Supplementary Material for \Data-driven selection of the number of change-points via error rate control"

Hui Chen[1], Haojie Ren[2], Fang Yao[3*] and Changliang Zou[1]

[1]*School of Statistics and Data Science, Nankai University, China*

[2]*School of Mathematical Sciences, Shanghai Jiao Tong University, China*

[3]*School of Mathematical Sciences, Peking University, China*

This supplementary material contains the lemmas used in the proof of Theorem 1 (Appendix C), the proofs of Proposition 1, Theorems 1–3 and Corollaries 1–2 (Appendix D–F), and some additional simulation results (Appendix G).

Appendix B: Equivalence of de nitions given by Eqs.(3) and (4)

- If there exists one ${}_k^* \in \left[\frac{1}{2}({}_{j-1} + {}_j); \frac{1}{2}({}_j + {}_{j+1})\right)$ as (3), we have ${}_k^* - {}_{j-1} \geq {}_j - {}_k^*$ when ${}_j \geq {}_k^*$ or ${}_{j+1} - {}_k^* > {}_k^* - {}_j$ when ${}_j < {}_k^*$, that is $|{}_j - {}_k^*| = \min_{l \in \mathcal{T}} |{}_l - {}_k^*|$ from which ${}_j$ follows (4);

- On the contrary, if ${}_j = \arg\min_{l \in \mathcal{T}} |{}_l - {}_k^*|$ as the definition of (4), we have ${}_k^* \geq \frac{1}{2}({}_{j-1} + {}_j)$ due to ${}_k^* - {}_{j-1} > {}_k^* - {}_j$ if ${}_k^* < {}_j$; Similarly, ${}_k^* < \frac{1}{2}({}_j + {}_{j-1})$ holds for ${}_k^* > {}_j$. Say, ${}_j$ follows the definition of (3).

*Corresponding author: fyao@math.pku.edu.cn

Appendix C: Auxiliary lemmas

Lemma S.1 *If the model (1) and Assumption 1 hold, $\ _n^{-1} = \ + O_p(K_n n^{-1/2})$, where $\ $ is some positive matrix depending on $\ _k^*$'s.*

This lemma can be proved using the similar arguments in the Proposition 1 of Zou et al. (2020), thus the details are omitted here.

Lemma S.2 *[Bernstein's inequality] Let $X_1, \ldots, X_n$ be independent centered random variables a.s. bounded by $A < \infty$ in absolute value. Let $\ ^2 = n^{-1}\sum_{i=1}^{n}\mathbb{E}(X_i^2)$. Then for all $x > 0$,*

$$\Pr\Big(\sum_{i=1}^{n} X_i \geq x\Big) \leq \exp\Big(-\frac{x^2}{2n\ ^2 + 2Ax/3}\Big).$$

The third one is a moderate deviation result for the mean; See Petrov (2002).

Lemma S.3 (Moderate Deviation for the Independent Sum)

Suppose that $X_1, \ldots, X_n$ are independent random variables with mean zero, satisfying $\mathbb{E}(|X_j|^{2+q}) < \infty$ $(j = 1, 2, \ldots)$ for some $q > 0$. Let $B_n = \sum_{i=1}^{n}\mathbb{E}(X_i^2)$. Then

$$\frac{\Pr\big(\sum_{i=1}^{n} X_i > x\sqrt{B_n}\big)}{1-\Phi(x)} \to 1 \quad \text{and} \quad \frac{\Pr\big(\sum_{i=1}^{n} X_i < -x\sqrt{B_n}\big)}{\Phi(-x)} \to 1,$$

as $n \to \infty$ uniformly in x in the domain $0 \leq x \lesssim \{\log(1/L_n)\}^{1/2}$, where $L_n = B_n^{-1-\frac{q}{2}}\sum_{i=1}^{n}\mathbb{E}(|X_i|^{2+q})$.

For notational convenience, we note that our estimation procedure can be reformulated as follows. Suppose we have two independent sets of d-dimensional observations $\{\mathbf{S}_1^O, \ldots, \mathbf{S}_n^O\}$ and $\{\mathbf{S}_1^E, \ldots, \mathbf{S}_n^E\}$ collected from the following multiple change-point model

$$\mathbf{S}_j^O = \boldsymbol{\mu}_k^* + \mathbf{U}_j, \ \mathbf{S}_j^E = \boldsymbol{\mu}_k^* + \mathbf{V}_j, \ j \in (\ _k^*, \ _{k+1}^*], \ k = 0, \ldots, K_n,$$

where $\mathbf{U}_1, \ldots, \mathbf{U}_n, \mathbf{V}_1, \ldots, \mathbf{V}_n$ are independent standardized noises satisfying $\mathbb{E}(\mathbf{U}_1) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{U}_1) = \mathrm{Cov}(\mathbf{V}_1) = \ _k^*$. Let $\underline{\varpi} = \min_{0\leq k\leq K_n}\mathrm{Eig}_{\min}(\ _k^*)$ and $\varpi = \max_{0\leq k\leq K_n}\mathrm{Eig}_{\max}(\ _k^*)$, where $\mathrm{Eig}_{\min}(\mathbf{A})$ and $\mathrm{Eig}_{\max}(\mathbf{A})$ denote the smallest and largest eigenvalues of a square matrix $\mathbf{A}$. By Assumption 1, we know that $0 < \underline{\varpi} < \varpi < \infty$. To keep the subscript consistent with the main body, we roughly let $\mathbf{S}^O_{2i}$, $\mathbf{U}_{2i}$, $\mathbf{S}^E_{2i-1}$, $\mathbf{V}_{2i-1}$ as 0 for $i = 1, \ldots, m$.

The next one establishes an uniform bound for $\| \sum_{i=k_1+1}^{k_2} \mathbf{U}_i \|$.

Lemma S.4 *Suppose Assumption 1 holds. Then we have as $n \to \infty$,*

$$\Pr\left(\max_{(k_1,k_2)\in\mathcal{T}(\omega_n)} (k_2 - k_1)^{-1} \left\| \sum_{i=k_1+1}^{k_2} \mathbf{U}_i \right\|^2 > C \log n \right) = O(n^{1-\ \ -\ \ });$$

for some large $C > 0$ and any $0 < \ \ < \ \ - 2\ \ ^{-1}$.

Proof. We shall show that the assertion holds when $d = 1$ and the case for $d > 1$ is straightforward by using the Bonferroni inequality. Denote $M_n = n^{1/(\ \ -\ \ )}$ for some $0 < \ \ < \ \ $, and observe that

$$\begin{aligned} U_i &= [U_i \mathbb{I}(|U_i| \le M_n) - \mathbb{E}\{U_i \mathbb{I}(|U_i| \le M_n)\}] + [U_i \mathbb{I}(|U_i| > M_n) - \mathbb{E}\{U_i \mathbb{I}(|U_i| > M_n)\}] \\ &=: U_{i1} + U_{i2}. \end{aligned}$$

It suffices to prove that the assertion holds with U_{i1} and U_{i2} respectively. Let $x = \sqrt{C \log n}$ with a sufficiently large C,

$$\begin{aligned} &\Pr\left(\max_{(k_1,k_2)\in\mathcal{T}(\omega_n)} (k_2 - k_1)^{-1} \left(\sum_{i=k_1+1}^{k_2} U_i \right)^2 > x^2 \right) \\ \le &\Pr\left(\max_{(k_1,k_2)\in\mathcal{T}(\omega_n)} (k_2 - k_1)^{-1/2} \left| \sum_{i=k_1+1}^{k_2} U_{i1} \right| > x/2 \right) \\ &+ \Pr\left(\max_{(k_1,k_2)\in\mathcal{T}(\omega_n)} (k_2 - k_1)^{-1/2} \left| \sum_{i=k_1+1}^{k_2} U_{i2} \right| > x/2 \right) \\ =: &P_1 + P_2. \end{aligned}$$

On one hand, by the Bernstein inequality in Lemma S.2, we have

$$P_1 \le n^2 \Pr\left((k_2 - k_1)^{-1/2} \left| \sum_{i=k_1+1}^{k_2} U_{i1} \right| > x/2 \right) \le 2n^2 \exp\left\{ -\frac{w_n x^2}{C_1 w_n + C_2 M_n w_n^{1/2} x} \right\} = o(n^{1-\ \ -\ \ });$$

where $C_1; C_2$ are some positive constants and we use the assumption that $\;\; < \;\; - 2 \;\; ^{-1}$.

On the other hand, according to Cauchy inequality and Markov inequality, we note that

$$\mathbb{E}^2\{|U_i|\mathbb{I}(|U_i| > M_n)\} \le \mathbb{E}(U_i^2)\Pr(|U_i| > M_n) \le C_3 n^{- \;\; \overline{\;\;\;\;}};$$

for some constant $C_3 > 0$. Further, it yields $\max_{(k_1;k_2)\in\mathcal{T}(\,!\,_n)}(k_2 - k_1)^{1=2}\mathbb{E}\{|U_i|\mathbb{I}(|U_i| > M_n)\} = o(1)$: Thus, by Assumption 1 and Markov inequality, we have

$$\begin{aligned}
P_2 \le &\Pr\Big(\max_{(k_1;k_2)\in\mathcal{T}(\,!\,_n)}(k_2 - k_1)^{-1=2}\sum_{i=k_1+1}^{k_2}|U_i|\mathbb{I}(|U_i| > M_n) > x=4\Big)\\
\le &\Pr\Big(\max_{(k_1;k_2)\in\mathcal{T}(\,!\,_n)}(k_2 - k_1)^{-1=2}\sum_{i=k_1+1}^{k_2}|U_i| > x=2 \mid \max_i |U_i| > M_n\Big)\Pr\Big(\max_i |U_i| > M_n\Big)\\
\le &n\Pr\Big(|U_i| \;\; > M_n\Big) \le C_4 n^{1- \;\; \overline{\;\;\;\;}};
\end{aligned}$$

for some positive constant C_4. The lemma is proved.

A direct corollary of Lemma S.4 is the following lemma. Denote $\mathsf{T}_{1j} = \sqrt{\frac{n_j n_{j+1}}{n_j + n_{j+1}}} \;\; _n(\bar{\mathsf{S}}_j^O - \bar{\mathsf{S}}_{j+1}^O)$ and $\mathsf{T}_{2j} = \sqrt{\frac{n_j n_{j+1}}{n_j + n_{j+1}}}(\bar{\mathsf{S}}_j^E - \bar{\mathsf{S}}_{j+1}^E)$.

Lemma S.5 *Suppose Assumptions 1-2 hold. For those $j \in \mathcal{I}_0$, then we have as $n \to \infty$,*

$$\Pr\left\{\|\mathsf{T}_{kj}\|^2 > C(\log n + \,!\,_n^{-1} \;\; _n^2)\right\} = O(n^{1- \;\; \overline{\;\;\;\;}});\; k = 1; 2$$

for some large $C > 0$ and any $0 < \;\; < \;\; - 2 \;\; ^{-1}$.

Proof. We take T_{2j} as example. By Assumption 2, if there exists a true change point $\;_k^*$ between $\widehat{\;}_{j-1}$ and $\widehat{\;}_{j+1}$, it can only be either close to $\widehat{\;}_{j-1}$ or $\widehat{\;}_{j+1}$, but not $\widehat{\;}_j$. Without loss of generality, assume $0 \leq \ _k^* - \widehat{\ }_{j-1} \leq \ _n$. Then we note that

$$
\begin{aligned}
\|\mathsf{T}_{2j}\| &= \left\| \sqrt{\frac{n_j n_{j+1}}{n_j + n_{j+1}}} \left\{ \frac{1}{n_j} \sum_{i=\widehat{\ }_{j\ 1}+1}^{\widehat{\ }_j} \mathsf{V}_i - \frac{1}{n_{j+1}} \sum_{i=\widehat{\ }_j+1}^{\widehat{\ }_{j+1}} \mathsf{V}_i + \frac{\ _k^* - \widehat{\ }_{j-1}}{n_j} (\boldsymbol{\mu}_k^* - \boldsymbol{\mu}_{k+1}^*) \right\} \right\| \\
&\leq \sqrt{\frac{n_j}{n_j + n_{j+1}}} \left\| n_j^{-1=2} \sum_{i=\widehat{\ }_{j\ 1}+1}^{\widehat{\ }_j} \mathsf{V}_i \right\| + \sqrt{\frac{n_{j+1}}{n_j + n_{j+1}}} \left\| n_{j+1}^{-1=2} \sum_{i=\widehat{\ }_j+1}^{\widehat{\ }_{j+1}} \mathsf{V}_i \right\| + \left\| \frac{\ _k^* - \widehat{\ }_{j-1}}{n_j} (\boldsymbol{\mu}_k^* - \boldsymbol{\mu}_{k+1}^*) \right\| \\
&\leq \sqrt{\frac{n_j}{n_j + n_{j+1}}} \left\| n_j^{-1=2} \sum_{i=\widehat{\ }_{j\ 1}+1}^{\widehat{\ }_j} \mathsf{V}_i \right\| + \sqrt{\frac{n_{j+1}}{n_j + n_{j+1}}} \left\| n_{j+1}^{-1=2} \sum_{i=\widehat{\ }_j+1}^{\widehat{\ }_{j+1}} \mathsf{V}_i \right\| + !_n^{-1=2}\ _n \left\| \boldsymbol{\mu}_k^* - \boldsymbol{\mu}_{k+1}^* \right\| \\
&\leq 2 \max_{(k_1;k_2)\in\mathcal{T}(!_n)} (k_2 - k_1)^{-1=2} \left\| \sum_{i=k_1+1}^{k_2} \mathsf{V}_i \right\| + !_n^{-1=2}\ _n \left\| \boldsymbol{\mu}_k^* - \boldsymbol{\mu}_{k+1}^* \right\| :
\end{aligned}
$$

The assertion is immediately verified by using Lemma S.4.

Appendix D: Proof of Proposition 1

The proof of this proposition follows similarly to Theorem 2 in Barber et al. (2020) which shows that the Model-X knockoff selection procedure incurs an inflation of the false discovery rate that is proportional to the errors in estimating the distribution of each feature conditional on the remaining features. Fix $\ > 0$ and for any threshold $t > 0$, define

$$
R\ (t) = \frac{\sum_{j\in\mathcal{I}_0} \mathbb{I}(W_j \geq t; \Delta_j \leq \ )}{1 + \sum_{j\in\mathcal{I}_0} \mathbb{I}(W_j \leq -t)} :
$$

Consider the event that $\mathcal{A} = \{\Delta := \max_{j\in\mathcal{I}_0} \Delta_j \leq \ \}$. Furthermore, for a threshold rule

L.173 Td [( )]TJ 0 -755203 f38 -21.669 T(-203 f38 -11978 0 Td [(T)]73 Td [( )] Tf8

$1 + \sum_j (W_j \leq Ri)$

It is crucial to get an upper bound for $\mathbb{E}\{R\ (L) \mid \mathcal{Z}_O\}$. In what follows, all the "$\mathbb{E}(\cdot)$" denote the expectations given $\mathcal{Z}_O$. We have

$$
\begin{aligned}
\mathbb{E}\{R\ (L)\} &= \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{ \frac{\mathbb{I}\left(W_j \geq L; \Delta_j \leq \ \right)}{1+\sum_{j\in\mathcal{I}_0}\mathbb{I}\left(W_j \leq -L\right)} \right\} \\
&= \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{ \frac{\mathbb{I}\left(W_j \geq L_j; \Delta_j \leq \ \right)}{1+\sum_{k\in\mathcal{I}_0;k\neq j}\mathbb{I}\left(W_k \leq -L_j\right)} \right\} \\
&= \sum_{j\in\mathcal{I}_0} \mathbb{E}\left[\mathbb{E}\left\{ \frac{\mathbb{I}\left(W_j \geq L_j; \Delta_j \leq \ \right)}{1+\sum_{k\in\mathcal{I}_0;k\neq j}\mathbb{I}\left(W_k \leq -L_j\right)} \mid |W_j|; \mathbf{W}_{-j} \right\} \right] \\
&= \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{ \frac{\Pr\left(W_j > 0 \mid |W_j|; W_{j-1}; W_{j+1}; \mathcal{Z}_O\right)\mathbb{I}\left(|W_j| \geq L_j; \Delta_j \leq \ \right)}{1+\sum_{k\in\mathcal{I}_0;k\neq j}\mathbb{I}\left(W_k \leq -L_j\right)} \right\}; \qquad \text{(S.1)}
\end{aligned}
$$

where the last step holds since the only unknown is the sign of W_j after conditioning on $(|W_j|; W_{j-1}; W_{j+1})$. By definition of Δ_j, we have $\Pr\left(W_j > 0 \mid |W_j|; W_{j-1}; W_{j+1}; \mathcal{Z}_O\right) \leq 1/2 + \Delta_j$.

Hence,

$$
\begin{aligned}
&\mathbb{E}\{R\ (L)\} \\
&\leq \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{ \frac{(\frac{1}{2}+\Delta_j)\mathbb{I}\left(|W_j| \geq L_j; \Delta_j \leq \ \right)}{1+\sum_{k\in\mathcal{I}_0;k\neq j}\mathbb{I}\left(W_k \leq -L_j\right)} \right\} \\
&\leq (\frac{1}{2}+\ ) \left[\sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{ \frac{\mathbb{I}\left(W_j \geq L_j; \Delta_j \leq \ \right)}{1+\sum_{k\in\mathcal{I}_0;k\neq j}\mathbb{I}\left(W_k \leq -L_j\right)} \right\} + \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{ \frac{\mathbb{I}\left(W_j \leq -L_j\right)}{1+\sum_{k\in\mathcal{I}_0;k\neq j}\mathbb{I}\left(W_k \leq -L_j\right)} \right\} \right] \\
&= (\frac{1}{2}+\ ) \left[\mathbb{E}\{R\ (L)\} + \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{ \frac{\mathbb{I}\left(W_j \leq -L_j\right)}{1+\sum_{k\in\mathcal{I}_0;k\neq j}\mathbb{I}\left(W_k \leq -L_j\right)} \right\} \right].
\end{aligned}
$$

Finally, the sum in the last expression can be simplified as: if for all null j, $W_j > -L_j$, then the sum is equal to zero, while otherwise,

$$
\sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{ \frac{\mathbb{I}\left(W_j \leq -L_j\right)}{1+\sum_{k\in\mathcal{I}_0;k\neq j}\mathbb{I}\left(W_k \leq -L_j\right)} \right\} = \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{ \frac{\mathbb{I}\left(W_j \leq -L_j\right)}{1+\sum_{k\in\mathcal{I}_0;k\neq j}\mathbb{I}\left(W_k \leq -L_k\right)} \right\} = 1;
$$

where the first step comes from the fact: for any $j; k$, if $W_j \leq -\min(L_j; L_k)$ and $W_k \leq -\min(L_j; L_k)$, then $L_j = L_k$; see Barber et al. (2020).

Accordingly, we have

$$\mathbb{E}\{R\ (L)\} \le \frac{1/2+}{1/2-} \le 1+5\ .$$

Consequently, the assertion of this proposition holds.

Appendix E: Proof of Lemmas A.1-A.2

Note that both the candidate change-points set $\widehat{\mathcal{T}}_{p_n}$ and the statistics W_j are dependent with $\mathcal{Z}_O$. In fact, we derive the following two lemmas on the basis of conditional probability on $\mathcal{Z}_O$. To be specific, conditional on $\mathcal{Z}_O$, $\widehat{\mathcal{T}}_{p_n}$ is fixed as well as $(\bar{\mathbf{S}}_j^O - \bar{\mathbf{S}}_{j+1}^O)^\top \ _n$. Due to the independence between $\mathcal{Z}_E$ and $\mathcal{Z}_O$, the standard results for independent sum such as Lemmas S.2-S.3 can be applied for $\bar{\mathbf{S}}_j^E - \bar{\mathbf{S}}_{j+1}^E$ in the following arguments.

Proof of Lemma A.1

Define $\ _n = \{C(\log n + !_n^{-1}\ _n^2)\}^{1/2}$ for a large $C > 0$ specified in Lemma S.5. Let $\mathcal{A}_n = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| \ge t/\ _n\}$. Then, we observe that

$$\frac{G(t)}{G_-(t)} - 1 = \frac{\sum_{j\in\mathcal{I}_0}\{\Pr(\mathbf{T}_{1j}^\top \mathbf{T}_{2j} \ge t \mid \mathcal{Z}_O) - \Pr(\mathbf{T}_{1j}^\top \mathbf{T}_{2j} \le -t \mid \mathcal{Z}_O)\}}{p_0 G_-(t)}$$

Conditional on $\mathcal{Z}_O$, we have two cases. Firstly, for the case $\mathbf{T}_{1j} \in \mathcal{A}_n^c$, by Lemma S.5 we obtain that

$$\frac{G(t)}{G_-(t)} - 1 \le \frac{\sum_{j\in\mathcal{I}_0}\Pr(\mathbf{T}_{1j}^\top \mathbf{T}_{2j} \ge t \mid \mathcal{Z}_O)}{p_0 G_-(t)} \le \frac{\sum_{j\in\mathcal{I}_0}\Pr(\|\mathbf{T}_{2j}\| > \ _n \mid \mathcal{Z}_O)}{p_0/p_n} = O_p(n^{1-\ \overline{\quad}}\, p_n);$$

where the first inequality is due to $t \le G_-^{-1}(1/p_n)$, and thus we claim that $\frac{G(t)}{G\ \ (t)} - 1 = O_p(n^{1-\ \overline{\quad}}\, p_n)$.

Next, we consider the case $\mathbf{T}_{1j} \in \mathcal{A}_n$. We introduce a new sequence of independent random variables $\{\mathbf{B}_i\}$ defined as follows:

$$\mathbf{B}_i = \begin{cases} \frac{\sqrt{n_j n_{j+1}}}{n_j\sqrt{n_j+n_{j+1}}}\mathbf{V}_i; & \hat{\ }_{j-1} < i \le \hat{\ }_j; \\ -\frac{\sqrt{n_j n_{j+1}}}{n_{j+1}\sqrt{n_j+n_{j+1}}}\mathbf{V}_i; & \hat{\ }_j < i \le \hat{\ }_{j+1}. \end{cases}$$

By Lemma S.3, we firstly verify that for any given $\mathsf{u} \in \mathcal{A}_n$,

$$\frac{\Pr\left\{\sum_{i=\widehat{\ }_{j}\ \ {}_{1}+1}^{\widehat{\ }_{j}+1} \mathsf{u}^\top \mathsf{B}_i \geq t \mid \mathcal{Z}_O\right\}}{1 - \Phi(t = \sqrt{}}$$

Proof of Lemma A.2

We only show the validity of the first formula and the second one hold similarly. Note that the $G(t)$ is a deceasing and continuous function. Let $z_0 < z_1 < \cdots < z_{d_n} \leq 1$ and $t_i = G^{-1}(z_i)$, where $z_0 = a_n/p_n, z_i = a_n/p_n + a_n i\ /p_n, d_n = [\{(p_n - a_n)/a_n\}^{1/\ }]$ with $\ > 1$. Note that $G(t_i)/G(t_{i+1}) = 1 + o(1)$ uniformly in i. It is therefore enough to obtain the convergence rate of

$$D_n = \sup_{0 \leq i \leq d_n} \left| \frac{\sum_{j \in \mathcal{I}_0} \{\mathbb{I}(W_j \geq t_i) - \Pr(W_j \geq t_i \mid \mathcal{Z}_O)\}}{p_0 G(t_i)} \right| .$$

Define $\mathcal{S}_j = \{k \in \mathcal{I}_0 : W_k \text{ is dependent with } W_j\}$ and further

$$D(t) = \mathbb{E}\left[\left\{ \sum_{j \in \mathcal{I}_0} \mathbb{I}(W_j \geq t) - \Pr(W_j \geq t \mid \mathcal{Z}_O) \right\}^2 \mid \mathcal{Z}_O \right].$$

It is noted that

$$D(t) = \sum_{j \in \mathcal{I}_0} \sum_{k \in \mathcal{S}_j} \mathbb{E}\left[\{\mathbb{I}(W_j \geq t) - \Pr(W_j \geq t \mid \mathcal{Z}_O)\} \{\mathbb{I}(W_k \geq t) - \Pr(W_k \geq t \mid \mathcal{Z}_O)\} \mid \mathcal{Z}_O \right] \leq 2 p_0 G(t).$$

Note that conditional on $\mathcal{Z}_O$, $W_1, \dots, W_{p_n}$ is a 1-dependent sequence and so is $\mathbb{I}(W_j \geq t_i)$. We can get

$$\begin{aligned} \Pr(D_n \geq \ ) &\leq \sum_{i=0}^{d_n} \Pr\left(\left| \frac{\sum_{j \in \mathcal{I}_0} \{\mathbb{I}(W_j \geq t_i) - \Pr(W_j \geq t_i \mid \mathcal{Z}_O)\}}{p_0 G(t_i)} \right| \geq \ \right) \\ &\leq \frac{1}{\ ^2} \sum_{i=0}^{d_n} \frac{1}{p_0^2 G^2(t_i)} D(t_i) \leq \frac{2}{\ ^2} \sum_{i=0}^{d_n} \frac{1}{p_0 G(t_i)}. \end{aligned}$$

Moreover, observe that

$$\begin{aligned} \sum_{i=0}^{d_n} \frac{1}{p_0 G(t_i)} &= \frac{p_n}{p_0} \left(\frac{1}{a_n} + \sum_{i=1}^{d_n} \frac{1}{a_n + a_n i\ } \right) \\ &\leq c \left(\frac{1}{a_n} + a_n^{-1} \sum_{i=1}^{d_n} \frac{1}{1 + i\ } \right) \leq c a_n^{-1} \{1 + O(1)\}. \end{aligned}$$

In sum, we can have $\Pr(D_n \geq \ ) \to 0$ provided that $a_n \to \infty$.

Appendix F: Proof of Theorems 1-3 and Corollaries 1-2

Proof of Corollary 1

(i) By Assumption 2, we know that the event that $|\mathcal{I}_1| = K_n$ and for each $\widehat{\ }_j \in \mathcal{I}_1$, $|\widehat{\ }_j - \ {}^*_j| \leq \ {}_n$ occur with probability approaching one as $n \to \infty$. Therefore, in what follows we always implicitly work with the occurrence of this event. From the proof of Theorem 1, we know that $L \ . \ {}^2_n$. Hence

$$
\begin{aligned}
&\Pr\left(W_j < L; \text{ for some } \widehat{\ }_j \in \mathcal{I}_1 \mid \mathcal{Z}_O\right) \\
\leq& K_n \Pr\left(\frac{n_j n_{j+1}}{n_j + n_{j+1}} (\bar{\mathsf{S}}_j^O - \bar{\mathsf{S}}_{j+1}^O)^\top \ {}_n (\bar{\mathsf{S}}_j^E - \bar{\mathsf{S}}_{j+1}^E) < L \mid \mathcal{Z}_O\right) \\
\leq& K_n \Pr\left(\frac{n_j n_{j+1}}{n_j + n_{j+1}} (\bar{\mathsf{U}}_j - \bar{\mathsf{U}}_{j+1})^\top \ {}_n (\bar{\mathsf{V}}_j - \bar{\mathsf{V}}_{j+1}) + O_p^+(!_n \min_{1 \leq k \leq K_n} \|\boldsymbol{\mu}_{k+1}^* - \boldsymbol{\mu}_k^*\|^2) < L \mid \mathcal{Z}_O\right) \\
\leq& K_n \Pr\left(O_p^+(!_n \min_{1 \leq k \leq K_n} \|\boldsymbol{\mu}_{k+1}^* - \boldsymbol{\mu}_k^*\|^2) \leq L\right) \\
&+ K_n \Pr\left(\frac{n_j n_{j+1}}{n_j + n_{j+1}} (\bar{\mathsf{U}}_j - \bar{\mathsf{U}}_{j+1})^\top \ {}_n (\bar{\mathsf{V}}_j - \bar{\mathsf{V}}_{j+1}) > O_p^+(!_n \min_{1 \leq k \leq K_n} \|\boldsymbol{\mu}_{k+1}^* - \boldsymbol{\mu}_k^*\|^2) \mid \mathcal{Z}_O\right) \to 0
\end{aligned}
$$

in probability, where we use Lemma S.4. The result immediately holds.

(ii) From (i), we have $\lim_{n\to\infty} \Pr(\mathcal{M} \supseteq \mathcal{I}_1) = 1$. Here, we only need to prove $\lim_{n\to\infty} \Pr(\mathcal{M} \subseteq \mathcal{I}_1) = 1$, which is equivalent to show that $\lim_{n\to\infty} \Pr(\mathcal{M} \cap \mathcal{I}_0 = \emptyset) = 1$.

It is noted that

$$
\Pr(W_j \geq L; \text{ for some } j \in \mathcal{I}_0 \mid \mathcal{Z}_O) \leq \sum_{j \in \mathcal{I}_0} \Pr(W_j \geq L \mid \mathcal{Z}_O) \sim p_0 \frac{n}{p_n} \ . \ K_n \ ;
$$

By using the condition $K_n \ \to 0$, the corollary is proved.

Proof of Theorem 1

Following the notations in Section 2, assume $\widehat{\ }_j \in \mathcal{M}$ is an informative point and $\ {}^*_{j^0}$ is its corresponding true change-point such that $|\widehat{\ }_j - \ {}^*_{j^0}| \leq \ {}_n$ by Assumption 2. Note that $\widetilde{\ }_k \in \widetilde{\mathcal{M}}$ is the selected one such that $|\widetilde{\ }_k - \widehat{\ }_j| = \min_{\widetilde{\ }_l \in \widetilde{\mathcal{T}}_{p_n}} |\widetilde{\ }_l - \widehat{\ }_j|$. Because $\mathcal{M}$ and $\widetilde{\mathcal{M}}$ have the same cardinality, we only need to show that $\tilde{\ }_k \in \mathcal{I}_1(\widetilde{\mathcal{T}}_{p_n})$, say

$$|\tilde{\ }_k - \ {}^{*}_{j^0}| = \min_{\tilde{\ }_l \in \widetilde{\mathcal{T}}_{p_n}} |\tilde{\ }_l - \ {}^{*}_{j^0}| \tag{S.2}$$

(ii) Let $a_n = (C \log n)^{1/2}$, where $C > 0$ is specified in Lemma S.5. Define $\mathcal{B}_n = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| \geq t/a_n\}$. Let $\mathcal{C} = \bigcap_{j \in \mathcal{I}_0} \{|\tilde{W}_j| \leq \ _j\}$, where $\ _j$ satisfies $\Pr(|\tilde{W}_j| > \ _j \mid \mathcal{Z}_O) = b_n$ and b_n be a sequence satisfies the conditions that $b_n \to 0$, $p_n b_n \to 0$ and $n^{\ /2} b_n \to \infty$. According to the condition $p_n n^{-\ /2} \to 0$ in the theorem, such b_n is well defined. By the definition of $\tilde{W}_j$, we know that $\mathbb{E}(\tilde{W}_j) = 0$ for all $\hat{\ }_j \in \mathcal{I}_0$. Moreover, by Lemma S.5, we have $\ _j \lesssim a_n^2$ uniformly in j.

According to Proposition 1, we have

$$\Pr\left(\max_{j \in \mathcal{I}_0} \Delta_j > \ \mid \mathcal{Z}_O\right) = \Pr\left(\max_{j \in \mathcal{I}_0} \Delta_j > \ \mid \mathcal{C}, \mathcal{Z}_O\right) \Pr(\mathcal{C} \mid \mathcal{Z}_O) + \Pr\left(\max_{j \in \mathcal{I}_0} \Delta_j > \ , \mathcal{C}^c \mid \mathcal{Z}_O\right)$$
$$\leq \Pr\left(\max_{j \in \mathcal{I}_0} \Delta_j > \ \mid \mathcal{C}, \mathcal{Z}_O\right) + \Pr(\mathcal{C}^c \mid \mathcal{Z}_O) := A_1 + A_2.$$

By the definition of b_n, $A_2 = o_p(1)$. It remains to handle A_1.

Notice that conditional on $\mathcal{C}$,

$$\max_{j \in \mathcal{I}_0} \Delta_j \leq \max_{j \in \mathcal{I}_0} \sup_{0 \leq t \leq \ _j} |f_j(-t)/f_j(t) - 1|, \tag{S.3}$$

where $f_j(\cdot)$ is the density of $\tilde{W}_j$ conditional on $\mathcal{Z}_O$. It remains to prove that the right-hand side of (S.3) goes to zero as $n \to \infty$.

Denote $\tilde{\mathbf{T}}_{1j} = \sqrt{\frac{n_j n_{j+1}}{n_j + n_{j+1}}}\ _n(\tilde{\mathbf{S}}^O_{Lj} - \tilde{\mathbf{S}}^O_{Rj}) = \mathbf{u}$ given $\mathcal{Z}_O$. In a similar way to the proof of Lemma A.1, we consider two cases for $\mathbf{u}$. As to the case $\mathbf{u} \in \mathcal{B}_n^c$, $\max_{j \in \mathcal{I}_0} \Delta_j = O_p\{(n^{\ /2} b_n)^{-1}\}$ by the definition of $\ _j$ and $0 \leq t \leq \ _j$. On the other hand, we consider the case $\mathbf{u} \in \mathcal{B}_n$. Then, for $j \in \mathcal{I}_0$ by Lemma S.3, we have

$$f_j(t) = \{\tilde{\Phi}(t/s) - \tilde{\Phi}(t/s-)\}\{1 + o_p(1)\} = \frac{1}{s}\ (t/s)\{1 + o_p(1)\},$$

where $s = \sqrt{\mathbf{u}^\top \ \mathbf{u}}$. Similarly, we also have $f_j(-t) = \frac{1}{s}\ (-t/s)\{1 + o_p(1)\}$, which yields that the right-hand side of (S.3) goes to zero since $\ (-t/s) = \ (t/s)$ and $\ (t/s)$ is bounded. Then, the result (ii) in the theorem holds.

Appendix G: Additional simulation results

Selection of p_n and ω_n

Table S1 reports the FDR, TPR and $\widehat{K}$ of MOPS in conjunction with OP, PELT and WBS detection algorithms with different p_n and ω_n under Example I. We consider the error from $N(0, 1)$ and fix $n = 4096$, $K_n = 15$ and SNR=0.5. We observe that different values of $c \in (1, 2]$ for $p_n = \lfloor cn^{2/5} \rfloor$ and $\gamma \in [0.3, 0.5]$ for $\omega_n = n^{\gamma}$ present similar results and their FDRs are not significantly different. Thus we recommend $p_n = \lfloor 2n^{2/5} \rfloor$ and $\omega_n = \min(\lfloor n^{0.5} \rfloor, 60)$ in the simulation studies.

Table S1: *FDR(%), TPR(%) and $\widehat{K}$ of MOPS in conjunction with OP, PELT and WBS detection algorithms when error follows $N(0, 1)$, $n = 4096$, $K_n = 15$ and SNR=0.5 under Example I. The p_n is chosen as $p_n = \lfloor cn^{2/5} \rfloor$ with $c = 1.2, 1.5, 2$ and $\omega_n = n^{\gamma}$ with $\gamma = 0.3, 0.4, 0.5$.*

		$\gamma = 0.3$			$\gamma = 0.4$			$\gamma = 0.5$		
p_n	Method	FDR	TPR	$\widehat{K}$	FDR	TPR	$\widehat{K}$	FDR	TPR	$\widehat{K}$
$1.2n^{2/5}$	M-OP	19.8	91.3	17.7	19.1	92.7	17.9	18.9	95.2	18.3
	M-PELT	19.5	91.2	17.5	19.4	93.1	18.0	19.9	95.7	18.8
	M-WBS	16.9	91.8	17.3	17.2	92.3	17.5	19.5	95.3	18.5
$1.5n^{2/5}$	M-OP	18.6	90.0	17.2	21.0	92.9	18.6	20.5	93.7	18.4
	M-PELT	16.6	89.3	16.7	20.8	93.1	18.6	21.2	94.1	18.9
	M-WBS	17.3	85.9	16.3	18.3	86.3	16.6	16.4	90.9	17.0
$2n^{2/5}$	M-OP	20.5	79.5	17.4	19.5	82.1	16.5	20.4	85.3	17.0
	M-PELT	20.2	80.3	17.1	20.0	82.7	16.7	19.7	85.5	17.0
	M-WBS	19.6	76.7	15.9	17.8	77.8	15.8	18.1	83.1	16.5

Next, we investigate the performance of our methods in the case that $p_n > 2n^{2/5}$. Figure S1 presents the FDR and TPR curves of MOPS, R-MOPS and M-MOPS when p_n varies in $(2n^{2/5}, n/10)$ and the WBS algorithm is employed under Example I. Here we fix $\omega_n = 10$ and the true change-point number $K_n = 30$ and consider the error comes from $N(0, 1)$ and standardized $\chi^2(3)$. The FDR values of MOPS vary in an acceptable range of the target level no matter the choice of p_n under normal error, but are slightly distorted under standardized $^2(3)$ error. The R-MOPS is able to improve TPR and yield smaller FDR levels than MOPS due to the use of full sample information. We also observe that the M-MOPS leads to more conservative FDR levels and smaller TPR than R-MOPS because of only using half of the observations around each candidate point. That is consistent with our theoretical analysis in Proposition 1 and Theorem 3. Similar results can also be found in Figure S2.

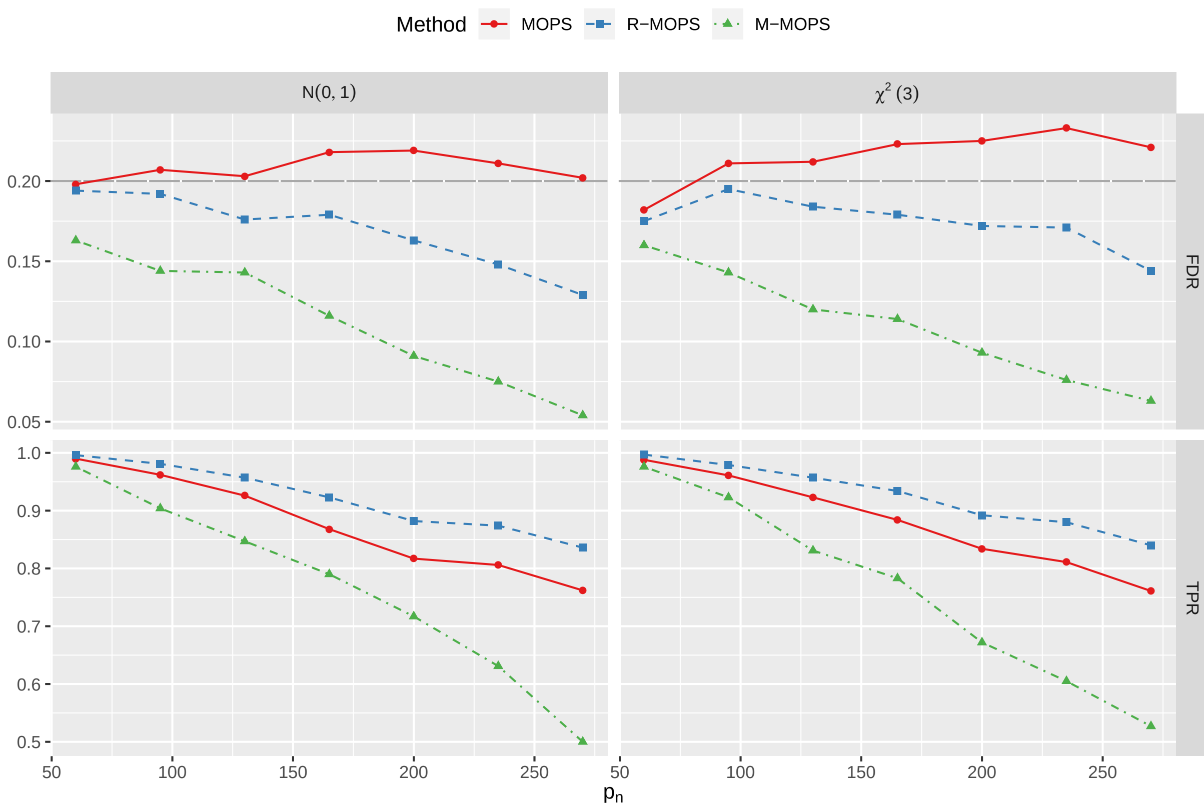

Figure S1: *FDR and TPR curves against* $p_n \in (2n^{2=5}; n{=}10)$ *of MOPS, R-MOPS and M-MOPS in conjunction with WBS algorithm when* n = 4096, K_n = 30 *and SNR=1 under Example I. The* $!_n$ *is xed as 10.*

Figure S2 shows the FDR and TPR curves against $!_n$ of the MOPS, R-MOPS and M-MOPS in conjunction with WBS algorithm when n = 4096, K_n = 10 and p_n is fixed as $\lfloor 2n^{2=5} \rfloor$ under Example I. It implies that all the procedures are not sensitive to the choice of $!_n$ in terms of FDR control. Meanwhile, a large $!_n$ could improve the detection power due to more observations in each segment.

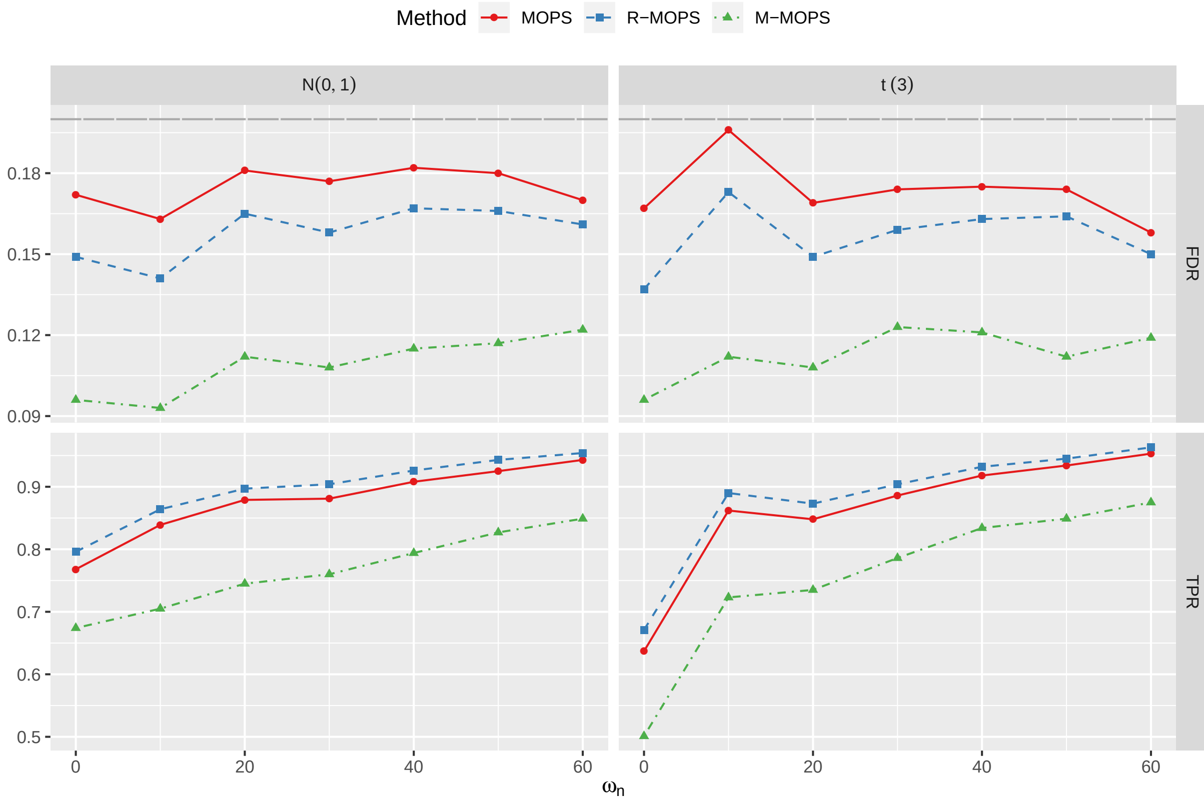

Figure S2: *FDR and TPR curves against ω_n of MOPS, R-MOPS and M-MOPS in conjunction with WBS algorithm when $n = 4096$, $K_n = 10$ and SNR=0.7 under Example I. The p_n is fixed as $\lfloor 2n^{2/5} \rfloor$.*

Comparison under other models

Three other MCP models are considered, reflecting changes in different aspects such as the location and scale. Table S2 gives a summary of all three simulated models along with the associated statistics $\bar{\mathsf{S}}_j^O$ in constructing W_j.

Under multivariate mean change model (Example III), we examine the performance of the refined MOPS in conjunction with the OP and PELT algorithms. For simplicity, each dimension of the signals $\boldsymbol{\mu}_i$'s is set as the same as the signals μ_i's in Example I. Two scenarios for the error distribution are considered: (i) $\boldsymbol{\varepsilon}_i \overset{\text{iid}}{\sim} N(\mathbf{0}, \Sigma)$ with $\Sigma = (0.5^{|i-j|})_{d\times d}$;

Table S2: *Preview of simulated models and the sample mean* $\mathbf{S}_j^O$ *of the* j*-th segment for the odd part. Change-points* $\widehat{\ }_j$*'s are estimated on the basis of* $\mathcal{Z}_O$*.*

NO.	Model	$\mathbf{S}_j^O$
III	$\mathbf{X}_i = \boldsymbol{\mu}_i + \quad \boldsymbol{\varepsilon}_i$	$\mathbf{X}^O_{\widehat{\ }_{j\ \ 1};\widehat{\ }_j}$
IV	$\mathbf{X}_i \sim \text{Multinomial}(m; \mathbf{q}_i)$	$\mathbf{X}^O_{\widehat{\ }_{j\ \ 1};\widehat{\ }_j}$
V	$X_i = \quad {}_i{}''_i$	$V^O_{\widehat{\ }_{j\ \ 1};\widehat{\ }_j}$, $V_i = \log X_i^2$

(ii) $\boldsymbol{\varepsilon}_i = (''_{i1}; \dots; ''_{id})^\top$, where $''_{i1}; \dots; ''_{id} \overset{\text{iid}}{\sim} (\ \ \frac{2}{5} - 5)/\sqrt{10}$. We consider the dimension $d = 5$, 10 and adjust the scale parameter to $\quad = 9\sqrt{d}$. Table S3 presents the results when the sample size $n = 3072$ and the number of change-points $K_n = 27$. The R-MOPS-based methods perform reasonably well in terms of FDR control and reliable TPR. In contrast, the CV-PELT results in overly conservative FDR levels across all the settings and its P_a's are much smaller than those of R-MOPS.

Table S3: *Comparison results of FDR(%), TPR(%),* P_a*(%) and* $\widehat{K}$ *when* $K_n = 27$ *and* $n = 3072$ *under Example III (multivariate mean shift).*

		$d = 5$				$d = 10$			
errors	Method	FDR	TPR	P_a	$\widehat{K}$	FDR	TPR	P_a	$\widehat{K}$
	RM-OP	18.5	97.1	53.5	32.9	18.9	92.9	35.0	32.1
$\boldsymbol{\varepsilon}_i \sim N(0; \quad)$	RM-PELT	18.7	97.5	54.0	32.8	18.5	92.5	33.5	31.9
	CV-PELT	0.9	91.3	18.0	24.9	0.6	86.7	4.0	23.6
	RM-OP	19.7	99.1	87.0	33.0	20.3	95.9	68.5	33.1
$''_{ij} \sim \frac{\frac{2}{5}-5}{\sqrt{10}}$	RM-PELT	19.5	99.0	85.5	32.7	20.8	96.2	69.5	33.3
	CV-PELT	0.8	85.9	4.5	23.4	1.7	84.1	0.0	23.1

Further, we consider the MCP problem for multinomial distributions (Example IV), i.e. $\mathbf{X}_i \sim \text{Multinom}(n_0; \mathbf{q}_i)$, where the variance of the observation relies on their mean. Braun et al. (2000) integrated the problem into quasi-likelihood framework in combination with BIC to determine the number of change-points. In particular, they aimed to identify the breaks in the probability vectors $\mathbf{q}_i$'s and recommended the BIC with a penalty $\ _n = 0.5n^{0.23}$, which will be seen as a benchmark for comparison in this example. To implement MOPS, we apply their algorithm in our training step, i.e., given a candidate model size p_n, we obtain the estimated change-points by constructing the statistics W_j in (5). We follow the same mechanism in Braun et al. (2000) to generate $\mathbf{q}_i$'s. To be specific, the initial mean vector $\mathbf{q} = (q_1, \ldots, q_d)^\top$ is given as $q_j = U_j / \sum_{l=1}^{d} U_l$ for $j = 1, \ldots, d$ where $U_j \sim \text{Uniform}(0, 1)$. The jump mean vector $\mathbf{q}_k^* = (q_1^*, \ldots, q_d^*)^\top$ for change point k is obtained by normalizing $\text{expit}(\text{logit} q_l^* + U_l^*)$ for $l = 1, \ldots, d$ where $U_l^* \sim \text{Uniform}(-J, J)$ with $J = 0.8/\sqrt{d}$. Table S4 reports the simulation results when $n = 2048$, $K_n = 20$, $n_0 \in (80, 100, 120)$ and d is chosen as 5 or 10. Again, our R-MOPS can successfully control the FDR at the nominal level in most cases. The BIC method appears to result in a slightly underfitting model on average. Accordingly, the BIC method delivers conservative FDR levels and it may miss some change-points due to relatively low P_a.

Table S4: *Comparison results of FDR(%), TPR(%), P_a(%) and $\widehat{K}$ between R-MOPS and BIC in conjunction with Braun et al. (2000)'s algorithm when $K_n = 20$ and $n = 2048$ under Example IV.*

		$d = 5$				$d = 10$			
n_0	Method	FDR	TPR	P_a	$\widehat{K}$	FDR	TPR	P_a	$\widehat{K}$
80	R-MOPS	20.2	98.1	85.5	25.2	17.1	92.8	45.0	23.1
	BIC	1.8	92.2	41.0	19.4	1.9	89.2	32.0	19.3
100	R-MOPS	21.1	99.2	92.0	26.0	20.1	98.3	75.5	25.2
	BIC	1.6	94.7	62.5	19.6	1.5	93.2	55.5	19.5
120	R-MOPS	21.5	99.8	97.5	26.2	21.2	99.0	85.0	26.0
	BIC	1.3	97.2	73.5	19.7	1.1	96.4	69.0	19.6

At last, we investigate the performance of R-MOPS in conjunction with PELT under Example V when the scale signal function of $\ _i$'s is chosen as a piecewise constant function with values alternating between 1 and 0.5. We fix $n = 4096$ and show the curves of FDR, TPR and P_a when $K_n \in [28, 35]$ in Figure S3. We observe that the FDRs of R-MOPS with PELT get closer to the target level as K_n increases, which is in accordance with the theoretical justification. Meanwhile, the CV-PELT method usually results in an underfitting model because some true change-points are not selected.

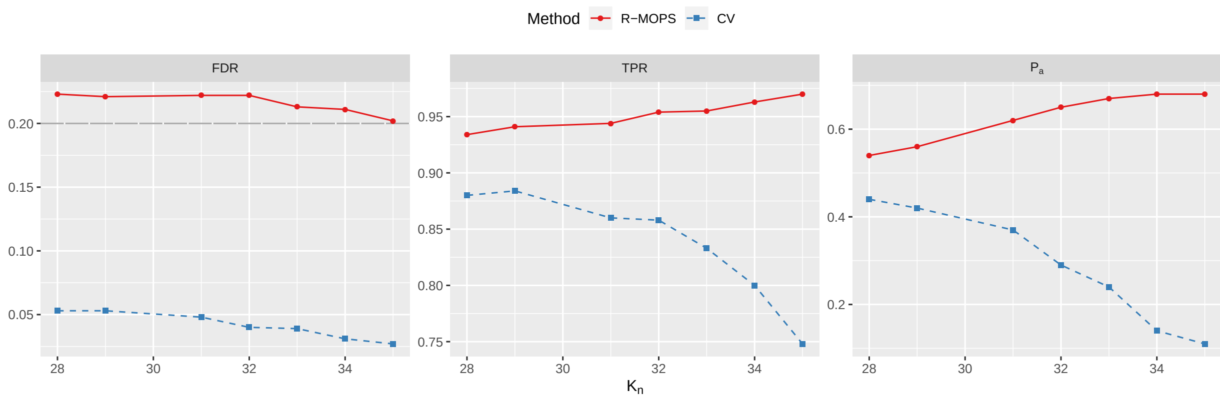

Figure S3: *FDR, TPR and P_a curves against K_n between R-MOPS and CV criterion based on PELT when n = 4096 and errors are i.i.d from standardized t_5 under Example V.*

Extension on controlling PFER

Table S5 reports some PFER results of the MOPS in conjunction with OP and PELT when the target PFER level $k_0 = 1, 5$ or 10. We fix the sample size $n = 4096$, the dimension $d = 5$ for multivariate data and consider that all errors are distributed from $N(0, 1)$. The validity of our MOPS approach in terms of PFER control is clear.

Others

Figure S4 displays the performance comparison under Example I with the same model setting as Section 5.1 when the target FDR level is $= 0.1$. The comparison results are analogous to those in nominal level $= 0.2$.

Table S6 presents the comparisons between our R-SaRa and dFDR-SaRa under Example I. Following the recommendation in Hao et al. (2013), we choose four thresholds $h_1 = \lfloor 3 \log n \rfloor$, $h_2 = \lfloor 5 \log n \rfloor$, $h_3 = \lfloor 7 \log n \rfloor$ and $h_4 = \lfloor 9 \log n \rfloor$ as simple competitors. It is

Table S5: *PFER performance of MOPS in conjunction with OP and PELT when the target PFER level* $k_0 = 1; 5$ *and* 10 *under Examples I-V.*

Example	Method \ k_0	$K_n = 5$: 1	$K_n = 5$: 5	$K_n = 5$: 10	$K_n = 10$: 1	$K_n = 10$: 5	$K_n = 10$: 10	$K_n = 15$: 1	$K_n = 15$: 5	$K_n = 15$: 10
I	M-OP	1.08	5.07	9.83	0.98	5.13	9.73	0.92	4.96	10.56
	M-PELT	0.86	4.94	9.86	0.91	5.23	10.18	1.06	5.07	10.90
II	M-OP	0.79	4.86	10.03	0.69	4.72	10.25	0.89	4.97	10.04
	M-PELT	0.74	4.14	9.57	0.77	4.93	10.36	0.66	5.05	8.58
III	M-OP	0.65	5.04	10.05	1.06	5.10	10.13	0.94	5.01	10.72
	M-PELT	0.67	4.78	9.83	0.83	4.87	10.27	0.72	4.91	10.60
IV	M-OP	0.81	4.13	9.18	1.01	5.16	9.93	0.97	5.13	9.75
	M-PELT	0.68	4.22	9.00	1.02	4.74	9.74	0.83	5.09	10.08
V	M-OP	0.78	5.10	9.93	0.89	5.09	10.08	1.13	5.07	10.89
	M-PELT	0.62	4.97	10.21	0.77	4.89	10.38	0.72	5.02	11.12

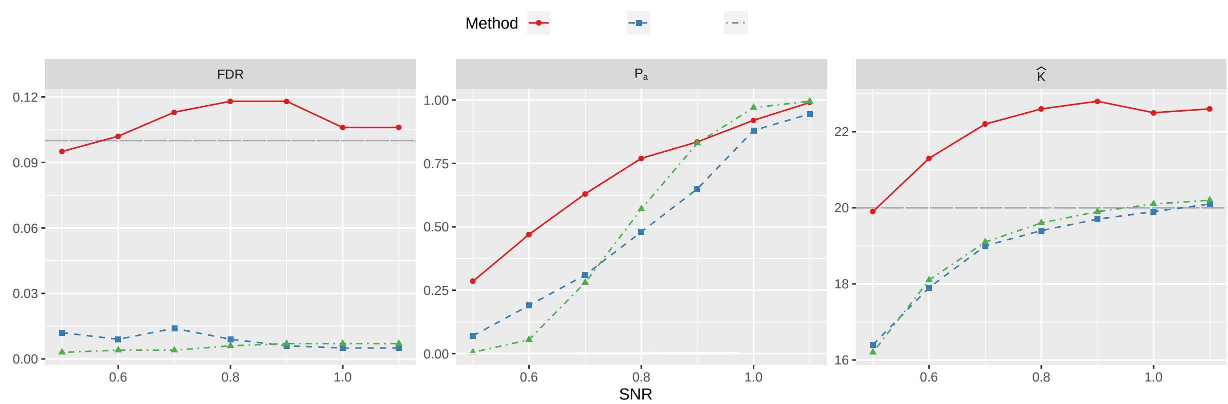

Figure S4: *FDR,* P_a *and the average number of estimated change-points* $\widehat{K}$ *curves against SNR among RM-PELT, CV-PELT and FDRseg when* $K_n = 20$, $n = 2048$ *and the target FDR level* $= 0{:}1$ *under Example I.*

clear that the R-MOPS performs well in terms of FDR control, but the performance of dFDR-SaRa depends on the choice of h to a large extent.

For the frequent change-point setting, Fryzlewicz (2020) proposed WBS2 detection algo-

Table S6: *Comparison results of FDR(%), TPR(%), P_a(%) and $\widehat{K}$ between RM-Sara and dFDR-SaRa-h in Hao et al. (2013) when n = 10240 and SNR=0.7 under Example I.*

		$K_n = 20$				$K_n = 40$			
Errors	Method	FDR	TPR	P_a	$\widehat{K}$	FDR	TPR	P_a	$\widehat{K}$
	RM-SaRa	19.5	99.2	84.0	25.4	22.2	99.8	92.0	52.2
	dFDR-SaRa-h_1	17.1	78.2	6.5	19.2	10.7	83.9	1.0	37.8
$N(0,1)$	dFDR-SaRa-h_2	10.2	94.3	44.0	21.0	3.0	95.9	29.0	39.6
	dFDR-SaRa-h_3	9.6	97.3	70.5	21.6	0.2	98.3	49.5	39.4
	dFDR-SaRa-h_4	3.4	99.1	90.0	20.5	0.0	95.8	1.0	38.3
	RM-SaRa	18.6	99.7	94.5	25.3	20.9	99.9	96.5	51.1
	dFDR-SaRa-h_1	16.8	89.2	18.5	21.6	11.0	92.8	13.5	41.9
$\chi^2(3)$	dFDR-SaRa-h_2	12.8	98.1	74.0	22.7	2.0	99.3	81.0	40.5
	dFDR-SaRa-h_3	7.2	99.7	96.5	21.6	0.3	99.8	92.0	40.0
	dFDR-SaRa-h_4	2.6	100.0	100.0	20.6	0.0	95.3	0.0	39.0

rithm with threshold-based model selection criterion "Steepest Drop to Low Levels" (SDLL). We compare our procedure R-MOPS in conjunction with WBS2 to the WBS2.SDLL criterion when the "extreme.teeth" example of the univariate changes in Fryzlewicz (2020) is considered. Specially, in the "extreme.teeth" example, the mean μ_i's for each observation are defined as follows: $\mu_i = 0$ if $1 \leq \mathrm{mod}(i,10) \leq 5$ and $\mu_i = 1$ if $\mathrm{mod}(i,10) \in \{0,6,7,8,9\}$, and the sample size n is 1000. Two values of SNR and three error distributions including $N(0,1)$, standardized $t(3)$ and standardized $\chi^2(3)$ are considered. We fix $\omega_n = 4$ and $p_n = 250$ for the R-MOPS. From Table S7, we can see that the FDRs of R-MOPS with WBS2 are still controlled, though they appear to be overly conservative. The WBS2.SDLL generally has better performances in terms of $\widehat{K}$ estimation in the most settings.

Another real-data example: OPEC oil price

We analyze the daily Organisation of the Petroleum Exporting Countries (OPEC) Reference Basket oil prices from Jan. 6, 2003 to Dec. 16, 2020 with sample size $n = 4610$, which is available from `https://www.quandl.com`. As the raw oil price series tend to ex-

Table S7: *Comparisons of $\widehat{K}$, FDR(%) and TPR(%) between R-MOPS and SDLL in conjunction with WBS2 Fryzlewicz (2020)'s "extreme.teeth" example when $n = 1000$, $K_n = 199$ and three error distributions are considered. The target FDR level is $\alpha = 0.2$ and σ^2 is the error variance.*

		$\sigma = 0.3$			$\sigma = 0.5$		
Error	Method	$\widehat{K}$	FDR	TPR	$\widehat{K}$	FDR	TPR
$N(0,1)$	RMOPS	193.7	7.1	90.4	160.6	10.0	72.6
	SDLL	199.4	3.8	96.4	71.6	9.0	29.3
$t(3)$	RMOPS	193.9	7.1	90.5	176.5	8.0	81.6
	SDLL	209.8	7.1	97.8	221.8	19.6	89.0
$\chi^2(3)$	RMOPS	193.1	7.1	90.2	167.9	8.9	76.8
	SDLL	211.0	8.1	97.2	200.5	22.8	77.3

hibit strong autocorrelation (Baranowski et al., 2019), we consider analyzing the log-returns $100\log(P_i/P_{i-1})$, where P_i is the daily oil price. Figure S5 presents the data sequence of log-returns and its autocorrelation, indicating the correlations of log-returns are relatively weak. As Baranowski et al. (2019) pointed out that both mean and scale changes exist in the sequence, we build $\mathsf{S}_i = (\mathbb{Z}_i, \log(\mathbb{Z}_i^2))^\top$ in W_j for the proposed MOPS procedure to detect changes in both the mean and variance when PELT algorithm is applied. In this study, we use the function `cpt.meanvar()` in R package `changepoint` to implement the PELT algorithm and also report change-points detected by the BIC for comparison.

The BIC results in 33 change-points, while the R-MOPS with PELT yields 36 and 55 change-points when the target FDR level is 0.05 and 0.1, respectively. The locations of the change-points identified by BIC and R-MOPS with $\alpha = 0.05$ are given in the left panel of Figure S5. The estimated change-points of both methods largely agree each other. However, the BIC does not indicate any changes in late 2004 and early 2005 and meanwhile R-MOPS has several estimated change-points in that period. This period could potentially be related to a noticeable expansion of the production volume in the late 2004, which leads to a significant change of oil price elasticity. Thus, Murray and King (2012) called the early

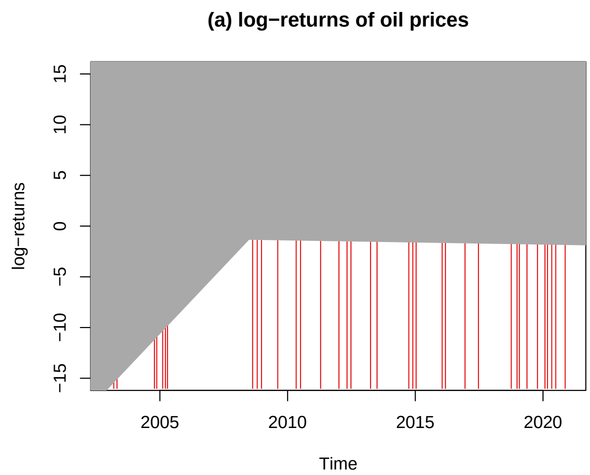

Figure S5: *(a): Scatter plots of the log-returns of daily OPEC oil prices, where the blue dash and red solid lines represent the estimated change-points detected by BIC and R-MOPS with PELT algorithm under* $= 0.05$*; (b) Autocorrelation of log-returns.*

2005 was oil's tipping point.

References

Baranowski, R., Chen, Y., and Fryzlewicz, P. (2019), “Narrowest-over-threshold detection of multiple change points and change-point-like features,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81, 649–672.

Barber, R. F., Candes, E. J., and Samworth, R. J. (2020), “Robust inference with knockoffs,” *The Annals of Statistics*, 48, 1409–1431.

Braun, J. V., Braun, R. K., and Müller, H. G. (2000), “Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation,” *Biometrika*, 87, 301–314.

Fryzlewicz, P. (2020), “Detecting possibly frequent change-points: Wild Binary Segmentation 2 and steepest-drop model selection,” *Journal of the Korean Statistical Society*, 1–44.

Hao, N., Niu, Y. S., and Zhang, H. (2013), “Multiple change-point detection via a screening and ranking algorithm,” *Statistica Sinica*, 23, 1553–1572.

Murray, J. and King, D. (2012), “Oil's tipping point has passed,” *Nature*, 481, 433–435.

Petrov, V. (2002), “On probabilities of moderate deviations,” *Journal of Mathematical Sciences*, 109, 2189–2191.

Zou, C., Wang, G., and Li, R. (2020), “Consistent selection of the number of change-points via sample-splitting,” *The Annals of Statistics*, 48, 413–439.