Data-Driven Determination of the Number of Jumps in Regression Curves

Guanghui Wang*
KLATASDS-MOE, Academy of Statistics and Interdisciplinary Sciences, East China Normal University
and
Changliang Zou†
School of Statistics and Data Science, Nankai University
and
Peihua Qiu
Department of Biostatistics, University of Florida

September 5, 2021

The supplementary material contains a list of assumptions and proofs of Theorems 1 and 2, along with some additional numerical results.

In the description below, the notation $A_n \sim B_n$ means that A_n/B_n and B_n/A_n are both bounded (in probability) as $n \to \infty$. The notations "$\gtrsim$" and "$\lesssim$" are similarly defined.

A Assumptions

In the discussion, the jump magnitudes $\{\delta_j^*\}$ and the true number of jumps J^* are allowed to be dependent on n, while we keep the variance σ^2 as fixed. If σ^2 is changing in n, we could replace δ_j^* by δ_j^*/σ in the following asymptotic analysis. For simplicity, the design

*Wang was supported by NNSF of China Grant 11901314. Zou was supported by NNSF of China Grants (11925106, 11690015, 11931001, 11971247) and NSF of Tianjin Grant 18JCJQJC46000. Qiu was supported in part by the NSF grant DMS-1914639.

†Corresponding author: nk.chlzou@gmail.com

points are assumed to be equidistant in $[0, 1]$, i.e., $t_i = i/n$ for $i = 1, \dots, n$. The required assumptions on the model and the smoothing techniques are listed below.

Assumption S.1 (*Noises*)**.** There exist some constants $\theta \geq 3$ and $C_\varepsilon > 0$ such that $\mathbb{E}(|\varepsilon_1/\sigma|^\theta) \leq C_\varepsilon$.

Assumption S.2 (*Kernel*)**.** $K(\cdot)$ is a nonnegative, bounded and Lipschitz-1 continuous function with support $[0, 1]$ such that $\int_0^1 K(u)\mathrm{d}u = 1$ and $K(0) > 0$.

Assumption S.3 (*Bandwidth*)**.** The bandwidth $h > 0$ satisfies the conditions that $h \to 0$, $nh \to \infty$ and $nh^5/\log n \to 0$, as $n \to \infty$.

Assumption S.4 (*Jump distance*)**.** The minimum distance between two successive jumps, $\min_{0\leq j\leq J^*}\{t^*_{j+1} - t^*_j\}$, satisfies the condition that $\min_{0\leq j\leq J^*}\{t^*_{j+1} - t^*_j\}/h$ tends to infinity as $n \to \infty$, where $t^*_0 = 0$ and $t^*_{J^*+1} = 1$.

Assumption S.5 (*Smoothness*)**.** The underlying regression curve $f(t)$ is twice differentiable and $f''(t)$ is uniformly bounded in its support except at the jump points $\{t^*_j : j = 1, \dots, J^*\}$, at which $f(t)$ has bounded left and right second-order derivatives.

Assumption S.6 (*Jump magnitudes*)**.** The jump magnitudes satisfy the conditions that (a) $\max_{1\leq j\leq J^*} |\delta^*_j| = O(1)$ and (b) $\min_{1\leq j\leq J^*} |\delta^*_j|/\sqrt{\log n/(nh)} \to \infty,$ as $n \to \infty$.

Assumption S.1 is used to control the supremum of the noise terms in the local linear smoothers. Assumptions S.2, S.3 and S.5 are standard in the kernel smoothing literature. Assumptions S.4 and S.6 are technical conditions required to achieve consistency. Assumption S.4 allows the minimum distance between two consecutive jumps to go to zero as n

increase, but the rate converging to zero should be slower than h to facilitate asymptotic analysis. Assumption S.6 requires that the minimum jump magnitude should dominate the noise level of the one-sided LLK estimate of $f(t)$. In a simplified scenario that there are two true jumps with equal jump magnitudes $\delta^* > 0$ and both jumps are precisely identified by a jump detection procedure, it can be checked that $\widehat{\delta}^O(\widehat{t}_j^O) = \delta^* + O_p\{(nh)^{-1/2}\}$ and $\widehat{\delta}^E(\widehat{t}_j^O) = \delta^* + O_p\{(nh)^{-1/2}\}$ for $j = 1, 2$. To avoid underfitting, it is required that $C(2; \mathcal{Z}^O, \mathcal{Z}^E) < C(1; \mathcal{Z}^O, \mathcal{Z}^E)$ with an overwhelming probability. A necessary condition is $\delta^*/(nh)^{-1/2} \to \infty$ such that the jump magnitudes δ^* are distinguishable from the noises. Assumption S.6 is close to this rate up to a logarithmic factor $\log n$, which is required since we need a uniform dominance over the noise level at multiple design points. If the locations $\{t_j^*\}$ and magnitudes $\{\delta_j^*\}$ of jumps are assumed to be fixed, then Assumptions S.4 and S.6 hold naturally (Xia and Qiu, 2015). In a closely-related problem that the signal is a piecewise constant function, Niu and Zhang (2012) proposed their detection statistic based on the difference between a right- and a left-sided local average, and the number of jumps was selected by a thresholding rule. They constructed the selection consistency under the normality assumption, the following assumption on the jump magnitudes

$$\min_{1 \le j \le J^*} |\delta_j^*/\sigma| \times \sqrt{\min_{0 \le j \le J^*} \{t_{j+1}^* - t_j^*\}/h} > \sqrt{32 \log n/(nh)}, \tag{S.1}$$

and some other requirements on the bandwidth h and the threshold parameter. Assumptions S.4 and S.6 can guarantee the assumption (S.1). For our proposed method, it requires a future study regarding whether the assumptions on both the minimum jump distance and the minimum jump magnitude can be weakened.

B Proofs

B.1 Proof of Theorem 1

For notional convenience, rewrite Model (1) as

$$Y_{2i-1} = f(t_i^O) + U_i \text{ and } Y_{2i} = f(t_i^E) + V_i,\ i = 1, \ldots, n,$$

where $t_i^O = (i-0.5)/n$, $t_i^E = i/n$ and U_i's and V_i's are i.i.d. with mean 0 and variance σ^2. Note that the original sample size is assumed to be $N = 2n$ for some integer n. Write for short that $t_i = t_i^O$, $f_i = f(t_i^O)$ for $i = 1, \ldots, n$ and $\widehat{t}_j = \widehat{t}_j^O$ (cf. Eq. (4)) for $j = 1, 2, \ldots$. For a sequence of quantities $\{g_1, \ldots, g_n\}$, we introduce

$$\widehat{g}_\pm(t) = \frac{\sum_{i=1}^n w_i^O(t; K_\pm)\, g_i}{\sum_{i=1}^n w_i^O(t; K_\pm)} \text{ and } \widehat{g}_\pm^E(t) = \frac{\sum_{i=1}^n w_i^E(t; K_\pm)\, g_i}{\sum_{i=1}^n w_i^E(t; K_\pm)},$$

and $\widehat{d}^g(t) = \widehat{g}_+(t) - \widehat{g}_-(t)$ and $\widehat{d}_E^g(t) = \widehat{g}_+^E(t) - \widehat{g}_-^E(t)$.

Our goal is to show that, with an overwhelming probability, $\mathcal{C}(J^*) < \mathcal{C}(J)$ in cases when either $1 \le J < J^*$ or $J^* < J \le \bar{J}$, where

$$\mathcal{C}(J) := \mathcal{C}(J; \mathcal{Z}^O, \mathcal{Z}^E) = \sum_{j=1}^{J} \left\{\widehat{\delta}^E(\widehat{t}_j) - \widehat{\delta}^O(\widehat{t}_j)\right\}^2 + \sum_{j=J+1}^{\bar{J}} \left\{\widehat{\delta}^E(\widehat{t}_j)\right\}^2.$$

The definitions of $\widehat{\delta}^O(t)$ (or $\widehat{\delta}^E(t)$), together with $w_i^O(t; K_\pm)$ (or $w_i^E(t; K_\pm)$) can be found in Eq. (3).

First, we give some lemmas that will be used in the proof of Theorem 1.

Recall that the two one-sided kernels are defined as $K_+(u) = K(u)\mathbb{I}(u \in [0,1])$ and $K_-(u) = K(-u)\mathbb{I}(u \in [-1, 0))$, where $K(u)$ is defined in Assumption S.2. Denote $\mu_r = \int_0^1 u^r K(u)\mathrm{d}u$, for $r = 0, 1, 2$. Let $s_{\pm r}(t)$ be either $s_r^O(t; K_\pm)$ or $s_r^E(t; K_\pm)$ (cf. Eq. (3)), and

$w_{\pm i}(t)$ be either $w_i^O(t;K_\pm)$ or $w_i^E(t;K_\pm)$. By the boundedness and Lipschitz-1 continuity of $K(\cdot)$, it is straightforward to have the following result.

Lemma S.1. *Under Assumption S.2, we have, for $t \in [h, 1-h]$,*

(i) $s_{\pm r}(t) = nh^{r+1}[\mu_{\pm r} + O\{(nh)^{-1}\}]$*, where* $\mu_{\pm r} = (\pm 1)^r \mu_r$*;*

(ii) $\sum_{i=1}^n w_{\pm i}(t) = n^2h^4[\mu_2\mu_0 - \mu_1^2 + O\{(nh)^{-1}\}]$.

Let

$$\widetilde{K}(u) = \frac{\mu_2 - \mu_1 u}{\mu_2\mu_0 - \mu_1^2} K(u), \text{ for } u \in [0,1],$$

and we introduce two one-sided kernels $\widetilde{K}_+(u) = \widetilde{K}(u)\mathbb{I}(u \in [0,1])$ and $\widetilde{K}_-(u) = \widetilde{K}(-u)\mathbb{I}(u \in [-1,0))$. For any sequence $\{g_1, \ldots, g_n\}$, define

$$\widetilde{g}_\pm(t) = (nh)^{-1}\sum_{i=1}^n \widetilde{K}_\pm\{(t_i - t)/h\}g_i \text{ and } \widetilde{g}_\pm^E(t) = (nh)^{-1}\sum_{i=1}^n \widetilde{K}_\pm\{(t_i^E - t)/h\}g_i.$$

Then, in cases when $nh \sim n^\eta$ for some $\eta > 0$, we have the following result.

Lemma S.2. *Under Assumptions S.1–S.2, we have*

$$\Pr\left\{\max_{t\in\mathcal{G}} |\widetilde{U}_+(t)| > C\left(\sqrt{\frac{\log n}{nh}}\right)\right\} = O(n^{1-\eta\theta/2}),$$

for some large $C > 0$.

Proof. Let $\mathcal{A} = \{\max_i |U_i| \le M_n\}$, where $M_n = n^{1/(\theta-\gamma)}$ for some small $\gamma > 0$. By Assumption S.1 and the Markov's inequality, we have

$$\Pr(\mathcal{A}) \ge 1 - n\Pr(|U_i|^\theta > M_n^\theta) \ge 1 - C_\varepsilon n^{1-\frac{\theta}{\theta-\gamma}}.$$

Conditional on the event $\mathcal{A}$, the Bernstein inequality yields

$$\begin{aligned}\Pr\left\{|\widetilde{U}_+(t)| > x/\sqrt{nh}\right\} \\ \leq 2\exp\left(-\frac{nhx^2}{2\sigma^2\sum_{i=1}^n \widetilde{K}_+^2\{(t_i-t)/h\} + 2/3M_n(nh)^{1/2}x}\right),\end{aligned} \tag{S.2}$$

for any $x > 0$. By Lemma S.1, $\sum_{i=1}^n \widetilde{K}_+^2\{(t_i-t)/h\} = O(nh)$ uniformly for all $t \in \mathcal{G}$. Taking $x = C(\log n)^{1/2}$ for some sufficiently large $C > 0$, (S.2) leads to

$$\begin{aligned}&\Pr\left(\max_{t\in\mathcal{G}} |\widetilde{U}_+(t)| > x/\sqrt{nh}\right) \\ &\leq \sum_{i=1}^n \Pr\left\{|\widetilde{U}_+(i/n)| > x/\sqrt{nh}\right\} \\ &\leq 2n\exp\left(-\frac{nh\log n}{C_1 nh + C_2 M_n (nh\log n)^{1/2}}\right) \\ &\leq 2n\exp(-C'\log n) = o(n^{1-\frac{\theta}{\theta-\gamma}}),\end{aligned}$$

where C_1, C_2 and C' are some constants and $0 < \gamma < \theta - 2\eta^{-1}$. □

Lemma S.3. *Under Assumption S.2, if $h \to 0$ and $nh \to \infty$, then we have*

$$\max_{t\in[h,1-h]} |\widetilde{U}_+(t) - \widetilde{U}_-(t)| \gtrsim \sqrt{\frac{\log n}{nh}}.$$

Proof. For a given $t \in [h, 1-h]$, let $\Delta(t) = \widetilde{U}_+(t) - \widetilde{U}_-(t)$, and G_n be the distribution of $\sqrt{nh}\Delta(t)/s$, where $s = \sigma\sqrt{2\int_0^1 \widetilde{K}^2(u)\mathrm{d}u}$. Then, it is easy to check that G_n converges to $N(0,1)$ in distribution. Let $M = \sup\{m : 2mh \leq 1-h\}$. Then, $M = O(h^{-1})$, and

$$\max_{t\in[h,1-h]} |\Delta(t)| \geq \max_{i=1,\ldots,M} |\Delta(2ih)|,$$

where the terms $\{\Delta(2ih)\}$ are independent. Let ξ_n be a sequence such that $\sqrt{nh}\xi_n \to \infty$.

Then,

$$\begin{aligned}
&\Pr\left(\max_{t\in[h,1-h]}|\Delta(t)|<\xi_n\right)\\
&\le \Pr\left(\max_{i=1,\ldots,M}|\Delta(2ih)|<\xi_n\right)\\
&=\prod_{i=1}^{M}\Pr\left(|\Delta(2ih)|<\xi_n\right)\\
&\le \left\{G_n\left(\sqrt{nh}\xi_n/s\right)\right\}^M.
\end{aligned}$$

By the Berry-Esseen bound that $\sup_x|G_n(x)-\Phi(x)|\le C(nh)^{-1/2}$ and by the fact that $\Phi(x)<1-x/(1+x^2)\phi(x)$, we have

$$\Pr\left(\max_{t\in[h,1-h]}|\Delta(t)|<\xi_n\right)\le\left\{\sup_x|G_n(x)-\Phi(x)|+\Phi\left(\sqrt{nh}\xi_n/s\right)\right\}^M\to 0,$$

where we can set $\xi_n=\sqrt{\log n/(\zeta nh)}s$ for some sufficiently large $\zeta>0$. □

Lemma S.4. *Under Assumptions S.1–S.6, there exists some $k_j\in\{1,\ldots,J^*\}$, for each $j=1,\ldots,J^*$, such that $\Pr(|\widehat{t}_j-t^*_{k_j}|\le d_{nj}/n)\to 1$, where d_{nj} is an integer satisfying $d_{nj}\sim\sqrt{nh\log n}/|\delta^*_{k_j}|\to\infty$.*

Proof. We prove the lemma only for $\widehat{t}_1$, since the result for other detected jumps could be proved similarly. To this end, we will show that

$$\Pr(|\widehat{t}_1-t^*_j|>d_{nj}/n,\ \forall j=1,\ldots,J^*)=o(1/J^*).$$

By Lemma S.1 and Assumption S.5, it is not hard to show that,

$$\widehat{f}_\pm(t)=f(t)+\delta^*_j(nh)^{-1}\sum\widetilde{K}_\pm\{(t_i-t)/h\}\left\{\mathbb{I}\big(t_i\ge t^*_j\big)-\mathbb{I}\big(t\ge t^*_j\big)\right\}[1+O\{(nh)^{-1}\}]+O(h^2),$$

where t_j^* is the change-point located in $[t, t+h]$ or $[t-h, t]$ (if either interval contains a jump point). Noting that

$$\Pr(|\widehat{t}_1 - t_j^*| > h,\ \forall j) \leq \Pr\left\{ \sup_{t:|t-t_j^*|>h,\ \forall j} |\widehat{\delta}^O(t)| > |\widehat{\delta}^O(t_j^*)|,\ \forall j \right\},$$

and by using arguments similar to those in the proof of Lemma S.2, we conclude that $\Pr(|\widehat{t}_1 - t_j^*| > h,\ \forall j) = o(1/J^*)$. Then, it suffices to show that $J^* \Pr(t_j^* + d_{nj}/n < \widehat{t}_1 \leq t_j^* + h,\ \forall j) \to 0$, since the other half follows similarly. Note that

$$\widehat{d}^f(t_j^* + d_{nj}/n) - \widehat{d}^f(t_j^*) = -\delta_j^*(nh)^{-1} \sum_{-d_{nj} \leq i < 0} \widetilde{K}_-\{i/(nh)\}[1 + O\{(nh)^{-1}\}] + O(h^2).$$

Again using arguments similar to those in the proof of Lemma S.2 and by the assumption that $K_+(0) > 0$, the conclusion follows. □

Proof of Theorem 1

Let $f_i^E := f(t_i^E)$ for $i = 1, \ldots, n$. First, assume that $1 \leq J < J^*$. In such cases, we have

$$C(J) - C(J^*) = -\sum_{j=J+1}^{J^*} \left\{ \widehat{d}_E^{f^E}(\widehat{t}_j) - \widehat{d}^f(\widehat{t}_j) + \widehat{d}_E^V(\widehat{t}_j) - \widehat{d}^U(\widehat{t}_j) \right\}^2 + \sum_{j=J+1}^{J^*} \left\{ \widehat{d}_E^{f^E}(\widehat{t}_j) + \widehat{d}_E^V(\widehat{t}_j) \right\}^2.$$

Noticing that $\max_{j=1,\ldots,J^*} |\delta_j^*| = O(1)$ and by the boundedness of K and f_{C}', it can be verified that $\widehat{d}_E^{f^E}(\widehat{t}_j) - \widehat{d}^f(\widehat{t}_j) \sim n^{-1}$, uniformly for all $j = J+1, \ldots, J^*$. By a similar proof to that of Lemma S.4, under the assumption that $nh^5/\log n \to 0$, we have

$$\begin{aligned} |\widehat{d}^f(\widehat{t}_j) - \delta_{k_j}^*| &= \left| \delta_{k_j}^*(nh)^{-1} \sum_{i:0 \leq |t_i - \widehat{t}_j| \leq |t_j^* - \widehat{t}_j|} \widetilde{K}_-\{(t_i - \widehat{t}_j)/h\}[1 + O\{(nh)^{-1}\}] + O(h^2) \right| \\ &= O\left(\sqrt{\frac{\log n}{nh}} \right), \end{aligned}$$

which holds with a probability approaching one. Then, by Lemma S.2, we have

$$\Pr\{C(J) - C(J^*) > 0\} = \Pr\left\{ (J^* - J) O_p\left(\frac{\log n}{nh} \right) + \sum_{j=J+1}^{J^*} \left(\delta_{k_j}^* \right)^2 > 0 \right\} \to 1.$$

Hence, $\Pr(\widehat{J} \geq J^*) \to 1$.

Now, in the case when $J^* < J \leq \bar{J}$, we have

$$C(J) - C(J^*) = \sum_{j=J^*+1}^{J} \left\{ \widehat{d}_E^{f^E}(\widehat{t}_j) - \widehat{d}^f(\widehat{t}_j) + \widehat{d}_E^V(\widehat{t}_j) - \widehat{d}^U(\widehat{t}_j) \right\}^2 - \sum_{j=J^*+1}^{J} \left\{ \widehat{d}_E^{f^E}(\widehat{t}_j) + \widehat{d}_E^V(\widehat{t}_j) \right\}^2.$$

Let $\mathcal{G}_J = \mathcal{G} \setminus \cup_{k=1}^{J-1} [\widehat{t}_k - h, \widehat{t}_k + h]$. Then, by Lemma S.4, $\mathcal{G}_J$ contains no jump points in probability approaching one. Once again, $\widehat{d}_E^{f^E}(\widehat{t}_j) - \widehat{d}^f(\widehat{t}_j) \sim n^{-1}$, uniformly for all $j > J^*$, and similar to the proof of Lemma S.4, we have $\widehat{d}^f(\widehat{t}_j) = O(h^2)$, for $j > J^*$. On one hand, by the independence between $\mathcal{Z}^O$ and $\mathcal{Z}^E$, we have $\widehat{d}_E^V(\widehat{t}_j) = O_p\{(nh)^{-1/2}\}$. On the other hand, by Lemma S.3, we have $\widehat{d}^U(\widehat{t}_j) \gtrsim \sqrt{\log n/(nh)}$. It then follows that $\Pr\{C(J) - C(J^*) > 0\} \to 1$, which completes the proof. $\square$

B.2 Proof of Theorem 2

Consider the following thresholding rule

$$L = \inf \left\{ s > 0 : \frac{1 + \#\{k : W_k \leq -s\}}{\#\{k : W_k \geq s\} \vee 1} \leq \alpha \right\}.$$

Lemma S.5. *For any $\alpha \in (0,1)$, we have*

$$\mathrm{FDR}(L) \equiv \mathbb{E}\left[\frac{\#\{k : W_k \geq L, \widehat{t}_k^O \in \mathcal{I}_0\}}{\#\{k : W_k \geq L\} \vee 1} \right] \leq \min_{\epsilon \geq 0} \left\{ \alpha(1 + 5\epsilon) + \Pr\left(\max_{\widehat{t}_k^O \in \mathcal{I}_0} \Delta_k > \epsilon \right) \right\},$$

where $\Delta_k = |\Pr(W_k > 0 \mid |W_k|, W_{k-1}, W_{k+1}) - 1/2|$, and $W_0 = W_{\bar{J}+1} = 0$.

Proof. Lemma S.5 can be proved similarly to Theorem 2 in Barber et al. (2019) which shows that the Model-X knockoff selection procedure has a FDR proportional to the errors in estimating the distribution of each feature conditional on the remaining features. Let $j \in \mathcal{I}_0$ denote $\widehat{t}_j^O \in \mathcal{I}_0$. Then, for given $\epsilon > 0$ and $t > 0$, define

$$R_\epsilon(t) = \frac{\sum_{j \in \mathcal{I}_0} \mathbb{I}(W_j \geq t, \Delta_j \leq \epsilon)}{1 + \sum_{j \in \mathcal{I}_0} \mathbb{I}(W_j \leq -t)}.$$

Consider the event that $\mathcal{A} = \{\Delta \equiv \max_{j\in\mathcal{I}_0} \Delta_j \le \epsilon\}$. For a given threshold rule $L = T(\mathbf{W})$ that maps statistics $\mathbf{W} = (W_1, \ldots, W_{\bar{J}})'$ to a threshold value $L \ge 0$, define

$$L_j = T\left(W_1, \ldots, W_{j-1}, |W_j|, W_{j+1}, \ldots, W_{\bar{J}}\right) \ge 0,$$

for each $j = 1, \ldots, \bar{J}$. Then, for the SOPS method with the threshold L, we can write

$$\begin{aligned}
\frac{\sum_{j\in\mathcal{I}_0} \mathbb{I}\left(W_j \ge L, \Delta_j \le \epsilon\right)}{1 \vee \sum_j \mathbb{I}(W_j \ge L)} &= \frac{1 + \sum_j \mathbb{I}\left(W_j \le -L\right)}{1 \vee \sum_j \mathbb{I}(W_j \ge L)} \times \frac{\sum_{j\in\mathcal{I}_0} \mathbb{I}\left(W_j \ge L, \Delta_j \le \epsilon\right)}{1 + \sum_j \mathbb{I}\left(W_j \le -L\right)} \\
&\le \alpha \times R_\epsilon(L).
\end{aligned}$$

It is crucial to obtain an upper bound for $\mathbb{E}\{R_\epsilon(L)\}$. To this end, we have

$$\begin{aligned}
\mathbb{E}\{R_\epsilon(L)\} &= \mathbb{E}\left\{\frac{\sum_{j\in\mathcal{I}_0} \mathbb{I}\left(W_j \ge L, \Delta_j \le \epsilon\right)}{1 + \sum_{j\in\mathcal{I}_0} \mathbb{I}\left(W_j \le -L\right)}\right\} &\text{(S.3)} \\
&= \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{\frac{\mathbb{I}\left(W_j \ge L_j, \Delta_j \le \epsilon\right)}{1 + \sum_{k\in\mathcal{I}_0, k\ne j} \mathbb{I}\left(W_k \le -L_j\right)}\right\} \\
&= \sum_{j\in\mathcal{I}_0} \mathbb{E}\left[\mathbb{E}\left\{\frac{\mathbb{I}\left(W_j \ge 0\right)\mathbb{I}\left(|W_j| \ge L_j, \Delta_j \le \epsilon\right)}{1 + \sum_{k\in\mathcal{I}_0, k\ne j} \mathbb{I}\left(W_k \le -L_j\right)} \mid |W_j|, \mathbf{W}_{-j}\right\}\right] \\
&= \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{\frac{\Pr\left(W_j > 0 \mid |W_j|, W_{j-1}, W_{j+1}\right)\mathbb{I}\left(|W_j| \ge L_j, \Delta_j \le \epsilon\right)}{1 + \sum_{k\in\mathcal{I}_0, k\ne j} \mathbb{I}\left(W_k \le -L_j\right)}\right\}, &\text{(S.4)}
\end{aligned}$$

where $\mathbf{W}_{-j} = \mathbf{W} \backslash W_j$.

By the definition of Δ_j, we have $\Pr\left(W_j > 0 \mid |W_j|, W_{j-1}, W_{j+1}\right) \le 1/2 + \Delta_j$. Hence,

$$\begin{aligned}
&\mathbb{E}\{R_\epsilon(L)\} \\
&\le \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{\frac{(\frac{1}{2} + \Delta_j)\mathbb{I}\left(|W_j| \ge L_j, \Delta_j \le \epsilon\right)}{1 + \sum_{k\in\mathcal{I}_0, k\ne j} \mathbb{I}\left(W_k \le -L_j\right)}\right\} \\
&\le (\frac{1}{2} + \epsilon)\left[\sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{\frac{\mathbb{I}\left(W_j \ge L_j, \Delta_j \le \epsilon\right)}{1 + \sum_{k\in\mathcal{I}_0, k\ne j} \mathbb{I}\left(W_k \le -L_j\right)}\right\} + \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{\frac{\mathbb{I}\left(W_j \le -L_j\right)}{1 + \sum_{k\in\mathcal{I}_0, k\ne j} \mathbb{I}\left(W_k \le -L_j\right)}\right\}\right] \\
&= (\frac{1}{2} + \epsilon)\left[\mathbb{E}\{R_\epsilon(L)\} + \sum_{j\in\mathcal{I}_0} \mathbb{E}\left\{\frac{\mathbb{I}\left(W_j \le -L_j\right)}{1 + \sum_{k\in\mathcal{I}_0, k\ne j} \mathbb{I}\left(W_k \le -L_j\right)}\right\}\right].
\end{aligned}$$

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

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

where the first equation comes from the fact that for any j,k, if $W_j\le -\min(L_j,L_k)$ and $W_k\le -\min(L_j,L_k)$, then $L_j=L_k$; see Barber et al. (2019). So, we have

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

Therefore, the result in Lemma S.5 is valid. □

For $j=1,\ldots,\bar{J}$, we write $\widehat{\delta}_{1j}=\widehat{\delta}^O(\widehat{t}_j^O)$ and $\widehat{\delta}_{2j}=\widehat{\delta}^E(\widehat{t}_j^O)$ for simplicity, and denote $T_{1j}=\sqrt{nh}\widehat{\delta}_{1j}$ and $T_{2j}=\sqrt{nh}\widehat{\delta}_{2j}$. Let $a_n=C\sqrt{\log n}$, where $C>0$ is specified in Lemma S.2. Let $\mathcal{C}=\{\cap_{j\in\mathcal{I}_0}|W_j|\le\lambda_n\}$, where $\lambda_n=\inf\{z:\Pr(\max_{j\in\mathcal{I}_0}|W_j|>z)\le b_n\}$, and b_n be a sequence satisfies the conditions that $b_n\to 0$ and $\bar{J}a_n^3/(n^{\eta/2}b_n)\to 0$.

Lemma S.6. *Under the conditions in Theorem 2, we have*

$$\sup_{0\le t\le\lambda_n}\left|\frac{\Pr(W_j\ge t)}{\Pr(W_j\le -t)}-1\right|=O(c_n),$$

uniformly for $j\in\mathcal{I}_0$, where $c_n=(a_n^3/\sqrt{nh})/b_n$.

Proof. Let $\mathcal{A}_{nt}=\{v\in\mathbb{R}:|v|\ge t/a_n\}$. Note that

$$\begin{aligned}
&\frac{\Pr(T_{1j}T_{2j}>t)}{\Pr(T_{1j}T_{2j}<-t)}-1\\
=&\frac{\Pr(T_{1j}T_{2j}>t,T_{1j}\in\mathcal{A}_{nt})-\Pr(T_{1j}T_{2j}<-t,T_{1j}\in\mathcal{A}_{nt})}{\Pr(T_{1j}T_{2j}<-t)}\\
&+\frac{\Pr(T_{1j}T_{2j}>t,T_{1j}\in\mathcal{A}_{nt}^c)-\Pr(T_{1j}T_{2j}<-t,T_{1j}\in\mathcal{A}_{nt}^c)}{\Pr(T_{1j}T_{2j}<-t)}\\
:=&A_{11}+A_{12}.
\end{aligned}$$

By Lemma S.2, we have

$$A_{12} \lesssim \frac{\Pr\left(|T_{2j}| \geq a_n\right)}{b_n} = o(c_n), \tag{S.5}$$

provided that $\eta > 2/(\theta - 1)$. Furthermore, we have

$$\begin{aligned}
\Pr(T_{1j}T_{2j} \geq t, T_{1j} \in \mathcal{A}_{nt}) &= \int_{v \in \mathcal{A}_{nt}} \Pr(T_{1j}T_{2j} > t \mid T_{1j} = v) f(v) dv \\
&= \int_{v \in \mathcal{A}_{nt}} \tilde{\Phi}(t/(vs)) f(v) dv + O(a_n^3/\sqrt{nh}) \\
&= \int_{v \in \mathcal{A}_{nt}} \Phi(-t/(vs)) f(v) dv + O(a_n^3/\sqrt{nh}) \\
&= \Pr(T_{1j}T_{2j} < -t, T_{1j} \in \mathcal{A}_{nt}) + O(a_n^3/\sqrt{nh}),
\end{aligned}$$

where the second equality holds due to the Berry-Esseen theorem. By the definition of $\mathcal{C}$, we have $\Pr(W_j \geq t) \geq b_n$, for $t \leq \lambda_n$. By combining this result with (S.5), we have $\Pr(W_j \geq t, T_{1j} \in \mathcal{A}_{nt}) \gtrsim b_n$. So,

$$\Pr(W_j \geq t, T_{1j} \in \mathcal{A}_{nt}) = \Pr(W_j \leq -t, T_{1j} \in \mathcal{A}_{nt})\{1 + o(1)\},$$

which is true uniformly in j and $0 \leq t \leq \lambda_n$, as long as $c_n \to 0$. Consequently, Lemma S.6 follows. $\square$

Proof of Theorem 2–(i)

We first show that the result is valid under the condition that W_i's are independent. According to Lemma S.5, we have

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

By definition, $A_2 = o(1)$. It remains to handle A_1.

Notice that

$$A_1 \le \sum_{j \in \mathcal{I}_0} \sup_{0 \le t \le \lambda_n} |f_j(-t)/f_j(t) - 1|, \tag{S.6}$$

where $f_j(\cdot)$ is the density of W_j. It remains to prove that the right-hand side of (S.6) goes to zero as $n \to \infty$. To this end, denote $H_j(t) = \Pr(W_j > t)$ and $H_{j-}(t) = \Pr(W_j < -t)$. Then,

$$\begin{aligned}
\sup_{0 \le t \le \lambda_n} |f_j(-t)/f_j(t) - 1| &\le \sup_{0 \le t \le \lambda_n} \{f_j(t)\}^{-1} \left(|H_j(t) - H_{j-}(t)| + |H_j(t-) - H_{j-}(t-)|\right) \\
&\lesssim c_n \sup_{0 \le t \le \lambda_n} \{f_j(t)\}^{-1} H_j(t),
\end{aligned}$$

where Lemma S.6 has been used. So, we need to study $\inf_{0 \le t \le \lambda_n}\{f_j(t)\}$. Note that

$$\begin{aligned}
f_j(t) &= \int_{v \in \mathcal{A}_{nt}} \tilde{\Phi}(t/(vs)) f(v) dv - \int_{v \in \mathcal{A}_{nt}} \tilde{\Phi}(t/(vs)-) f(v) dv + o(b_n) \\
&= \int_{v \in \mathcal{A}_{nt}} \frac{1}{vs} \phi(t/(vs)) f(v) dv + o(b_n) \\
&\ge \int_{v \in \mathcal{A}_{nt}} \frac{t}{(vs)^2} \tilde{\Phi}(t/(vs)) f(v) dv + o(b_n) \\
&\ge \frac{a_n^2}{ts^2} \int_{v \in \mathcal{A}_{nt}} \tilde{\Phi}(t/(vs)) f(v) dv + o(b_n) \\
&= \frac{a_n^2}{ts^2} H_j(t)\{1 + o(b_n)\} + o(b_n).
\end{aligned}$$

Because $\lambda_n \lesssim a_n^2$, $f_j(t) \gtrsim H_j(t)$. Hence, (S.6) holds if $\bar{J} c_n \to 0$, from which the result (i) in the theorem holds. □

To prove the result (ii) in the theorem, we need to show that

$$\left| \frac{J_0^{-1} \sum_{j \in \mathcal{I}_0} \mathbb{I}(W_j \ge t)}{J_0^{-1} \sum_{j \in \mathcal{I}_0} \mathbb{I}(W_j \le -t)} - 1 \right| = o_p(1),$$

which is true uniformly in $t \in (0, \infty)$. To this end, we first build Lemmas S.7 and S.8 below, in which $G(t) = J_0^{-1} \sum_{j \in \mathcal{I}_0} \Pr(W_j \geq t)$, $G_-(t) = J_0^{-1} \sum_{j \in \mathcal{I}_0} \Pr(W_j \leq -t)$ and $J_0 = |\mathcal{I}_0|$.

Lemma S.7. *Under the conditions in Theorem 2, for any* $0 \leq t \leq G_-^{-1}(1/\bar{J})$, *we have*

$$\frac{\sum_{j \in \mathcal{I}_0} \Pr(W_j \geq t)}{\sum_{j \in \mathcal{I}_0} \Pr(W_j \leq -t)} - 1 \xrightarrow{\mathcal{P}} 0.$$

The proof of this Lemma is similar to that of Lemma S.6, and thus omitted here.

Lemma S.8. *Under the conditions in Theorem 2, for any* $a_n \to \infty$, *we have*

$$\sup_{0 \leq t \leq G^{-1}(a_n/\bar{J})} \left| \{J_0 G(t)\}^{-1} \sum_{j \in \mathcal{I}_0} \mathbb{I}\left(W_j \geq t\right) - 1 \right| = o_p(1),$$
$$\sup_{0 \leq t \leq G_-^{-1}(a_n/\bar{J})} \left| \{J_0 G_-(t)\}^{-1} \sum_{j \in \mathcal{I}_0} \mathbb{I}\left(W_j \leq -t\right) - 1 \right| = o_p(1).$$

Proof. We only prove the first equation here, and the second one can be proved similarly. Note that $G(t)$ is a deceasing continuous function. Let $z_0 < z_1 < \cdots < z_{d_n} \leq 1$, $t_i = G^{-1}(z_i)$, where $z_0 = a_n/\bar{J}$ and $z_i = a_n/\bar{J} + a_n i^{\delta}/\bar{J}$, and $d_n = [\{(\bar{J} - a_n)/a_n\}^{1/\delta}]$, where $\delta > 1$. Then, $G(t_i)/G(t_{i+1}) = 1 + o(1)$ uniformly in i. It is therefore enough to obtain the convergence rate of

$$D_n = \sup_{0 \leq i \leq d_n} \left| \frac{\sum_{j \in \mathcal{I}_0} \{\mathbb{I}(W_j > t_i) - \Pr(W_j > t_i)\}}{J_0 G(t_i)} \right|.$$

To this end, define $\mathcal{S}_j = \{k \in \mathcal{I}_0 : \ W_k \text{ is dependent with } W_j\}$ and

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

Then,

$$D(t) = \sum_{j \in \mathcal{I}_0} \sum_{k \in \mathcal{S}_j} \mathbb{E}\left[\{\mathbb{I}(W_j > t) - \Pr(W_j > t)\}\{\mathbb{I}(W_k > t) - \Pr(W_k > t)\}\right] \leq 2 J_0 G(t).$$

Since $W_1, \ldots, W_{\bar{J}}$ is a 1-dependent sequence, so is $\mathbb{I}(W_j > t_i)$. Then, we have

$$
\begin{aligned}
\Pr(D_n \geq \epsilon) &\leq \sum_{i=0}^{d_n} \Pr\left(\left| \frac{\sum_{j \in \mathcal{I}_0} [\mathbb{I}(W_j > t_i) - \Pr(W_j > t_i)]}{J_0 G(t_i)} \right| \geq \epsilon \right) \\
&\leq \frac{1}{\epsilon^2} \sum_{i=0}^{d_n} \frac{1}{p_0^2 G^2(t_i)} D(t_i) \leq \frac{2}{\epsilon^2} \sum_{i=0}^{d_n} \frac{1}{J_0 G(t_i)}.
\end{aligned}
$$

Moreover, we have

$$
\begin{aligned}
\sum_{i=0}^{d_n} \frac{1}{J_0 G(t_i)} &= \frac{\bar{J}}{J_0} \left(\frac{1}{a_n} + \sum_{i=1}^{d_n} \frac{1}{a_n + a_n i^\delta} \right) \\
\leq \quad c \left(\frac{1}{a_n} + a_n^{-1} \sum_{i=1}^{d_n} \frac{1}{1 + i^\delta} \right) &\leq c a_n^{-1} \{1 + o(1)\}.
\end{aligned}
$$

Therefore, we can have the result that $\Pr(D_n \geq \epsilon) \to 0$ when $a_n \to \infty$. □

Proof of Theorem 2–(ii)

Next, we show that the result (ii) of Theorem 2 is valid. By definition, our thresholding rule is equivalent to select j if $W_j > L$, where

$$
L = \inf \left\{ t \geq 0 : 1 + \sum_j \mathbb{I}(W_j < -t) \leq \alpha \max \left(\sum_j \mathbb{I}(W_j > t), 1 \right) \right\}.
$$

In order to use the results of Lemmas S.7-S.8, we need to establish an asymptotic bound for L. Let $t^* = G_-^{-1}(\alpha \beta_n / \bar{J})$. By Assumption 1, we have

$$
\Pr\left(\sum_j \mathbb{I}(W_j > C a_n^2) \geq \beta_n \right) \to 1. \tag{S.7}
$$

Thus, we have $\Pr(\sum_j \mathbb{I}(W_j > t^*) \geq \beta_n) \to 1$, which implies that

$$
1 + \sum_j \mathbb{I}(W_j < -t^*) \lesssim \alpha \beta_n \leq \alpha \sum_j \mathbb{I}(W_j > t^*).
$$

Therefore, it is true that $G_-(L) \gtrsim G_-(t^*)$. By Lemmas S.7-S.8,

$$
\frac{\sum_{j \in \mathcal{I}_0} \mathbb{I}(W_j > L)}{\sum_{j \in \mathcal{I}_0} \mathbb{I}(W_j < -L)} = 1 + o(1). \tag{S.8}
$$

Then, we have

$$\begin{aligned}\mathrm{FDP} &= \frac{\sum_{j\in\mathcal{I}_0}\mathbb{I}\left(W_j \geq L\right)}{1\vee\sum_j \mathbb{I}(W_j\geq L)} = \frac{1+\sum_j \mathbb{I}\left(W_j\leq -L\right)}{1\vee\sum_j\mathbb{I}(W_j\geq L)}\times\frac{\sum_{j\in\mathcal{I}_0}\mathbb{I}\left(W_j\geq L\right)}{1+\sum_j\mathbb{I}\left(W_j\leq -L\right)}\\ &\leq \alpha\times R(L).\end{aligned}$$

Note that $R(L) \leq \sum_{j\in\mathcal{I}_0}\mathbb{I}\left(W_j\geq L\right)/\sum_{j\in\mathcal{I}_0}\mathbb{I}\left(W_j\leq -L\right)$. Thus, by (S.8), $\limsup_{n\to\infty}\mathrm{FDP}\leq \alpha$ in probability. Then, for any $\epsilon>0$,

$$\mathrm{FDR}\leq(1+\epsilon)\alpha R(L)+\Pr\left(\mathrm{FDP}\geq(1+\epsilon)\alpha R(L)\right),$$

from which the result (ii) of the theorem is proved. □

Under all the conditions in Theorem 2 and Assumption S.6, we have $\lim_{n\to\infty}\Pr\left\{\mathcal{S}(L)\supseteq\mathcal{I}_1\right\}=1$ due to the fact that $L\lesssim a_n^2$.

C Additional numerical results

Figures S1–S2 depict the estimation precision of the detected jumps by different methods in Example I when the bandwidth h changes. Suppose the set of true jumps is $\{t_1^*,\dots,t_{J^*}^*\}$ and the set of estimated jumps is $\{\widehat{t}_1,\dots,\widehat{t}_{\widehat{J}^*}\}$ returned by any of the considered methods. To measure the estimation precision, we define two indices

$$d_1=\sup_{k=1,\dots,\widehat{J}^*}\inf_{j=1,\dots,J^*}|\widehat{t}_k-t_j^*| \text{ and } d_2=\sup_{j=1,\dots,J^*}\inf_{k=1,\dots,\widehat{J}^*}|\widehat{t}_k-t_j^*|,$$

and the Hausdorff distance $d_H=\max\{d_1,d_2\}$. Intuitively, d_1 is small if each detected jump is close to some true jump, and d_2 is small if each true jump is close to some estimated jump. An estimate with well detection precision should have both small values of d_1 and

d_2. From Figures S1–S2, we observe that the ability of correctly identifying the number of jumps by these methods seems to be revived for some large h values (e.g., near 0.16 for the JIC method when $\sigma = 0.4$). In fact, this is not the whole story if we also check the estimation precision of the detected jumps. Figures S3–S5 present the estimation precision of the detected jumps by different methods in Example II when the bandwidth h changes.

Figure S6 shows the probabilities of correct, under- and over-estimation of the number of true jumps against the sample sizes under relatively large and small noise levels for three different COPS*-type procedures in Example I, respectively. The three procedures differ in the selection of the bandwidth. Recall that $\hat{h}_u = \arg\min_{h\in\mathcal{H}} \mathrm{RSS}^u(h)$ for $u = 1, \ldots, U$. The proposed COPS* procedure uses $\widehat{h} = \max_{u=1,\ldots,U} \hat{h}_u$ as the bandwidth, which is termed as "COPS*-min-max". A new procedure "COPS*-min-med" selects the bandwidth as the the median of all $\hat{h}_u$'s. And another procedure "COPS*-ave-min" is based on the bandwidth given by $\arg\min_{h\in\mathcal{H}} U^{-1}\sum_{u=1}^{U} \mathrm{RSS}^u(h)$. From the plots we can see that, either "COPS*-min-med" or 'COPS*-ave-min" would result in underestimation under low signal-to-noise ratio (SNR) scenarios such as those with small sample sizes or large noise levels. This may be due to the inaccurate estimation of some (here 1) jump magnitudes, and consequently ignoring unidentified true jumps could even give a better curve fitting (or a smaller value of RSS). A slightly larger value $\widehat{h}$ in "COPS*-min-max" helps to mitigate that effect.

Figure S7 depicts the probabilities of correct, under- and over-estimation of the number of true jumps against the number of random sample-splittings in the COPS* method in Example I, which suggests that the proposed approaches perform quite robustly to the choice of the number of random sample-splitting U if it is not selected too small. For practical implementation, we recommend using $U = 20$.

Figure S8 shows the FDR and TPR values of different methods when the bandwidth h changes, under different combinations of the sample size and the noise distribution considered in Example III when the noise level $\sigma = 0.2$.

Figure S9 shows that the estimated number of jumps against the value of the bandwidth for three versions of the JIC, COPS and SOPS methods, for the real data example. For the SOPS procedure, we use $\alpha = 20\%$. It can be seen that the detected numbers of jumps are close to each other among all methods for each bandwidth, while they differ significantly as the bandwidth varies for each method. Table S1 lists the specific dates of the detected jumps obtained by the COPS* and SOPS* methods, with the corresponding jump magnitudes.

Table S1: Dates of the detected jumps obtained by the COPS* and SOPS* methods, with the corresponding jump magnitudes.

Date	COPS*	SOPS*
1986-07-14	-247.6670	NA
1986-07-21	NA	-212.01796
1987-10-12	-741.2939	-681.73730
1988-08-08	NA	-97.38503
1989-05-08	176.8476	NA
1989-10-09	NA	-159.11230
1990-08-13	-435.3975	-457.32177
1991-12-30	256.1716	266.18602
1993-02-08	NA	120.12357

1994-03-21	NA	-235.39690
1994-03-28	-240.8167	NA
1995-04-03	221.6169	227.58850
1996-04-29	-345.0389	NA
1996-06-03	NA	-430.46053
1997-05-05	428.9300	NA
1997-10-06	NA	-874.96441
1998-07-27	NA	-1464.81459
1998-08-03	-1497.7883	NA
2000-01-24	-752.1800	NA
2000-02-14	NA	-773.51320
2001-07-09	-699.4715	NA
2001-08-20	NA	-1079.50397
2002-06-17	NA	-1404.62870
2002-07-08	-1576.2382	NA
2003-06-02	NA	644.59422
2004-04-26	-724.6147	NA
2004-05-03	NA	-660.29931
2005-03-21	NA	-484.11798
2005-11-07	369.2984	NA
2006-05-08	NA	-647.17828
2007-04-23	821.0891	696.33857
2008-10-06	-2780.6403	-2184.25277
2009-07-27	NA	959.54301
2010-05-03	-1258.6945	NA

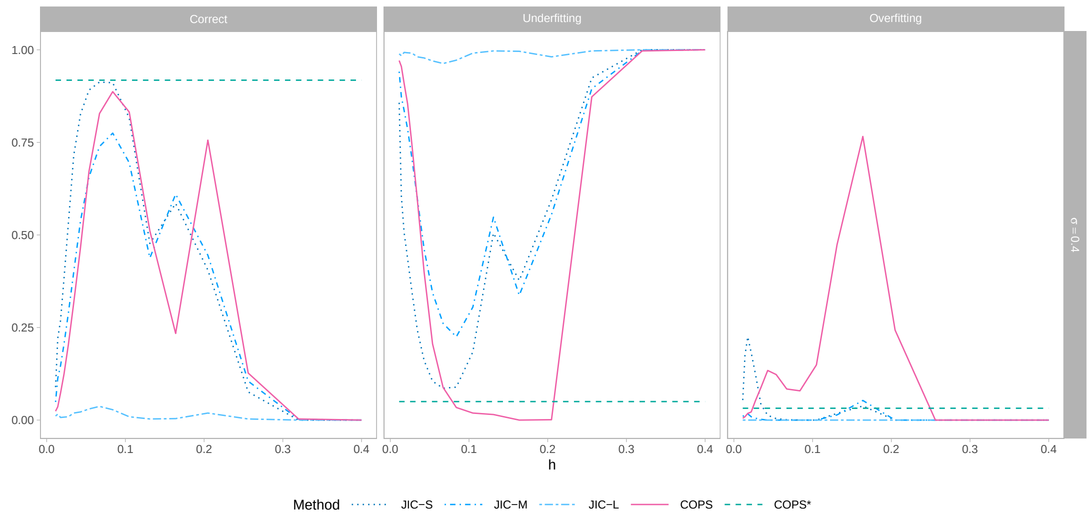
2010-05-17	NA	-1077.73956
2011-07-25	-1633.2045	-1733.46844
2012-05-21	NA	-818.14233
2013-01-21	835.6474	NA
2013-03-11	NA	799.57119
2014-01-06	NA	-159.22462
2014-01-20	-272.5445	NA
2014-10-27	NA	630.30186
2015-08-03	-1320.4034	NA
2015-08-17	NA	-965.25943
2016-09-05	-333.8651	NA
2016-12-12	NA	1087.59420
2017-10-16	1541.0521	NA
2018-03-19	NA	-2128.76793
2018-11-19	-1587.4973	NA
2019-02-18	NA	1268.42514
2020-02-24	-6397.8115	-7227.39968

References

Barber, R. F., Candès, E. J., and Samworth, R. J. (2019), "Robust inference with knockoffs," *Ann. Statist.*, to appear.

Niu, Y. S., and Zhang, H. (2012), "The screening and ranking algorithm to detect DNA copy number variations," *The Annals of Applied Statistics*, 6, 1306–1326.

Xia, Z., and Qiu, P. (2015), "Jump information criterion for statistical inference in estimating discontinuous curves," *Biometrika*, 102, 397–408.



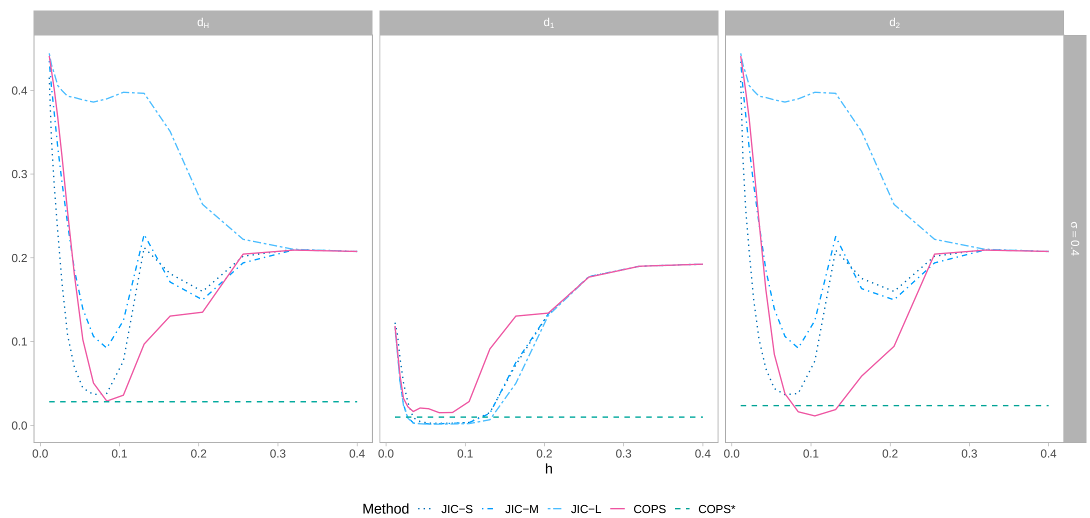


Figure S1: Estimation precision of the detected jumps by different methods in Example I when the bandwidth h changes, under a relatively low SNR with $\sigma = 0.4$.

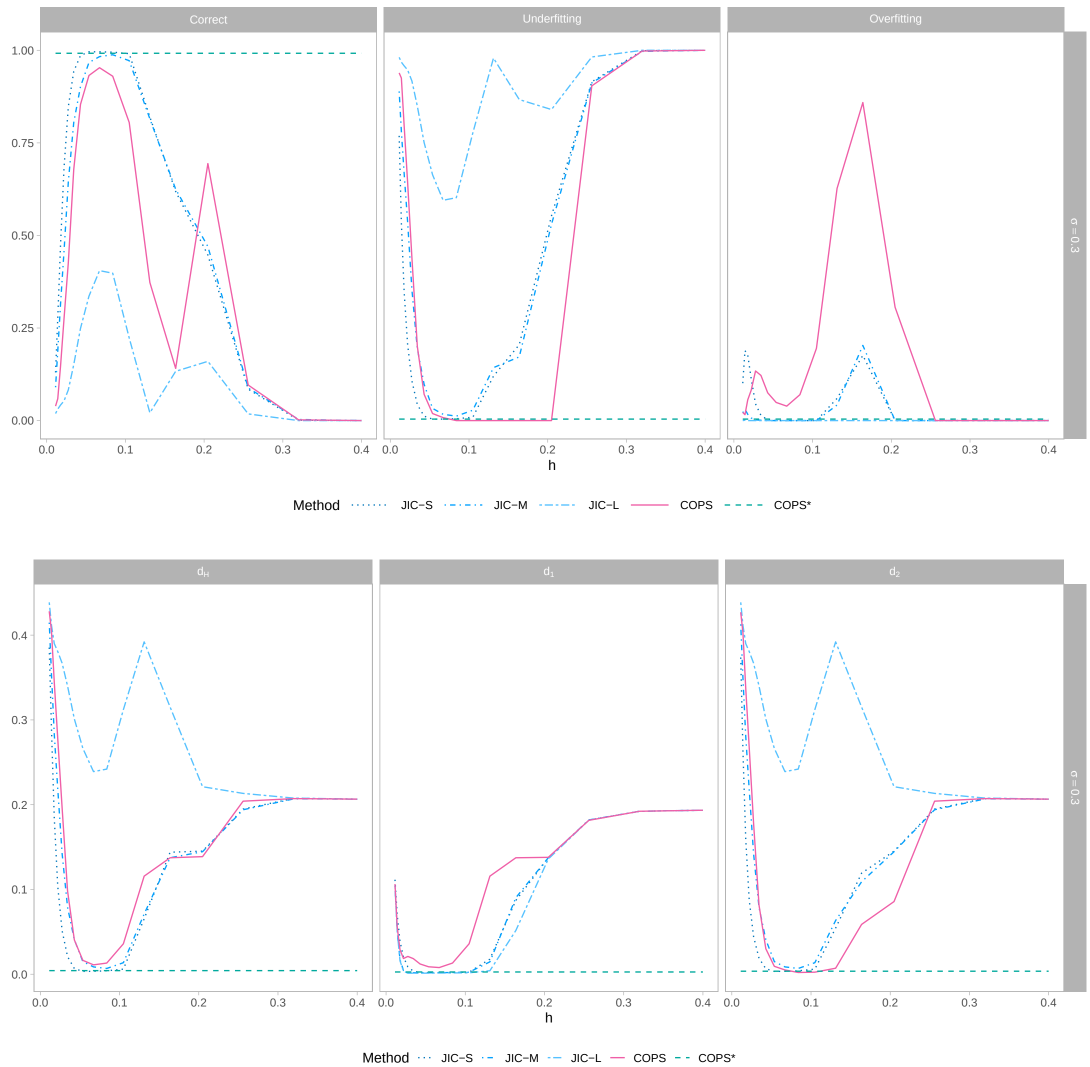


Figure S2: Estimation precision of the detected jumps by different methods in Example I when the bandwidth h changes, under a relatively high SNR with $\sigma = 0.3$.

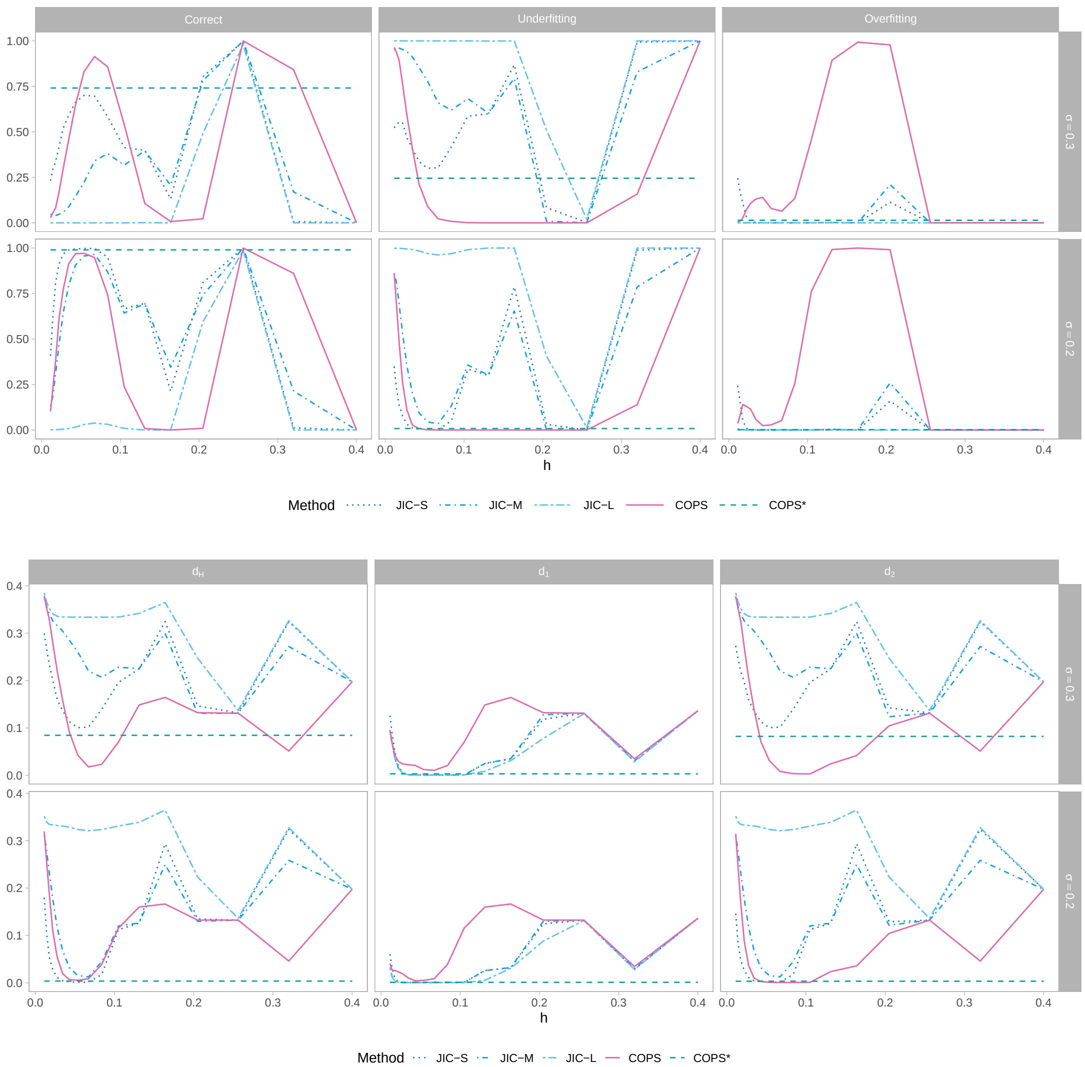
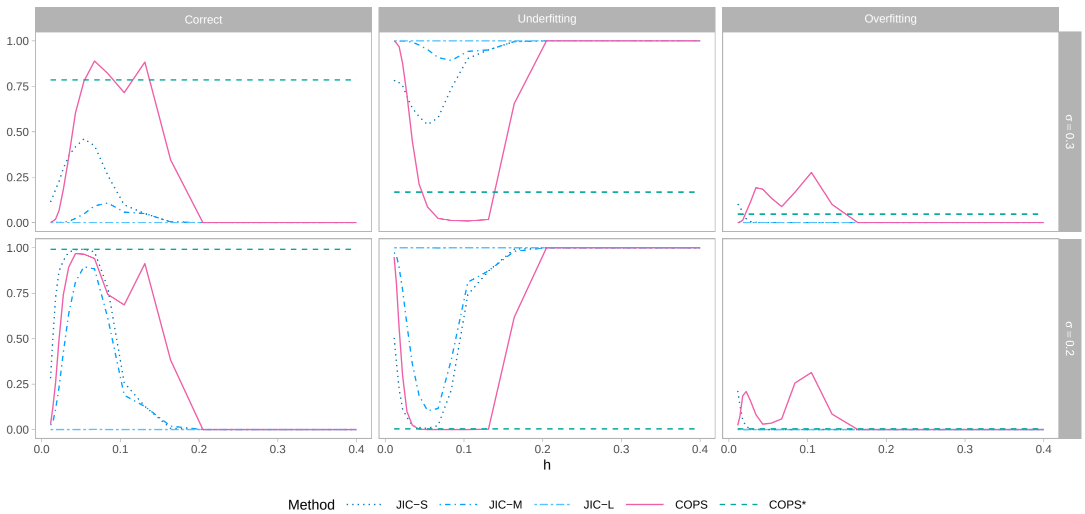


Figure S3: Estimation precision of the detected jumps by different methods in Example II with $\zeta = 0.5$ when the bandwidth h changes.



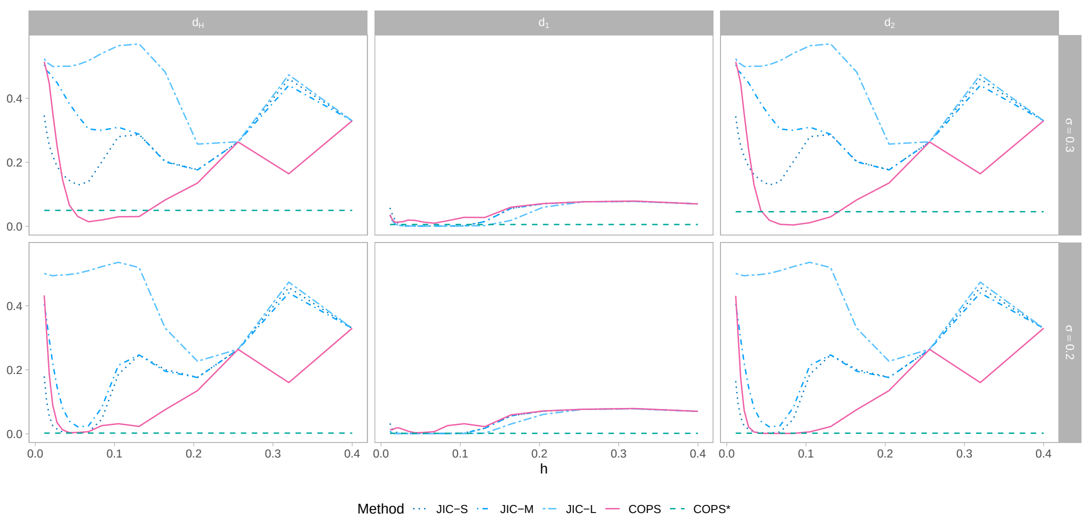


Figure S4: Estimation precision of the detected jumps by different methods in Example II with $\zeta = 0.75$ when the bandwidth h changes.

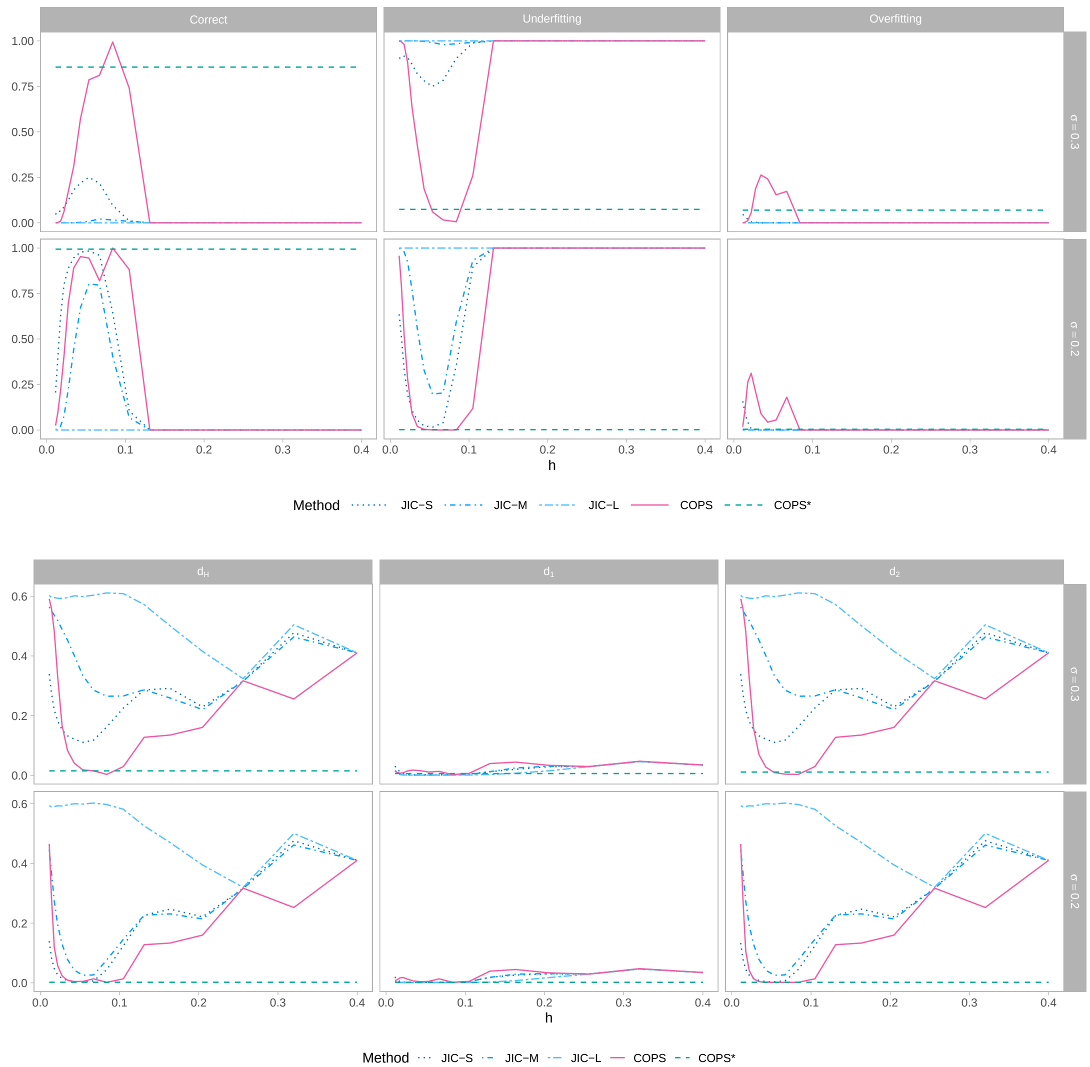


Figure S5: Estimation precision of the detected jumps by different methods in Example II with $\zeta = 1$ when the bandwidth h changes.

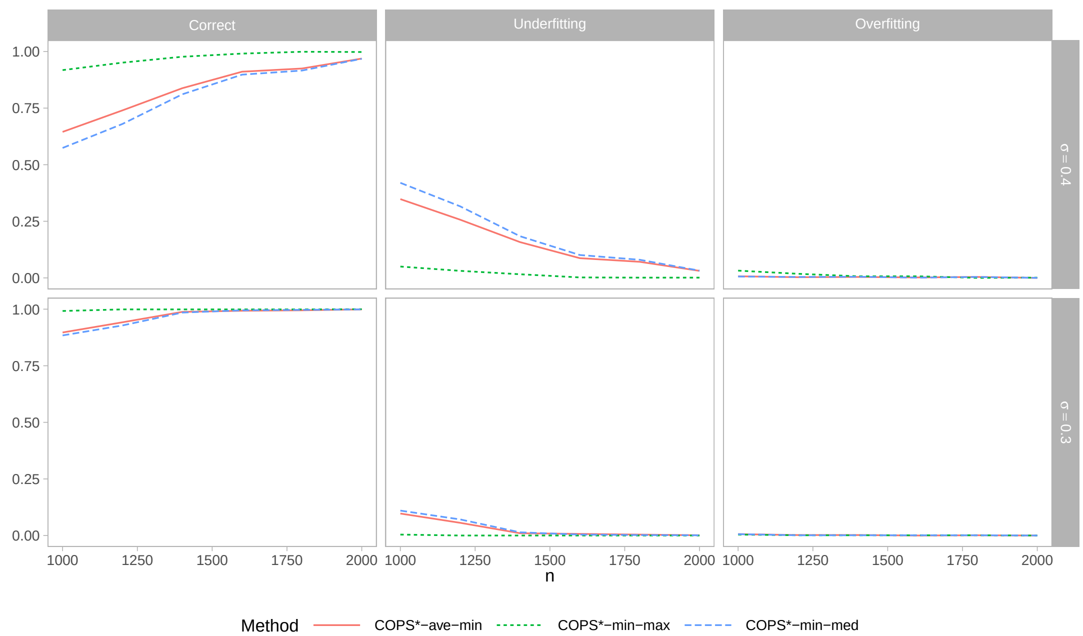


Figure S6: Probabilities of correct, under- and over-estimation of the number of true jumps against the sample sizes under relatively larger and smaller noise levels for three different COPS*-type procedures in Example I.

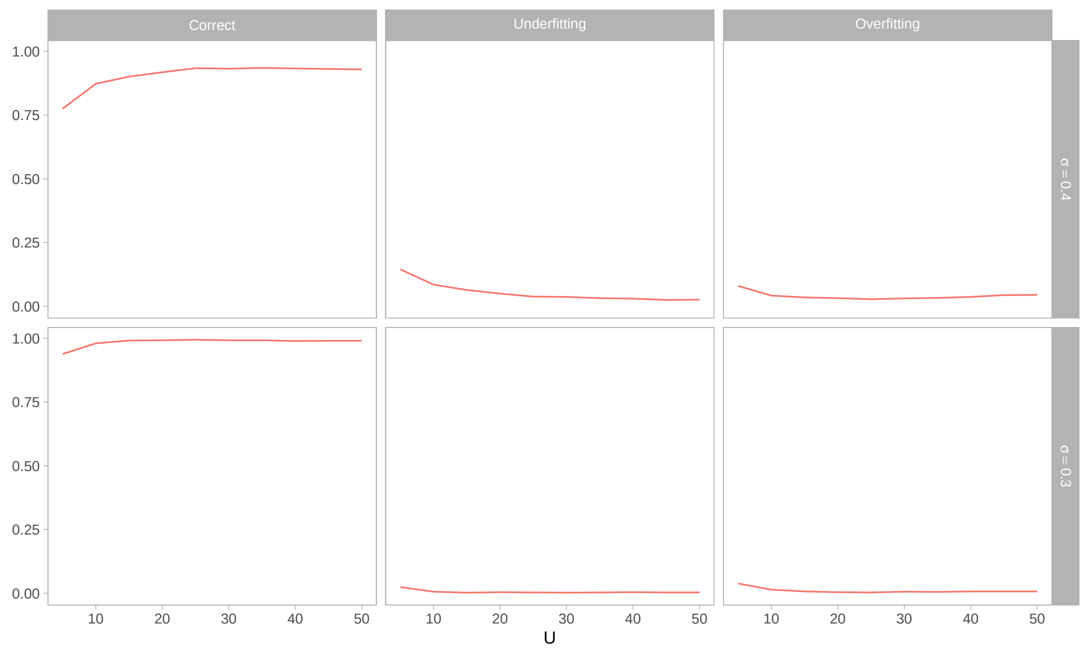


Figure S7: Probabilities of correct, under- and over-estimation of the number of true jumps against the number of random sample-splittings in the COPS* method in Example I.

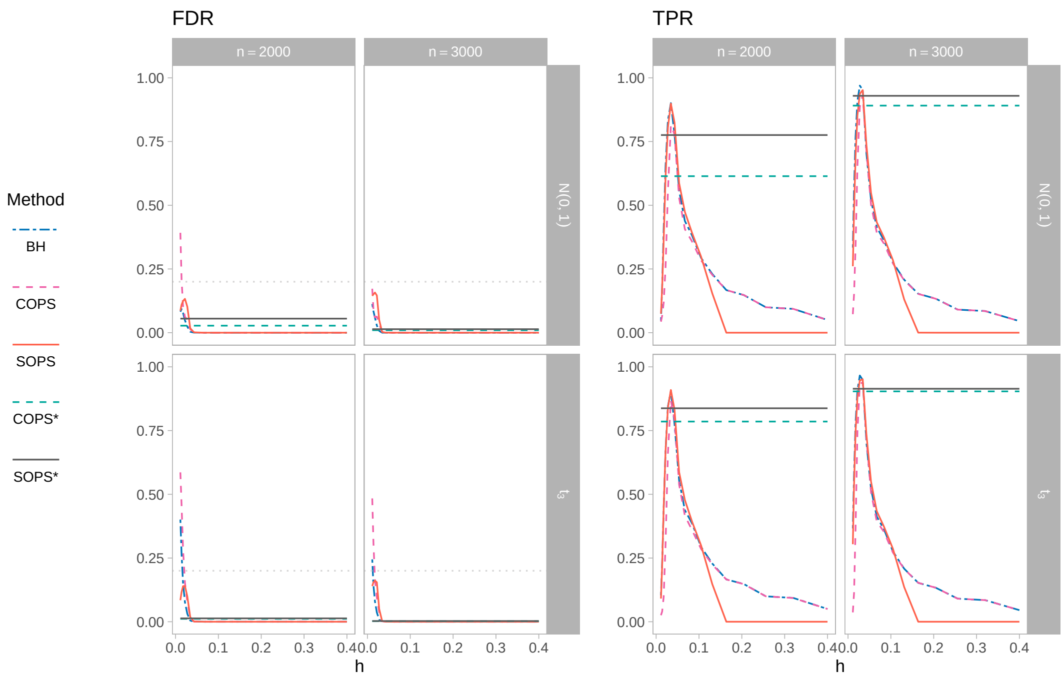


Figure S8: FDR and TPR values of different methods when the bandwidth h changes, under different combinations of the sample size and the noise distribution considered in Example III when the noise level $\sigma = 0.2$.

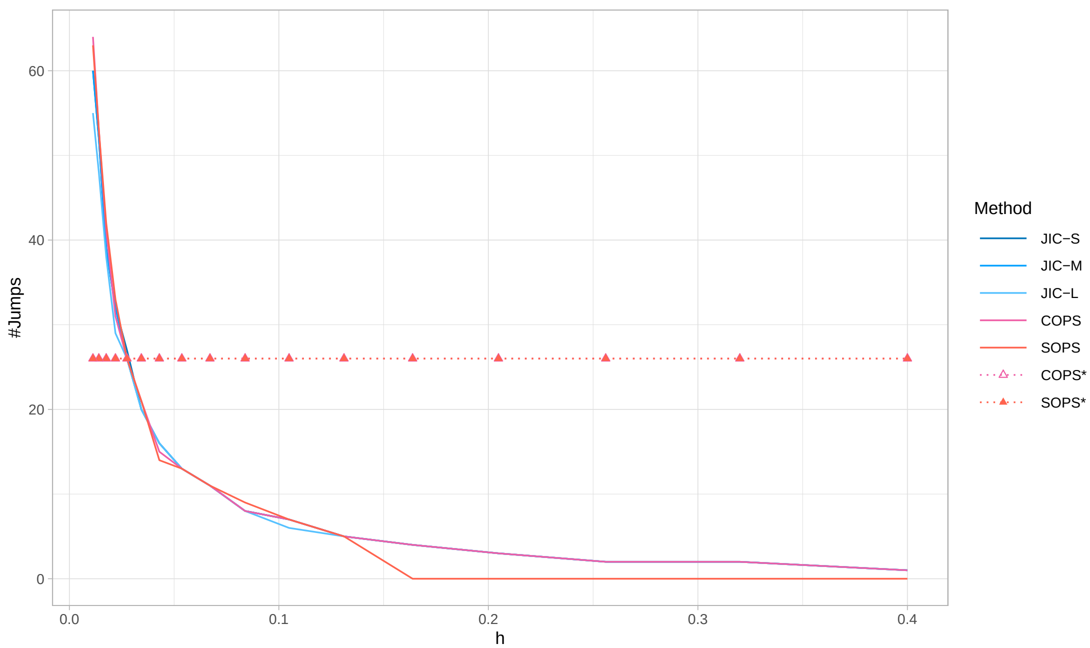


Figure S9: Estimated number of jumps against the value of the bandwidth for three versions of the JIC, COPS and SOPS methods with their bandwidth-adaptive versions, for the real data example.