a detectable change point within the interval of consideration of this particular iteration due to (a), which guarantees that no further estimator for θ_j is added to $\widehat{\Theta}$.

When there exists $\theta_j \in \bar{\Theta}^{(s,e)} \setminus \Theta^{(s,e)}$ at some iteration, Theorem 1 indicates that it may or may not get detected at this iteration. If it does, an acceptable estimator of θ_j is added to $\widehat{\Theta}$ and the same argument as above applies. If not, without loss of generality, suppose $\theta_j - s \le e - \theta_j$. By construction, $c^*\xi_n \le d_j^2(\theta_j - s) < C^*\xi_n$ and thus from (b), we have

$$d_{j-1}^2(s - \theta_{j-1}) = d_{j-1}^2(\theta_j - \theta_{j-1}) \left\{ 1 - \frac{d_j^2(\theta_j - s)}{d_j^2(\theta_j - \theta_{j-1})} \right\} \ge D_n - C^*\xi_n > \rho_n \nu_n,$$

i.e. the boundary point s cannot be an acceptable estimator for either θ_{j-1} or θ_j. Consequently, it cannot have already been added to $\widehat{\Theta}$ in the previous iterations by Theorem 1. Therefore, all acceptable estimators for θ_j, with the possible exception of $k_\circ$ identified in Step 1, remain in $\mathcal{C}$ by (a)–(b) and how it is reduced in Step 4 of `LocAlg`.

Next, we justify the removal of $k_\circ$ from $\mathcal{C}$ at each iteration. Clearly, if $k_\circ$ is not acceptable for any change point, it can be safely removed from the future consideration. Next, suppose that $k_\circ$ is an acceptable estimator of θ_j and $\theta_j - s \le e - \theta_j$.

(a) When $\theta_j \in \Theta^{(s,e)}$, we have either $k_\circ$ or another acceptable estimator of θ_j accepted by `PrunAlg`, and therefore $k_\circ$ can be removed.

(b) When $\theta_j \in \bar{\Theta}^{(s,e)} \setminus \Theta^{(s,e)}$, if θ_j is detected at the current iteration, the same argument as in (a) applies. If not, as shown above, s has not been added to $\widehat{\Theta}$ yet and by construction of the interval of consideration in Step 2, it follows that

$$d_j^2 G_L(k_\circ) = d_j^2(k_\circ - s) \le C^*\xi_n + \rho_n \nu_n = C^*\xi_n(1 + o(1)),$$

which shows that $k_\circ$ cannot fulfil (8) for θ_j (nor any other change point as it is acceptable for θ_j). Consequently, $k_\circ$ can safely be removed from $\mathcal{C}$ since by Assumption 5 and the construction of $\mathcal{R}$ in Step 4, there remains at least one acceptable estimator for θ_j that fulfils (8) in $\mathcal{C}$ after the current iteration.

(c) When $\theta_j \notin \bar{\Theta}^{(s,e)}$ (which is not necessarily situated within (s, e)), we first consider the case where s has already been accepted. Then by Theorem 1, s is acceptable for some change point, say $\theta_{j'}$, such that

$$
\begin{aligned}
&d^2_{j'+1}(\theta_{j'+1}-s) = d^2_{j'+1}(\theta_{j'+1}-\theta_{j'})\left\{1-\frac{d^2_{j'}(s-\theta_{j'})}{d^2_{j'}(\theta_{j'+1}-\theta_{j'})}\right\}\\
&\geq D_n\left(1-\frac{\rho_n\nu_n}{D_n}\right) > C^*\xi_n,
\end{aligned}
$$

i.e. $\theta_{j'+1}$ is either surely detectable within (s,e), too close to e, or $\theta_{j'+1}\notin(s,e)$ to have been detected by $k_\circ$. Therefore, $j=j'$ and $k_\circ$ can safely be removed as in (a) since there already exists an acceptable estimator s in $\widehat{\Theta}$. If s has not been accepted, the argument analogous to that in (b) applies.

The case when $\theta_j - s > e - \theta_j$ is similarly handled.
The above (b)–(c) show that under Assumptions 4 and 5, for each $j=1,\ldots,q_n$, acceptable estimators of θ_j remain in $\mathcal{C}$ until its detection and at least one of them, when set as $k_\circ$ in Step 1 of `LocAlg`, leads θ_j to belong to $\Theta^{(s,e)}$ at some iteration, from which we conclude that all $\theta_j\in\Theta$ are eventually detected by acceptable estimators. Finally, $|\mathcal{R}|\geq 1$ at all iterations since $\mathcal{R}$ contains $k_\circ$ at least, which ensures that `LocAlg` terminates eventually.

C.3 Auxiliary lemmas

In this section, we list some auxiliary lemmas that will be used in the proof of Theorem 1. Unless stated otherwise, we assume that the conditions made in Theorem 1 are met throughout.
Recall the definition of the CUSUM statistic computed on X_t as

$$
\mathcal{X}_{k_-,k_\circ,k_+} \equiv \mathcal{X}_{k_-,k_\circ,k_+}(X) := \sqrt{\frac{k_+-k_-}{(k_+-k_\circ)(k_\circ-k_-)}}\sum_{t=k_-+1}^{k_\circ}(X_t-\bar{X}_{(k_-+1):k_+})
$$

for any $1\leq k_-+1\leq k_\circ<k_+\leq n$, and analogously define $\mathcal{X}_{k_-,k_\circ,k_+}(f)$ and $\mathcal{X}_{k_-,k_\circ,k_+}(\varepsilon)$ with f_t and ε_t in place of X_t, respectively. Also, we use the notation $\bar{f}_{u:v}=(v-u+1)^{-1}\sum_{t=u}^v f_t$ for any $1\leq u\leq v\leq n$ and $\bar{\varepsilon}_{u:v}$ is defined analogously.

Lemma C.1. For $\max(k_-,\theta_-)<k<\theta_\circ<\min(k_+,\theta_+)$, it holds with $r_+ := \max(0,k_+-\theta_+)$ and $r_- := \max(0,\theta_- - k_-)$,

$$
\mathcal{F}_k = \sum_{t=k_-+1}^{k}(f_t-\bar{f}_{(k_-+1):k_+}) = -\frac{(k-k_-)\,(k_+-\theta_\circ)}{k_+-k_-}\,d_\circ - \frac{k-k_-}{k_+-k_-}\,d_+r_+ - \frac{k_+-k}{k_+-k_-}\,d_-r_-
$$

as well as

$$
\mathcal{X}_{k_-,k,k_+}(f) = -\sqrt{\frac{(k-k_-)(k_+-k)}{k_+-k_-}}\left(\frac{(k_+-\theta_\circ)\,d_\circ}{k_+-k}+\frac{r_+\,d_+}{k_+-k}+\frac{r_-\,d_-}{k-k_-}\right).
$$

Similarly, for $\max(k_-, \theta_-) < \theta_\circ \le k < \min(c_+, \theta_+)$, it holds

$$\mathcal{F}_k = -\frac{(k_+ - k)\,(\theta_\circ - k_-)}{k_+ - k_-}\, d_\circ - \frac{k - k_-}{k_+ - k_-}\, d_+ r_+ - \frac{k_+ - k}{k_+ - k_-}\, d_- r_-$$

as well as

$$\mathcal{X}_{k_-,k,k_+}(f) = -\sqrt{\frac{(k - k_-)(k_+ - k)}{k_+ - k_-}} \left(\frac{(\theta_\circ - k_-)\, d_\circ}{k - k_-} + \frac{r_+\, d_+}{k_+ - k} + \frac{r_-\, d_-}{k - k_-} \right).$$

Proof. The results follow from straightforward calculations. □

Lemma C.2. For an arbitrary set $\mathcal{A} \subset \mathcal{K}$ and $k_\circ \in \mathcal{A}$, let $k_\pm \in \mathcal{A} \cup \{0, n\}$ satisfy $k_- < k_\circ < k_+$ with $\mathcal{A} \cap (k_-, k_+) = \emptyset$. Then,

$$\mathrm{RSS}(\mathcal{A} \setminus \{k_\circ\}) - \mathrm{RSS}(\mathcal{A}) = |\mathcal{X}_{k_-,k_\circ,k_+}|^2.$$

Proof.

$$\begin{aligned}
&\mathrm{RSS}(\mathcal{A} \setminus \{k_\circ\}) - \mathrm{RSS}(\mathcal{A}) \\
=& \sum_{t=k_-+1}^{k_+} (X_t - \bar{X}_{(k_-+1):k_+})^2 - \left\{ \sum_{t=k_-+1}^{k_\circ} (X_t - \bar{X}_{(k_-+1):k_\circ})^2 + \sum_{t=k_\circ+1}^{k_+} (X_t - \bar{X}_{(k_\circ+1):k_+})^2 \right\} \\
=&(k_\circ - k_-)\bar{X}^2_{(k_-+1):k_\circ} + (k_+ - k_\circ)\bar{X}^2_{(k_\circ+1):k_+} \\
&\quad - \frac{1}{k_+ - k_-} \left\{ (k_\circ - k_-)\bar{X}_{(k_-+1):k_\circ} + (k_+ - k_\circ)\bar{X}_{(k_\circ+1):k_+} \right\}^2 \\
=& \left\{ \sqrt{\frac{(k_\circ - k_-)(k_+ - k_\circ)}{k_+ - k_-}} \left(\bar{X}_{(k_-+1):k_\circ} - \bar{X}_{(k_\circ+1):k_+} \right) \right\}^2 = |\mathcal{X}_{k_-,k_\circ,k_+}|^2.
\end{aligned}$$

□

Lemma C.3. Under Assumptions 1, 2 and 4 (b), there exist fixed $C', C'' > 0$ for which we have $C'n \le \mathrm{RSS}(\mathcal{A}) \le C''n$ for any $\mathcal{A} \subset \mathcal{K}$ on $\mathcal{M}_n^{(11)}$.

Proof. Firstly, by ergodicity and that $0 < \mathsf{Var}(\varepsilon_t) < \infty$, there exist $c_l, c_u \in (0, \infty)$ such that

$$0 < c_l \le \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2 \le c_u < \infty \quad \text{a.s.}$$

From $\sum_{t=s}^{e} (X_t - \bar{X}_{s:e})^2 = \min_{a \in \mathbb{R}} \sum_{t=s}^{e} (X_t - a)^2$, it holds that $\mathrm{RSS}(\mathcal{A}) \ge \mathrm{RSS}(\mathcal{A}')$ for any $\mathcal{A} \subset \mathcal{A}'$. Thus we can find $C'' \in (0, \infty)$ such that for any $\mathcal{A} \subset \mathcal{K}$,

$$\mathrm{RSS}(\mathcal{A}) \le \mathrm{RSS}(\emptyset) = \sum_{t=1}^{n} (X_t - \bar{X}_{1:n})^2 \le 2 \sum_{t=1}^{n} (\varepsilon_t - \bar{\varepsilon}_{1:n})^2 + 2 \sum_{t=1}^{n} (f_t - \bar{f}_{1:n})^2$$

$$\leq 2 \sum_{t=1}^{n} \varepsilon_t^2 + 2n \bar{f}^2 \leq n \left(2c_u + 2\bar{f}^2\right) \leq C'' n,$$

where $\bar{f} = \max_{1 \leq j \leq q_n} |f_j - \bar{f}_{1:n}|$ is bounded by that $\max_{1 \leq j \leq q_n} |d_j| = O(1)$.
Next, let $\tilde{\mathcal{K}} := \mathcal{K} \cup \Theta = \{\tilde{k}_1 < \ldots < \tilde{k}_{A_n}\}$ with $\tilde{k}_0 = 0$ and $\tilde{k}_{A_n+1} = n$, where $A_n \leq Q_n + q_n$. Then, we can find $C' \in (0, \infty)$ such that for any $\mathcal{A} \subset \mathcal{K}$ and n large enough,

$$\mathrm{RSS}(\mathcal{A}) \geq \mathrm{RSS}(\tilde{\mathcal{K}}) \geq \sum_{j=0}^{A_n} \sum_{t=\tilde{k}_j+1}^{\tilde{k}_{j+1}} (\varepsilon_t - \bar{\varepsilon}_{(\tilde{k}_j+1):\tilde{k}_{j+1}})^2 = \sum_{t=1}^{n} \varepsilon_t^2 - \sum_{j=0}^{A_n} \left(\frac{\sum_{t=\tilde{k}_j+1}^{\tilde{k}_{j+1}} \varepsilon_t}{\sqrt{\tilde{k}_{j+1} - \tilde{k}_j}} \right)^2$$
$$\geq n \left(c_l - \frac{(Q_n + q_n)\omega_n^2}{n} \right) \geq C' n,$$

where the last inequality follows from that $(\min_{1 \leq j \leq q_n} \delta_j)^{-1} \omega_n^2 \to 0$ under Assumption 2 and thus $n^{-1} \omega_n^2 q_n \to 0$, and from Assumption 4 (b). □

Lemma C.4. Let the conditions in Lemma C.3 hold. Then, there exist fixed $\underline{C}, \bar{C} > 0$ such that we have

$$\underline{C} \, |\mathcal{X}_{k_-, k_\circ, k_+}|^2 - \xi_n \leq \mathrm{SC}(\mathcal{A} \setminus \{k_\circ\}) - \mathrm{SC}(\mathcal{A}) \leq \bar{C} \, |\mathcal{X}_{k_-, k_\circ, k_+}|^2 - \xi_n \tag{C.2}$$

for any $k_\circ \in \mathcal{A} \subset \mathcal{K}$.

Proof. From Lemmas C.2–C.3 and that $\log(1+x) \leq x$ for all $x \geq 0$, we obtain

$$\mathrm{SC}(\mathcal{A} \setminus \{k_\circ\}) - \mathrm{SC}(\mathcal{A}) = \frac{n}{2} \log \left\{ \frac{\mathrm{RSS}(\mathcal{A} \setminus \{k_\circ\})}{\mathrm{RSS}(\mathcal{A})} \right\} - \xi_n = \frac{n}{2} \log \left\{ 1 + \frac{|\mathcal{X}_{k_-, k_\circ, k_+}|^2}{\mathrm{RSS}(\mathcal{A})} \right\} - \xi_n$$
$$\leq \frac{|\mathcal{X}_{k_-, k_\circ, k_+}|^2}{2C'} - \xi_n$$

hence the RHS of (C.2) holds with $\bar{C} = 1/(2C')$.
Furthermore, by Lemmas C.2–C.3 it holds

$$1 \leq \frac{\mathrm{RSS}(\mathcal{A} \setminus \{k_\circ\})}{\mathrm{RSS}(\mathcal{A})} \leq \frac{C''}{C'}.$$

Let $g(x) = \log(x)/(x-1)$. Since $\lim_{x \downarrow 1} g(x) \to 1$ and from its continuity, there exists a constant $C''' > 0$ such that $\inf_{1 \leq x \leq C''/C'} g(x) \geq C'''$. Hence by Lemma C.3

$$\mathrm{SC}(\mathcal{A} \setminus \{k_\circ\}) - \mathrm{SC}(\mathcal{A}) = \frac{n}{2} \log \left\{ \frac{\mathrm{RSS}(\mathcal{A} \setminus \{k_\circ\})}{\mathrm{RSS}(\mathcal{A})} \right\} - \xi_n \geq \frac{C'''}{2C''} \, |\mathcal{X}_{k_-, k_\circ, k_+}|^2 - \xi_n,$$

so that $\underline{C} = C'''/(2C'')$ meets (C.2). □

C.4 Proofs of the Propositions C.1–C.3

Within the proofs of the propositions, the o-notation always refers to M in (C.1) being large enough, which in turn follows for large enough n, and precise bounds can be given in each instance.

C.4.1 Proof of Proposition C.1

Choose $C^* > \max(1, 2/\underline{C})$ and $c^* < \min(1, 1/\bar{C})$ for $\underline{C}$ and $\bar{C}$ defined in Lemma C.4. The tighter the choice is, the larger M in (C.1) is required to be.
Firstly, in the situation of (a), we have $d_\circ^2 \min(\theta_\circ - k_-, k_+ - \theta_\circ) \geq C^* \xi_n$ and $\max(d_+^2 r_+, d_-^2 r_-) \leq C^* \xi_n$. Then, when $k'_\circ \geq \theta_\circ$, it holds from Lemma C.1,

$$|\mathcal{X}_{k_-, k'_\circ, k_+}| \geq \sqrt{\frac{(k'_\circ - k_-)(k_+ - k'_\circ)}{k_+ - k_-}} \left\{ \frac{(\theta_\circ - k_-)|d_\circ|}{k'_\circ - k_-} - \frac{r_-|d_-|}{k'_\circ - k_-} - \frac{r_+|d_+|}{k_+ - k'_\circ} \right\} - |\mathcal{X}_{k_-, k'_\circ, k_+}(\varepsilon)|.$$

For the first summand, note that

$$\sqrt{\frac{(k'_\circ - k_-)(k_+ - k'_\circ)}{k_+ - k_-}} \frac{(\theta_\circ - k_-)|d_\circ|}{k'_\circ - k_-} = |d_\circ| \sqrt{\frac{(k_+ - \theta_\circ)(\theta_\circ - k_-)}{k_+ - k_-}} \left(\frac{k_+ - k'_\circ}{k_+ - \theta_\circ} \right)^{1/2} \left(\frac{\theta_\circ - k_-}{k'_\circ - k_-} \right)^{1/2}$$
$$\geq |d_\circ| \sqrt{\frac{1}{2} \min(\theta_\circ - k_-, k_+ - \theta_\circ)} \left(1 - \frac{\rho_n \nu_n}{\xi_n} \right) \geq \sqrt{\frac{C^* \xi_n}{2}} (1 + o(1)).$$

For the second summand,

$$\sqrt{\frac{(k'_\circ - k_-)(k_+ - k'_\circ)}{k_+ - k_-}} \frac{r_- \, |d_-|}{k'_\circ - k_-} = \sqrt{\frac{k_+ - k'_\circ}{k_+ - k_-}} \frac{r_- \, |d_-|^2}{\sqrt{k'_\circ - k_-}|d_-|} \leq \frac{C^* \xi_n}{\sqrt{D_n}} = o(\sqrt{\xi_n}),$$

where the inequality follows from noting that when $r_- > 0$, we have $k'_\circ - k_- \geq \theta_\circ - \theta_-$ as well as $d_-^2(\theta_\circ - \theta_-) \geq D_n$ under Assumption 2. An analogous argument applies to the third summand, noting that $k_+ - k'_\circ \geq (\theta_+ - \theta_\circ)(1 - \rho_n \nu_n / D_n)$ when $\theta_+ < k_+$ (hence $r_+ > 0$). Finally, on $\mathcal{M}_n^{(11)}$, the fourth summand satisfies

$$|\mathcal{X}_{k_-, k'_\circ, k_+}(\varepsilon)| = \sqrt{\frac{(k_\circ - k_-)(k_+ - k_\circ)}{k_+ - k_-}} \left| \bar{X}_{(k_- + 1):k_\circ} - \bar{X}_{(k_\circ + 1):k_+} \right| \leq 2\omega_n = o(\sqrt{\xi_n}). \quad \text{(C.3)}$$

Putting the above together, for M large enough (whose exact value depends on the choice of C^*),

$$|\mathcal{X}_{k_-, k'_j, k_+}| \geq \sqrt{\frac{C^* \xi_n}{2}} \left(1 + o(1) + \sqrt{2\, C^*}\, o(1) \right) + o(\sqrt{\xi_n}) \geq \sqrt{\frac{\xi_n}{\underline{C}}}, \quad \text{(C.4)}$$

which holds uniformly for any $s, e, \theta_\circ, k_\pm$ and $k'_\circ$ meeting the conditions of the proposition. By symmetric arguments (reversing time), the same holds when $k'_\circ < \theta_\circ$. The assertion of (a) now follows from Lemma C.4.

Next, we suppose that $d_\circ^2(\theta_\circ - k_-) \le c^* \xi_n$ as in the case of (b). Recalling the decomposition of $\mathcal{X}_{k_-,k,k_+}(f)$ from Lemma C.1, the term that does not depend on r_+ (note that $r_- = 0$ in the situation of (b)) satisfies for $k \ge \theta_\circ$,

$$\sqrt{\frac{(k-k_-)(k_+-k)}{k_+-k_-}}\frac{(\theta_\circ - k_-)|d_\circ|}{k-k_-} \le \sqrt{d_\circ^2 \min(\theta_\circ - k_-, k_+ - k)}$$
$$\le \sqrt{d_\circ^2 \min(\theta_\circ - k_-, k_+ - \theta_\circ)} \le \sqrt{c^* \xi_n}$$

and analogously for $k \le \theta_\circ$, that

$$\sqrt{\frac{(k-k_-)(k_+-k)}{k_+-k_-}}\frac{(k_+ - \theta_\circ)|d_\circ|}{k_+-k} \le \sqrt{c^* \xi_n}.$$

Also, when $r_+ > 0$,

$$\sqrt{\frac{(k-k_-)(k_+-k)}{k_+-k_-}}\frac{r_+\,|d_+|}{k_+-k} \le \frac{r_+\,|d_+|^2}{\sqrt{|d_+|^2(k_+-k)}} \le \frac{\sqrt{2}\,C^*\xi_n}{\sqrt{D_n}}, \tag{C.5}$$

by noting that $k - k_- \le k_+ - k_-$ and $k_+ - k \ge (\theta_+ - \theta_\circ)/2$ when $r_+ > 0$, because k is closer to $\theta_\circ$ than to θ_+. Together with (C.3), this leads to

$$|\mathcal{X}_{k_-,k,k_+}| \le \sqrt{c^*\,\xi_n} + \frac{\sqrt{2}C^*\xi_n}{\sqrt{D_n}} + 2\omega_n = \sqrt{c^*\xi_n}\left(1 + \frac{C^*}{\sqrt{c^*}}\,o(1) + o(1)\right) < \sqrt{\frac{\xi_n}{\bar{C}}}, \tag{C.6}$$

for M sufficiently large (depending on $\bar{C}, \underline{C}$), with the inequality holding uniformly for any $s, e, \theta_\circ, k_\pm$ and k meeting the conditions. Hence the conclusion of (b) follows from Lemma C.4. The proof of (c) follows by symmetry (reversing time).

C.4.2 Proof of Proposition C.2

Under (a), i.e. when there is no change point contained within this interval, by (C.3) we get

$$\left|\mathcal{X}_{k_-,k_\circ,k_+}\right| = \left|\mathcal{X}_{k_-,k_\circ,k_+}(\varepsilon)\right| \le 2\omega_n = o(\sqrt{\xi_n})$$

on $\mathcal{M}_n^{(11)}$, so that assertion (a) follows from Lemma C.4.

In the case of (b.i), w.l.o.g., we assume that $d_\circ^2\,|k_- - \theta_\circ| \le \rho_n \nu_n$, which in particular implies that $r_- = 0$ (for M large enough); otherwise consider the series in reversed time. Then, by assumption, $d_+^2 r_+ \le C^* \xi_n$ as well as $k_+ - k_\circ \ge (\theta_+ - \theta_\circ)/2$ when $k_+ > \theta_+$, since $k_\circ$ is closer to $\theta_\circ$ than any other change point within (k_-, k_+). We now distinguish the two cases: (I)

$k_- < \theta_\circ \le k_\circ$ and (II) $k_- < k_\circ < \theta_\circ$. If (I) holds, Lemma C.1 leads to

$$\begin{aligned}|\mathcal{X}_{k_-,k_\circ,k_+}| &\le \sqrt{\frac{(k_\circ - k_-)(k_+ - k_\circ)}{k_+ - k_-}} \left\{ \frac{(\theta_\circ - k_-)|d_\circ|}{k_\circ - k_-} + \frac{r_+ \, |d_+|}{k_+ - k_\circ} \right\} + |\mathcal{X}_{k_-,k_\circ,k_+}(\varepsilon)| \\ &\le \sqrt{\theta_\circ - k_-} \, |d_\circ| + \frac{\sqrt{2}C^* \xi_n}{\sqrt{D_n}} + 2\omega_n = \sqrt{\xi_n} \, (o(1) + C^* o(1)) < \sqrt{\frac{\xi_n}{\bar{C}}},\end{aligned}$$

for M large enough. The assertion follows from Lemma C.4. The case of (II) can be dealt with analogously.

Similarly, in the case of (b.ii), w.l.o.g., suppose $d_\circ^2 \, |k_- - \theta_\circ| \le C^* \xi_n$ such that in particular, $r_- = 0$ (for M large enough). Also, as in (b.i), it holds $k_+ - k_\circ \ge (\theta_+ - \theta_\circ)/2$ when $k_+ > \theta_+$. In this case, necessarily $k_- < \theta_\circ \le k_\circ$ and thus by Lemma C.1,

$$\begin{aligned}|\mathcal{X}_{k_-,k_\circ,k_+}| &\le \sqrt{\frac{(k_\circ - k_-)(k_+ - k_\circ)}{k_+ - k_-}} \left\{ \frac{(\theta_\circ - k_-)|d_\circ|}{k_\circ - k_-} + \frac{r_+ \, |d_+|}{k_+ - k_\circ} \right\} + |\mathcal{X}_{k_-,k_\circ,k_+}(\varepsilon)| \\ &\le \frac{(\theta_\circ - k_-)|d_\circ|}{\sqrt{k_\circ - \theta_\circ}} + \frac{\sqrt{2}C^* \xi_n}{\sqrt{D_n}} + 2\omega_n = \sqrt{\xi_n} \left(\frac{C^*}{\sqrt{\widetilde{C}}} + o(1) + C^* o(1) \right) < \sqrt{\frac{\xi_n}{\bar{C}}},\end{aligned}$$

for M large enough, completing the proof.

C.4.3 Proof of Proposition C.3

We start with some preliminary numerical calculations that will be used throughout the proof. We use the notations

$$\mathcal{W}_k = \frac{k_+ - k_-}{(k - k_-)(k_+ - k)}, \; \mathcal{F}_k = \sum_{t=k_-+1}^{k} (f_t - \bar{f}_{(k_-+1):k_+}) \text{ and } \mathcal{E}_k = \sum_{t=k_-+1}^{k} (\varepsilon_t - \bar{\varepsilon}_{(k_-+1):k_+});$$

we suppress the dependence of the above definitions on $k_\pm$ for brevity.

From Lemma C.1, we get $\mathcal{F}_k = \widetilde{\mathcal{F}}_k - R_k^+ - R_k^-$ with

$$\widetilde{\mathcal{F}}_k = -d_\circ \begin{cases} \frac{(k-k_-)(k_+-\theta_\circ)}{k_+-k_-}, & k \le \theta_\circ, \\ \frac{(k_+-k)(\theta_\circ-k_-)}{k_+-k_-}, & k \ge \theta_\circ, \end{cases}$$

$$R_k^+ = \frac{k - k_-}{k_+ - k_-} d_+ r_+, \qquad R_k^- = \frac{k_+ - k}{k_+ - k_-} d_- r_-.$$

Note that

$$\mathcal{W}_k \widetilde{\mathcal{F}}_k = -d_\circ \begin{cases} \frac{k_+-\theta_\circ}{k_+-k}, & k \le \theta_\circ, \\ \frac{\theta_\circ-k_-}{k-k_-}, & k \ge \theta_\circ, \end{cases} \quad \mathcal{W}_k \widetilde{\mathcal{F}}_k^2 = d_\circ^2 \begin{cases} \frac{(k_+-\theta_\circ)^2 \, (k-k_-)}{(k_+-k)(k_+-k_-)} & k \le \theta_\circ, \\ \frac{(\theta_\circ-k_-)^2 \, (k_+-k)}{(k-k_-) \, (k_+-k_-)}, & k \ge \theta_\circ, \end{cases} \tag{C.7}$$

which yields

$$\mathcal{W}_k \widetilde{\mathcal{F}}_k - \mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ} = d_\circ \begin{cases} \frac{\theta_\circ - k}{k_+ - k}, & k \le \theta_\circ, \\ \frac{k - \theta_\circ}{k - k_-}, & k \ge \theta_\circ, \end{cases} \tag{C.8}$$

$$\mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}^2_{\theta_\circ} - \mathcal{W}_k \widetilde{\mathcal{F}}^2_k = d_\circ^2 \begin{cases} \frac{(\theta_\circ - k)(k_+ - \theta_\circ)}{k_+ - k}, & k \le \theta_\circ, \\ \frac{(k - \theta_\circ)(\theta_\circ - k_-)}{k - k_-}, & k \ge \theta_\circ. \end{cases} \tag{C.9}$$

Concerning the remainder term, we get

$$\mathcal{W}_k R_k^+ = \frac{d_+ r_+}{k_+ - k}, \qquad \mathcal{W}_k R_k^- = \frac{d_- r_-}{k - k_-}, \qquad \mathcal{W}_k R_k^+ R_k^- = \frac{d_+ r_+ \cdot d_- r_-}{k_+ - k_-}, \tag{C.10}$$

as well as

$$\begin{aligned} \mathcal{W}_k R_k^+ - \mathcal{W}_{\theta_\circ} R_{\theta_\circ}^+ &= d_+ r_+ \frac{k - \theta_\circ}{(k_+ - k)(k_+ - \theta_\circ)}, \\ \mathcal{W}_k R_k^- - \mathcal{W}_{\theta_\circ} R_{\theta_\circ}^- &= d_- r_- \frac{\theta_\circ - k}{(k - k_-)(\theta_\circ - k_-)}. \end{aligned} \tag{C.11}$$

Furthermore,

$$\mathcal{W}_k (R_k^+)^2 = d_+^2\, r_+^2 \, \frac{k - k_-}{(k_+ - k_-)(k_+ - k)}, \qquad \mathcal{W}_k (R_k^-)^2 = d_-^2\, r_-^2 \, \frac{k_+ - k}{(k_+ - k_-)(k - k_-)}$$

and thus

$$\begin{aligned} \mathcal{W}_k (R_k^+)^2 - \mathcal{W}_{\theta_\circ} (R_{\theta_\circ}^+)^2 &= d_+^2 r_+^2 \, \frac{k - \theta_\circ}{(k_+ - k)(k_+ - \theta_\circ)}, \\ \mathcal{W}_{\theta_\circ} (R_{\theta_\circ}^-)^2 - \mathcal{W}_k (R_k^-)^2 &= d_-^2 r_-^2 \, \frac{k - \theta_\circ}{(k - k_-)(\theta_\circ - k_-)}. \end{aligned} \tag{C.12}$$

Finally, for the terms involving both $\widetilde{\mathcal{F}}_k$ and $\mathcal{R}_k$, we get

$$\begin{aligned} \mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ} R_{\theta_\circ}^+ - \mathcal{W}_{k_\circ} \widetilde{\mathcal{F}}_{k_\circ} R_{k_\circ}^+ &= d_+ r_+ d_\circ \frac{\theta_\circ - k_\circ}{k_+ - k_\circ} \, \mathbb{I}_{k_\circ \le \theta_\circ} \\ \mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ} R_{\theta_\circ}^- - \mathcal{W}_{k_\circ} \widetilde{\mathcal{F}}_{k_\circ} R_{k_\circ}^- &= d_- r_- d_\circ \frac{k_\circ - \theta_\circ}{k_\circ - k_-} \, \mathbb{I}_{k_\circ \ge \theta_\circ} \end{aligned} \tag{C.13}$$

Concerning the error terms, on $\mathcal{M}_n^{(12)} \cap \mathcal{M}_n^{(13)}$, it holds uniformly in $k_\circ \in \mathcal{V}_\circ$ and $k_\circ^* \in \mathcal{V}_\circ^*$,

$$\begin{aligned} |\mathcal{E}_{k_\circ} - \mathcal{E}_{k_\circ^*}| &\le \left| \sum_{t=k_\circ+1}^{\theta_\circ} \varepsilon_t \right| + \left| \sum_{t=\min(k_\circ^*, \theta_\circ)+1}^{\max(k_\circ^*, \theta_\circ)} \varepsilon_t \right| + \frac{k_\circ^* - k_\circ}{k_+ - k_-} \left| \sum_{t=k_-+1}^{k_+} \varepsilon_t \right| \\ &\le \frac{(\theta_\circ - k_\circ)\omega_n^{(1)}}{\sqrt{d_\circ^{-2} \rho_n \nu_n}} + \sqrt{d_\circ^{-2} \rho_n}\, \omega_n^{(2)} + \frac{2(\theta_\circ - k_\circ)\omega_n}{\sqrt{k_+ - k_-}} \end{aligned} \tag{C.14}$$

as well as

$$|\mathcal{E}_k| \le \left| \sum_{t=k_-+1}^{k} \varepsilon_t \right| + \left| \frac{k-k_-}{k_+-k_-} \sum_{t=k_-+1}^{k_+} \varepsilon_t \right| \le \sqrt{k-k_-}\,\omega_n + \frac{k-k_-}{\sqrt{k_+-k_-}}\,\omega_n \le 2\sqrt{k-k_-}\,\omega_n$$

and by symmetry of $\mathcal{E}_k$ also $|\mathcal{E}_k| \le 2\sqrt{k_+-k}\,\omega_n$, such that

$$|\mathcal{E}_k| \le 2\sqrt{\min(k-k_-, k_+-k)}\,\omega_n. \tag{C.15}$$

In what follows, we consider the following two cases: When $k_\circ$ is closer to one of the boundary points than to $\theta_\circ$, i.e. $|\theta_\circ - k_\circ| \ge \min(k_\circ - k_-, k_+ - k_\circ)$, and when this is not so.

Case 1: $|\theta_\circ - k_\circ| \ge \min(k_\circ - k_-, k_+ - k_\circ)$**.**

We further distinguish the following two cases:

(a) $|\theta_\circ - k_\circ| \ge k_\circ - k_-$, which can occur only if $k_\circ < \theta_\circ$ and $r_- = 0$ (otherwise $k_\circ$ is closer to θ_- than $\theta_\circ$ which contradicts that $k_\circ$ detects $\theta_\circ$). In particular, this implies that

$$k_\circ - k_- < (\theta_\circ - k_-)/2. \tag{C.16}$$

(b) $|\theta_\circ - k_\circ| \ge k_+ - k_\circ$, which can occur only if $k_\circ < k_+$ and $r_+ = 0$.

We detail the proof of (a) below; the assertion under (b) follow by symmetry (reversing time). First, by (C.7) and (C.15), it holds for any $k_- < k < k_+$

$$\begin{aligned}
\mathcal{W}_k \mathcal{E}_k^2 &\le 2\,\mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ}^2 \,\frac{(k_+-k_-)^2 \min(k-k_-, k_+-k)\,\omega_n^2}{d_\circ^2\,(k_+-\theta_\circ)(\theta_\circ-k_-)(k_+-k)(k-k_-)} \\
&\le 8\,\mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ}^2 \,\frac{\omega_n^2}{d_\circ^2 \min(k_+-\theta_\circ, \theta_\circ-k_-)} \le 8\,\mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ}^2 \,\frac{\omega_n^2}{c^*\xi_n} = o\left(\mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ}^2\right).
\end{aligned}$$

Concerning the remainder term (keeping in mind that in this situation, $r_- = 0$), we get by (C.12)

$$\begin{aligned}
\mathcal{W}_{k_\circ} (R_{k_\circ}^+)^2 &\le \mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ}^2 \,\frac{d_+^2 r_+^2}{d_\circ^2 (k_+-\theta_\circ)^2}\,\frac{(k_\circ-k_-)(k_+-\theta_\circ)}{(\theta_\circ-k_-)\,(k_+-k_\circ)} \le \mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ}^2 \,\frac{(C^*)^2\xi_n^2}{D_n^2} \\
&= o\left(\mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ}^2\right).
\end{aligned} \tag{C.17}$$

Furthermore, by (C.7) it holds for all $k_- < k < k_+$

$$\mathcal{W}_k \widetilde{\mathcal{F}}_k^2 \le \mathcal{W}_{\theta_\circ} \widetilde{\mathcal{F}}_{\theta_\circ}^2,$$

which is used to deal with the mixed terms to arrive at

$$|\mathcal{X}_{k_-,k_\circ,k_+}|^2 = \mathcal{W}_{k_\circ}\widetilde{\mathcal{F}}_{k_\circ}^2 + o\left(\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\right).$$

For $k_\circ$ replaced by $k_\circ^*$ in (C.17) we get

$$\mathcal{W}_{k_\circ^*}(R_{k_\circ^*}^+)^2 \le \mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2 \frac{(C^*)^2\xi_n^2}{D_n^2}\frac{1+\frac{\rho_n}{c^*\xi_n}}{1-\frac{\rho_n}{c^*\xi_n}} = o\left(\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\right),$$

resulting in

$$|\mathcal{X}_{k_-,k_\circ^*,k_+}|^2 = \mathcal{W}_{k_\circ^*}\widetilde{\mathcal{F}}_{k_\circ^*}^2 + o\left(\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\right).$$

By (C.7) and (C.9) we get

$$\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2 - \mathcal{W}_{k_\circ}\widetilde{\mathcal{F}}_{k_\circ}^2 = \mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2 \frac{\theta_\circ - k_\circ}{\theta_\circ - k_-}\frac{k_+ - k_-}{k_+ - k_\circ} \ge \mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\left(1 - \frac{k_\circ - k_-}{\theta_\circ - k_-}\right) \ge \frac{1}{2}\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2,$$

where the last inequality follows from (C.16). Similarly,

$$\begin{aligned}
&\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2 - \mathcal{W}_{k_\circ^*}\widetilde{\mathcal{F}}_{k_\circ^*}^2 = \mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\,|\theta_\circ - k_\circ^*|\,\frac{k_+ - k_-}{(k_+ - k_\circ^*)(\theta_\circ - k_-)} \\
&\le \mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\left(\frac{d_\circ^2|\theta_\circ - k_\circ^*|}{d_\circ^2\min(k_+ - \theta_\circ, \theta_\circ - k_-)}\right)(2 + o(1)) \\
&\le \mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\frac{\rho_n}{c^*\xi_n}(2 + o(1)) = o\left(\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\right).
\end{aligned} \tag{C.18}$$

Putting the above together and by Lemma C.2,

$$\begin{aligned}
&\mathrm{RSS}(\mathcal{A}\cup\{k_\circ\}) - \mathrm{RSS}(\mathcal{A}\cup\{k_\circ^*\}) = |\mathcal{X}_{k_-,k_\circ^*,k_+}|^2 - |\mathcal{X}_{k_-,k_\circ,k_+}|^2 \\
&= \mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2 - \mathcal{W}_{k_\circ}\widetilde{\mathcal{F}}_{k_\circ}^2 + o\left(\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\right) \ge \mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2\left(\frac{1}{2} + o(1)\right) > 0,
\end{aligned}$$

which proves the claim.

Case 2: $\min(k_\circ - k_-, k_+ - k_\circ) > |k_\circ - \theta_\circ|$**.**

In this case, we have

$$k_\circ - k_- > (\theta_\circ - k_-)/2 \quad \text{and} \quad k_+ - k_\circ > (k_+ - \theta_\circ)/2. \tag{C.19}$$

By Lemma C.2, the following decomposition holds:

$$\begin{aligned}
&\mathrm{RSS}(\mathcal{A}\cup\{k_\circ\}) - \mathrm{RSS}(\mathcal{A}\cup\{k_\circ^*\}) = |\mathcal{X}_{k_-,k_\circ^*,k_+}|^2 - |\mathcal{X}_{k_-,k_\circ,k_+}|^2 \\
&= \mathcal{W}_{k_\circ^*}(\mathcal{F}_{k_\circ^*} + \mathcal{E}_{k_\circ^*})^2 - \mathcal{W}_{k_\circ}(\mathcal{F}_{k_\circ} + \mathcal{E}_{k_\circ})^2
\end{aligned}$$

$$
\begin{aligned}
&= \big(\mathcal{W}_{\theta_\circ}\mathcal{F}_{\theta_\circ}^2 - \mathcal{W}_{k_\circ}\mathcal{F}_{k_\circ}^2\big) + \big(\mathcal{W}_{k_\circ^*}\mathcal{F}_{k_\circ^*}^2 - \mathcal{W}_{\theta_\circ}\mathcal{F}_{\theta_\circ}^2\big) + 2\mathcal{W}_{k_\circ}\mathcal{F}_{k_\circ}(\mathcal{E}_{k_\circ^*} - \mathcal{E}_{k_\circ}) \\
&\quad + 2(\mathcal{W}_{k_\circ^*}\mathcal{F}_{k_\circ^*} - \mathcal{W}_{\theta_\circ}\mathcal{F}_{\theta_\circ})\mathcal{E}_{k_\circ^*} + 2(\mathcal{W}_{\theta_\circ}\mathcal{F}_{\theta_\circ} - \mathcal{W}_{k_\circ}\mathcal{F}_{k_\circ})\mathcal{E}_{k_\circ^*} \\
&\quad + \mathcal{W}_{k_\circ}(\mathcal{E}_{k_\circ^*}^2 - \mathcal{E}_{k_\circ}^2) + (\mathcal{W}_{k_\circ^*} - \mathcal{W}_{k_\circ})\mathcal{E}_{k_\circ^*}^2 \\
&=: A_1(\mathcal{F}) + A_2(\mathcal{F}) + A_3(\mathcal{F}) + A_4(\mathcal{F}) + A_5(\mathcal{F}) + A_6 + A_7. \qquad \text{(C.20)}
\end{aligned}
$$

We now show that for n large enough,

$$
\begin{aligned}
&A_1(\widetilde{\mathcal{F}}) > 0, \qquad \frac{|A_1(\widetilde{\mathcal{F}}) - A_1(\mathcal{F})|}{A_1(\widetilde{\mathcal{F}})} = o(1), \\
&\frac{|A_j(\mathcal{F})|}{A_1(\widetilde{\mathcal{F}})} = o(1) \quad \text{for } j = 2, \ldots, 5, \quad \text{and} \quad \frac{|A_j|}{A_1(\widetilde{\mathcal{F}})} = o(1) \quad \text{for } j = 6, 7
\end{aligned}
$$

on $\mathcal{M}_n$, uniformly in $k_\pm, k_\circ$ and $k_\circ^*$ meeting the conditions of the proposition. Consequently, $\mathrm{RSS}\,(\mathcal{A} \cup \{k_\circ\}) > \mathrm{RSS}\,(\mathcal{A} \cup \{k_\circ^*\})$, which proves the assertion.

W.l.o.g., let $k_\circ < \theta_\circ$ (otherwise consider the time series in reverse). In what follows, all the inequalities are uniform in the sense that they hold provided that the conditions of the proposition are met.

Firstly, from (C.9),

$$
\begin{aligned}
A_1(\widetilde{\mathcal{F}}) &= \mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}^2 - \mathcal{W}_{k_\circ}\widetilde{\mathcal{F}}_{k_\circ}^2 = \frac{d_\circ^2(\theta_\circ - k_\circ)(k_+ - \theta_\circ)}{k_+ - k_\circ} \ge \frac{d_\circ^2}{2}\min(\theta_\circ - k_\circ, k_+ - \theta_\circ) \\
&\ge \begin{cases} \frac{\rho_n \nu_n}{2} > 0 & \text{when } \theta_\circ - k_\circ \le k_+ - \theta_\circ, \\ \frac{c^*\xi_n}{2} > 0 & \text{when } \theta_\circ - k_\circ > k_+ - \theta_\circ. \end{cases} \qquad \text{(C.21)}
\end{aligned}
$$

Next, under the conditions imposed on $k_\pm$ and from (C.12), when $\theta_+ < k_+$ such that $r_+ \neq 0$,

$$
\begin{aligned}
\big|\mathcal{W}_{k_\circ}(R_{k_\circ}^+)^2 - \mathcal{W}_{\theta_\circ}(R_{\theta_\circ}^+)^2\big| &= \frac{d_+^2 r_+^2(\theta_\circ - k_\circ)}{(k_+ - k_\circ)(k_+ - \theta_\circ)} \le A_1(\widetilde{\mathcal{F}}) \cdot \frac{d_+^4 r_+^2}{d_\circ^2 d_+^2 (k_+ - \theta_\circ)^2} \\
&\le A_1(\widetilde{\mathcal{F}}) \cdot \left(\frac{C^*\xi_n}{D_n}\right)^2 = o(A_1(\widetilde{\mathcal{F}})). \qquad \text{(C.22)}
\end{aligned}
$$

Similarly, when $k_- < \theta_-$ such that $r_- \neq 0$, by (C.19)

$$
\begin{aligned}
\big|\mathcal{W}_{k_\circ}(R_{k_\circ}^-)^2 - \mathcal{W}_{\theta_\circ}(R_{\theta_\circ}^-)^2\big| &= \frac{d_-^2 r_-^2(\theta_\circ - k_\circ)}{(k_\circ - k_-)(\theta_\circ - k_-)} = A_1(\widetilde{\mathcal{F}}) \cdot \frac{d_-^4 r_-^2}{d_\circ^2(k_\circ - k_-)\, d_-^2(\theta_\circ - k_-)} \frac{k_+ - k_\circ}{k_+ - \theta_\circ} \\
&\le 2A_1(\widetilde{\mathcal{F}}) \cdot \left(\frac{C^*\xi_n}{D_n}\right)^2 \left(1 + \frac{d_\circ^2(\theta_\circ - k_\circ)}{d_\circ^2(k_+ - \theta_\circ)}\right) = o\left(A_1(\widetilde{\mathcal{F}})\right) \qquad \text{(C.23)}
\end{aligned}
$$

since $d_\circ^2(\theta_\circ - k_\circ) \le \widetilde{C}\xi_n$. Furthermore, from (C.13), if $r_+ \neq 0$,

$$
\left|\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}R_{\theta_\circ}^+ - \mathcal{W}_{k_\circ}\widetilde{\mathcal{F}}_{k_\circ}R_{k_\circ}^+\right| + \left|\mathcal{W}_{\theta_\circ}\widetilde{\mathcal{F}}_{\theta_\circ}R_{\theta_\circ}^- - \mathcal{W}_{k_\circ}\widetilde{\mathcal{F}}_{k_\circ}R_{k_\circ}^-\right| = \frac{|d_+|r_+ \cdot |d_\circ|(\theta_\circ - k_\circ)}{k_+ - k_\circ}
$$

$$= A_1(\widetilde{\mathcal{F}}) \cdot \frac{d_+^2 r_+}{|d_\circ d_+| \, (k_+ - \theta_\circ)} \leq A_1(\widetilde{\mathcal{F}}) \cdot \frac{C^* \xi_n}{D_n} = o\left(A_1(\widetilde{\mathcal{F}})\right). \tag{C.24}$$

Together, (C.21)–(C.24) establish that $|A_1(\widetilde{\mathcal{F}}) - A_1(\mathcal{F})| = o\left(A_1(\widetilde{\mathcal{F}})\right)$.
For $A_2(\mathcal{F})$, first note that by (C.9) it holds

$$A_2(\widetilde{\mathcal{F}}) \leq d_\circ^2 |k_\circ^* - \theta_\circ| \leq \rho_n = o\left(A_1(\widetilde{\mathcal{F}})\right).$$

By analogous arguments to those adopted in (C.22)–(C.24), we also obtain $A_2(\mathcal{F}) = o\left(A_1(\widetilde{\mathcal{F}})\right)$.
From (C.7), (C.21) and (C.14), we yield

$$\begin{aligned}
&|A_3(\widetilde{\mathcal{F}})| = 2|\mathcal{W}_{k_\circ} \widetilde{\mathcal{F}}_{k_\circ}| \, |\mathcal{E}_{k_\circ^*} - \mathcal{E}_{k_\circ}| = \frac{2(k_+ - \theta_\circ)|d_\circ|}{k_+ - k_\circ} \, |\mathcal{E}_{k_\circ^*} - \mathcal{E}_{k_\circ}| \\
&\leq 2A_1(\widetilde{\mathcal{F}}) \left(\frac{\omega_n^{(1)}}{\sqrt{\rho_n \nu_n}} + \frac{\sqrt{\rho_n} \omega_n^{(2)}}{d_\circ^2(\theta_\circ - k_\circ)} + \frac{2\omega_n}{|d_\circ| \sqrt{k_+ - k_-}} \right) \leq 2A_1(\widetilde{\mathcal{F}}) \left(\frac{\omega_n^{(1)}}{\sqrt{\rho_n \nu_n}} + \frac{\omega_n^{(2)}}{\nu_n \sqrt{\rho_n}} + \frac{2\omega_n}{\sqrt{c^* \xi_n}} \right) \\
&= o\left(A_1(\widetilde{\mathcal{F}})\right).
\end{aligned}$$

Also, by (C.10) and because $k_+ - \theta_\circ \geq \theta_+ - \theta_\circ$ when $r_+ > 0$, we get

$$\begin{aligned}
|A_3(R^+)| &= \frac{2r_+|d_+|}{k_+ - k_\circ} \, |\mathcal{E}_{k_\circ^*} - \mathcal{E}_{k_\circ}| \\
&\leq 2A_1(\widetilde{\mathcal{F}}) \left(\frac{\omega_n^{(1)} \, r_+|d_+|}{\sqrt{\rho_n \nu_n} \, |d_\circ| \, (k_+ - \theta_\circ)} + \frac{\sqrt{\rho_n} \omega_n^{(2)} \, r_+ \, |d_+|}{|d_\circ|^3 \, (\theta_\circ - k_\circ) \, (k_+ - \theta_\circ)} + \frac{2\omega_n \, r_+ \, |d_+|}{d_\circ^2 \sqrt{k_+ - k_-} \, (k_+ - \theta_\circ)} \right) \\
&\leq 2 \, C^* \, A_1(\widetilde{\mathcal{F}}) \, \frac{\xi_n}{D_n} \left(\frac{\omega_n^{(1)}}{\sqrt{\rho_n \nu_n}} + \frac{\omega_n^{(2)}}{\nu_n \sqrt{\rho_n}} + \frac{2\omega_n}{\sqrt{c^* \xi_n}} \right) = o\left(A_1(\widetilde{\mathcal{F}})\right).
\end{aligned}$$

Similarly, using (C.19) for the last inequality,

$$\begin{aligned}
&|A_3(R^-)| = \frac{2r_-|d_-|}{k_\circ - k_-} \, |\mathcal{E}_{k_\circ^*} - \mathcal{E}_{k_\circ}| \\
&\leq 2A_1(\widetilde{\mathcal{F}}) \, \frac{k_+ - k_\circ}{k_+ - \theta_\circ} \left(\frac{|d_-| \, r_- \, \omega_n^{(1)}}{|d_\circ| \, (k_\circ - k_-) \sqrt{\rho_n \nu_n}} + \frac{|d_-| \, r_- \sqrt{\rho_n} \, \omega_n^{(2)}}{|d_\circ|^3 (k_\circ - k_-) \, (\theta_\circ - k_\circ)} + \frac{2 \, |d_-| \, r_- \, \omega_n}{d_\circ^2 (k_\circ - k_-) \sqrt{k_+ - k_-}} \right) \\
&\leq 4A_1(\widetilde{\mathcal{F}}) \, C^* \left(1 + \frac{\widetilde{C}}{c^*} \right) \frac{\xi_n}{D_n} \left(\frac{\omega_n^{(1)}}{\sqrt{\rho_n \nu_n}} + \frac{\omega_n^{(2)}}{\nu_n \sqrt{\rho_n}} + \frac{2\omega_n}{\sqrt{c^* \xi_n}} \right) = o\left(A_1(\widetilde{\mathcal{F}})\right).
\end{aligned}$$

As for $A_4(\mathcal{F})$, (C.8), (C.21) and (C.15) lead to

$$|A_4(\widetilde{\mathcal{F}})| \leq \frac{2|d_\circ||k_\circ^* - \theta_\circ|}{\min(k_+ - k_\circ^*, k_\circ^* - k_-)} \, |\mathcal{E}_{k_\circ^*}| \leq \frac{4 \, |d_\circ| \, |k_\circ^* - \theta_\circ| \, \omega_n}{\sqrt{\min(k_+ - k_\circ^*, k_\circ^* - k_-)}}$$

$$\leq 8\, A_1(\widetilde{\mathcal{F}})\, \frac{\omega_n}{\sqrt{c^*\xi_n\left(1-\frac{\rho_n}{c^*\xi_n}\right)}}\, \frac{\rho_n}{\min(\rho_n\nu_n, c^*\xi_n)} = o\left(A_1(\widetilde{\mathcal{F}})\right),$$

while from (C.11),

$$\begin{aligned}
|A_4(R^+)| &\leq \frac{4|d_+|r_+|k_\circ^*-\theta_\circ|\,\omega_n\sqrt{\min(k_+-k_\circ^*, k_\circ^*-k_-)}}{(k_+-k_\circ^*)(k_+-\theta_\circ)} \\
&\leq 4\, A_1(\widetilde{\mathcal{F}})\, \frac{|d_+|\, r_+}{|d_\circ|\,(k_+-\theta_\circ)}\, \frac{\omega_n\,|k_\circ^*-\theta_\circ|}{|d_\circ|\,\sqrt{k_+-k_\circ^*}\,\min(\theta_\circ-k_\circ, k_+-\theta_\circ)} \\
&\leq 4\, A_1(\widetilde{\mathcal{F}})\, \frac{C^*\xi_n}{D_n}\, \frac{\omega_n}{\sqrt{c^*\xi_n\left(1-\frac{\rho_n}{c^*\xi_n}\right)}}\, \frac{\rho_n}{\min(\rho_n\nu_n, c^*\xi_n)} = o\left(A_1(\widetilde{\mathcal{F}})\right).
\end{aligned}$$

Analogously, we obtain the bound of the same order for $|A_4(R^-)|$.
For $A_5(\mathcal{F})$, from (C.8), (C.11) and (C.19),

$$\begin{aligned}
|A_5(\widetilde{\mathcal{F}})| &= 2\,|\mathcal{E}_{k_\circ^*}|\frac{|d_\circ|\,(\theta_\circ-k_\circ)}{k_+-k_\circ} \leq 4\, A_1(\widetilde{\mathcal{F}})\, \frac{\sqrt{\min(k_+-k_\circ^*, k_\circ^*-k_-)}\,\omega_n}{|d_\circ|(k_+-\theta_\circ)} \\
&\leq 4\, A_1(\widetilde{\mathcal{F}})\, \frac{\omega_n}{|d_\circ|\sqrt{k_+-\theta_\circ}}\sqrt{\frac{k_+-k_\circ^*}{k_+-\theta_\circ}} \leq 4\, A_1(\widetilde{\mathcal{F}})\, \frac{\omega_n}{\sqrt{c^*\xi_n}}\sqrt{1+\frac{\rho_n}{c^*\xi_n}} = o\left(A_1(\widetilde{\mathcal{F}})\right).
\end{aligned}$$

Furthermore, by (C.11), (C.19), (C.21) and (C.15) it holds

$$\begin{aligned}
|A_5(R^-)| &= 2\,\frac{|d_-|\,r_-\,(\theta_\circ-k_\circ)}{(k_\circ-k_-)\,(\theta_\circ-k_-)}\,|\mathcal{E}_{k_\circ^*}| \\
&\leq 4\, A_1(\widetilde{\mathcal{F}})\, \frac{|d_-|\,r_-}{d_\circ^2\,(k_\circ-k_-)}\, \frac{k_+-k_\circ}{(k_+-\theta_\circ)\,(\theta_\circ-k_-)}\,\omega_n\sqrt{\min(k_+-k_\circ^*, k_\circ^*-k_-)} \\
&\leq 16\, A_1(\widetilde{\mathcal{F}})\, \frac{|d_-|\,r_-}{|d_\circ|\,(\theta_\circ-k_-)}\, \frac{\omega_n}{|d_\circ|\sqrt{\min(k_+-\theta_\circ, \theta_\circ-k_-)}}\sqrt{\frac{\min(k_+-k_\circ^*, k_\circ^*-k_-)}{\min(k_+-\theta_\circ, \theta_\circ-k_-)}} \\
&\leq 16\, A_1(\widetilde{\mathcal{F}})\, \frac{C^*\xi_n}{D_n}\, \frac{\omega_n}{\sqrt{c^*\xi_n}}\sqrt{1+\frac{\rho_n}{c^*\xi_n}} = o\left(A_1(\widetilde{\mathcal{F}})\right).
\end{aligned}$$

Similar but slightly easier arguments give the same bound (with the factor 16 replaced by 4) for $|A_5(R^+)|$. Since

$$\begin{aligned}
&\frac{k_+-k_-}{(k_+-\theta_\circ)(\theta_\circ-k_-)}\,|\mathcal{E}_{k_\circ^*}+\mathcal{E}_{k_\circ}| \leq 2\,\omega_n\,\frac{\sqrt{\min(k_+-k_\circ, k_\circ-k_-)+\min(k_+-k_\circ^*, k_\circ^*-k_-)}}{\min(k_+-\theta_\circ, \theta_\circ-k_-)} \\
&\leq 2\,\omega_n\,\frac{\sqrt{2+\frac{\widetilde{C}}{c^*}+\frac{\rho_n}{c^*\xi_n}}}{\sqrt{\min(k_+-\theta_\circ, \theta_\circ-k_-)}},
\end{aligned}$$

we yield from (C.19) and (C.14)–(C.15)

$$
\begin{aligned}
|A_6| &= \frac{k_+ - k_-}{(k_\circ - k_-)(k_+ - k_\circ)} \, |\mathcal{E}_{k_\circ^*} + \mathcal{E}_{k_\circ}| \, |\mathcal{E}_{k_\circ^*} - \mathcal{E}_{k_\circ}| \\
&\le 8 A_1(\widetilde{\mathcal{F}}) \sqrt{2 + \frac{\widetilde{C}}{c^*} + \frac{\rho_n}{c^* \xi_n}} \, \frac{\omega_n}{|d_\circ| \sqrt{\min(k_+ - \theta_\circ, \theta_\circ - k_-)}} \\
&\quad \times \left(\frac{\omega_n^{(1)}}{\sqrt{\rho_n \nu_n}} + \frac{\sqrt{\rho_n}\, \omega_n^{(2)}}{|d_\circ|^2 (\theta_\circ - k_\circ)} + 2 \frac{\omega_n}{|d_\circ| \sqrt{k_+ - k_-}} \right) \\
&\le 8 A_1(\widetilde{\mathcal{F}}) \sqrt{2 + \frac{\widetilde{C}}{c^*} + \frac{\rho_n}{c^* \xi_n}} \, \frac{\omega_n}{\sqrt{c^* \xi_n}} \left(\frac{\omega_n^{(1)}}{\sqrt{\rho_n \nu_n}} + \frac{\omega_n^{(2)}}{\nu_n \sqrt{\rho_n}} + \frac{2\omega_n}{\sqrt{c^* \xi_n}} \right) = o\left(A_1(\widetilde{\mathcal{F}}) \right).
\end{aligned}
$$

Finally, noting that by (C.19)

$$
\begin{aligned}
|\mathcal{W}_{k_\circ^*} - \mathcal{W}_{k_\circ}| &\le \frac{(k_+ - k_-)}{(k_+ - k_\circ^*)(k_\circ^* - k_-)} \, \frac{|k_\circ^* - k_\circ| \left\{ (k_+ - k_\circ^*) + (k_\circ - k_-) \right\}}{(k_+ - k_\circ)(k_\circ - k_-)} \\
&\le 2\, A_1(\widetilde{\mathcal{F}}) \, \frac{1}{\min(k_+ - k_\circ^*, k_\circ^* - k_-)} \, \frac{|k_\circ - k_\circ^*|}{|\theta_\circ - k_\circ|} \left(\frac{1}{d_\circ^2\, (k_\circ - k_-)} \, \frac{k_+ - k_\circ^*}{k_+ - \theta_\circ} + \frac{1}{d_\circ^2 (k_+ - \theta_\circ)} \right) \\
&\le 2\, A_1(\widetilde{\mathcal{F}}) \, \frac{1}{\min(k_+ - k_\circ^*, k_\circ^* - k_-)} \left(1 + \frac{1}{\nu_n} \right) \left(3 + \frac{2\rho_n}{c^* \xi_n} \right) \frac{1}{c^* \xi_n}
\end{aligned}
$$

we bound A_7 as

$$
|A_7| \le 8\, A_1(\widetilde{\mathcal{F}}) \left(1 + \frac{1}{\nu_n} \right) \left(3 + \frac{2\rho_n}{c^* \xi_n} \right) \frac{\omega_n^2}{c^* \xi_n} = o\left(A_1(\widetilde{\mathcal{F}}) \right),
$$

which concludes the proof.

D Proof of the results in Section 4

D.1 Proof of Proposition 4

The following lemma is used for the proofs of Proposition 4 and Corollary D.1.

Lemma D.1. (a) Under the assumption of Proposition 4, consider

$$
\mathcal{S}_n(j) = \left\{ \left| T_{\theta_j, n}(G(j)) \right| \ge \max \left(\max_{|k - \theta_j| > (1-\eta) G(j)} |T_{k,n}(G(j))| \,, \tau\, D_n(G(j), \alpha) \right) \right\},
$$

and $\mathcal{S}_n = \bigcap_{1 \le j \le q_n} \mathcal{S}_n(j)$. Then for any $\alpha, \eta \in (0, 1)$, we have

$$
\mathsf{P}(\mathcal{S}_n(j)) \to 1 \text{ for any } j = 1, \ldots, q_n \quad \text{and} \quad \mathsf{P}(\mathcal{S}_n) \to 1.
$$

(b) Under the assumptions of Corollary D.1 below, analogous assertions hold with $\widetilde{\mathcal{S}}_n(j)$

replacing $\mathcal{S}_n(j)$, where

$$\widetilde{\mathcal{S}}_n(j) = \bigcap_{0 \le r \le 2/\eta - 2} \left[\left\{ \left| T_{\theta_j + r\eta G/2, n}(G) \right| > \max_{k \in [\theta_j + (r+1)\eta G/2, \theta_j + (r+2)\eta G/2]} |T_{k,n}(G)| \right\} \right.$$
$$\left. \bigcap \left\{ \left| T_{\theta_j - r\eta G/2, n}(G) \right| > \max_{k \in [\theta_j - (r+2)\eta G/2, \theta_j - (r+1)\eta G/2} |T_{k,n}(G)| \right\} \right].$$

Proof. Adopting the arguments analogous to those used in the proof of Lemma 5.1 (a) of Eichinger and Kirch (2018), we get

$$\sqrt{2}\,|T_{\theta_j,n}(G(j))| \ge |d_j|\sqrt{G(j)} + O_P(\omega_n) = |d_j|\sqrt{G(j)}\,(1 + o_P(1)),$$
$$\max_{|k-\theta_j| > (1-\eta)G(j)} \sqrt{2}\,|T_{k,n}(G(j))| \le \eta\,|d_j|\sqrt{G(j)} + O_P(\omega_n) = \eta\,|d_j|\sqrt{G(j)}\,(1 + o_P(1)).$$

Also, noting that $D_n(G(j), \alpha) = O(\sqrt{\log(n)})$ and $D_n/\sqrt{\log(n)} \to \infty$, the 'significance' of $|T_{\theta_j,n}(G(j))|$ follows, and so does the assertion for $S_n(j)$. For the set S_n, the assertion follows because all O_P-terms hold uniformly in j. The assertion of (b) follows analogously. $\square$

With the help of Lemma D.1, we can now prove Proposition 4 by adopting the arguments of the proof of Theorem 3.2 of Eichinger and Kirch (2018). Therefore we only sketch the proof by emphasizing the differences using the notations adopted therein. In particular the quantities $V_{l,n}^{(j)}(G(j))$ and $A_i(l, n; G(j)) = A_i(l, n)$, $i = 1, 2$ are defined as in that proof.
On $S_n(j)$ defined in Lemma D.1 (a), the maximiser of $|T_{b,n}(G(j))|$ over b satisfying $|b - \theta_j| \le (1-\eta)G(j)$, fulfils the η-criterion and as such is a candidate produced by the MOSUM procedure which we denote by k_j in the following. For this candidate, it holds:

$$\left\{ k_j - \theta_j < -C_M(\omega_n^{(1)}/d_j)^2 \right\} \subset$$
$$\left\{ \max_{\theta_j - G(j) + 1 \le l < \theta_j - C_M(\omega_n^{(1)}/d_j)^2} V_{l,n}^{(j)}(G(j)) \ge \max_{\theta_j - C_M(\omega_n^{(1)}/d_j)^2 \le l \le \theta_j + G(j)} V_{l,n}^{(j)}(G(j)) \right\}.$$

Furthermore, by Condition (b) we obtain

$$\max_{1 \le j \le q_n} \frac{1}{|d_j|\sqrt{G(j)}} \max_{|l - \theta_j| < G(j)} |A_2(l, n; G(j))| = o_P(1).$$

Also by Condition (c), we can find a suitable constant $\widetilde{C}_M > 0$ such that for all $C_M > 0$, it holds

$$\mathsf{P}\left(\max_{1 \le j \le q_n} \sqrt{2G(j)} \max_{\theta_j - G(j) \le l \le \theta_j - C_M(\omega_n^{(1)}/d_j)^2} \frac{\sqrt{C_M(\omega_n^{(1)}/d_j)^2}}{|\theta_j - l|} |A_1(l, n; G(j))| > \widetilde{C}_M\,\omega_n^{(1)} \right) \to 0,$$

from which we can find a suitable choice of C_M depending only on $\widetilde{C}_M$ such that

$$\mathsf{P}\left(\max_{1\le j\le q_n}\frac{\sqrt{2G(j)}}{|d_j|}\max_{\theta_j-G(j)\le l\le\theta_j-C_M(\omega_n^{(1)}/d_j)^2}\frac{|A_1(l,n;G(j))|}{|\theta_j-l|}\ge\frac{1}{3}\right)\to 0.$$

Consequently,

$$\mathsf{P}\left(\min_{1\le j\le q_n} d_j^2(k_j-\theta_j)<-C_M(\omega_n^{(1)})^2,\mathcal{S}_n\right)=o(1).$$

The case $k_j-\theta_j>C_M(\omega_n^{(1)}/d_j)^2$ can be dealt with analogously, which concludes the proof.

D.2 Proof of Proposition 5

Firstly note that the η-criterion employed by the MOSUM procedure implicitly imposes an upper bound on the number of estimators returned: At bandwidths $\mathbf{G}=(G_\ell,G_r)$, for each local maximiser k of the MOSUM detector, it is checked whether the local maximum corresponds to the maximum absolute MOSUM value within the interval $(k-\eta\,G_\ell,k+\eta\,G_r]$ and, if so, k is marked as a candidate change point. Therefore, the maximal number of possible candidates detectable at scale (G_ℓ,G_r) is $(\eta\min(G_\ell,G_r))^{-1}n$. Then, by (9), it holds

$$\min(G_\ell,G_r)\ge\frac{\max(G_\ell,G_r)}{C_{\text{asym}}}\ge\frac{G_\ell+G_r}{2\,C_{\text{asym}}}.$$

From this and by the construction of $\mathcal{G}$ with $G_\ell=F_\ell\,G_0$, it holds

$$|\mathcal{K}(\mathcal{H},\alpha)|\le\sum_{\ell,r=1}^{H_n}|\mathcal{K}(G_\ell,G_r,\alpha)|\le 2\,C_{\text{asym}}\frac{n}{\eta G_0}\sum_{\ell,r=1}^{H_n}\frac{1}{F_\ell+F_r}\le 2\,C_{\text{asym}}\frac{\psi}{\eta}\frac{n}{G_0},$$

for some universal constant ψ satisfying

$$\sum_{\ell,r=1}^{\infty}\frac{1}{F_\ell+F_r}\le\psi<\infty.$$

This holds as the Fibonacci numbers are asymptotically bounded from below by an exponentially decreasing sequence, i.e. $F_\ell\ge(3/2)^\ell$ for all $\ell\ge 10$ which is easily seen by induction. Then, the conclusion follows from $\omega_n^2/G_0\to 0$.

D.3 Single-scale MOSUM procedure

As a corollary, we show that the single-bandwidth MOSUM procedure yields consistent estimators with optimal localisation rate either under sub-Gaussianity, or when there are finitely many change points, but only under the assumption that the change points are *homogeneous* as defined in Definition 1 (a). It improves upon Theorem 3.2 of Eichinger and Kirch (2018) where the optimal rate is obtained only in the case when q_n is finite. By construction, when the change points are heterogeneous, the single-bandwidth MOSUM procedure cannot produce consistent estimators.

Corollary D.1. Let $\mathcal{K}(G, \alpha_n) = \{k_{G,j} : 1 \le j \le \widehat{q}_G\}$ denote the set of estimated change points from a single-bandwidth MOSUM procedure, obtained according to either the η- or ϵ-criterion (see Meier et al. (2021b) for their description) with $\eta, \epsilon \in (0, 1)$, where the bandwidth G and the significance level α_n satisfy

$$\min_{0 \le j \le q_n} (\theta_{j+1} - \theta_j) > 2G, \quad \min_{1 \le j \le q_n} \frac{d_j^2 G}{\log(n/G)} \to \infty, \quad \text{and}$$
$$\alpha_n \to 0 \quad \text{with} \quad D_n(G; \alpha_n) = O(\sqrt{\log(n/G)}).$$

We further assume that the invariance principle holds as in Proposition 1 (c.i) with

$$\frac{\lambda_n^2 \log(n/G)}{G} \to 0.$$

Then, there exists a universal constant $C_M > 0$ such that

$$\mathsf{P}\left(\widehat{q}_G = q_n; \max_{1 \le j \le q_n} d_j^2 |k_{G,j} - \theta_j| \le C_M (\omega_n^{(1)})^2\right) \to 1.$$

Proof. First, we need to show that asymptotically, (i) there is exactly one significant local maximum in the G-environment of each change point, and (ii) there are no other significant local maxima. The second assertion follows by Lemma 5.1 (b) of Eichinger and Kirch (2018). Concerning (i), by Lemma D.1 (b), there is only one (and significant by (a)) local maximum within a G-environment of every change point on an asymptotic one set, which also fulfils the ϵ-criterion by Lemma 5.1 (a) in Eichinger and Kirch (2018). Then, the localisation rate follows by the same arguments as in the proof of Proposition 4, completing the proof. □

E CUSUM-based candidate generation

The CUSUM statistic in (3) is designed to test the null hypothesis of no change point ($H_0 : q_n = 0$) against the at-most-one-change alternative ($H_1 : q_n = 1$). It corresponds to the likelihood ratio statistic under i.i.d. Gaussian errors and as such, is particularly appropriate for single change point estimation.

For multiple change point detection, Vostrikova (1981) and Venkatraman (1992) establish the consistency of the Binary Segmentation algorithm that makes recursive use of CUSUM-based estimation. However, its sub-optimality, both in terms of the conditions required for the consistency and the rate of change point localisation, has been noted in Fryzlewicz (2014). As an alternative, he proposes the Wild Binary Segmentation (WBS) which aims at isolating the change points by drawing a large number of random intervals. When a sufficient number of random intervals are drawn, with large probability, there exists at least one interval which is well-suited for the detection and localisation of each θ_j, $j = 1, \ldots, q_n$. Since then, Fryzlewicz (2020) proposes its variation (WBS2) that draws random intervals in a more systematic fashion and generates a complete solution path, while Kovács et al. (2020) propose a 'seeded' version of WBS that constructs the background intervals in a deterministic fashion. In the WBS and its variants, the candidates are generated by scanning the data multiple times over a large number of (randomly drawn) intervals, and various pruning methods have been proposed including thresholding, sequential application of an information criterion (Fryzlewicz, 2014) and the steepest-drop to low levels (SDLL) method (Fryzlewicz, 2020).

We propose the following version of WBS2 as a candidate generating mechanism. It requires the tuning parameters R_n, the maximal number of random intervals to be drawn at each iteration, and $\widetilde{Q}_n$, which relates to the maximal depth of recursion L_n as $L_n = \lfloor \log_2(\widetilde{Q}_n + 1) \rfloor$. The step-by-step description of the WBS2 is provided below.

Step 0: Initialise the input arguments: The set of candidates $\mathcal{K}(R_n, \widetilde{Q}_n) = \emptyset$, $s = 0$, $e = n$ and the recursion depth $\ell = 1$.

Step 1: Quit the routine if $e - s = 1$ or $\ell > L_n$; if not, let $\widetilde{R} = \min\{R_n, (e-s)(e-s-1)/2\}$. If $\widetilde{R} \le R_n$, let $\mathcal{R}_{s,e} = \{(l, r) \in \mathbb{Z}^2 : s \le l < r \le e \text{ and } r - l > 1\}$ serve as $[s_m, e_m]$, $m = 1, \ldots, \widetilde{R}$. If not, draw $\widetilde{R}$ intervals $[s_m, e_m]$, $m = 1, \ldots, \widetilde{R}$, uniformly at random from the set $\mathcal{R}_{s,e}$.

Step 2: Identify $(m_\circ, k_\circ) = \arg\max_{(m,b):\, 1 \le m \le \widetilde{R},\, s_m < b < e_m} |\mathcal{X}_{s_m,b,e_m}|$.

Step 3: Update $\mathcal{K}(R_n, \widetilde{Q}_n)$ by adding $k_\circ$ and store its natural detection interval $\mathcal{I}_N(k_\circ) = (s_\circ, e_\circ]$.

Step 4: Repeat Steps 1–3 separately with $(s, k_\circ, \ell + 1)$ and $(k_\circ, e, \ell + 1)$.

Through implementing the maximal recursion depth into the procedure, it trivially holds that the size of candidate set satisfies $|\mathcal{K}(R_n, \widetilde{Q}_n)| \le \widetilde{Q}_n$. We propose to apply the localised pruning to the thus-generated set of candidates $\mathcal{K}(R_n, \widetilde{Q}_n)$, which satisfies Assumptions 4 and 5 on the set of candidate estimators.

Proposition E.1.

(a) Let $\mathsf{P}(\mathcal{M}_n^{(11)}) \to 1$ where $\mathcal{M}_n^{(11)}$ is defined in Assumption 1 (a). Also, suppose that there exist some $\beta \in (0, 1]$ and $c_\delta \in (0, 1)$ satisfying

$$\min_{1 \le j \le q_n} \delta_j \ge c_\delta n^\beta \quad \text{and} \quad \frac{\omega_n^2}{\min_{1 \le j \le q_n} d_j^2 n^{5\beta - 4}} \to 0 \tag{E.1}$$

where, as before, $\delta_j = \min(\theta_j - \theta_{j-1}, \theta_{j+1} - \theta_j)$ and ω_n is as in Assumption 1 (a). In addition, suppose that

$$\frac{n^{2-2\beta}\log(n)}{R_n} \to 0, \qquad \frac{q_n}{\widetilde{Q}_n} \to 0 \tag{E.2}$$

and let $\rho_n^{(\mathrm{W})} = c_W n^{4-4\beta}\omega_n^2$ for some $c_W \in (0, \infty)$. Then, it holds

$$\mathsf{P}\Big(\max_{1\le j\le q_n} \min_{k\in\mathcal{K}(R_n,\widetilde{Q}_n)} d_j^2|k - \theta_j| \le \rho_n^{(\mathrm{W})}\Big) \to 1.$$

(b) Suppose $n^{-1}\omega_n^2\widetilde{Q}_n \to 0$. Then, for any realisation of the random intervals, we have $n^{-1}\omega_n^2|\mathcal{K}(R_n, \widetilde{Q}_n)| \to 0$.

(c) Suppose that conditions in (a) hold. Then, for each $j = 1, \ldots, q_n$, there exists $\check{k} \in \{k \in \mathcal{K}(R_n, \widetilde{Q}_n) : d_j^2|k - \theta_j| \le \rho_n^{(\mathrm{W})}\}$ such that $\min(\check{k} - \check{s}, \check{e} - \check{k}) \ge c\delta_j$, where $\mathcal{I}_N(\check{k}) = (\check{s}, \check{e}]$ represents the natural detection interval of $\check{k}$ and c is a universal constant satisfying $c \in (0, 1]$.

Remark E.1. For each $k \in \mathcal{K}(R_n, \widetilde{Q}_n)$, the natural detection interval $\mathcal{I}_N(k) = (s, e]$ can serve as its detection interval $\mathcal{I}(k)$ in which case the detection distances are given by $G_L(k) = k - s$ and $G_R(k) = e - k$. Proposition E.1 (c) indicates that $\mathcal{K}(R_n, \widetilde{Q}_n)$ fulfils Assumption 5 under Assumption 2. Besides, by construction, each estimator in $\mathcal{K}(R_n, \widetilde{Q}_n)$ is distinct and therefore $\mathcal{K}(R_n, \widetilde{Q}_n)$ bypasses the issue discussed in Remark 7 (b).

Compared to the condition (a) of Proposition 4 on the minimal size of change, measured by the jump size d_j and spacing δ_j, for the MOSUM-based candidate generating mechanism, the corresponding condition in (E.1) is considerably stronger. Also, the rate of localisation reported in Proposition 4 is always tighter than $\rho_n^{(\mathrm{W})}$ given in the above theorem. The bottleneck in our theoretical analysis of the CUSUM-based candidate generation procedure is the following: The WBS-type procedures looking for the largest CUSUM at each iteration, do not rule out that a change point θ_j is detected by its estimator $k_\circ$ within an interval $(s_\circ, e_\circ)$ which also contains θ_{j-1} or θ_{j+1} (and more) well within the interval. In such a case, the localisation rate $|k_\circ - \theta_j|$ depends not only on d_j but also on the minimum spacing $\min_{1\le j\le q_n}\delta_j$, which results in the sub-optimal localisation rate as well as the detection lower bound given in Proposition E.1. Besides, the theoretical guarantee therein is for the homogeneous change points only. An analogous result is reported in Wang and Samworth (2018a) where the WBS is adopted for high-dimensional change point detection which, to the best of our knowledge, is the best available result on the detection lower bound and the localisation rate of the WBS.

The maximum number of intervals to be drawn at each iteration, R_n, is required to increase as the minimal spacing $\min_{1\le j\le q_n}\delta_j$ decreases (see (E.2)), thus increasing the total computational complexity of the candidate generating procedure as $O(R_n n)$.

The consistency of the localised pruning algorithm in combination with the CUSUM-based candidate generating mechanism follows from Proposition E.1 and Theorem 2.

Theorem E.1. Let Assumptions 1–2 and 3 hold and additionally, let $\xi_n^{-1} n^{4-4\beta} \omega_n^2 \to 0$. Also suppose that the conditions in Proposition E.1 are satisfied. Then, the localised pruning algorithm `LocAlg` applied to $\mathcal{K}(R_n, \widetilde{Q}_n)$ yields $\widehat{\Theta} = \{\widehat{\theta}_j,\, 1 \le j \le \widehat{q}_n : \widehat{\theta}_1 < \ldots < \widehat{\theta}_{\widehat{q}_n}\}$ which consistently estimates Θ, i.e.,

$$\mathsf{P}\left\{ \widehat{q}_n = q_n;\, \max_{1 \le j \le q_n} d_j^2 |\widehat{\theta}_j - \theta_j| \le \rho_n^{(\mathrm{W})} \nu_n \right\} \to 1,$$

with $\rho_n^{(\mathrm{W})}$ as in Proposition E.1 and $\nu_n \to \infty$ arbitrarily slow.

The additional requirement on the penalty ξ_n is necessary due to the localisation rate achieved by the CUSUM-based candidate generation always dominating ω_n^2, such that the penalty needs to be chosen accordingly larger; see also the discussion following (C.1). In view of the discussion below Proposition E.1, we believe that such a requirement on the penalty term cannot be lifted when performing model selection on the candidates generated by a WBS-type method using an information criterion, unless some modification of the WBS such as that proposed in Baranowski et al. (2019) is adopted.
In practice, it is not straightforward to select $\widetilde{Q}_n$ which effectively imposes an upper bound on the number of candidates. For numerical studies in Section 5, instead of selecting $\widetilde{Q}_n$, we choose a weak threshold ζ_n as a multiple of $\sqrt{\log(n)}$, and keep only those candidates for which the corresponding CUSUM statistics (after standardisation) exceed ζ_n. This approach provides more flexibility to deal with heavy-tailedness or serial dependence present in the error sequence.

E.1 Proof of Proposition E.1

Firstly, (b) follows directly from the construction of WBS2 and the condition on $\widetilde{Q}_n$, since

$$|\mathcal{K}(R_n, \widetilde{Q}_n)| \le \sum_{j=0}^{L_n} 2^j = 2^{L_n} - 1 \le \widetilde{Q}_n.$$

The following proof of (a) is an adaptation of the proof of Theorem 3.1 (iii) of Fryzlewicz (2020) and that of Theorem 2 of Wang and Samworth (2018a). Throughout the proof, we adopt C_i, $i \ge 1$ to denote positive constants. Also, $\mathcal{X}_{s,b,e}(f)$ (resp. $\mathcal{X}_{s,b,e}(\varepsilon)$) denotes the CUSUM statistic analogously defined as $\mathcal{X}_{s,b,e}$ in (3) with f_t (ε_t) replacing X_t.
We define the following intervals for $j = 0, \ldots, q_n$,

$$I_j = [r_j, \ell_{j+1}] \qquad \text{where} \quad r_j = \theta_j + \lceil (\theta_{j+1} - \theta_j)/3 \rceil, \quad \ell_{j+1} = \theta_{j+1} - \lceil (\theta_{j+1} - \theta_j)/3 \rceil$$

and for $1 \le u+1 < v \le q_n + 1$,

$$I_{u,v}^{t_1,t_2} = [\max(0, \theta_u + t_1), \min(\theta_v + t_2, n)] \quad \text{with} \quad t_1, t_2 \in [-\underline{\Delta}_n, \underline{\Delta}_n], \text{ where } \underline{\Delta}_n = \frac{\rho_n^{(W)}}{\min_{1 \le j \le n} d_j^2}.$$

Suppose that on each interval $I_{u,v}^{t_1,t_2}$, we draw R_n intervals $\{[s_m, e_m], m = 1, \ldots, R\}$ randomly and uniformly from $\left\{(l, r) \in I_{u,v}^{t_1,t_2} \times I_{u,v}^{t_1,t_2} : l + 1 < r\right\}$. When $R_n \ge |I_{u,v}^{t_1,t_2}|(|I_{u,v}^{t_1,t_2}| - 1)/2$, we use $\{[s_m, e_m], m = 1, \ldots, \widetilde{R}\}$ with $\widetilde{R} = |I_{u,v}^{t_1,t_2}|(|I_{u,v}^{t_1,t_2}| - 1)/2$ which contains all feasible sub-intervals of $I_{u,v}^{t_1,t_2}$. For notational convenience, we do not specify the (stochastic) dependence of (s_m, e_m) on u, v, t_1 or t_2.
For each interval $I_{u,v}^{t_1,t_2}$, consider the event $\mathcal{A}_{u,v}^{t_1,t_2} = \bigcap_{j=u+1}^{v-1} \bigcup_m \{(s_m, e_m) \in I_{j-1} \times I_j\}$. If $\widetilde{R} \le R_n$, we have $\mathsf{P}((\mathcal{A}_{u,v}^{t_1,t_2})^c) = 0$; if not,

$$\mathsf{P}\left((\mathcal{A}_{u,v}^{t_1,t_2})^c\right) \le q_n \prod_{m=1}^{R_n} \max_{u+1 \le j \le v-1} \{1 - \mathsf{P}((s_m, e_m) \in I_{j-1} \times I_j)\} \le q_n \left(1 - \frac{c_\delta^2}{9n^{2-2\beta}}\right)^{R_n}$$

such that for $\Omega_n := \bigcap_{t_1,t_2,u,v} \mathcal{A}_{u,v}^{t_1,t_2}$, by $\log(1-x) \le -x$ for $x \in [0, 1)$,

$$\mathsf{P}(\Omega_n) \ge 1 - \sum_{t_1,t_2,u,v} \mathsf{P}((\mathcal{A}_{u,v}^{t_1,t_2})^c) \ge 1 - \frac{1}{2} q_n(q_n + 1)(q_n + 2)(2\underline{\Delta}_n + 1)^2 \exp\left(-\frac{c_\delta^2 R_n}{9n^{2-2\beta}}\right) \to 1$$

under (E.2). We claim that on $\Omega_n \cap \mathcal{M}_n^{(11)}$,

(W1) at some iteration, if there exist $1 \le u+1 < v \le q_n + 1$ such that s and e satisfy $\max\{d_u^2|s - \theta_u|, d_v^2|e - \theta_v|\} \le \rho_n^{(\mathrm{W})}$,

(W2) the call of Steps 1–3 of WBS2 with such s and e as its arguments adds $k_\circ$ which satisfies $d_j^2|k_\circ - \theta_j| \le \rho_n^{(W)}$ for some $j \in \{u+1, \ldots, v-1\}$.

The condition in (W1) trivially holds at the very first iteration of WBS2 with $s = \theta_0 = 0$ and $e = \theta_{q_n+1} = n$. Then by induction, each θ_j, $j = 1, \ldots, q_n$ is detected by an estimator within $(d_j^{-2}\rho_n^{(\mathrm{W})})$-distance before the depth exceeds $\lceil \log_2(q_n + 1) \rceil + 1$ thanks to (W2), since we add (at most) $2^{\ell-1}$ elements to $\mathcal{K}(R_n, \widetilde{Q}_n)$ at each depth ℓ, which completes the proof of (a).
It remains to show that (W2) holds given that (W1) is met by some s and e. Let

$$(s_\circ, k_\circ, e_\circ) = \arg \max_{(s_m, b, e_m):\, s_m < b < e_m, 1 \le m \le R_n} |\mathcal{X}_{s_m,b,e_m}|.$$

On the event Ω_n, there exists at least one interval $(s_{m(j)}, e_{m(j)}] \in \{(s_m, e_m] \subset (s, e], m = 1, \ldots, \widetilde{R}\}$ satisfying $(s_{m(j)}, e_{m(j)}) \in I_{j-1} \times I_j$ for each $j \in \{u+1, \ldots, v-1\}$, which is non-empty by (W1). Denoting by $k_j^* = \arg\max_{s_{m(j)} < b < e_{m(j)}} |\mathcal{X}_{s_{m(j)},b,e_{m(j)}}|$, we have

$$|\mathcal{X}_{s_\circ,k_\circ,e_\circ}| \ge \max_{u+1 \le j \le v-1} |\mathcal{X}_{s_{m(j)},k_j^*,e_{m(j)}}| \ge \max_{u+1 \le j \le v-1} |\mathcal{X}_{s_{m(j)},\theta_j,e_{m(j)}}|. \tag{E.3}$$

On $\mathcal{M}_n^{(11)}$, it holds as in (C.3)

$$|\mathcal{X}_{s,b,e}(\varepsilon)| \le 2\,\omega_n. \tag{E.4}$$

Also, under (E.1), it follows straightforwardly that

$$(d_j^2\delta_j)^{-1}\omega_n^2 \to 0. \tag{E.5}$$

Then, we have

$$\begin{aligned}|\mathcal{X}_{s_\circ,k_\circ,e_\circ}(f)| &\ge \max_{u+1\le j\le v-1} |\mathcal{X}_{s_{m(j)},\theta_j,e_{m(j)}}| - 2\omega_n \ge \max_{u+1\le j\le v-1} |\mathcal{X}_{s_{m(j)},\theta_j,e_{m(j)}}(f)| - 4\omega_n \\ &\ge \min_{1\le j\le q_n} \frac{\sqrt{d_j^2\delta_j}}{\sqrt{6}} - 4\omega_n > \min_{1\le j\le q_n} \frac{\sqrt{d_j^2\delta_j}}{2\sqrt{6}}\end{aligned} \tag{E.6}$$

by (E.5) for n large enough, which shows in particular that there is at least one change point within $(s_\circ, e_\circ)$.

Let $\theta_\pm$ denote the two change points $\theta_- < k_\circ \le \theta_+$ satisfying $(\theta_-, k_\circ)\cap\Theta = \emptyset$ and $(k_\circ, \theta_+)\cap\Theta = \emptyset$. From (E.6), at least one of $\theta_\pm$ belongs to $(s_\circ, e_\circ)$. If $\theta_+ \notin (s_\circ, e_\circ)$, then by Lemma 8 (b) of Wang and Samworth (2018b), $\mathcal{X}_{s_\circ,b,e_\circ}(f)$ does not change sign and has strictly decreasing absolute values for $\theta_- \le b \le k_\circ$. In this case, we set $\theta_j = \theta_-$. If $\theta_- \notin (s_\circ, e_\circ)$, similarly, $\mathcal{X}_{s_\circ,b,e_\circ}(f)$ does not change sign and has strictly increasing absolute values for $k_\circ \le b \le \theta_+$ (their Lemma 8 (a)), and we set $\theta_j = \theta_+$. If both $\theta_\pm \in (s_\circ, e_\circ)$, by Lemma 8 (c)–(d) of Wang and Samworth (2018b) and Lemma 2.2 of Venkatraman (1992), $\mathcal{X}_{s_\circ,b,e_\circ}(f)$ is either strictly decreasing in modulus without sign change for $\theta_- \le b \le k_\circ$, or strictly increasing in modulus without sign change for $k_\circ \le b \le \theta_+$. In the first case, we set $\theta_j = \theta_-$ while in the latter, we set $\theta_j = \theta_+$.

If the thus-identified $\theta_j \ge k_\circ$, we consider the time series in reverse such that w.l.o.g., we suppose that $k_\circ \ge \theta_j$ and $|\mathcal{X}_{s_\circ,b,e_\circ}(f)|$ is strictly decreasing between θ_j and $k_\circ$. In addition, we assume that $\mathcal{X}_{s_\circ,k_\circ,e_\circ} > 0$; otherwise, consider $-X_t$ (resp. $-f_t$ and $-\varepsilon_t$) in place of X_t (f_t and ε_t).

Then, by (E.4), (E.5) and the arguments analogous to those adopted in (E.6), we yield $|\mathcal{X}_{s_\circ,k_\circ,e_\circ}(\varepsilon)|/\mathcal{X}_{s_\circ,k_\circ,e_\circ} = o(1)$ and in particular,

$$\mathcal{X}_{s_\circ,k_\circ,e_\circ}(f) > 0 \tag{E.7}$$

for large enough n. Also from (E.3)–(E.4) and by the construction of $(s_{m(j)}, e_{m(j)}]$, we yield

$$\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) \ge \mathcal{X}_{s_\circ,k_\circ,e_\circ}(f) \ge \mathcal{X}_{s_\circ,k_\circ,e_\circ} - 2\omega_n \ge \mathcal{X}_{s_{m(j)},\theta_j,e_{m(j)}}(f) - 4\omega_n$$

$$\geq \frac{\sqrt{d_j^2 \delta_j}}{\sqrt{6}} - 4\omega_n \geq \frac{\sqrt{d_j^2 \delta_j}}{2\sqrt{6}} \tag{E.8}$$

under (E.5). Besides, since $\mathcal{X}_{s_\circ,k_\circ,e_\circ} \geq \mathcal{X}_{s_\circ,\theta_j,e_\circ}$, it holds

$$\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(f) \leq \mathcal{X}_{s_\circ,k_\circ,e_\circ}(\varepsilon) - \mathcal{X}_{s_\circ,\theta_j,e_\circ}(\varepsilon) \tag{E.9}$$

and further, the positivity of the LHS of (E.9) implies

$$1 \leq \frac{|\mathcal{X}_{s_\circ,\theta_j,e_\circ}(\varepsilon) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(\varepsilon)|}{\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(f)}. \tag{E.10}$$

Using the notations adopted in the proof of Proposition C.3, we denote $\mathcal{X}_{s,b,e}(\varepsilon) = \sqrt{\mathcal{W}_b}\mathcal{E}_b$ (suppressing the dependence on s and e). Then,

$$|\mathcal{X}_{s_\circ,k_\circ,e_\circ}(\varepsilon) - \mathcal{X}_{s_\circ,\theta_j,e_\circ}(\varepsilon)| \leq \left|\sqrt{\mathcal{W}_{k_\circ}} - \sqrt{\mathcal{W}_{\theta_j}}\right| \, |\mathcal{E}_{\theta_j}| + \sqrt{\mathcal{W}_{k_\circ}}|\mathcal{E}_{\theta_j} - \mathcal{E}_{k_\circ}|.$$

By the mean value theorem,

$$\left|\sqrt{\mathcal{W}_{k_\circ}} - \sqrt{\mathcal{W}_{\theta_j}}\right| \leq \frac{\sqrt{2}(k_\circ - \theta_j)}{\min(\theta_j - s_\circ, e_\circ - \theta_j)^{3/2}}.$$

Also, on $\mathcal{M}_n^{(11)}$,

$$|\mathcal{E}_{\theta_j}| = \left| \frac{e_\circ - \theta_j}{e_\circ - s_\circ} \sum_{t=s_\circ+1}^{\theta_j} \varepsilon_t - \frac{\theta_j - s_\circ}{e_\circ - s_\circ} \sum_{t=\theta_j+1}^{e_\circ} \varepsilon_t \right| \leq \sqrt{2\min(\theta_j - s_\circ, e_\circ - \theta_j)}\, \omega_n,$$

$$|\mathcal{E}_{\theta_j} - \mathcal{E}_{k_\circ}| = \left| \sum_{t=\theta_j+1}^{k_\circ} \varepsilon_t - \frac{k_\circ - \theta_j}{e_\circ - s_\circ} \sum_{t=s_\circ+1}^{e_\circ} \varepsilon_t \right| \leq \sqrt{k_\circ - \theta_j}\, \omega_n + \frac{k_\circ - \theta_j}{\sqrt{e_\circ - s_\circ}}\, \omega_n \leq 2\sqrt{k_\circ - \theta_j}\, \omega_n.$$

Combining the above, we arrive at

$$|\mathcal{X}_{s_\circ,k_\circ,e_\circ}(\varepsilon) - \mathcal{X}_{s_\circ,\theta_j,e_\circ}(\varepsilon)| \leq \frac{2\,(k_\circ - \theta_j)}{\min(\theta_j - s_\circ, e_\circ - \theta_j)}\, \omega_n + \sqrt{\frac{8(k_\circ - \theta_j)}{\min(k_\circ - s_\circ, e_\circ - k_\circ)}}\, \omega_n. \tag{E.11}$$

The proof proceeds by considering the following possible scenarios.

Case 1: There is at least one change point to the right of θ_j in $(s_\circ, e_\circ)$, i.e. $\theta_{j+1} < e_\circ$, and $\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) \geq \mathcal{X}_{s_\circ,\theta_{j+1},e_\circ}(f)$. Adopting the arguments in the proof of Theorem 2 in Wang and Samworth (2018a) under their Case 2 (b), we can show that $\theta_j - s_\circ \geq c_1\delta_j$ for some universal constant $c_1 \in (0, 1]$; otherwise, we cannot have $\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) \geq \mathcal{X}_{s_\circ,\theta_{j+1},e_\circ}(f)$. This ensures that

$j \in \{u+1, \ldots, v-1\}$. Then,

$$\begin{aligned}\frac{c_1\delta_j \mathcal{X}_{s_\circ,\theta_j,e_\circ}(f)(k_\circ - \theta_j)}{2n^2} &\le \mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(f) \\ &\le |\mathcal{X}_{s_\circ,\theta_j,e_\circ}(\varepsilon) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(\varepsilon)| \le 4\omega_n\end{aligned}$$

where the first inequality follows from Lemma 9 of Wang and Samworth (2018b) (with $\mathcal{X}_{s_\circ,s_\circ+t,e_\circ}(f)$, $e_\circ - s_\circ$, $\theta_j - s_\circ$, $\theta_{j+1} - s_\circ$ and $c_1\delta_j/n$ taking the roles of $g(t)$, n, z, z' and τ therein, respectively), the second from (E.9) and the last from (E.4). Together with (E.8) and that $\min(\theta_j - s_\circ, e_\circ - \theta_j) \ge c_1\delta_j$, we obtain

$$k_\circ - \theta_j \le 24\sqrt{6}(c_1|d_j|\delta_j^{3/2})^{-1}n^2\omega_n < \frac{c_1}{2}\delta_j \tag{E.12}$$

under (E.1) for n large enough which, together with (E.11), leads to

$$|\mathcal{X}_{s_\circ,k_\circ,e_\circ}(\varepsilon) - \mathcal{X}_{s_\circ,\theta_j,e_\circ}(\varepsilon)| \le 8\sqrt{\frac{k_\circ - \theta_j}{\min(\theta_j - s_\circ, e_\circ - \theta_j)}}\,\omega_n. \tag{E.13}$$

Then, combining this with (E.8)–(E.10), we yield

$$1 \le \frac{|\mathcal{X}_{s_\circ,\theta_j,e_\circ}(\varepsilon) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(\varepsilon)|}{\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(f)} \le \frac{8\omega_n\sqrt{(k_\circ - \theta_j)/\min(\theta_j - s_\circ, e_\circ - \theta_j)}}{c_1\delta_j\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f)(k_\circ - \theta_j)/(2n^2)} \le \frac{32\sqrt{6}n^2\omega_n}{\sqrt{c_1^3 d_j^2 \delta_j^4(k_\circ - \theta_j)}}$$

such that under (E.1), we can find some fixed c_W for $\rho_n^{(\mathrm{W})} = c_W n^{4-4\beta}(\omega_n)^2$ satisfying $d_j^2(k_\circ - \theta_j) \le 6144 c_1^{-3}\delta_j^{-4}n^4\omega_n^2 \le \rho_n^{(\mathrm{W})}$.

Case 2: $\theta_{j+1} < e_\circ$ and $\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) < \mathcal{X}_{s_\circ,\theta_{j+1},e_\circ}(f)$. In this case, from Lemma 8 (d) of Wang and Samworth (2018b), $\mathcal{X}_{s_\circ,b,e_\circ}(f)$ strictly decreases and then increases for $\theta_j \le b \le \theta_{j+1}$ without changing sign, and thus we can find $\tau := \max\{\theta_j + 1 \le b \le \theta_{j+1} : \mathcal{X}_{s_\circ,b,e_\circ}(f) \le \mathcal{X}_{s_\circ,\theta_{j+1},e_\circ}(f) - 4\omega_n\}$. Adopting the arguments in the proof of Theorem 2 in Wang and Samworth (2018a) under their Case 2 (c), we have $e_\circ - \theta_{j+1} \ge c_1\delta_{j+1}$, which in turn leads to $\theta_{j+1} - \tau + 1 < c_1\delta_{j+1}/2$. Since by construction and the first line of (E.8) we get

$$\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) \ge \mathcal{X}_{s_\circ,\theta_{j+1},e_\circ}(f) - 4\omega_n \ge \mathcal{X}_{s_\circ,\tau,e_\circ}(f),$$

we can then adopt the same argument as in Case 1 and prove the claim, by applying Lemma 9 of Wang and Samworth (2018b) with $\mathcal{X}_{s_\circ,s_\circ+t,e_\circ}(f)$, $e_\circ - s_\circ$, $\theta_j - s_\circ$, $\tau - s_\circ$ and $c_2\delta_j/n$ for some $c_2 \in (0, c_1/2]$ taking the roles of $g(t)$, n, z, z' and τ in the lemma, respectively.

Case 3: There is no change point to the right of θ_j in $(s_\circ, e_\circ)$, i.e. $\theta_{j+1} \ge e_\circ$. We first establish that $\min(\theta_j - s_\circ, e_\circ - \theta_j) \ge \min(c_2, c_3)\delta_j$ for c_2 introduced under Case 2 and some $c_3 \in (0, 1/24]$, which ensures that $j \in \{u+1, \ldots, v-1\}$. To this end, consider the following two cases: (a) $\theta_{j-1} \le s_\circ$ and (b) $\theta_{j-1} > s_\circ$. Under (a), if $\min(\theta_j - s_\circ, e_\circ - \theta_j) < c_3\delta_j$, by construction and from (E.4)–(E.5) we yield

$$\mathcal{X}_{s_\circ,k_\circ,e_\circ} \le \sqrt{c_3 d_j^2 \delta_j} + 2\omega_n < |\mathcal{X}_{s_{m(j)},\theta_j,e_{m(j)}}(f)| - 2\omega_n \le |\mathcal{X}_{s_{m(j)},k_j^*,e_{m(j)}}|$$

for large enough n, which contradicts (E.3). Under (b), we have either $\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) \ge \mathcal{X}_{s_\circ,\theta_{j-1},e_\circ}(f)$ or not. In either situations, applying the arguments borrowed from Wang and Samworth (2018b) under Cases 1–2 in the reverse direction, we can establish that $e_\circ - \theta_j \ge c_2\delta_j$. Next, define $\vartheta = \frac{1}{\theta_j - s_\circ}\sum_{t=s_\circ+1}^{\theta_j} f_t - f_{\theta_j+1}$. Then,

$$\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) \le \vartheta\sqrt{\min(\theta_j - s_\circ, e_\circ - \theta_j)}.$$

Applying Lemma 7 of Wang and Samworth (2018b) with $e_\circ - s_\circ$ and $\theta_j - s_\circ$ taking the roles of n and z in the lemma, respectively, we obtain

$$\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(f) \ge \frac{2\vartheta(k_\circ - \theta_j)}{3\sqrt{6\min(\theta_j - s_\circ, e_\circ - \theta_j)}} \ge \frac{2\,\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f)(k_\circ - \theta_j)}{3\sqrt{6}\min(\theta_j - s_\circ, e_\circ - \theta_j)}. \quad \text{(E.14)}$$

Combining (E.4), (E.5), (E.8), (E.9) and (E.14),

$$k_\circ - \theta_j \le 72(d_j^2\delta_j)^{-1/2}\omega_n \min(\theta_j - s_\circ, e_\circ - \theta_j) \le \frac{1}{2}\min(\theta_j - s_\circ, e_\circ - \theta_j)$$

for large enough n. Then, (E.8) and (E.10) with (E.13) yields

$$\begin{aligned}1 &\le \frac{|\mathcal{X}_{s_\circ,\theta_j,e_\circ}(\varepsilon) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(\varepsilon)|}{\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f) - \mathcal{X}_{s_\circ,k_\circ,e_\circ}(f)} \le \frac{4\omega_n\sqrt{(k_\circ - \theta_j)/\min(\theta_j - s_\circ, e_\circ - \theta_j)}}{\mathcal{X}_{s_\circ,\theta_j,e_\circ}(f)(k_\circ - \theta_j)/\{3\sqrt{6}\min(\theta_j - s_\circ, e_\circ - \theta_j)\}} \\ &\le \frac{144\omega_n\sqrt{\min(\theta_j - s_\circ, e_\circ - \theta_j)}}{\sqrt{d_j^2\delta_j(k_\circ - \theta_j)}} \le \frac{144\,\omega_n}{\sqrt{d_j^2(k_\circ - \theta_j)}}\end{aligned}$$

by noting that $\delta_j \ge e_\circ - \theta_j \ge \min(\theta_j - s_\circ, e_\circ - \theta_j)$ under Case 3. Therefore, there exists some large $c_W > 0$ such that $d_j^2(k_\circ - \theta_j) \le 144^2\,\omega_n^2 \le \rho_n^{(\mathrm{W})}$.

In all Cases 1–3, we have established that $\min(\theta_j - s_\circ, e_\circ - \theta_j) \ge \min(c_2, c_3)\delta_j$ (recalling that $c_2 < c_1/2$). From that $d_j^2|k_\circ - \theta_j| \le \rho_n^{(\mathrm{W})}$ and (E.1), we yield

$$\frac{|k_\circ - \theta_j|}{\delta_j} \le \frac{c_W d_j^{-2} n^{4-4\beta}\omega_n^2}{c_\beta n^\beta} \to 0$$

as $n \to \infty$. Hence, $\min(k_\circ - s_\circ, e_\circ - k_\circ) \ge \{\min(c_2, c_3) + o(1)\}\delta_j$ and we conclude that (c) holds with some $c \in (0, \min(c_2, c_3))$.

F Computational complexity

Analysing the computational complexity of the localised pruning is challenging without further assumption on the number of the candidates analysed at each iteration (denoted by $\mathcal{D}$ in Step 2 of `LocAlg`). In implementing the algorithm, we impose a fixed upper bound of $N = 24$ on $|\mathcal{D}|$; if $|\mathcal{D}| > N$ at a particular iteration, we modify the order in which the candidates remaining in $\mathcal{C}$ are processed which often resolves the issue. Theorem 1 holds irrespective of which candidate is chosen in Step 1 of `LocAlg`, and thus this step does not harm the theoretical guarantee. To guard against the contingency where all other candidates in $\mathcal{C}$ also have more than N conflicting candidates, a manual thinning step for the set $\mathcal{D}$ is implemented in the R package `mosum` which triggers a warning message, see Appendix A of Meier et al. (2021b) for further details. In practice, this manual thinning step is rarely activated; for example, for the **dense** test signals with frequent change points and $n \ge 2 \times 10^4$ considered in simulations (see Section G.1), we did not encounter a single occurrence over 1000 realisations for each test signal. Since there are at most $O(n \log^{-1}(n))$ candidates in total (see Assumption 4 (b)), the localised pruning requires $O(n + 2^N n \log^{-1}(n))$ operations in the worst case.

The MOSUM-based candidate generating procedure discussed in Section 4 requires $O(n|\mathcal{H}|)$ operations, where $\mathcal{H}$ denotes the set of asymmetric bandwidths. In Section 4, we propose a scheme for bandwidth generation which ensures the adaptivity of the multiscale MOSUM procedure while bounding the total number of bandwidths to be considered at $|\mathcal{H}| = O(\log(n))$ through a condition on the balancedness of asymmetric bandwidths (see (9)), which amounts to the computation time of $O(n \log(n))$ for the multiscale MOSUM procedure. The computational complexity of the CUSUM-based candidate generation depends on the number R_n of the random intervals drawn as $O(R_n n)$, which in turn needs to increase in $(\min_j \delta_j)^{-1}$ as $n^2 (\min_j \delta_j)^{-2} \log(n)/R_n \to 0$ (see (E.2)) for adaptivity.

In summary, with the MOSUM-based candidate generating mechanism, the combined two-stage methodology requires $O(n \log(n) + 2^N n \log^{-1}(n))$ operations in total, which is much faster than many competitors requiring dynamic programming-type solutions (such as those proposed in Frick et al. (2014), Wang et al. (2020) and Fromont et al. (2020)) whose computational complexity is $O(n^2)$, see also Table 1 for the summary of the computational complexity of various methods for univariate data segmentation.

G Complete simulation results

G.1 Set-up

G.1.1 Models

We consider the five test signals from Fryzlewicz (2014) referred to as `blocks`, `fms`, `mix`, `teeth10` and `stairs10`, see Appendix B therein for further details. In addition, we include the following test signals extending the original ones in order to investigate the scalability of the localised pruning algorithm:

Dense test signals. Each test signal is concatenated until the length of the resultant signal exceeds 2×10^4.

Sparse test signals: Each test signal is embedded in the series of i.i.d. random variables of length $n = 2 \times 10^4$ at $t = 500$.

For ε_t, we consider

($\mathcal{E}1$) independent Gaussian random variables as in Fryzlewicz (2014),

($\mathcal{E}2$) independent random variables following the t_5 distribution, and

($\mathcal{E}3$) AR(1) processes with Gaussian innovations and the AR parameter $\varrho \in \{0.3, 0.9\}$,

while keeping the signal-to-noise ratio defined by $\{\mathsf{Var}(\varepsilon_t)\}^{-1/2} \min_{1 \le j \le q_n} |d_j|$ constant across different error distributions. Under ($\mathcal{E}3$), in order to account for the information loss due to the serial dependence, the length of each segment between adjacent change points is increased by the factor of $\lfloor 1/(1-\varrho) \rfloor$.

G.1.2 Tuning parameter selection

We apply the localised pruning algorithm outlined in Section 3.1.1 together with the two candidate generating mechanisms described in Section 4 and Section E.

For the localised pruning procedure, the main tuning parameter is ξ_n for the penalty of SC. To investigate the sensitivity of the localised pruning with respect to its choice, we consider $\xi_n \in \{\log^{1.01}(n), \log^{1.1}(n)\}$ for ($\mathcal{E}1$); $\xi_n \in \{\log^{1.1}(n), n^{2/4.99}\}$ for ($\mathcal{E}2$) as well as $\xi_n \in \{\log^{1.1}(n), \log^2(n)\}$ for ($\mathcal{E}3$), referred to as the 'light' and 'heavy' penalties respectively. These penalty terms are chosen in line with the discussion below Assumption 3 on how the choice of ξ_n should reflect the behaviour of $\{\varepsilon_t\}_{t=1}^n$. As a candidate sorting function for the `LocAlg`; we consider $h_{\mathcal{J}}(k)$ in (5), which is readily available for any multiscale candidate generation methods, and $h_{\mathcal{P}}(k)$ that is compatible with the MOSUM-based candidate generation method. Our theoretical results do not depend on the choice of a particular sorting function and this is confirmed by the simulation studies. In practice, the use of $h_{\mathcal{P}}$ may slow down the pruning algorithm by generating many ties when many of the p-values are artificially set to zero by the machine (see Appendix B of Meier et al. (2021b)).

Each candidate generation method requires additional tuning parameters.

MOSUM-based candidate generation ('MoLP'): We use the multiscale extension of the MOSUM procedure with the asymmetric bandwidths $\mathcal{H}$ selected by the automatic bandwidth generation procedure described in Section 4, setting $G_0 = 10$ in the case of ($\mathcal{E}1$)–($\mathcal{E}2$) and $G_0 = \max(10, \lfloor 8/(1-\varrho) \rfloor)$ in the case of ($\mathcal{E}3$), and $C_{\text{asym}} = 4$. We justify the choice

$G_0 = 10$ by noting that our interest lies in change point detection rather than outlier detection, working under the assumption that $\min_{1 \le j \le q_n} \delta_j \to \infty$ as $n \to \infty$. Since asymmetric bandwidths are adopted for their small sample performance only, we do not anticipate different values for C_{asym} would change the result greatly, provided that the bandwidths are not too unbalanced.

In deriving the asymptotic critical value, we consider $\alpha \in \{0.1, 0.2\}$ except when the change points are dense, we consider $\alpha \in \{0.2, 0.4\}$ to ensure that $\mathcal{K}(\mathcal{H}, \alpha)$ meets Assumption 4 (a). Later, we observe that the choice of α does not influence the results greatly with one exception (`teeth10` requires a slightly more generous choice for higher true positive rate (TPR) when there are dense change points, see Table 4); as a rule of thumb, we recommend $\alpha = 0.2$ as a generous enough choice to ensure that the candidate set includes estimators for all the change points, while any spurious estimators can be removed by the pruning. We set $\eta = 0.4$ for locating the change points according to the η-criterion; while not reported here, we have considered different values of η and found that its choice has little influence on the results when used in combination with the pruning step.

For variance estimation, we adopt the scale-dependent, MOSUM variance estimator $\widehat{\sigma}^2_{k,G} = (2G)^{-1}\{\sum_{t=k-G+1}^{k}(X_t - \bar{X}_{(k-G+1):k})^2 + \sum_{t=k+1}^{k+G}(X_t - \bar{X}_{(k+1):(k+G)})^2\}$ proposed in Eichinger and Kirch (2018), in the case of the independent errors in ($\mathcal{E}1$)–($\mathcal{E}2$). When serial dependence is present under ($\mathcal{E}3$), we use the MOSUM variance estimator inflated by the factor of $(1 + \widehat{\varrho})/(1 - \widehat{\varrho})$ with an estimator of the AR parameter $\widehat{\varrho}$. For this, we first generate candidates via the multiscale MOSUM procedure. Here, through using the MOSUM variance estimator without any correction, the procedure is expected not to under-estimate the number of change points. Then, $\widehat{\varrho}$ is obtained as the Yule-Walker estimator from the resultant residuals, which is fed into correct the MOSUM variance estimator as above.

MoLP is implemented in the R package `mosum` (Meier et al., 2021a).

CUSUM-based candidate generation ('CuLP'): We select the number of random intervals for each recursion of WBS2 as recommended in Fryzlewicz (2020). Instead of selecting an upper bound $\widetilde{Q}_n$ on its cardinality, we use the following subset of $\mathcal{K}(R_n, \widetilde{Q}_n = n)$

$$\mathcal{K}(R_n, n, \zeta_n) = \left\{k_\circ : k_\circ \in \mathcal{K}(R_n, \widetilde{Q}_n = n) \text{ with the corresponding } |\mathcal{X}_{s_\circ, k_\circ, e_\circ}| \ge \zeta_n\right\},$$

which provides more flexibility with respect to the choice of the threshold ζ_n. In addition, for numerical stability in local variance estimation (described below), we consider only those $k_\circ \in \mathcal{K}(R_n, n, \zeta_n)$ with $\min(k_\circ - s_\circ, e_\circ - k_\circ) \ge 5$. As the thresholding is intended only to remove 'obviously' spurious estimators (corresponding to small $|\mathcal{X}_{s_\circ, k_\circ, e_\circ}|$), we can choose ζ_n quite small even in the presence of heavy-tailed or serially dependent errors. We use $\zeta_n = C_\zeta \cdot K \widehat{\tau}_n \sqrt{2\log(n)}$, where K is a constant chosen as per Fryzlewicz (2020). The deflation factor C_ζ is set at $C_\zeta = 0.9$, except when change points are dense, in which case we also consider stronger deflation by $C_\zeta = 0.5$ to ensure that $\mathcal{K}(R_n, n, \zeta_n)$ meets Assumption 4 (a).

As in MoLP, we estimate the (long-run) variance using a local estimator extending the MOSUM variance estimator of Eichinger and Kirch (2018): for ($\mathcal{E}1$)–($\mathcal{E}2$), it is obtained as the sample variance of the residuals over $\mathcal{I}(k)$ after fitting a stump function with a break at the candidate change point k; for ($\mathcal{E}3$), we inflate the local variance estimator by the factor of $(1+\widehat{\varrho})/(1-\widehat{\varrho})$.

G.1.3 Competing methods

The following competitors are considered for the comparative study.

1. The multiscale MOSUM procedure with the 'bottom-up' merging (`bottom.up`) implemented in the R package `mosum` (Meier et al., 2021a) (see Messer et al. (2014) and also Meier et al. (2021b)).

2. WBS (Fryzlewicz, 2014) applied with the strengthened Bayesian information criterion (WBS.sBIC, implemented in the R package `breakfast` (Anastasiou et al., 2021)). For the generation of random intervals, the same approach as that in CuLP is taken.

3. WBS2.SDLL proposed in Fryzlewicz (2020), whose implementation is available on `https://github.com/pfryz/wild-binary-segmentation-2.0`).

4. Pruned exact linear time (PELT) algorithm of Killick et al. (2012) (R package `changepoint` (Killick et al., 2016)).

5. The dynamic programming algorithm based on functional pruning (S3IB) proposed in Rigaill (2015) (R package `Segmentor3IsBack` (Cleynen et al., 2016)).

6. Tail-greedy unbalanced Haar (TGUH) algorithm of Fryzlewicz (2018) (R package `breakfast` (Anastasiou et al., 2021)).

7. FDRSeg (Li et al., 2016), the multiscale segmentation method controlling the false discovery rate (R package `FDRSeg` (Li and Sieling, 2017)).

8. cumSeg (Muggeo and Adelfio, 2010), the method based on transforming the data and iteratively fitting a linear model (R package `cumSeg` (Muggeo, 2012)).

Unless stated otherwise, we apply the above methods with default choices of parameters recommended by the authors. Additionally, we consider:

9. Functional pruning optimal partitioning (FPOP) algorithm of Maidstone et al. (2017) (R package `FPOP` (Rigaill and Hocking, 2019)) with the penalty set at $\sqrt{2\log(n)}$ is considered for the test signals with $n \ge 2 \times 10^4$.

10. Jump segmentation for dependent data (JUSD) of Tecuapetla-Gómez and Munk (2017) and DepSMUCE of Dette et al. (2020) are considered for ($\mathcal{E}3$). the latter extending the simultaneous multiscale change point estimator (SMUCE) (Frick et al., 2014) to the dependent case (both implemented using the R package `stepR` (Pein et al., 2019)). For JUSD, the estimator of the long-run variance relies on the assumption of m-dependence yet there does not exist an automatic way of determining m; instead we use $m = [\log(0.1)/\log(\varrho)]$ utilising the typically unavailable knowledge of ϱ; For DepSMUCE, the recommended choice of block length $K = 10$ often severely under-estimates the long-run variance, and thus we supply $K = [\log(0.1)/\log(\varrho)]$.

Note that PELT, S3IB and FPOP set out to solve the ℓ_0-penalised least squares estimation problem; theoretical investigation into the performance of such an approach is provided in Wang et al. (2020). Many of the algorithms mentioned above are specifically tailored for the data with i.i.d. innovations following sub-Gaussian distributions, with the exception of cumSeg, JUSD and DepSMUCE.

G.2 Results

All simulations are based on 1000 replications.
We define that a change point θ_j is detected if there exists at least one estimator that falls between $\max\{(\theta_j+\theta_{j-1})/2, \theta_j - \bar{\delta}\}$ and $\min\{(\theta_j+\theta_{j+1})/2, \theta_j+\bar{\delta}\}$, where $\bar{\delta} = \min_{1\le j\le q_n-1}(\theta_{j+1} - \theta_j)$. Based on this, we report the true positive rate (TPR, the proportion of the correctly identified change point out of the q_n true change points) and false positive rate (FPR, the proportion of the spurious estimators out of the $\widehat{q}$ estimated change points). Also reported are the Adjusted Rand Index (ARI) measuring the similarity between the estimated and true segmentations (Rand, 1971; Hubert and Arabie, 1985), the relative mean squared error (MSE) of the estimated piecewise constant signal to that of the signals estimated using the true change points, Bayesian information criterion (BIC) with the penalty term $\log(n)$, and the weighted average of trimmed distances $\delta_{\text{trim}} = (\sum_{j=1}^{q_n} d_j^2)^{-1} \sum_{j=1}^{q_n} d_j^2 \cdot \delta_{\text{trim},j}$ where

$$\delta_{\text{trim},j} = \min\left\{\frac{\theta_{j+1}-\theta_j}{2}, \frac{\theta_j-\theta_{j-1}}{2}, \min_{1\le j'\le\widehat{q}} |\widehat{\theta}_{j'} - \theta_j|\right\}, \tag{G.1}$$

averaged over 1000 replications. Also, we provide $v_{\text{trim}} = q_n^{-1}\sum_{j=1}^{q_n} \text{MAD}(\delta_{\text{trim},j})$, where the MAD operator is taken over 1000 replications for each change point θ_j. Finally, for the dense and sparse test signals, we report the average execution time.

($\mathcal{E}1$) Independent Gaussian errors

Tables 3–5 report the simulation results in the presence of independent Gaussian errors for the original five test signals and their dense and sparse versions. Figures 3–12 visualise the performance of various methods by plotting the weighted densities of estimated change point(s) falling between two adjacent change points $[(\theta_{j-1} + \theta_j)/2 + 1, (\theta_j + \theta_{j+1})/2]$ for $j = 1, \ldots, q_n$. Table 3 indicate that choices of α for the MoLP or the sorting function h and the penalty ξ_n for the localised pruning algorithm do not greatly influence the results. In particular, with n relatively small (≤ 2048), the choice of penalty ξ_n does not alter the results much. Difference in performance due to these choices are more apparent when n is large ($\ge 2 \times 10^4$), see Tables 4–5.

When change points are dense, a lighter penalty ξ_n and a generous choice of the critical value for the candidate generation method (larger α for the MoLP, smaller C for the CuLP) are preferable for some test signals such as `teeth10`, which ensures that the candidate set contains at least one valid estimator for each θ_j (Assumption 4 (a)). On the other hand, when the change points are sparse, a heavier penalty ξ_n is successful in removing spurious false positives over a long stretch of stationary observations without harming the TPR much. Between the two methods equipped with different candidate generating methods CuLP tends to incur more false positives than the MoLP. Overall, the MoLP produces estimators of better localisation accuracy, possibly benefiting from the systematic approach to candidate generation adopted by the multiscale MOSUM procedure. This is also reflected on the execution time of the two methods when the change points are dense.

`bottom.up`, compared to the MoLP, tends to return many false positives. This reflects the corresponding theoretical requirements on the MOSUM procedure, that the significance level is small ($\alpha = \alpha_n \to 0$, Eichinger and Kirch (2018)) and that the bandwidths are in the order of n (Messer et al., 2014), and the problem is further amplified with increasing n (see Table 5) and heavy-tailed errors as observed under ($\mathcal{E}2$). An interesting phenomenon is observed in Figure 4 which plots the weighted densities of estimated change points for the `fms` test signal, where `bottom.up` incurs several false positives systematically. This is attributed to spurious estimators detected with large bandwidths between the first and the second change points.

There is no single method that outperforms the rest universally for all test signals and evaluation criteria. While S3IB marginally outperforms other competitors in terms of TPR, it is at the price of larger FPR. FPOP, another functional pruning algorithm, is computationally fast and generally performs well, but fails at handling the teeth-like jump structure of `teeth10` (see Tables 4–5). WBS2.SDLL shows its strength in handling frequent changes, although returning marginally more false positives compared to other methods achieving comparable TPR. Both PELT and cumSeg tend to under-estimate the number of change points across all test signals and so does WBS.sBIC. The latter result indicates that minimisation of an information criterion along a solution path is not as efficient as the pruning criteria (C1)–(C2) adopted by `PrunAlg`, both computationally or empirical performance-wise. Interestingly, when the frequent changes in `teeth10` are repeated over $n \ge 2 \times 10^4$ observations, the BIC is minimised at the null model (Table 4), further suggesting that the sequential minimisation of

BIC often leads to less favourable results compared to `PrunAlg`.

In terms of computation time, FPOP, PELT and `bottom.up` take less than 0.1 seconds to process a long signal. It is followed by the MoLP and TGUH, demonstrating that the localised pruning is scalable to long signals. While CuLP tends to be slower than MoLP, it still surpasses WBS.sBIC and WBS2.SDLL in this respect (except for the dense `block` signal), which demonstrates the computational gain achievable by the localised exhaustive search adopted in the proposed methodology. FDRSeg and S3IB, while showing good performance for short test signals, are computationally too expensive for long signals and, along with cumSeg, are omitted in these situations.

Meier et al. (2021b) observed that for the MoLP, the ordering function $h_{\mathcal{P}}$ incurs many ties as the p-values associated with candidates detected at larger bandwidths are set exactly to be zero by the machine, which increases the search space for the inner algorithm `PrunAlg` and consequently slows down the pruning procedure. As there is no meaningful difference in terms of change point detection accuracy, we recommend the use of $h_{\mathcal{J}}$.

Table 3: Summary of change point estimation over 1000 realisations for the test signals with Gaussian errors: we use $\xi_n \in \{\log^{1.01}(n), \log^{1.1}(n)\}$ as the 'light' and 'heavy' penalties for the localised pruning.

model	α	penalty	method	TPR	FPR	ARI	MSE	BIC	δ_{trim}	v_{trim}
`blocks`	0.1	light	MoLP-$h_{\mathcal{P}}$	0.954	0.009	0.977	5.155	4784.242	351.119	262.916
			MoLP-$h_{\mathcal{J}}$	0.955	0.009	0.978	5.003	4783.629	332.611	246.996
		heavy	MoLP-$h_{\mathcal{P}}$	0.944	0.004	0.977	5.231	4784.409	377.367	195.167
			MoLP-$h_{\mathcal{J}}$	0.945	0.004	0.978	5.1	4783.634	347.61	179.247
	0.2	light	MoLP-$h_{\mathcal{P}}$	0.96	0.014	0.975	5.217	4784.947	365.521	258.931
			MoLP-$h_{\mathcal{J}}$	0.961	0.014	0.977	4.911	4783.747	327.638	227.091
		heavy	MoLP-$h_{\mathcal{P}}$	0.949	0.006	0.977	5.08	4784.11	356.655	195.167
			MoLP-$h_{\mathcal{J}}$	0.949	0.005	0.978	5.01	4783.597	338.395	163.327
		light	CuLP	0.934	0.095	0.919	10.091	4805.107	1001.623	284.352
		heavy	CuLP	0.936	0.033	0.95	7.631	4794.349	708.949	197.160
	0.2	-	`bottom.up`	0.958	0.278	0.877	6.308	4812.993	372.152	309.686
	-	-	WBS.sBIC	0.938	0.032	0.962	7.304	4795.326	694.854	264.426
	-	-	WBS2.SDLL	0.94	0.027	0.971	5.51	4785.457	359.36	195.167
	-	-	PELT	0.878	0.001	0.961	6.413	4785.848	588.644	220.480
	-	-	S3IB	0.974	0.019	0.979	4.773	4782.984	306.041	186.930
	-	-	cumSeg	0.772	0.002	0.914	13.119	4818.988	1743.155	555.444
	-	-	TGUH	0.948	0.023	0.967	6.589	4788.875	488.462	342.805
	0.2	-	FDRSeg	0.975	0.081	0.956	5.367	4788.694	328.394	235.020
`fms`	0.1	light	MoLP-$h_{\mathcal{P}}$	0.982	0.015	0.954	4.402	-564.883	0.175	0.168
			MoLP-$h_{\mathcal{J}}$	0.981	0.015	0.955	4.356	-564.894	0.175	0.151
		heavy	MoLP-$h_{\mathcal{P}}$	0.98	0.009	0.955	4.407	-564.878	0.178	0.168
			MoLP-$h_{\mathcal{J}}$	0.979	0.01	0.955	4.354	-564.897	0.178	0.168
	0.2	light	MoLP-$h_{\mathcal{P}}$	0.99	0.02	0.958	4.138	-565.219	0.148	0.151

			MoLP-$h_{\mathcal{J}}$	0.99	0.021	0.957	4.119	-565.183	0.148	0.151
		heavy	MoLP-$h_{\mathcal{P}}$	0.989	0.012	0.958	4.129	-565.219	0.152	0.151
			MoLP-$h_{\mathcal{J}}$	0.988	0.012	0.959	4.064	-565.258	0.15	0.151
		light	CuLP	0.997	0.149	0.905	5.379	-562.971	0.137	0.033
		heavy	CuLP	0.997	0.074	0.937	4.446	-565.116	0.139	0.033
	0.2	-	`bottom.up`	0.976	0.32	0.836	6.266	-548.575	0.312	0.151
	-	-	WBS.sBIC	0.975	0.014	0.96	4.747	-564.272	0.235	0.103
	-	-	WBS2.SDLL	0.995	0.032	0.955	4.187	-566.031	0.139	0.033
	-	-	PELT	0.934	0.001	0.954	5.016	-565.769	0.389	0.033
	-	-	S3IB	0.999	0.1	0.944	4.98	-566.096	0.101	0.033
	-	-	cumSeg	0.754	0.012	0.918	14.05	-549.512	1.841	0.103
	-	-	TGUH	0.995	0.04	0.945	4.822	-565.036	0.15	0.067
	0.2	-	FDRSeg	0.998	0.086	0.953	4.441	-564.844	0.113	0.033
`mix`	0.1	light	MoLP-$h_{\mathcal{P}}$	0.911	0.007	0.738	4.195	842.617	30.552	18.818
			MoLP-$h_{\mathcal{J}}$	0.913	0.007	0.74	4.178	842.562	29.944	18.818
		heavy	MoLP-$h_{\mathcal{P}}$	0.9	0.003	0.717	4.262	842.654	30.849	23.436
			MoLP-$h_{\mathcal{J}}$	0.901	0.004	0.716	4.238	842.554	30.361	23.436
	0.2	light	MoLP-$h_{\mathcal{P}}$	0.929	0.009	0.772	4.096	842.634	30.252	16.765
			MoLP-$h_{\mathcal{J}}$	0.93	0.009	0.772	4.083	842.564	29.729	16.765
		heavy	MoLP-$h_{\mathcal{P}}$	0.916	0.005	0.749	4.178	842.633	30.459	17.791
			MoLP-$h_{\mathcal{J}}$	0.916	0.005	0.748	4.14	842.547	29.933	18.818
		light	CuLP	0.937	0.054	0.788	4.844	845.765	43.738	17.905
		heavy	CuLP	0.926	0.026	0.77	4.609	844.302	40.569	15.738
	0.2	-	`bottom.up`	0.951	0.064	0.805	4.326	848.366	32.832	19.160
	-	-	WBS.sBIC	0.817	0.034	0.638	9.916	869.485	131.693	18.533
	-	-	WBS2.SDLL	0.91	0.021	0.735	4.562	843.571	35.944	18.304
	-	-	PELT	0.771	0.002	0.461	6.148	846.354	48.85	12.659
	-	-	S3IB	0.96	0.074	0.815	4.774	843.513	33.146	20.642
	-	-	cumSeg	0.333	0	0.273	25.195	904.25	752.167	87.473
	-	-	TGUH	0.902	0.026	0.702	5.374	845.653	47.727	30.336
	0.2	-	FDRSeg	0.936	0.075	0.775	4.951	846.699	36.313	16.765
`teeth10`	0.1	light	MoLP-$h_{\mathcal{P}}$	0.95	0.001	0.92	2.337	-73.202	0.333	0.000
			MoLP-$h_{\mathcal{J}}$	0.95	0.001	0.92	2.337	-73.202	0.333	0.000
		heavy	MoLP-$h_{\mathcal{P}}$	0.944	0	0.912	2.421	-73.173	0.362	0.000
			MoLP-$h_{\mathcal{J}}$	0.944	0	0.912	2.421	-73.173	0.362	0.000
	0.2	light	MoLP-$h_{\mathcal{P}}$	0.97	0.001	0.945	1.986	-73.584	0.235	0.000
			MoLP-$h_{\mathcal{J}}$	0.97	0.001	0.945	1.986	-73.584	0.235	0.000
		heavy	MoLP-$h_{\mathcal{P}}$	0.965	0.001	0.938	2.077	-73.552	0.263	0.000
			MoLP-$h_{\mathcal{J}}$	0.965	0.001	0.938	2.077	-73.552	0.263	0.000
		light	CuLP	0.985	0.017	0.904	3.65	-76.47	0.463	0.000
		heavy	CuLP	0.979	0.011	0.899	3.702	-76.476	0.488	0.000
	0.2	-	`bottom.up`	0.983	0.004	0.965	1.813	-73.084	0.164	0.000
	-	-	WBS.sBIC	0.644	0.02	0.579	9.065	-71.534	2.029	1.140

	-	-	WBS2.SDLL	0.977	0.023	0.896	3.879	-76.254	0.501	0.000
	-	-	PELT	0.391	0.007	0.287	13.038	-69.041	3.194	0.342
	-	-	S3IB	0.997	0.101	0.902	4.039	-76.144	0.392	0.000
	-	-	cumSeg	0.001	0	0	18.287	-63.097	4.995	0.000
	-	-	TGUH	0.961	0.018	0.867	4.385	-75.328	0.631	0.000
	0.2	-	FDRSeg	0.958	0.061	0.859	4.511	-75.015	0.623	0.000
`stairs10`	0.1	light	MoLP-$h_{\mathcal{P}}$	0.998	0.002	0.979	2.097	-120.634	0.103	0.000
			MoLP-$h_{\mathcal{J}}$	0.998	0.002	0.979	2.097	-120.634	0.103	0.000
		heavy	MoLP-$h_{\mathcal{P}}$	0.998	0.001	0.979	2.091	-120.63	0.103	0.000
			MoLP-$h_{\mathcal{J}}$	0.998	0.001	0.979	2.096	-120.629	0.103	0.000
	0.2	light	MoLP-$h_{\mathcal{P}}$	0.998	0.002	0.979	2.097	-120.634	0.103	0.000
			MoLP-$h_{\mathcal{J}}$	0.998	0.002	0.979	2.097	-120.634	0.103	0.000
		heavy	MoLP-$h_{\mathcal{P}}$	0.998	0.001	0.979	2.091	-120.63	0.103	0.000
			MoLP-$h_{\mathcal{J}}$	0.998	0.001	0.979	2.096	-120.629	0.103	0.000
		light	CuLP	0.999	0.021	0.961	2.924	-120.123	0.172	0.000
		heavy	CuLP	0.999	0.012	0.963	2.874	-120.195	0.174	0.000
	0.2	-	`bottom.up`	0.997	0.005	0.978	2.094	-119.81	0.104	0.000
	-	-	WBS.sBIC	1	0.034	0.959	2.95	-120.052	0.165	0.000
	-	-	WBS2.SDLL	0.998	0.014	0.958	3.085	-119.333	0.196	0.000
	-	-	PELT	0.993	0.001	0.966	2.729	-120.597	0.175	0.000
	-	-	S3IB	1	0.09	0.953	3.165	-120.419	0.134	0.000
	-	-	cumSeg	0.986	0.006	0.878	7.533	-95.203	0.639	0.424
	-	-	TGUH	0.999	0.009	0.963	2.93	-120.113	0.178	0.000
	0.2	-	FDRSeg	1	0.059	0.957	3.013	-119.874	0.146	0.000

Table 4: Summary of change point estimation over 1000 realisations for the test signals with **dense** change points and Gaussian errors; we use $h = h_{\mathcal{J}}$ and $\xi_n \in \{\log^{1.01}(n), \log^{1.1}(n)\}$ for the localised pruning.

model	α/C_ζ	penalty	method	TPR	FPR	ARI	MSE	BIC	δ_{trim}	v_{trim}	speed
`blocks`	0.2	light	MoLP	0.935	0.005	0.981	5.011	48093.8	2.313	220.54	0.660
	0.2	heavy	MoLP	0.91	0.001	0.979	5.473	48095.6	2.743	235.905	0.675
	0.4	light	MoLP	0.937	0.007	0.98	5.051	48098.09	2.308	218.629	0.778
	0.4	heavy	MoLP	0.912	0.002	0.979	5.472	48098.12	2.674	228.336	0.808
	0.5	light	CuLP	0.863	0.018	0.899	15.378	48513.02	9.978	290.71	16.819
	0.5	heavy	CuLP	0.873	0.003	0.934	11.312	48333.58	7.139	246.124	15.557
	0.9	light	CuLP	0.933	0.007	0.977	5.408	48105.5	2.528	200.072	4.978
	0.9	heavy	CuLP	0.904	0.002	0.97	6.285	48124.13	3.362	211.04	5.090
	0.2	-	`bottom.up`	0.914	0.206	0.887	6.468	48360.28	2.76	274.792	0.050
	-	-	WBS.sBIC	0.908	0.034	0.955	7.892	48242.17	4.946	212.22	77.772
	-	-	WBS2.SDLL	0.951	0.063	0.962	5.319	48143.77	2.101	204.006	5.028
	-	-	PELT	0.81	0	0.955	8.098	48128.61	6.336	293.799	0.029
	-	-	TGUH	0.919	0.005	0.974	6.32	48131.39	3.241	287.898	1.497

	-	-	FPOP	0.931	0.002	0.983	4.782	48076.92	2.331	198.34	0.010
`fms`	0.2	light	MoLP	0.98	0.003	0.97	4.222	-22251.66	0.539	0.124	1.186
	0.2	heavy	MoLP	0.968	0.001	0.968	4.563	-22248.33	0.676	0.155	1.187
	0.4	light	MoLP	0.985	0.003	0.971	4.032	-22257.81	0.48	0.093	1.301
	0.4	heavy	MoLP	0.973	0.001	0.969	4.392	-22253.62	0.623	0.142	1.325
	0.5	light	CuLP	0.973	0.006	0.962	4.291	-22238.88	0.533	0.055	4.571
	0.5	heavy	CuLP	0.941	0.001	0.95	5.403	-22175.93	0.823	0.086	4.568
	0.9	light	CuLP	0.986	0.002	0.976	3.625	-22300.42	0.41	0.046	3.992
	0.9	heavy	CuLP	0.973	0.001	0.973	3.988	-22294.87	0.547	0.049	3.990
	0.2	-	`bottom.up`	0.906	0.28	0.728	7.563	-21260.04	1.615	0.392	0.054
	-	-	WBS.sBIC	0.923	0.004	0.965	15.452	-21006.43	1.892	0.079	74.857
	-	-	WBS2.SDLL	0.997	0.017	0.973	3.605	-22278.99	0.321	0.053	6.443
	-	-	PELT	0.74	0	0.945	9.446	-22159.09	2.756	0.057	0.026
	-	-	TGUH	0.986	0.002	0.963	4.099	-22258.01	0.459	0.103	1.433
	-	-	FPOP	0.958	0.001	0.977	3.859	-22335.57	0.627	0.033	0.010
`mix`	0.2	light	MoLP	0.879	0.002	0.678	4.211	31852.09	0.786	11.113	1.004
	0.2	heavy	MoLP	0.852	0.001	0.634	4.431	31835.61	0.871	12.482	1.014
	0.4	light	MoLP	0.887	0.002	0.695	4.154	31863.58	0.76	10.792	1.105
	0.4	heavy	MoLP	0.858	0.001	0.647	4.401	31843.89	0.855	12.38	1.134
	0.5	light	CuLP	0.906	0.005	0.733	4.364	31905.35	0.786	19.431	8.807
	0.5	heavy	CuLP	0.868	0.003	0.667	4.779	31878.13	0.941	11.417	9.170
	0.9	light	CuLP	0.837	0.003	0.631	6.9	32114.51	1.671	12.574	7.334
	0.9	heavy	CuLP	0.739	0.002	0.511	11.338	32547.75	3.198	21.751	11.187
	0.2	-	`bottom.up`	0.887	0.025	0.705	4.385	32016.71	0.784	21.159	0.060
	-	-	WBS.sBIC	0.676	0	0.464	11.937	32905.97	3.445	22.272	73.173
	-	-	WBS2.SDLL	0.908	0.013	0.735	4.338	31917.1	0.751	18.889	9.400
	-	-	PELT	0.625	0	0.34	9.266	32013.18	2.766	5.743	0.032
	-	-	TGUH	0.821	0.002	0.564	5.593	31877.43	1.212	29.959	1.420
	-	-	FPOP	0.803	0.002	0.541	5.105	31766.6	1.142	31.329	0.011
`teeth10`	0.2	light	MoLP	0.784	0	0.694	5.785	-2301.625	1.187	0	1.068
	0.2	heavy	MoLP	0.592	0	0.492	9.512	-3259.876	2.177	0.047	1.054
	0.4	light	MoLP	0.821	0	0.743	5.124	-2138.238	1.004	0	1.121
	0.4	heavy	MoLP	0.639	0	0.542	8.704	-2987.338	1.936	0	1.108
	0.5	light	CuLP	0.903	0.004	0.814	4.904	-1962.88	0.84	0	4.319
	0.5	heavy	CuLP	0.751	0.002	0.631	7.505	-2478.257	1.56	0	4.349
	0.9	light	CuLP	0.688	0.003	0.532	8.439	-2617.247	1.885	0.414	4.156
	0.9	heavy	CuLP	0.438	0.001	0.309	12.302	-4266.959	3.111	0.234	4.185
	0.2	-	`bottom.up`	0.847	0	0.799	5.01	-1478.868	0.879	0	0.091
	-	-	WBS.sBIC	0	0	0	16.854	-8928.946	5.382	0	69.941
	-	-	WBS2.SDLL	0.932	0.005	0.859	4.612	-1831.672	0.712	0	10.689
	-	-	PELT	0	0.001	0	16.851	-8927.371	5.379	0	0.018
	-	-	TGUH	0.594	0.003	0.329	9.436	-3818.443	2.331	1.594	1.389

	-	-	FPOP	0.114	0.004	0.036	15.588	-7955.71	4.747	0	0.012
`stairs10`	0.2	light	MoLP	0.998	0	0.977	3.103	-6871.387	0.13	0	8.376
	0.2	heavy	MoLP	0.997	0	0.976	2.339	-6902.29	0.139	0	8.566
	0.4	light	MoLP	0.998	0	0.977	3.103	-6871.408	0.13	0	8.416
	0.4	heavy	MoLP	0.997	0	0.976	2.339	-6902.319	0.139	0	8.574
	0.5	light	CuLP	0.989	0.001	0.948	3.966	-6282.889	0.311	0	9.660
	0.5	heavy	CuLP	0.983	0.001	0.944	4.222	-6265.687	0.34	0	9.607
	0.9	light	CuLP	0.988	0.001	0.948	3.982	-6279.911	0.313	0	9.768
	0.9	heavy	CuLP	0.982	0.001	0.944	4.239	-6262.338	0.342	0	9.777
	0.2	-	`bottom.up`	0.994	0.004	0.976	2.243	-6683.881	0.126	0	0.170
	-	-	WBS.sBIC	0.985	0.011	0.946	4.016	-5946.003	0.303	0	70.928
	-	-	WBS2.SDLL	0.992	0.009	0.95	3.775	-6354.621	0.293	0	7.340
	-	-	PELT	0.87	0	0.866	9.5	-5624.565	0.888	0	0.028
	-	-	TGUH	0.991	0	0.961	3.165	-6725.094	0.225	0	1.377
	-	-	FPOP	0.99	0	0.968	2.803	-6892.929	0.189	0	0.010

Table 5: Summary of change point estimation over 1000 realisations for the test signals with **sparse** change points and Gaussian errors; we set $\alpha = 0.2$ (for MoLP and `bottom.up`) and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n \in \{\log^{1.01}(n), \log^{1.1}(n)\}$ for the localised pruning.

model	penalty	method	TPR	FPR	ARI	MSE	BIC	δ_{trim}	v_{trim}	speed
`blocks`	light	MoLP	0.93	0.019	0.928	5.501	46,137.59	2.361	204.58	0.262
	heavy	MoLP	0.906	0.004	0.986	5.909	46,137.37	2.882	204.58	0.264
	light	CuLP	0.936	0.043	0.862	5.85	46,139.66	2.25	178.918	1.399
	heavy	CuLP	0.913	0.006	0.977	5.796	46,137.67	2.561	178.918	1.435
	-	`bottom.up`	0.918	0.454	0.146	8.264	46,218.74	2.831	262.171	0.042
	-	WBS.sBIC	0.91	0.004	0.998	6.512	46,141.24	4.047	178.918	62.783
	-	WBS2.SDLL	0.915	0.018	0.945	5.982	46,139.19	2.38	178.918	8.163
	-	PELT	0.811	0.001	0.999	8.588	46,140.96	6.272	615.151	0.022
	-	TGUH	0.92	0.007	0.998	6.849	46,141.3	3.236	324.322	1.361
	-	FPOP	0.931	0.002	0.999	5.1	46,135.74	2.336	188.331	0.013
`fms`	light	MoLP	0.954	0.031	0.907	5.569	-24,026.74	0.805	0.151	0.272
	heavy	MoLP	0.941	0.007	0.981	5.527	-24,027.52	0.949	0.151	0.272
	light	CuLP	0.984	0.063	0.801	5.205	-24,025.83	0.465	0.033	1.145
	heavy	CuLP	0.969	0.008	0.973	4.658	-24,028.14	0.618	0.033	1.295
	-	`bottom.up`	0.909	0.635	0.063	11.633	-23,932.78	1.595	0.372	0.047
	-	WBS.sBIC	0.757	0.013	0.933	50.855	-23,928.6	4.764	0.338	70.233
	-	WBS2.SDLL	0.978	0.023	0.92	4.857	-24,026.59	0.552	0.103	9.007
	-	PELT	0.751	0	0.999	10.924	-24,025.52	2.722	0.068	0.025
	-	TGUH	0.965	0.002	0.995	5.548	-24,024.01	0.652	0.136	1.560
	-	FPOP	0.96	0.001	0.998	4.066	-24,029.76	0.619	0.033	0.013
`mix`	light	MoLP	0.885	0.015	0.883	4.724	27,834.89	0.797	12.431	0.218
	heavy	MoLP	0.862	0.004	0.922	4.934	27,834.06	0.882	12.431	0.218

	light	CuLP	0.879	0.041	0.763	5.434	27,837.56	0.906	12.431	2.057
	heavy	CuLP	0.845	0.006	0.901	5.554	27,834.88	1.04	14.256	2.759
	-	`bottom.up`	0.905	0.372	0.063	6.041	27,902.7	0.781	16.252	0.043
	-	WBS.sBIC	0.638	0.002	0.863	14.56	27,873.07	3.43	21.099	62.190
	-	WBS2.SDLL	0.847	0.021	0.835	6.169	27,840.44	1.087	16.081	8.049
	-	PELT	0.665	0	0.844	9.118	27,839.65	2.255	13.8	0.039
	-	TGUH	0.535	0.006	0.682	48.116	28,025.52	12.786	35.24	1.377
	-	FPOP	0.834	0.002	0.922	5.133	27,832.15	0.992	12.431	0.017
`teeth10`	light	MoLP	0.738	0.02	0.914	6.857	-18,210.56	1.745	0	0.213
	heavy	MoLP	0.639	0.006	0.981	8.82	-18,212.2	2.601	0.114	0.217
	light	CuLP	0.783	0.045	0.788	7.588	-18,207.1	1.848	0	0.981
	heavy	CuLP	0.671	0.011	0.957	9.464	-18,210.74	3.091	1.026	1.294
	-	`bottom.up`	0.848	0.386	0.048	6.646	-18,139.38	2.621	0	0.041
	-	WBS.sBIC	0.42	0.006	0.995	13.294	-18,215.21	3.806	0	63.106
	-	WBS2.SDLL	0.825	0.038	0.811	6.99	-18,206.21	1.483	0	7.996
	-	PELT	0.164	0.015	0.999	16.11	-18,213.93	4.682	0	0.022
	-	TGUH	0.8	0.007	0.994	6.922	-18,210.9	1.652	0	1.402
	-	FPOP	0.444	0.01	0.999	12.511	-18,219.81	3.354	0.342	0.013
`stairs10`	light	MoLP	0.996	0.016	0.898	2.599	-23,948.07	0.139	0	0.227
	heavy	MoLP	0.989	0.003	0.973	2.707	-23,948.82	0.174	0	0.229
	light	CuLP	0.974	0.035	0.792	6.56	-23,934.51	0.628	0	0.990
	heavy	CuLP	0.966	0.006	0.964	6.536	-23,936.69	0.668	0	1.055
	-	`bottom.up`	0.994	0.32	0.05	3.375	-23,886.67	0.161	0	0.042
	-	WBS.sBIC	0.988	0.01	0.987	4.655	-23,943.36	0.385	0	63.537
	-	WBS2.SDLL	0.981	0.018	0.906	5.169	-23,941.01	0.524	0	8.097
	-	PELT	0.955	0	1	4.625	-23,947.28	0.396	0	0.013
	-	TGUH	0.943	0.002	0.991	9.605	-23,915.32	7.433	0	1.382
	-	FPOP	0.998	0	1	2.521	-23,949.57	0.152	0	0.013

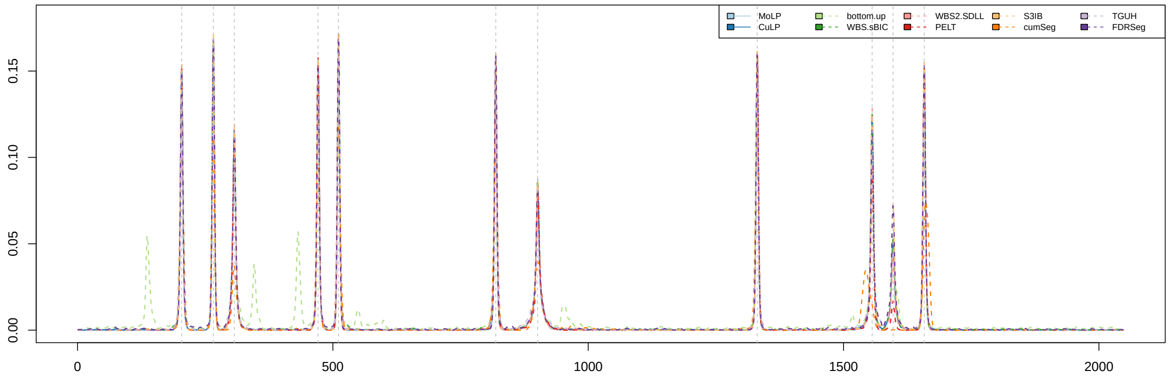

Figure 3: Test signal `blocks` with Gaussian errors: weighted density of estimated change points over $[(\theta_{j-1} + \theta_j)/2, (\theta_j + \theta_{j+1})/2]$, $j = 1, \ldots, q_n$, with the vertical lines indicating the locations of true change points. We set $\alpha = 0.2$ for MoLP and `bottom.up` and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n = \log^{1.01}(n)$ for the localised pruning.

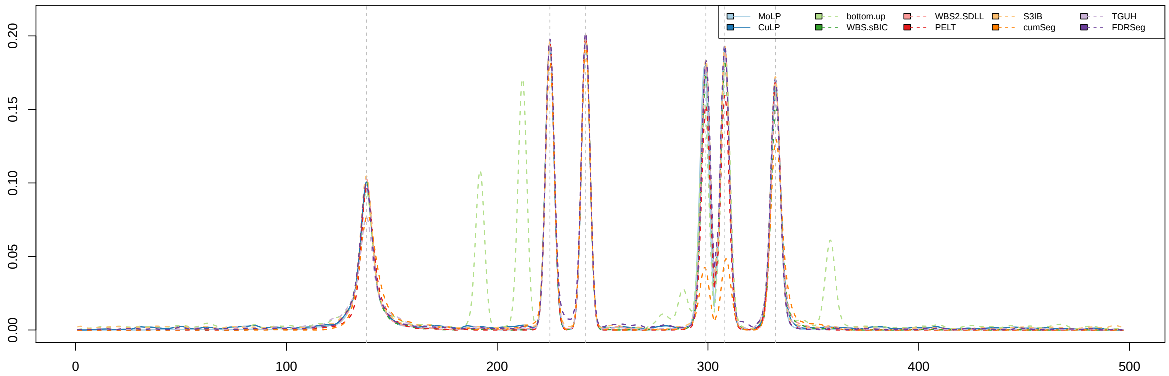

Figure 4: Test signal `fms` with Gaussian errors: weighted density of estimated change points.

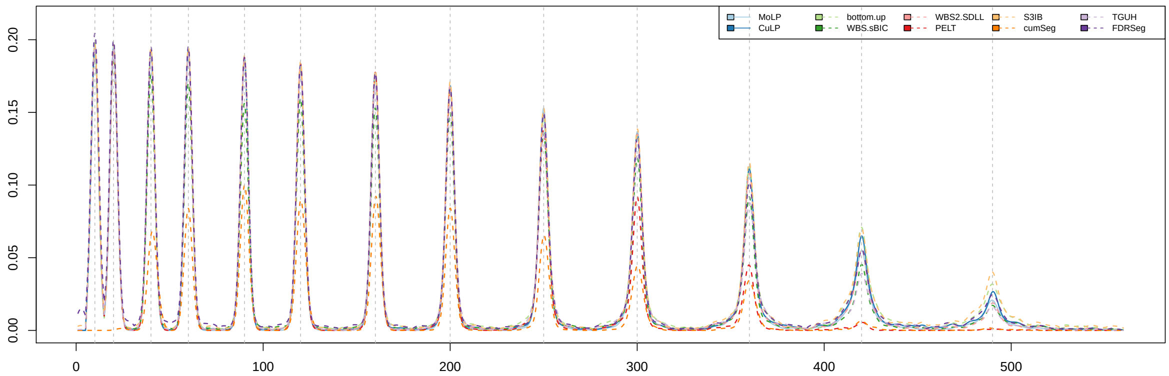

Figure 5: Test signal `mix` with Gaussian errors: weighted density of estimated change points.

($\mathcal{E}2$) Independent heavy-tailed errors

Tables 6–7 report the results when $\varepsilon_t \sim_{\text{iid}} t_5$. Figures 13–17 visualise the performance of various methods by plotting the weighted densities of estimated change point(s).

The heavy penalty $\xi_n = n^{2/4.99}$ is a theoretically valid choice conforming to Assumption 3

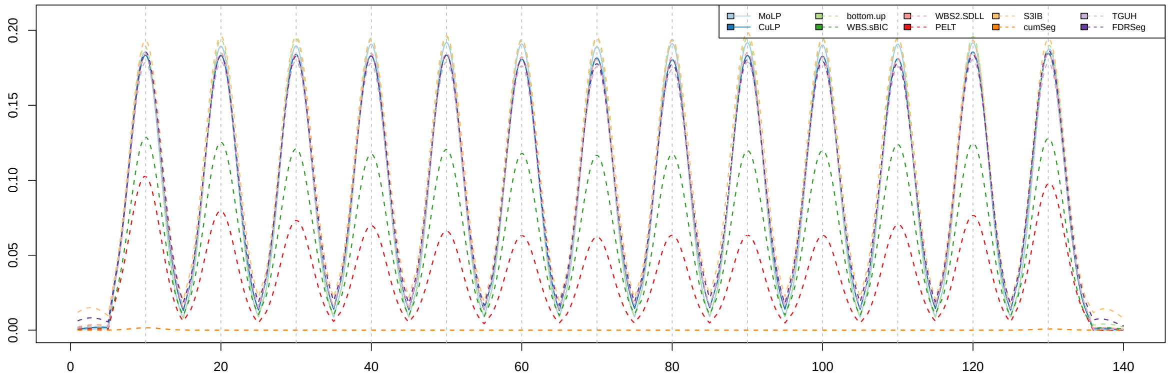

Figure 6: Test signal `teeth10` with Gaussian errors: weighted density of estimated change points.

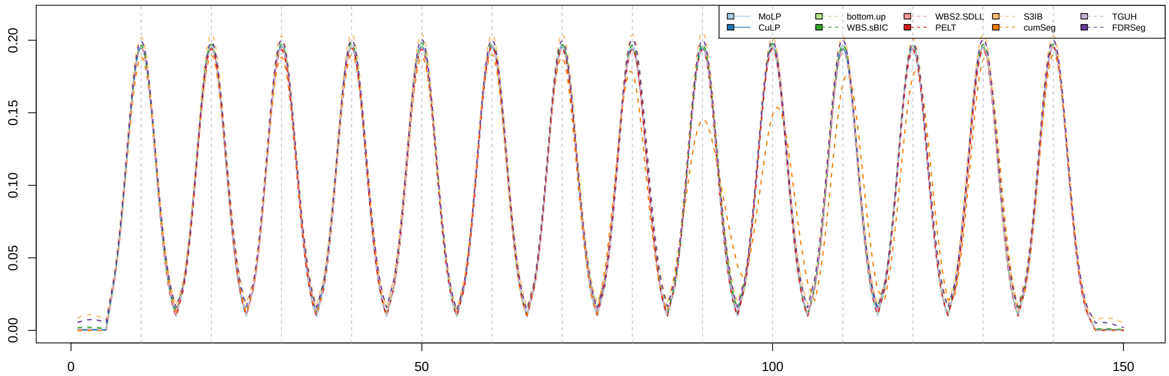

Figure 7: Test signal `stairs10` with Gaussian errors: weighted density of estimated change points.

in light of Remark 2 (b). When n is small (Table 6), this penalty successfully prevents false positives but the resulting procedure lacks power. The light penalty $\xi_n = \log^{1.1}(n)$

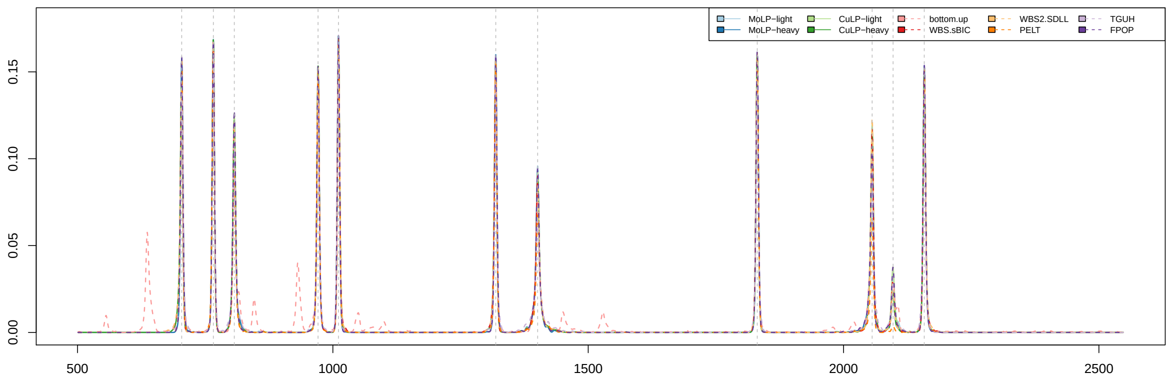

Figure 8: Long test signal `blocks` with sparse change points and Gaussian errors: weighted density of estimated change points with the vertical lines indicating the locations of true change points. We set $\alpha = 0.2$ for MoLP and `bottom.up` and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n \in \{\log^{1.01}(n), \log^{1.1}(n)\}$ for the localised pruning.

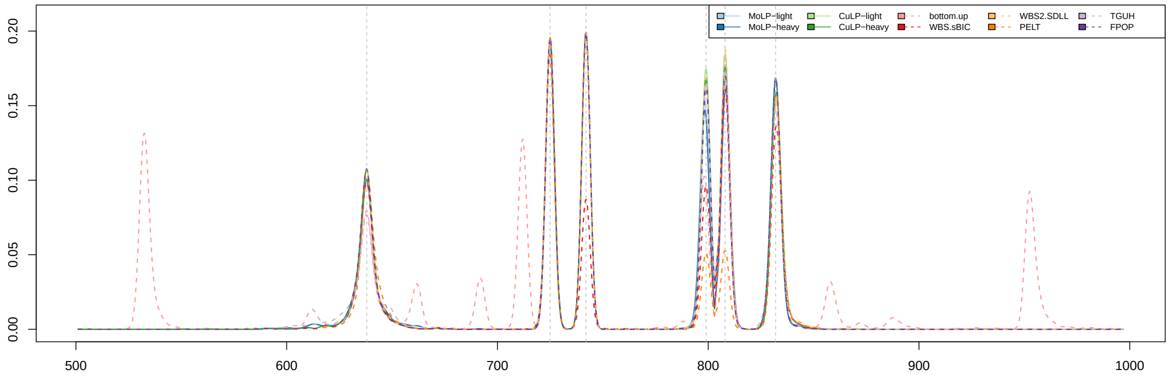

Figure 9: Long test signal `fms` with sparse change points and Gaussian errors: weighted density of estimated change points.

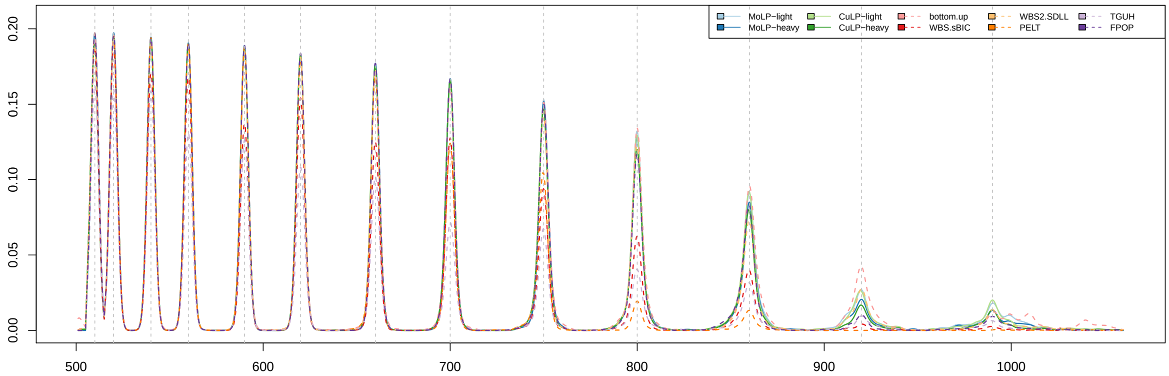

Figure 10: Long test signal `mix` with **sparse** change points and Gaussian errors: weighted density of estimated change points.

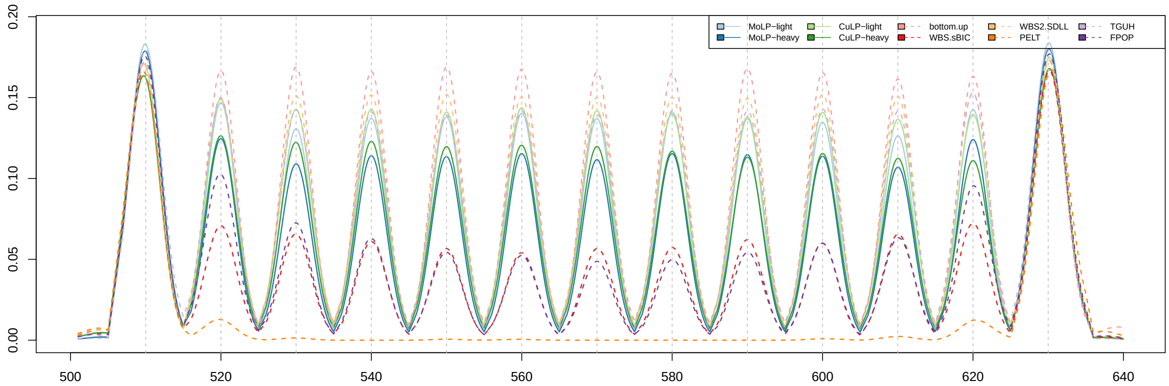

Figure 11: Long test signal `teeth10` with sparse change points and Gaussian errors.

works reasonably well in not causing false positives while attaining high TPR, yielding the performance comparable to that observed with Gaussian errors (see Table 3). When n is large and change points are sparse (Table 7), the localised pruning under-estimates the number of change points for some test signals such as `fms`, `teeth10` and `stairs10`. This, in part, is due to that the candidate generating method fails to produce at least one valid estimator for each true

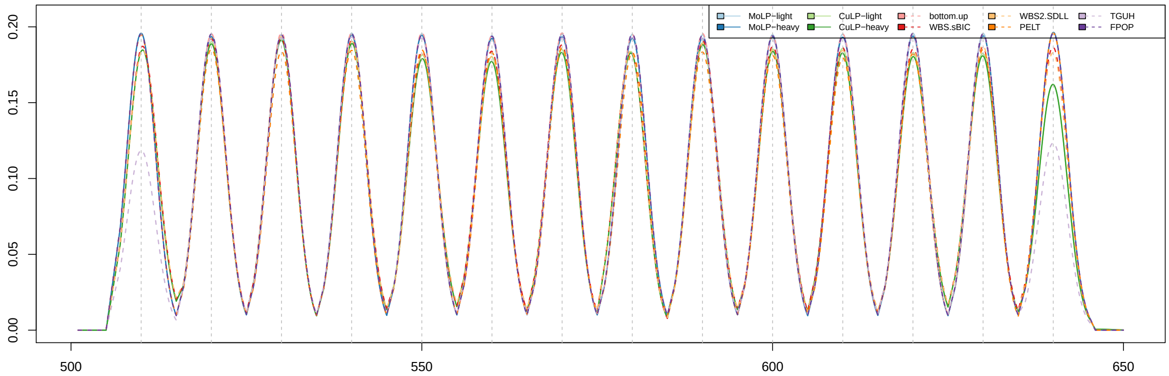

Figure 12: Long test signal `stairs10` with sparse change points and Gaussian errors.

change point, e.g. compare the TPR for MoLP (resp. CuLP) and `bottom.up` (WBS2.SDLL), thus failing Assumption 4 (a). In addition, for theoretical consistency, Assumption 2 requires the magnitude of changes to be larger for their detection in the presence of heavy-tailed errors whereas it is kept at the same level as in ($\mathcal{E}1$) with Gaussian errors. Most of the competitors are tailored for sub-Gaussian errors, and they incur considerable false positives, a phenomenon that is amplified in Table 7 as n is large and change points sparse.

Table 6: Summary of change point estimation over 1000 realisations for the test signals with t_5 errors; we set $\alpha = 0.2$ for MoLP and `bottom.up` and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n \in \{\log^{1.1}(n), n^{2/4.99}\}$ for the localised pruning.

model	penalty	method	TPR	FPR	ARI	MSE	BIC	δ_{trim}	v_{trim}
`blocks`	light	MoLP	0.948	0.005	0.98	4.76	4781.907	312.528	49.512
	heavy	MoLP	0.775	0	0.928	10.072	4796.438	808.116	261.309
	light	CuLP	0.943	0.013	0.974	5.255	4783.241	366.154	110.24
	heavy	CuLP	0.743	0	0.908	11.676	4800.744	1360.405	208.862
	-	`bottom.up`	0.95	0.217	0.921	6.295	4805.876	390.389	297.751
	-	WBS.sBIC	0.902	0.248	0.891	16.965	4785.934	852.116	195.167
	-	WBS2.SDLL	0.974	0.611	0.694	28.436	4808.934	288.258	111.779
	-	PELT	0.927	0.17	0.935	13.346	4761.393	398.671	57.462
	-	S3IB	0.965	0.265	0.913	15.858	4758.723	303.6	121.246
	-	cumSeg	0.768	0.001	0.915	13.462	4817.388	1758.934	592.281
	-	TGUH	0.961	0.35	0.885	15.58	4794.246	467.44	353.441
	-	FDRSeg	0.992	0.678	0.713	33.873	4822.753	312.756	252.471
`fms`	light	MoLP	0.984	0.009	0.961	4.429	-567.376	0.204	0.151
	heavy	MoLP	0.948	0.003	0.937	5.543	-566.093	0.312	0.168
	light	CuLP	0.988	0.012	0.963	3.871	-568.467	0.175	0.033
	heavy	CuLP	0.967	0.001	0.956	4.239	-568.043	0.249	0.033
	-	`bottom.up`	0.983	0.302	0.869	6.099	-550.932	0.325	0.237
	-	WBS.sBIC	0.978	0.174	0.91	10.382	-572.498	0.198	0.033
	-	WBS2.SDLL	0.999	0.464	0.749	17.344	-561.768	0.108	0.033
	-	PELT	0.984	0.108	0.941	8.202	-575.528	0.185	0.033

	-	S3IB	1	0.427	0.813	16.341	-572.761	0.111	0.033
	-	cumSeg	0.757	0.012	0.916	13.595	-552.024	1.777	0.120
	-	TGUH	0.998	0.306	0.866	11.119	-565.622	0.141	0.067
	-	FDRSeg	1	0.523	0.777	18.759	-562.495	0.116	0.050
`mix`	light	MoLP	0.916	0.005	0.753	3.874	839.603	30.371	12.659
	heavy	MoLP	0.854	0.001	0.634	4.428	841.166	34.099	6.501
	light	CuLP	0.905	0.007	0.736	4.202	840.603	38.31	13.686
	heavy	CuLP	0.822	0.001	0.577	5.118	843.018	46.415	14.028
	-	`bottom.up`	0.941	0.034	0.795	4.37	845.202	39.285	19.274
	-	WBS.sBIC	0.835	0.14	0.652	11.853	860.71	123.284	23.722
	-	WBS2.SDLL	0.961	0.303	0.769	9.536	843.276	29.806	15.624
	-	PELT	0.837	0.065	0.582	6.918	833.81	37.444	13.8
	-	S3IB	0.973	0.261	0.812	9.019	833.356	29.335	13.8
	-	cumSeg	0.346	0	0.286	23.734	900.692	742.379	75.442
	-	TGUH	0.93	0.229	0.724	8.453	842.304	40.007	27.257
	-	FDRSeg	0.971	0.409	0.775	11.321	850.546	32.429	19.274
`teeth10`	light	MoLP	0.937	0.001	0.905	2.49	-75.183	0.394	0
	heavy	MoLP	0.908	0	0.873	2.941	-74.966	0.531	0
	light	CuLP	0.843	0.001	0.76	5.507	-73.049	1.044	0
	heavy	CuLP	0.781	0.001	0.696	6.431	-72.824	1.33	0
	-	`bottom.up`	0.986	0.003	0.969	1.687	-75.091	0.147	0
	-	WBS.sBIC	0.722	0.064	0.655	8.248	-74.897	1.63	0
	-	WBS2.SDLL	0.988	0.092	0.903	4.353	-80.031	0.387	0
	-	PELT	0.643	0.036	0.539	9.114	-75.593	1.994	0.912
	-	S3IB	0.998	0.172	0.904	4.802	-81.069	0.317	0
	-	cumSeg	0.001	0	0	18.377	-63.278	4.996	0
	-	TGUH	0.985	0.082	0.896	4.628	-80.295	0.457	0
	-	FDRSeg	0.99	0.187	0.88	5.184	-78.135	0.408	0
`stairs10`	light	MoLP	0.994	0.001	0.972	2.352	-122.885	0.14	0.000
	heavy	MoLP	0.993	0.001	0.971	2.382	-122.851	0.145	0
	light	CuLP	0.99	0.001	0.952	3.424	-120.553	0.248	0
	heavy	CuLP	0.989	0.001	0.951	3.461	-120.514	0.255	0
	-	`bottom.up`	0.686	0.099	0.555	36.307	-44.435	3.031	2.859
	-	WBS.sBIC	1	0.082	0.953	3.643	-125.895	0.155	0
	-	WBS2.SDLL	0.999	0.07	0.951	3.682	-124.76	0.172	0
	-	PELT	0.997	0.015	0.967	2.905	-125.858	0.154	0
	-	S3IB	1	0.158	0.943	4.088	-126.209	0.124	0
	-	cumSeg	0.981	0.007	0.881	7.23	-98.274	0.625	0.424
	-	TGUH	0.999	0.074	0.956	3.636	-125.944	0.157	0
	-	FDRSeg	1	0.199	0.929	4.504	-122.972	0.138	0

Table 7: Summary of change point estimation over 1000 realisations for the test signals with **sparse** change points and t_5 errors; we set $\alpha = 0.2$ for MoLP and `bottom.up` and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n \in \{\log^{1.1}(n), n^{2/4.99}\}$ for the localised pruning.

model	penalty	method	TPR	FPR	ARI	MSE	BIC	δ_{trim}	υ_{trim}	speed
`blocks`	light	MoLP	1	0.007	0.979	2.271	23,128.27	369.841	0	0.174
	heavy	MoLP	0.995	0	1	2.749	23,130.49	3.024	0	0.174
	light	CuLP	1	0.02	0.934	3	23,129.62	1,047.866	0	5.167
	heavy	CuLP	0.996	0	1	3.05	23,130.73	2.753	0	5.170
	-	`bottom.up`	1	0.383	0.326	5.17	23,189.64	11,697.4	0	0.021
	-	WBS2.SDLL	1	0.899	0.016	155.253	23,393.5	17,507.5	0	3.735
	-	PELT	1	0.509	0.156	53.467	22,995.83	14,698.61	0	0.007
	-	TGUH	1	0.647	0.174	47.33	23,151.79	14,031.65	0	0.816
	-	FPOP	1	0.755	0.053	97.408	22,997.7	16,684.8	0	0.080
`fms`	light	MoLP	0.525	0.019	0.955	9.429	-11,996.67	497.164	0.208	0.174
	heavy	MoLP	0.113	0	0.588	29.304	-11,959.72	64.733	0	0.179
	light	CuLP	0.525	0.05	0.893	9.741	-11,996.39	1,314.255	0.208	5.075
	heavy	CuLP	0.012	0	0.056	29.968	-11,962.25	6.072	0	7.484
	-	`bottom.up`	0.644	0.523	0.183	12.73	-11,953.4	12,939.04	0.865	0.021
	-	WBS2.SDLL	0.73	0.958	0.006	300.3	-11,724.1	18,837.29	0.54	3.774
	-	PELT	0.493	0.784	0.067	107.746	-12,130.48	16,040.72	0.243	0.007
	-	TGUH	0.501	0.874	0.048	114.887	-11,972.02	15,887.4	1.211	0.813
	-	FPOP	0.634	0.899	0.02	188.625	-12,128.99	18,015.36	0.634	0.081
`mix`	light	MoLP	1	0.005	0.969	2.468	13,982.47	425.41	0.456	0.205
	heavy	MoLP	0.806	0	0.908	14.593	14,033.51	157.413	16.423	0.234
	light	CuLP	0.999	0.106	0.535	4.964	13,986.6	6,802.449	0.456	12.851
	heavy	CuLP	0.819	0.001	0.909	12.7	14,022.57	201.298	1.825	19.848
	-	`bottom.up`	1	0.246	0.205	4.795	14,020.46	12,786.47	1.939	0.023
	-	WBS2.SDLL	0.999	0.884	0.007	128.533	14,248.03	18,671.8	0.456	4.028
	-	PELT	0.998	0.469	0.08	44.492	13,849.79	15,860.93	0.456	0.007
	-	TGUH	0.951	0.637	0.072	52.071	14,027.56	15,578.4	3.307	0.874
	-	FPOP	1	0.723	0.024	80.491	13,851.49	17,852.09	0.456	0.088
`teeth10`	light	MoLP	0.165	0.088	0.963	7.109	-9,105.34	480.773	0.228	0.188
	heavy	MoLP	0.075	0.028	0.953	12.442	-9,088.252	106.865	0	0.191
	light	CuLP	0.177	0.35	0.493	9.382	-9,100.71	7,072.111	0.228	15.962
	heavy	CuLP	0.077	0.009	0.954	12.195	-9,090.402	140.83	0	21.479
	-	`bottom.up`	0.286	0.474	0.161	8.533	-9,069.533	13,124.67	0.228	0.022
	-	WBS2.SDLL	0.398	0.952	0.004	133.288	-8,820.831	19,022.91	0.228	3.851
	-	PELT	0.143	0.85	0.054	49.521	-9,239.104	16,239.11	0.342	0.007
	-	TGUH	0.257	0.853	0.065	47.879	-9,086.805	15,828	0.342	0.857
	-	FPOP	0.183	0.934	0.015	85.667	-9,236.122	18,214.47	0.228	0.083
`stairs10`	light	MoLP	0.59	0.007	0.966	10.283	-11,936.32	436.29	2.012	0.481
	heavy	MoLP	0.377	0	0.993	23.415	-11,883.82	12.535	0.212	0.625
	light	CuLP	0.599	0.025	0.894	10.521	-11,932.84	1,353.719	2.012	4.987

	heavy	CuLP	0.396	0	0.994	22.688	-11,886.22	13.098	0.424	5.221
	-	`bottom.up`	0.585	0.314	0.162	20.418	-11,861.37	13,113.7	1.906	0.023
	-	WBS2.SDLL	0.725	0.907	0.004	124.732	-11,662.26	19,024.46	1.059	4.216
	-	PELT	0.591	0.576	0.055	49.277	-12,069.15	16,222.22	2.012	0.007
	-	TGUH	0.651	0.68	0.086	47.761	-11,902.81	15,607.59	2.012	0.889
	-	FPOP	0.691	0.777	0.016	80.377	-12,067.29	18,207.36	1.589	0.085

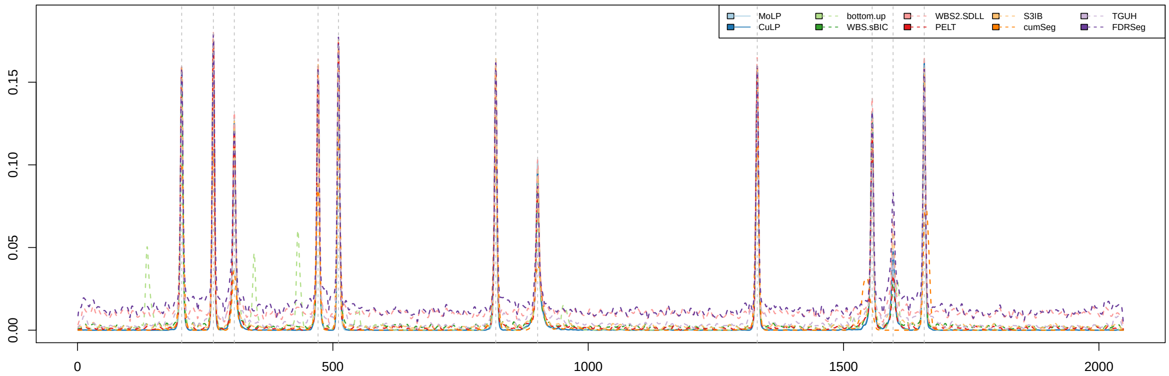

Figure 13: Test signal `blocks` with t_5 errors: weighted density of estimated change points with the vertical lines indicating the locations of true change points. We set $\alpha = 0.2$ for MoLP and `bottom.up` and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n = \log^{1.1}(n)$ for the localised pruning.

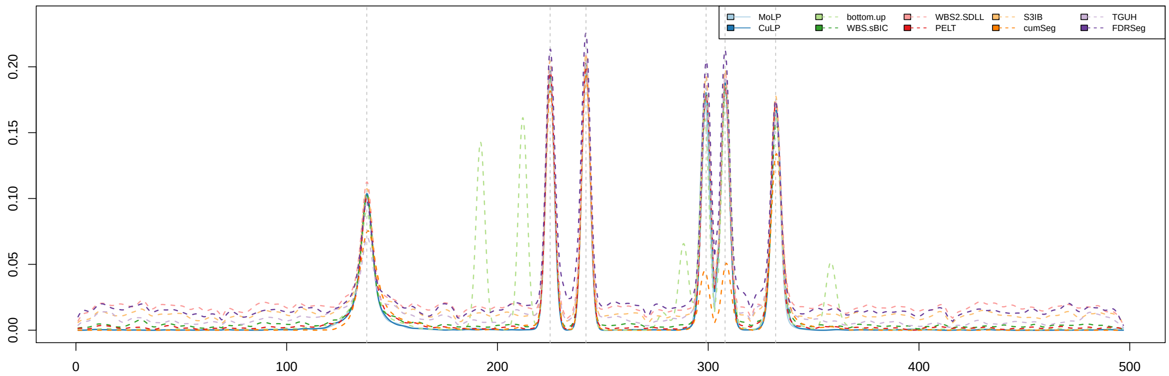

Figure 14: Test signal `fms` with t_5 errors: weighted density of estimated change points.

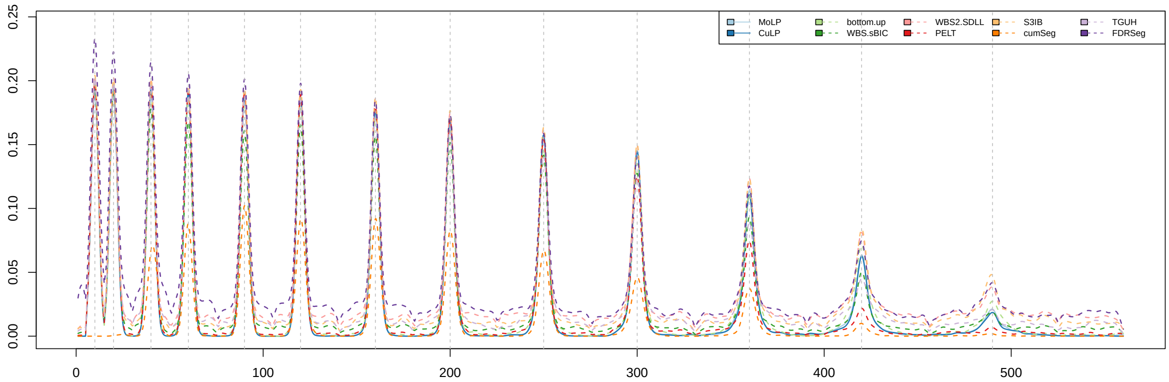

Figure 15: Test signal `mix` with t_5 errors: weighted density of estimated change points.

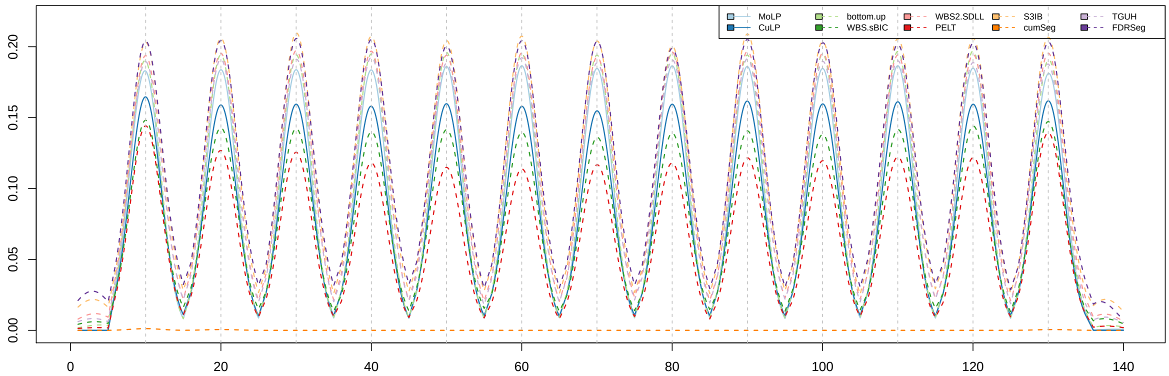

Figure 16: Test signal `teeth10` with t_5 errors: weighted density of estimated change points.

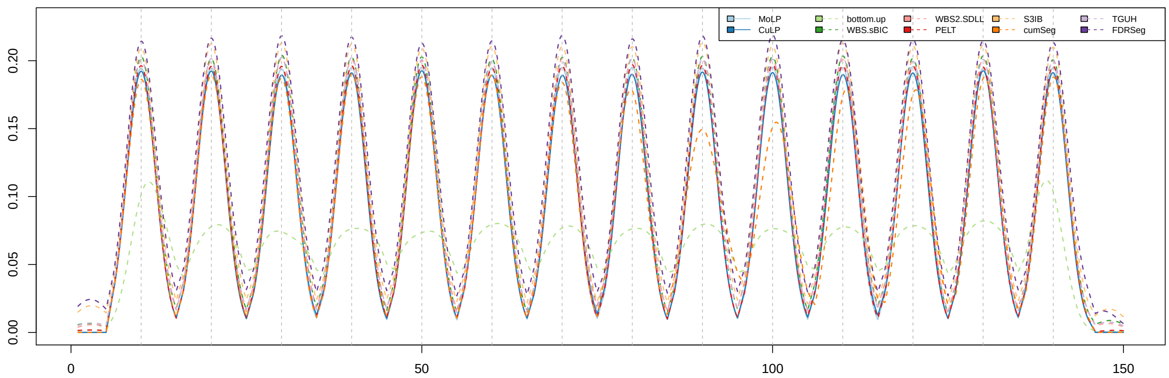

Figure 17: Test signal `stairs10` with t_5 errors: weighted density of estimated change point.

($\mathcal{E}3$) Serially correlated errors

Tables 8 –9 report the simulation results obtained from the test signals generated with serially correlated errors following AR(1) processes. Figures 18–27 visualise the performance of various methods by plotting the weighted densities of estimated change point(s).

In the presence of week serial dependence (AR parameter $\varrho = 0.3$), the choice of light penalty $\xi_n = \log^{1.1}(n)$ is observed to be effective in suppressing the false positives in the localised

pruning procedure, while attaining the TPR close to 90%. When the serial dependence is strong ($\varrho = 0.9$), a heavier penalty of $\xi_n = \log^2(n)$ is required to control the FPR. Overall the proposed localised pruning is successful in handling serial dependence.
JUSD tends to over-estimate τ^2 even with an informed choice of the parameter for its estimation. DepSMUCE shows weakness in detecting frequent jumps as in `teeth10`, whether the serial correlations are small or large, due to the block-based approach to the estimation of τ^2. We also consider those methods that do not require an explicit estimation of τ^2 (WBS.sBIC, cumSeg), or use a threshold involving its estimator only as a secondary check (WBS2.SDLL), to which we supply the estimator of τ^2 used by DepSMUCE; for `bottom.up`, we supplied the true τ^2. WBS.sBIC tends to over-estimate the number of change points due to the inadequacy of the chosen penalty when the serial dependence is strong, which is confirmed by that this set of many spurious estimators returns the minimum BIC. Although the final model returned by WBS2.SDLL does not critically depend on the estimator of τ^2, its performance appears to be heavily dependent on its estimator in some settings. The cumSeg, when the serial correlations are weak, tends to under-estimate the number of change points as in ($\mathcal{E}1$) whereas when the AR parameter is large, it returns many false positives.

Table 8: Summary of change point estimation over 1000 realisations for the test signals with Gaussian AR(1) process as ε_t where $\varrho = 0.3$ is used as the AR parameter; we set $\alpha = 0.2$ for MoLP, `bottom.up`, JUDS and DepSMUCE and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n \in \{\log^{1.1}(n), \log^2(n)\}$ for the localised pruning.

model	penalty	method	TPR	FPR	ARI	MSE	BIC	δ_{trim}	v_{trim}
`blocks`	light	MoLP	0.887	0.027	0.943	5.479	4774.904	740.203	319.633
	heavy	MoLP	0.348	0	0.66	20.411	4884.454	6922.511	444.873
	light	CuLP	0.878	0.04	0.931	6.193	4779.81	995.415	434.106
	heavy	CuLP	0.316	0	0.61	20.57	4885.94	7516.739	128.913
	-	`bottom.up`	0.851	0.148	0.903	6.75	4795.408	986.112	598.441
	-	WBS.sBIC	0.909	0.094	0.92	6.921	4785.532	1059.468	414.596
	-	WBS2.SDLL	0.943	0.226	0.864	6.832	4781.659	606.197	332.621
	-	cumSeg	0.742	0.007	0.887	9.407	4814.161	2347.276	961.464
	-	JUSD	0.757	0.007	0.921	8.575	4800.773	1889.446	658.822
	-	DepSMUCE	0.804	0.013	0.932	7.46	4792.994	1461.211	506.747
`fms`	light	MoLP	0.892	0.064	0.88	5.812	-566.721	0.576	0.287
	heavy	MoLP	0.411	0	0.595	17.571	-527.843	2.096	0
	light	CuLP	0.95	0.084	0.903	4.729	-570.251	0.392	0.136
	heavy	CuLP	0.42	0.001	0.605	16.863	-530.367	1.917	0
	-	`bottom.up`	0.834	0.261	0.825	7.536	-551.757	0.975	1.692
	-	WBS.sBIC	0.962	0.104	0.892	5.166	-569.772	0.33	0.136
	-	WBS2.SDLL	0.975	0.103	0.885	4.702	-570.251	0.287	0.136
	-	cumSeg	0.74	0.029	0.884	9.685	-551.414	2.167	0.465
	-	JUSD	0.397	0.001	0.586	17.694	-525.871	2.151	0.000
	-	DepSMUCE	0.824	0.009	0.912	7.265	-558.459	1.15	0.225

`mix`	light	MoLP	0.864	0.023	0.637	3.915	833.581	56.885	24.52
	heavy	MoLP	0.24	0	0.123	15.707	907.876	706.508	34.214
	light	CuLP	0.851	0.042	0.618	4.499	836.491	75.568	29.994
	heavy	CuLP	0.157	0	0.082	17.166	918.815	842.511	0
	-	`bottom.up`	0.863	0.014	0.649	4.173	840.582	72.021	30.907
	-	WBS.sBIC	0.857	0.101	0.658	6.551	856.824	128.398	35.810
	-	WBS2.SDLL	0.89	0.043	0.677	4.2	833.53	59.276	36.837
	-	cumSeg	0.399	0	0.33	13.36	893.913	703.169	184.755
	-	JUSD	0.423	0.004	0.299	15.69	919.363	695.591	290.019
	-	DepSMUCE	0.587	0.006	0.409	11.989	895.764	484.526	167.534
`teeth10`	light	MoLP	0.873	0.002	0.83	2.611	-79.075	0.718	0
	heavy	MoLP	0.084	0	0.075	10.437	-63.117	4.581	0
	light	CuLP	0.874	0.035	0.773	3.882	-80.887	1.042	0
	heavy	CuLP	0.081	0	0.067	10.4	-62.294	4.615	0
	-	`bottom.up`	0.78	0.003	0.736	3.979	-74.989	1.163	0
	-	WBS.sBIC	0.8	0.07	0.704	4.737	-79.29	1.37	0.000
	-	WBS2.SDLL	0.098	0.004	0.087	10.081	-65.161	4.558	0.000
	-	cumSeg	0.01	0	0.005	10.693	-63.908	4.956	0.000
	-	JUSD	0	0	0	10.727	-63.434	5	0.000
	-	DepSMUCE	0.002	0	0.001	10.724	-63.469	4.992	0.000
`stairs10`	light	MoLP	0.989	0.005	0.966	1.977	-127.028	0.174	0
	heavy	MoLP	0.616	0	0.669	17.122	-71.541	2.04	0.318
	light	CuLP	0.994	0.041	0.944	2.568	-127.349	0.235	0
	heavy	CuLP	0.688	0	0.709	13.441	-82.097	1.864	0
	-	`bottom.up`	0.651	0.083	0.543	23.182	-48.558	3.155	2.012
	-	WBS.sBIC	0.998	0.084	0.94	2.556	-128.074	0.193	0.000
	-	WBS2.SDLL	0.984	0.022	0.936	2.798	-125.513	0.319	0.000
	-	cumSeg	0.968	0.008	0.84	6.048	-98.312	0.862	0.847
	-	JUSD	0.524	0	0.616	19.1	-63.866	2.651	0.741
	-	DepSMUCE	0.551	0	0.627	18.169	-67.326	2.576	1.906

Table 9: Summary of change point estimation over 1000 realisations for the test signals with Gaussian AR(1) process as ε_t where $\varrho = 0.9$ is used as the AR parameter; we set $\alpha = 0.2$ for MoLP, `bottom.up`, JUDS and DepSMUCE and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n \in \{\log^{1.1}(n), \log^2(n)\}$ for the localised pruning.

model	penalty	method	TPR	FPR	ARI	MSE	BIC	δ_{trim}	υ_{trim}
`blocks`	light	MoLP	0.965	0.555	0.738	8.856	46624.3	4598.286	2825.82
	heavy	MoLP	0.942	0.14	0.905	5.792	46941.57	6350.733	3242.88
	light	CuLP	0.941	0.316	0.849	6.529	46908.09	6314.565	2814.783
	heavy	CuLP	0.916	0.059	0.94	5.445	47029.13	7342.356	3070.561
	-	`bottom.up`	0.824	0.067	0.918	6.786	47293.08	10658.44	5986.544
	-	WBS.sBIC	1	0.985	0.023	78.455	41629.19	764.912	176.043

	-	WBS2.SDLL	0.97	0.443	0.723	9.548	46582.41	5485.552	3300.133
	-	cumSeg	0.929	0.343	0.722	7.914	47207.89	10499.69	6610.925
	-	JUSD	0.782	0.009	0.928	7.675	47399.17	16699.38	5104.590
	-	DepSMUCE	0.917	0.137	0.876	5.602	47045.95	7352.367	3336.377
`fms`	light	MoLP	0.962	0.398	0.758	6.1	-6128.199	1.448	0.993
	heavy	MoLP	0.951	0.116	0.856	4.743	-6050.056	2.128	1.288
	light	CuLP	0.976	0.302	0.815	5.463	-6111.884	2.702	0.862
	heavy	CuLP	0.962	0.079	0.906	4.312	-6060.462	3.078	0.965
	-	`bottom.up`	0.74	0.04	0.771	9.335	-5761.415	10.171	15.326
	-	WBS.sBIC	1	0.973	0.039	40.62	-7663.268	0.133	0.050
	-	WBS2.SDLL	0.982	0.176	0.84	5.383	-6122.909	2.406	0.923
	-	cumSeg	0.959	0.565	0.518	11.825	-6044.72	12.896	8.774
	-	JUSD	0.762	0.01	0.878	8.301	-5811.716	12.849	1.678
	-	DepSMUCE	0.876	0.041	0.902	6.363	-5920.303	8.368	8.838
`mix`	light	MoLP	0.902	0.3	0.671	4.467	7585.826	341.03	284.26
	heavy	MoLP	0.868	0.046	0.64	4.034	7695.666	500.375	256.946
	light	CuLP	0.927	0.179	0.718	4.323	7598.369	464.118	306.1
	heavy	CuLP	0.886	0.038	0.673	3.934	7669.327	509.42	337.063
	-	`bottom.up`	0.739	0.013	0.461	5.539	7981.962	1040.016	406.518
	-	WBS.sBIC	1	0.948	0.104	19.779	5946.152	88.406	14.256
	-	WBS2.SDLL	0.937	0.093	0.742	4.128	7586.685	464.132	291.616
	-	cumSeg	0.864	0.28	0.738	8.866	8083.526	4268.618	728.527
	-	JUSD	0.624	0.007	0.454	11.194	8450.092	4347.482	1298.358
	-	DepSMUCE	0.828	0.009	0.643	5.659	7906.877	1559.279	447.061
`teeth10`	light	MoLP	0.924	0.003	0.877	2.75	-1301.465	6.007	0
	heavy	MoLP	0.898	0.001	0.843	2.901	-1290.286	7.262	0
	light	CuLP	0.887	0.113	0.803	3.878	-1262.316	9.108	0
	heavy	CuLP	0.807	0.02	0.726	4.243	-1215.827	12.82	0
	-	`bottom.up`	0.689	0.002	0.632	5.783	-1086.38	17.678	6.273
	-	WBS.sBIC	1	0.844	0.345	5.45	-1793.232	0.468	0.000
	-	WBS2.SDLL	0.651	0.017	0.605	5.568	-1138.749	19.842	4.505
	-	cumSeg	0.997	0.292	0.801	3.976	-1375.372	6.059	3.878
	-	JUSD	0.151	0.003	0.156	10.515	-680.996	44.887	0.000
	-	DepSMUCE	0.079	0.002	0.081	10.607	-658.432	47.344	0.000
`stairs10`	light	MoLP	0.99	0.008	0.968	1.959	-1854.062	1.632	0
	heavy	MoLP	0.99	0.002	0.97	1.943	-1851.876	1.65	0
	light	CuLP	0.991	0.128	0.936	2.507	-1893.471	1.895	0
	heavy	CuLP	0.99	0.024	0.958	2.314	-1862.168	2.083	0
	-	`bottom.up`	0.771	0.083	0.708	13.097	-1263.68	19.22	5.825
	-	WBS.sBIC	1	0.843	0.351	5.299	-2399.815	0.011	0.000
	-	WBS2.SDLL	0.864	0.01	0.871	6.744	-1612.956	8.71	0.000
	-	cumSeg	1	0.262	0.848	4.329	-1793.323	3.954	2.700
	-	JUSD	0.535	0	0.648	18.334	-975.106	25.541	8.419

	-	DepSMUCE	0.469	0	0.59	25.603	-772.494	29.519	5.242

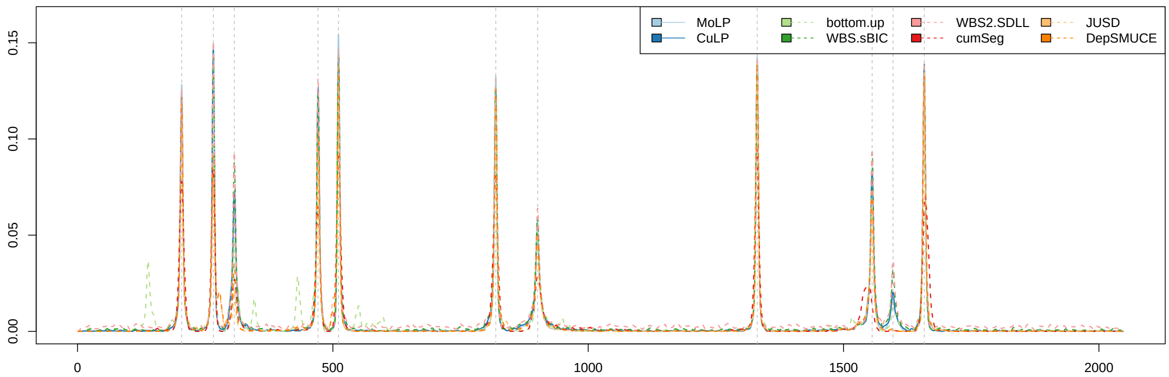

Figure 18: Test signal `blocks` with AR(1) process as ε_t where $\varrho = 0.3$: weighted density of estimated change points. We set $\alpha = 0.2$ for MoLP, `bottom.up`, JUSD and DepSMUCE and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n = \log^{1.1}(n)$ for the localised pruning.

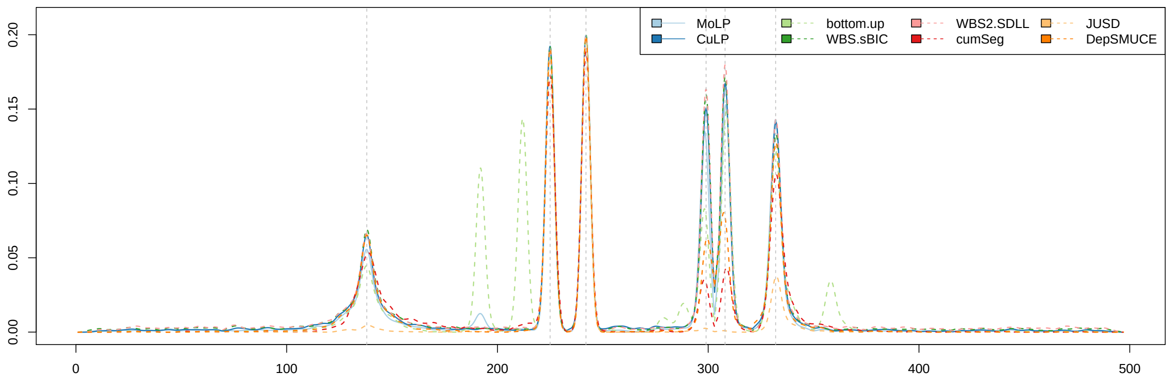

Figure 19: Test signal `fms` with AR(1) process as ε_t where $\varrho = 0.3$: weighted density of estimated change points.

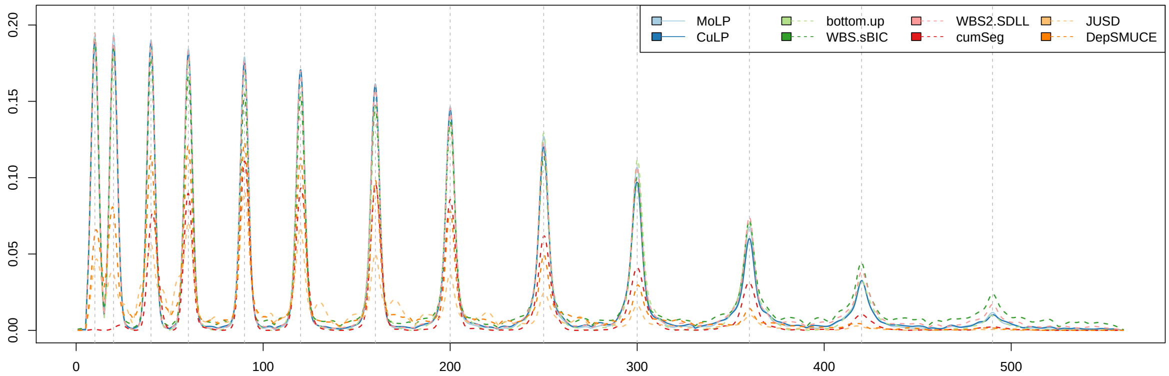

Figure 20: Test signal `mix` with AR(1) process as ε_t where $\varrho = 0.3$: weighted density of estimated change points.

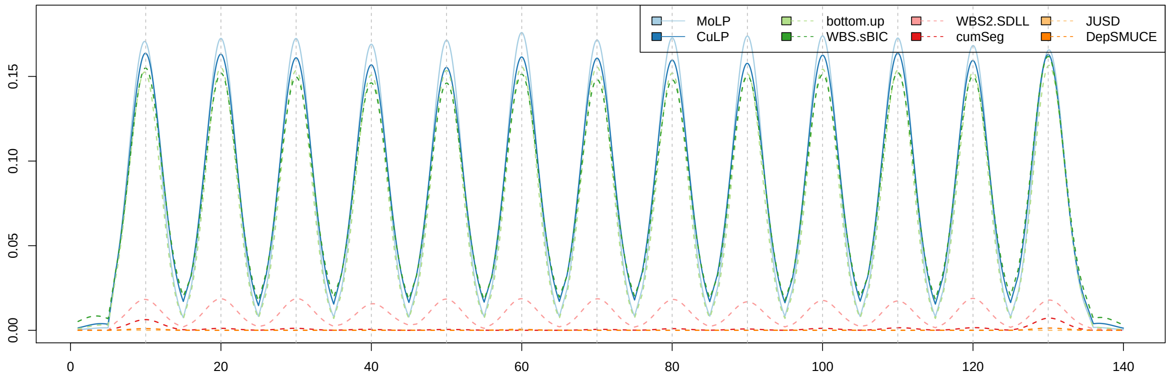

Figure 21: Test signal `teeth10` with AR(1) process as ε_t where $\varrho = 0.3$: weighted density of estimated change points.

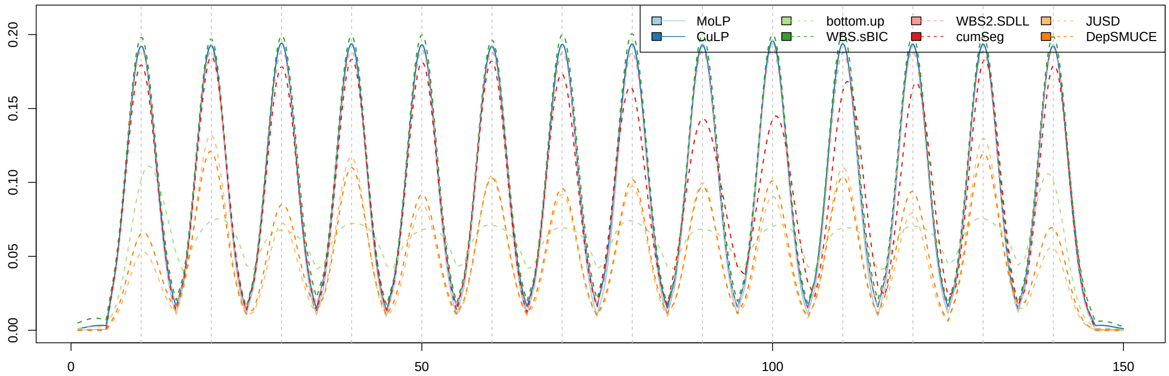

Figure 22: Test signal `stairs10` with AR(1) process as ε_t where $\varrho = 0.3$: weighted density of estimated change points.

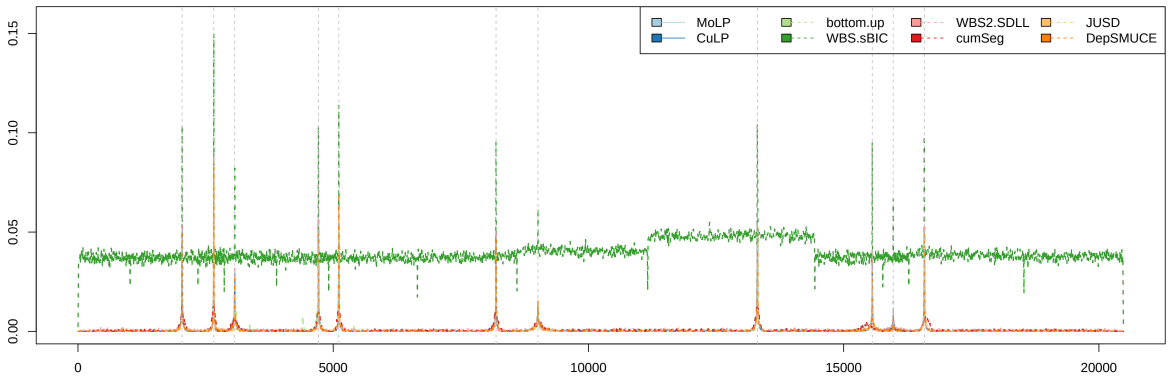

Figure 23: Test signal `blocks` with AR(1) process as ε_t where $\varrho = 0.9$: weighted density of estimated change points. We set $\alpha = 0.2$ for MoLP, `bottom.up`, JUSD and DepSMUCE and $C_\zeta = 0.9$ for CuLP, and use $h = h_{\mathcal{J}}$ and $\xi_n = \log^2(n)$ for the localised pruning.

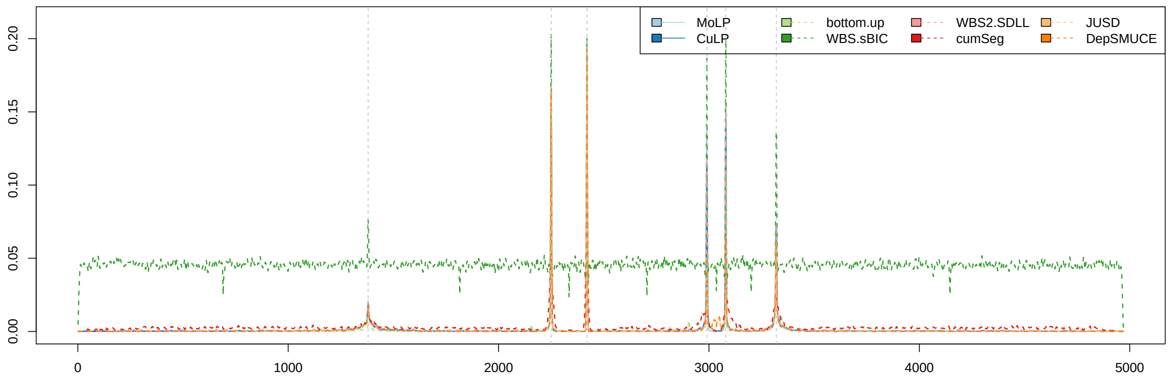

Figure 24: Test signal `fms` with AR(1) process as ε_t where $\varrho = 0.9$: weighted density of estimated change points.

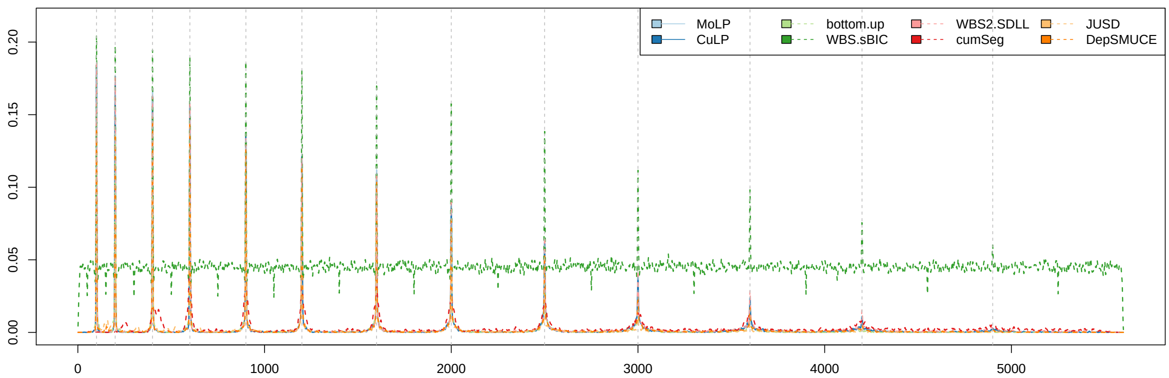

Figure 25: Test signal `mix` with AR(1) process as ε_t where $\varrho = 0.9$: weighted density of estimated change points.

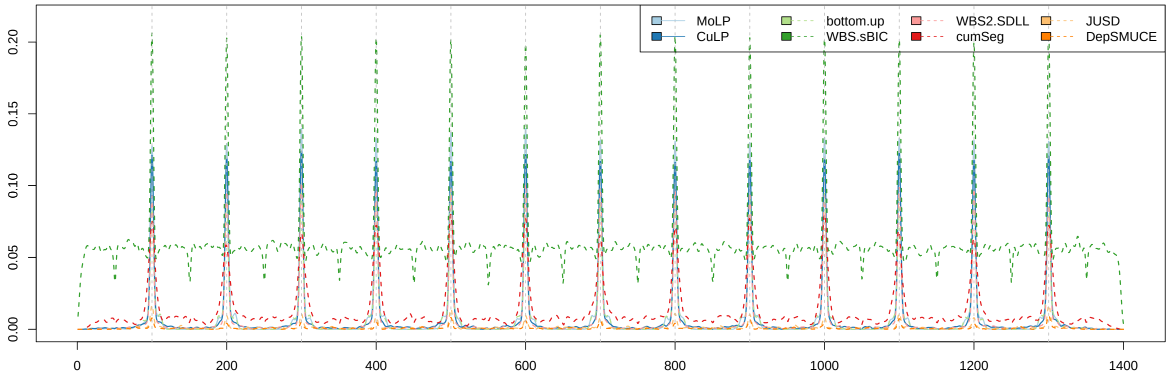

Figure 26: Test signal `teeth10` with AR(1) process as ε_t where $\varrho = 0.9$: weighted density of estimated change points.

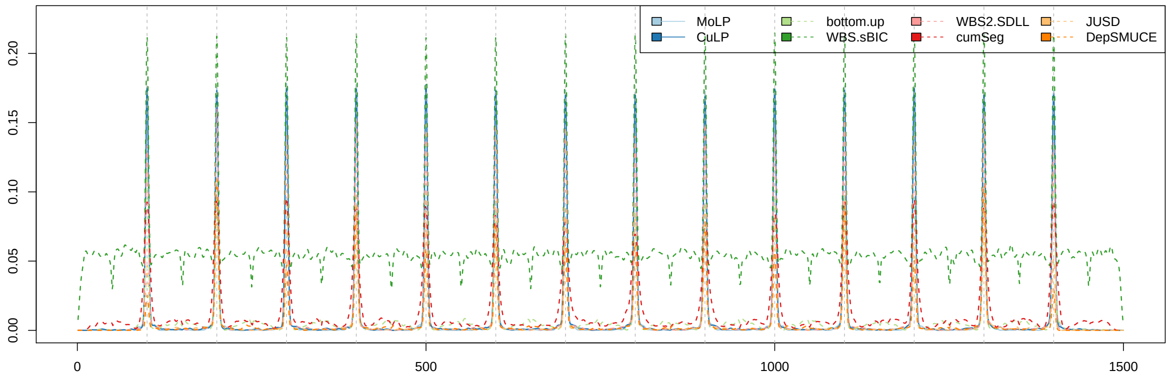

Figure 27: Test signal `stairs10` with AR(1) process as ε_t where $\varrho = 0.9$: weighted density of estimated change points.

H Algorithms

Algorithm 1 provides the pseudo code for the outer algorithm of the proposed localised pruning methodology, which iteratively identifies the local interval over which pruning is to be performed.

Algorithm 1: Outer algorithm for localisation (`LocAlg`)

Input: Data $\{X_t\}_{t=1}^n$, a set of candidate change point estimators $\mathcal{K}$, a candidate sorting function $h(\cdot)$

Step 0: set $\widehat{\Theta} = \emptyset$ and $\mathcal{C} \leftarrow \mathcal{K}$

repeat

  Step 1: find $\mathcal{C}_\circ$ as $\mathcal{C}_\circ \leftarrow \{k \in \mathcal{C} : h(k) = \max_{k' \in \mathcal{C}} h(k')\}$

    if $|\mathcal{C}_\circ| = 1$ **then** $k_\circ \leftarrow \mathcal{C}_\circ$

    else $k_\circ \leftarrow \arg\min_{k \in \mathcal{C}_\circ} |\mathcal{I}(k)|$

  Step 2: find

$$k_L \leftarrow \max\{k < k_\circ : k \in \widehat{\Theta} \cup \{0\} \text{ or } (k \in \mathcal{C} \text{ and } \mathcal{I}(k) \cap \mathcal{I}(k_\circ) = \emptyset)\},$$
$$k_R \leftarrow \min\{k > k_\circ : k \in \widehat{\Theta} \cup \{n\} \text{ or } (k \in \mathcal{C} \text{ and } \mathcal{I}(k) \cap \mathcal{I}(k_\circ) = \emptyset)\}$$

  and set $\mathcal{D} \leftarrow (k_L, k_R) \cap \mathcal{C}$

  Step 3: $\widehat{\mathcal{A}} \leftarrow$ `PrunAlg`$(\mathcal{D}, \mathcal{C}, \widehat{\Theta}, k_L, k_R)$

  Step 4: set $\mathcal{R} \leftarrow \{k_\circ\} \cup (\mathcal{D} \cap [\min \widehat{\mathcal{A}}, \max \widehat{\mathcal{A}}])$

    if $k_L \in \widehat{\Theta} \cup \{0\}$ **then** $\mathcal{R} \leftarrow \mathcal{R} \cup \{\mathcal{D} \cap (k_L, \min \widehat{\mathcal{A}})\}$

    if $k_R \in \widehat{\Theta} \cup \{n\}$ **then** $\mathcal{R} \leftarrow \mathcal{R} \cup \{\mathcal{D} \cap (\max \widehat{\mathcal{A}}, k_R)\}$

  Step 5: set $\widehat{\Theta} \leftarrow \widehat{\Theta} \cup \widehat{\mathcal{A}}$ and $\mathcal{C} \leftarrow \mathcal{C} \setminus \mathcal{R}$

until $\mathcal{C}$ *is empty*

Output: $\widehat{\Theta}$

Algorithm 2 outlines the efficient implementation of the inner algorithm employed in Step 3 of the outer algorithm (Algorithm 1). For further details on its implementation, see Meier et al. (2021b).

Algorithm 2: Inner algorithm for pruning (`PrunAlg`)

Function `PrunAlg`$(\mathcal{D}, \mathcal{C}, \widehat{\Theta}, s, e)$**:**

  Enumerate all $M = 2^{|\mathcal{D}|}$ subsets of $\mathcal{D}$ (including $\emptyset$) denoted by $\mathcal{D}_i$, $i = 1, \ldots, M$.
  Set $\mathcal{F} \leftarrow \emptyset$, $\widehat{\mathcal{A}} \leftarrow \emptyset$, $\ell \leftarrow |\mathcal{D}|$, and assign $\mathsf{flag}_i \leftarrow \mathsf{true}$ for all $i = 1, \ldots, M$.
  repeat
    for $\mathcal{D}_i$ with $|\mathcal{D}_i| = \ell$ and $\mathsf{flag}_i = \mathsf{false}$ **do**
      identify

$$\mathsf{child}(\mathcal{D}_i) = \{j : \mathcal{D}_j \subset \mathcal{D}_i \text{ with } |\mathcal{D}_j| = \ell - 1 \text{ and } \mathsf{flag}_j = \mathsf{true}\}$$

      for $j \in \mathsf{child}(\mathcal{D}_i)$ **do** $\mathsf{flag}_j \leftarrow \mathsf{false}$
    end
    for $\mathcal{D}_i$ with $|\mathcal{D}_i| = \ell$ and $\mathsf{flag}_i = \mathsf{true}$ **do**
      update $\mathcal{F} \leftarrow \mathcal{F} \cup \{i\}$ and identify $\mathsf{child}(\mathcal{D}_i)$
      for $j \in \mathsf{child}(\mathcal{D}_i)$ **do**
        if $\mathrm{SC}(\mathcal{D}_i|\mathcal{C}, \widehat{\Theta}, s, e) < \mathrm{SC}(\mathcal{D}_j|\mathcal{C}, \widehat{\Theta}, s, e)$ **then** $\mathsf{flag}_j \leftarrow \mathsf{false}$
      end
    end
    $\ell \leftarrow \ell - 1$
  until $\ell = 0$
  if $\mathcal{F} \neq \emptyset$ **then**
    find $m^* \leftarrow \min_{i \in \mathcal{F}} |\mathcal{D}_i|$
    identify $i^* \leftarrow \arg\min_{i:\, \mathcal{D}_i \subset_R \mathcal{D}_{i'}, i' \in \mathcal{F}, m^* \leq |\mathcal{D}_{i'}| \leq m^* + 2} \mathrm{SC}(\mathcal{D}_i|\mathcal{C}, \widehat{\Theta}, s, e)$
    set $\widehat{\mathcal{A}} \leftarrow \mathcal{D}_{i^*}$
  end
  return $\widehat{\mathcal{A}}$

References

Anastasiou, A., Chen, Y., Cho, H., and Fryzlewicz, P. (2021). *breakfast: Methods for Fast Multiple Change-Point Detection and Estimation.* R package version 2.2.

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*, 81:649–672.

Cleynen, A., Rigaill, G., and Koskas, M. (2016). *Segmentor3IsBack: A fast segmentation algorithm.* R package version 2.0.

Dette, H., Schüler, T., and Vetter, M. (2020). Multiscale change point detection for dependent data. *To appear in Scandinavian Journal of Statistics.*

Eichinger, B. and Kirch, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli*, 24:526–564.

Fisch, A. T. M., Eckley, I. A., and Fearnhead, P. (2018). A linear time method for the detection of point and collective anomalies. *arXiv preprint arXiv:1806.01947.*

Fisch, A. T. M., Grose, D. J., Eckley, I. A., and Fearnhead, P. (2018). *anomaly: An R package for detecting anomalies in data.* R package version 1.0.0.

Frick, K., Munk, A., and Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B*, 76:495–580.

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

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42:2243–2281.

Fryzlewicz, P. (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *The Annals of Statistics*, pages 3390–3421.

Fryzlewicz, P. (2020). Detecting possibly frequent change-points: Wild binary segmentation 2 and steepest-drop model selection. *To appear in the Journal of the Korean Statistical Society.*

Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-dimensional Statistical Models*, volume 40. Cambridge University Press, Cambridge.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2:193–218.

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.

Killick, R., Haynes, K., and Eckley, I. A. (2016). *changepoint: An R package for changepoint analysis.* R package version 2.2.2.

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

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 arXiv:2002.06633.*

Li, H., Munk, A., and Sieling, H. (2016). FDR-control in multiscale change-point segmentation. *Electronic Journal of Statistics*, 10:918–959.

Li, H. and Sieling, H. (2017). *FDRSeg: FDR-control in multiscale change-point segmentation.* R package version 1.0-3.

Maidstone, R., Hocking, T., Rigaill, G., and Fearnhead, P. (2017). On optimal multiple changepoint algorithms for large data. *Statistics and Computing*, 27:519–533.

Meier, A., Cho, H., and Kirch, C. (2021a). *mosum: Moving sum based procedures for changes in the mean.* R package version 1.2.5.

Meier, A., Kirch, C., and Cho, H. (2021b). mosum: A package for moving sums in change point analysis. *Journal of Statistical Software*, 97(8):1 – 42.

Messer, M., Kirchner, M., Schiemann, J., Roeper, J., Neininger, R., and Schneider, G. (2014). A multiple filter test for the detection of rate changes in renewal processes with varying variance. *The Annals of Applied Statistics*, 8:2027–2067.

Mikosch, T. and Račkauskas, A. (2010). The limit distribution of the maximum increment of a random walk with regularly varying jump size distribution. *Bernoulli*, 16:1016–1038.

Muggeo, V. M. (2012). *cumSeg: Change point detection in genomic sequences*. R package version 1.1.

Muggeo, V. M. and Adelfio, G. (2010). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics*, 27:161–166.

Pein, F., Hotz, T., Sieling, H., and Aspelmeier, T. (2019). *stepR: Multiscale change-point inference*. R package version 2.0-3.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66:846–850.

Rein, H. (2018). Open Exoplanet Catalogue. Online; accessed August 30, 2018.

Rigaill, G. (2015). A pruned dynamic programming algorithm to recover the best segmentations with 1 to k_max change-points. *Journal de la Société Frankcaise de Statistique*, 156(4):180–205.

Rigaill, G. and Hocking, T. D. (2019). *fpop: Segmentation using Optimal Partitioning and Function Pruning*. R package version 2019.01.22/r56.

Sartoretti, P. and Schneider, J. (1999). On the detection of satellites of extrasolar planets with the method of transits. *Astronomy and Astrophysics Supplement Series*, 134:553–560.

Stout, W. F. (1974). *Almost Sure Convergence*, volume 24. Academic Press, New York.

Tecuapetla-Gómez, I. and Munk, A. (2017). Autocovariance estimation in regression with a discontinuous signal and m-dependent errors: A difference-based approach. *Scandinavian Journal of Statistics*, 44:346–368.

Venkatraman, E. S. (1992). Consistency results in multiple change-point problems. *Technical Report No. 24, Department of Statistics, Stanford University.*

Vershynin, R. (2018). *High-dimensional Probability: An Introduction with Applications in Data Science*, volume 47. Cambridge University Press, Cambridge.

Vostrikova, L. J. (1981). Detecting 'disorder' in multidimensional random processes. *Soviet Doklady Mathematics*, 24:55–59.

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

Wang, T. and Samworth, R. J. (2018a). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B*, 80:57–83.

Wang, T. and Samworth, R. J. (2018b). Supplementary material for "high dimensional change point estimation via sparse projection". *Available from* *https://rss.onlinelibrary.wiley.com/doi/full/10.1111/rssb.12243*.