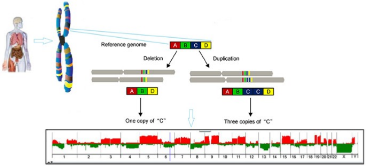
Inference for Change Points in Piecewise Polynomials

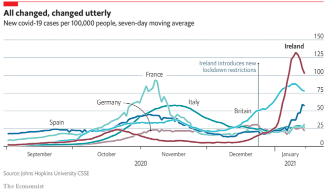
Shakeel Gavioli-Akilagun Piotr Fryzlewicz

London School of Economics
Department of Statistics

Change Point Regression

- In applications data can often be modelled as noise fluctuating around a piecewise parametric trend

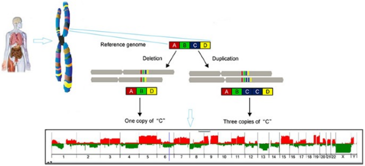


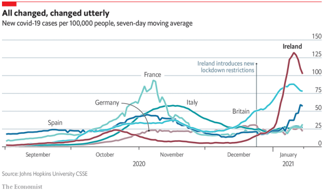


- Between break locations the parametric trend is easily interpretable
- Break locations themselves can be of interest to practitioners

Change Point Regression

- In applications data can often be modelled as noise fluctuating around a piecewise parametric trend

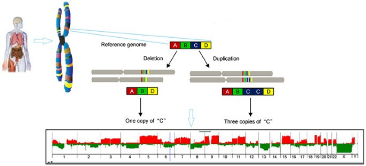


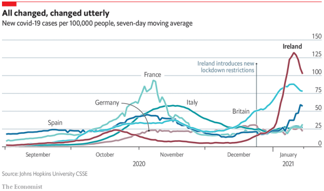


- Between break locations the parametric trend is easily interpretable
- Break locations themselves can be of interest to practitioners

Change Point Regression

- In applications data can often be modelled as noise fluctuating around a piecewise parametric trend





- Between break locations the parametric trend is easily interpretable
- Break locations themselves can be of interest to practitioners

Problem Statement

- We consider data $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ generated by a 'signal + noise' model with **piecewise polynomial** signal

$$Y_t = f_\circ(t/n) + \zeta_t, \qquad t = 1, \ldots, n$$

- Specifically there are N **integer valued change point locations** $1 < \eta_1 < \cdots < \eta_N < n$ between which $f_\circ$ is a polynomial of degree p
- **Goal:** to make unconditional inference statements about the unknown change point locations

Problem Statement

- We consider data $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ generated by a 'signal + noise' model with **piecewise polynomial** signal

$$Y_t = f_\circ(t/n) + \zeta_t, \qquad t = 1, \ldots, n$$

- Specifically there are N **integer valued change point locations** $1 < \eta_1 < \cdots < \eta_N < n$ between which $f_\circ$ is a polynomial of degree p
- **Goal:** to make unconditional inference statements about the unknown change point locations

Problem Statement

- We consider data $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ generated by a 'signal + noise' model with **piecewise polynomial** signal

$$Y_t = f_\circ (t/n) + \zeta_t, \qquad t = 1, \ldots, n$$

- Specifically there are N **integer valued change point locations** $1 < \eta_1 < \cdots < \eta_N < n$ between which $f_\circ$ is a polynomial of degree p
- **Goal:** to make unconditional inference statements about the unknown change point locations

Our Methods: piecewise constant signal

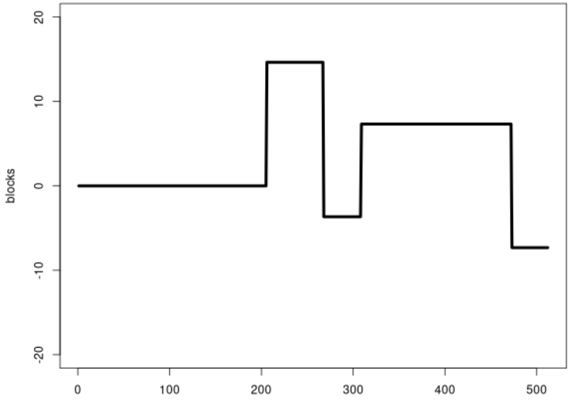


Figure 2: first 512 values of the `blocks` signal

Our Methods: piecewise constant signal

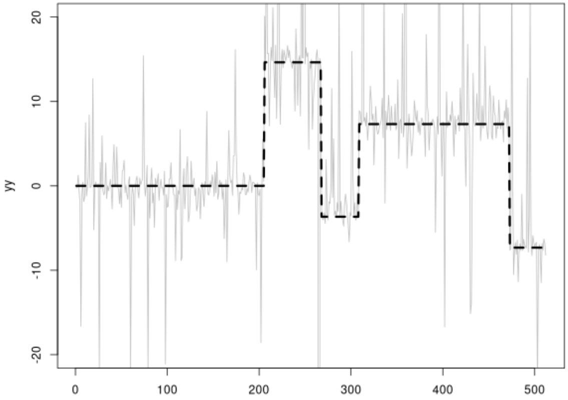


Figure 2: first 512 values of the `blocks` signal contaminated with Cauchy noise

Our Methods: piecewise constant signal

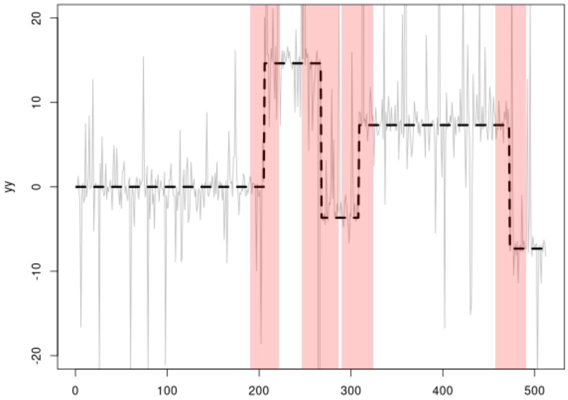


Figure 2: intervals of significance returned by our procedure; **interpretation** with probability at least $1 - \alpha$ each interval must contain a change point

Our Methods: piecewise constant signal

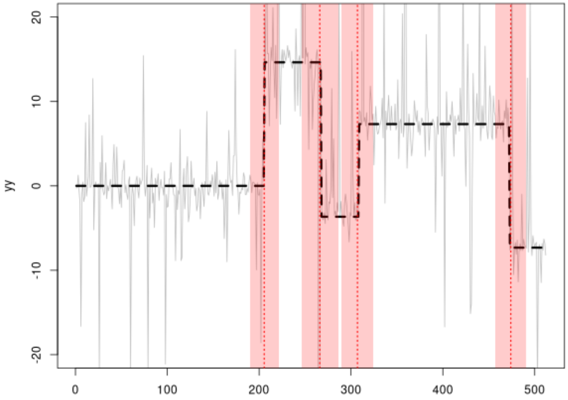


Figure 2: intervals of significance returned by our procedure and their midpoints

Our Methods: piecewise linear signal

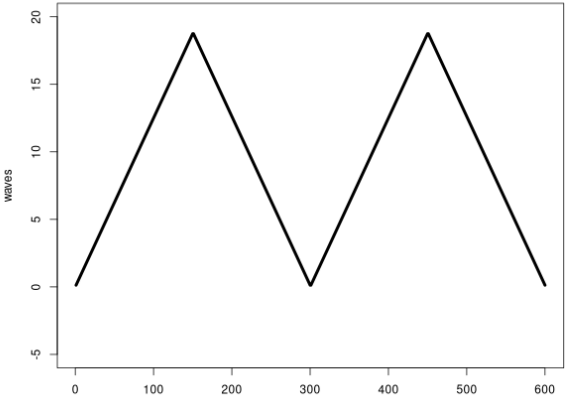


Figure 3: first 600 values of the `waves` signal

Our Methods: piecewise linear signal

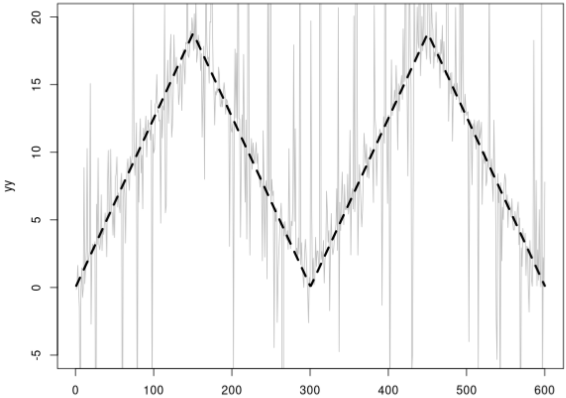


Figure 3: first 600 values of the `waves` signal contaminated with Cauchy noise

Our Methods: piecewise linear signal

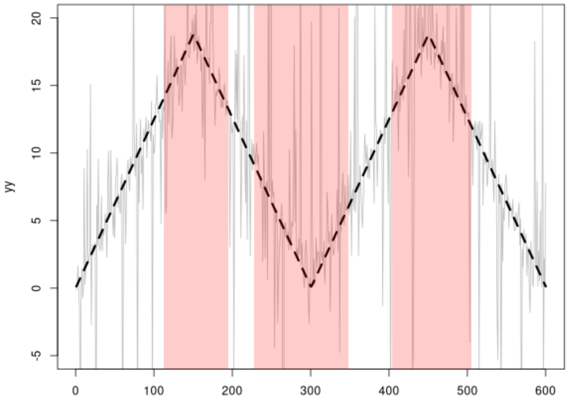


Figure 3: intervals of significance returned by our procedure; **interpretation** with probability at least $1 - \alpha$ each interval must contain a change point

Our Methods: piecewise linear signal

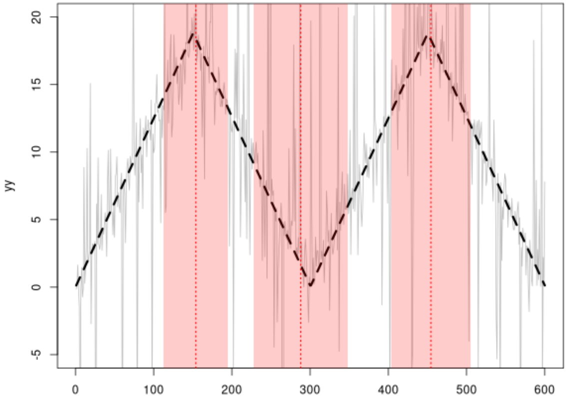


Figure 3: intervals of significance returned by our procedure and their midpoints

Main Idea

- Let $\mathcal{G}_n$ be a collection of sub-intervals $\{s, \ldots, e\} \subset \{1, \ldots, n\}$
- Let $\mathcal{T}_{\mathcal{G}_n} = \{T_{s:e}\}$ with $T_{s:e} : (Y_s, \ldots, Y_e)' \mapsto \{0, 1\}$ be a collection of local tests indexed on $\mathcal{G}_n$ for testing the local nulls

$$\mathcal{H}_{s:e} : f_\circ \in \mathscr{P}(p) \text{ 'locally'}$$

- If $\mathrm{FWE}(\mathcal{T}_{\mathcal{G}_n}) \leqslant \alpha$ whenever $T_{s:e} = 1$ the interval $\{s, \ldots, e\}$ must contain a change point with probability at least $1 - \alpha$

Main Idea

- Let $\mathcal{G}_n$ be a collection of sub-intervals $\{s, \ldots, e\} \subset \{1, \ldots, n\}$
- Let $\mathcal{T}_{\mathcal{G}_n} = \{T_{s:e}\}$ with $T_{s:e} : (Y_s, \ldots, Y_e)' \mapsto \{0, 1\}$ be a collection of local tests indexed on $\mathcal{G}_n$ for testing the local nulls

$$\mathcal{H}_{s:e} : f_\circ \in \mathscr{P}(p) \text{ 'locally'}$$

- If $\text{FWE}(\mathcal{T}_{\mathcal{G}_n}) \leqslant \alpha$ whenever $T_{s:e} = 1$ the interval $\{s, \ldots, e\}$ must contain a change point with probability at least $1 - \alpha$

Main Idea

- Let $\mathcal{G}_n$ be a collection of sub-intervals $\{s, \ldots, e\} \subset \{1, \ldots, n\}$
- Let $\mathcal{T}_{\mathcal{G}_n} = \{T_{s:e}\}$ with $T_{s:e} : (Y_s, \ldots, Y_e)' \mapsto \{0, 1\}$ be a collection of local tests indexed on $\mathcal{G}_n$ for testing the local nulls

$$\mathcal{H}_{s:e} : f_\circ \in \mathscr{P}(p) \text{ `locally'}$$

- If $\mathrm{FWE}(\mathcal{T}_{\mathcal{G}_n}) \leqslant \alpha$ whenever $T_{s:e} = 1$ the interval $\{s, \ldots, e\}$ must contain a change point with probability at least $1 - \alpha$

Generic Algorithm for Change Point Inference

- With this basic idea we can greedily search for the narrowest interval on which each change point is detectable using a procedure similar to binary segmentation

Algorithm 1: interval binary segmentation

```
function IntBinSeg(Y, s, e):
    if e − s < minimum segment then
        STOP
    end
    for (i, j) in G_{s:e} do
        if T_{i:j} = 1 then
            RecordInterval(i, j)
            IntBinSeg(Y, s, i)
            IntBinSeg(Y, j, e)
        end
    end
return
```

Fast Inference via Differencing

- We present a procedure for inference in the presence of independent $\mathcal{N}\left(0, \sigma^2\right)$ contaminating noise with σ know or estimated consistently
- **Advantage # 1:** computational complexity is always $\mathcal{O}\left(n \log(n)\right)$ due to new use of local averaging on a sparse grids
- **Advantage # 2:** thresholds on locals tests are sharpest possible due to asymptotic independence of scales in our grid
- **Advantage # 3:** thresholds are adaptive to the density of the grid used

Fast Inference via Differencing

- We present a procedure for inference in the presence of independent $\mathcal{N}\left(0, \sigma^2\right)$ contaminating noise with σ know or estimated consistently
- **Advantage # 1:** computational complexity is always $\mathcal{O}\left(n \log(n)\right)$ due to new use of local averaging on a sparse grids
- **Advantage # 2:** thresholds on locals tests are sharpest possible due to asymptotic independence of scales in our grid
- **Advantage # 3:** thresholds are adaptive to the density of the grid used

Fast Inference via Differencing

- We present a procedure for inference in the presence of independent $\mathcal{N}\left(0, \sigma^2\right)$ contaminating noise with σ know or estimated consistently
- **Advantage # 1:** computational complexity is always $\mathcal{O}\left(n \log(n)\right)$ due to new use of local averaging on a sparse grids
- **Advantage # 2:** thresholds on locals tests are sharpest possible due to asymptotic independence of scales in our grid
- **Advantage # 3:** thresholds are adaptive to the density of the grid used

Fast Inference via Differencing

- We present a procedure for inference in the presence of independent $\mathcal{N}\left(0, \sigma^2\right)$ contaminating noise with σ know or estimated consistently
- **Advantage # 1:** computational complexity is always $\mathcal{O}\left(n \log(n)\right)$ due to new use of local averaging on a sparse grids
- **Advantage # 2:** thresholds on locals tests are sharpest possible due to asymptotic independence of scales in our grid
- **Advantage # 3:** thresholds are adaptive to the density of the grid used

Local Tests

Local Tests Based on Differencing

- If it was know that the interval $\{s, \dots, e\}$ contained a change we could compute the local sums

$$X^j_{s:e} = Y_{s+\frac{jw_{s,e}}{p+2}} + \cdots + Y_{s+\frac{(j+1)w_{s,e}}{p+2}-1} \qquad w_{s,e} = (p+2)\left\lfloor \frac{e-s+1}{p+2} \right\rfloor$$

- Then if the change is sufficiently large the following test statistic should be large in absolute value

$$T^{(p)}_{s:e}(\mathbf{Y}) = \left\{ \left[\frac{w_{s,e}}{p+2} \right] \sum_{i=0}^{p+1} \binom{p+1}{i}^2 \right\} \sum_{j=0}^{p+1} (-1)^{p+1-j} \binom{p+1}{j} X^j_{s,e}$$

- For some $\lambda > 0$ we can define local tests as follows

$$T_{s,e} = \mathbf{1}\left\{ \left| T^{(p)}_{s:e}(\mathbf{Y}) \right| > \lambda \right\}$$

Local Tests Based on Differencing

- If it was know that the interval $\{s, \ldots, e\}$ contained a change we could compute the local sums

$$X_{s:e}^{j} = Y_{s+\frac{jw_{s,e}}{p+2}} + \cdots + Y_{s+\frac{(j+1)w_{s,e}}{p+2}-1} \qquad w_{s,e} = (p+2)\left\lfloor \frac{e-s+1}{p+2} \right\rfloor$$

- Then if the change is sufficiently large the following test statistic should be large in absolute value

$$T_{s:e}^{(p)}(\mathbf{Y}) = \left\{ \left[\frac{w_{s,e}}{p+2} \right] \sum_{i=0}^{p+1} \binom{p+1}{i}^{2} \right\} \sum_{j=0}^{p+1} (-1)^{p+1-j} \binom{p+1}{j} X_{s,e}^{j}$$

- For some $\lambda > 0$ we can define local tests as follows

$$T_{s,e} = \mathbf{1}\left\{ \left| T_{s:e}^{(p)}(\mathbf{Y}) \right| > \lambda \right\}$$

Local Tests Based on Differencing

- If it was know that the interval $\{s, \ldots, e\}$ contained a change we could compute the local sums

$$X^j_{s:e} = Y_{s+\frac{jw_{s,e}}{p+2}} + \cdots + Y_{s+\frac{(j+1)w_{s,e}}{p+2}-1} \qquad w_{s,e} = (p+2)\left\lfloor \frac{e-s+1}{p+2} \right\rfloor$$

- Then if the change is sufficiently large the following test statistic should be large in absolute value

$$T^{(p)}_{s:e}(\mathbf{Y}) = \left\{ \left\lfloor \frac{w_{s,e}}{p+2} \right\rfloor \sum_{i=0}^{p+1} \binom{p+1}{i}^2 \right\} \sum_{j=0}^{p+1} (-1)^{p+1-j} \binom{p+1}{j} X^j_{s,e}$$

- For some $\lambda > 0$ we can define local tests as follows

$$T_{s,e} = \mathbf{1}\left\{ \left| T^{(p)}_{s:e}(\mathbf{Y}) \right| > \lambda \right\}$$

Theorem (sup behaviour on sparse grids)

With tuning parameters $a > 1$ and $A \in \mathbb{N}$ introduce the a-adic grid which excludes scales of width $o(\log(n))$ as follows

$$\mathcal{G}_n(a, A) = \left\{(s, e) \in \mathbb{N}^2 \mid 1 \leqslant s < e \leqslant n, (e - s + 1) = \left\lfloor a^k \right\rfloor k \in \mathcal{K}_n(a, A)\right\}$$
$$\mathcal{K}_n(a, A) = \{\log_a \log(n) - A, \ldots, \log_a(n/2)\}$$

Introduce the sup of the (standardised) test statistic over $\mathcal{G}_n(a, A)$ as follows

$$M_n^{a,A} = \max_{(s,e) \in \mathcal{G}_n(a,A)} \left\{\frac{1}{\sigma} T_{s:e}^{(p)}(\mathbf{Y})\right\}$$

Then putting $a_n = \sqrt{2\log(n)}$ and $b_n = 2\log(n) - \frac{1}{2}\log\log(n) - \log(2\sqrt{\pi})$ there are constants H_1 and H_2 depending only on a, A, and p such that for fixed $\tau \in \mathbb{R}$ the following holds

$$o(1) + \exp\left(-H_1 e^{-\tau}\right) \leqslant \mathbb{P}\left(a_n M_n^{a,L} - b_n \leqslant \tau\right) \leqslant \exp\left(-H_2 e^{-\tau}\right) + o(1)$$

Performance of the Algorithm

Characterising the Change Point Problem

- The **changes in derivative** are given by $\Delta_{j,k} = \alpha_{j,k} - \beta_{j,k}$ and between change points the **signal** is parametrised as

$$f_\circ\left(t/n\right) = \begin{cases} \sum_{j=0}^{p} \alpha_{j,k}\left(t/n - \eta_k/n\right)^j & \eta_{k-1} < t \leqslant \eta_k \\ \sum_{j=0}^{p} \beta_{j,k}\left(t/n - \eta_k/n\right)^j & \eta_k < t < \eta_{k+1} \end{cases}$$

- The **most prominent** change in derivative at the k-th change point is uniquely defined as

$$p_k^* = \underset{0 \leqslant j \leqslant p}{\text{arg-max}} \left\{ \left|\Delta_{j,k}\right| \left(\frac{\delta_k}{n}\right)^j \right\}$$

- Assume that for each $k = 1, \ldots, N$ the following holds

$$\delta_k > C_1 \left(\frac{\log\left(n\right)}{a^A} \vee n^{\frac{2p_k^*}{2p_k^*+1}} \left(\frac{\sigma^2 \log\left(n\right)}{\Delta_{p_k^*}^2} \right)^{\frac{1}{2p_k^*+1}} \right)$$

Characterising the Change Point Problem

- The **changes in derivative** are given by $\Delta_{j,k} = \alpha_{j,k} - \beta_{j,k}$ and between change points the **signal** is parametrised as

$$f_\circ(t/n) = \begin{cases} \sum_{j=0}^{p} \alpha_{j,k} \left(t/n - \eta_k/n\right)^j & \eta_{k-1} < t \leqslant \eta_k \\ \sum_{j=0}^{p} \beta_{j,k} \left(t/n - \eta_k/n\right)^j & \eta_k < t < \eta_{k+1} \end{cases}$$

- The **most prominent** change in derivative at the k-th change point is uniquely defined as

$$p_k^* = \underset{0 \leqslant j \leqslant p}{\text{arg-max}} \left\{ \left|\Delta_{j,k}\right| \left(\frac{\delta_k}{n}\right)^j \right\}$$

- Assume that for each $k = 1, \ldots, N$ the following holds

$$\delta_k > C_1 \left(\frac{\log(n)}{a^A} \vee n^{\frac{2p_k^*}{2p_k^*+1}} \left(\frac{\sigma^2 \log(n)}{\Delta_{p_k^*}^2} \right)^{\frac{1}{2p_k^*+1}} \right)$$

Characterising the Change Point Problem

- The **changes in derivative** are given by $\Delta_{j,k} = \alpha_{j,k} - \beta_{j,k}$ and between change points the **signal** is parametrised as

$$f_\circ\left(t/n\right) = \begin{cases} \sum_{j=0}^{p} \alpha_{j,k}\left(t/n - \eta_k/n\right)^j & \eta_{k-1} < t \leqslant \eta_k \\ \sum_{j=0}^{p} \beta_{j,k}\left(t/n - \eta_k/n\right)^j & \eta_k < t < \eta_{k+1} \end{cases}$$

- The **most prominent** change in derivative at the k-th change point is uniquely defined as

$$p_k^* = \underset{0 \leqslant j \leqslant p}{\text{arg-max}} \left\{ \left|\Delta_{j,k}\right| \left(\frac{\delta_k}{n}\right)^j \right\}$$

- Assume that for each $k = 1, \ldots, N$ the following holds

$$\delta_k > C_1 \left(\frac{\log\left(n\right)}{a^A} \vee n^{\frac{2p_k^*}{2p_k^*+1}} \left(\frac{\sigma^2 \log\left(n\right)}{\Delta_{p_k^*}^2} \right)^{\frac{1}{2p_k^*+1}} \right)$$

Theorem (change point detection guarantees)

Let $\left\{\hat{I}_1, \ldots, \hat{I}_{\hat{N}}\right\}$ *be intervals returned the generic algorithm with locals tests* $\mathcal{T}_{\mathcal{G}_n}$ *on a grid* $\mathcal{G}_n(a, A)$ *using threshold* λ *defined below for some* $\alpha \in (0,1)$

$$\lambda = \sqrt{2\log(n)} + \frac{-\frac{1}{2}\log\log(n) - \log\left(2\sqrt{\pi}/H_2\right) + \log\left(-2\log^{-1}(1-\alpha)\right)}{\sqrt{2\log(n)}}$$

Then with probability $1 - \alpha + o(1)$ *the following events occurs simultaneously*

$$E_1^* = \left\{\hat{N} = N\right\}$$

$$E_2^* = \left\{\forall k = 1, \ldots, N \ \hat{I}_k \cap \Theta = \eta_k\right\}$$

$$E_3^* = \left\{\left|\hat{I}_k\right| \leqslant C_2\left(\frac{\log(n)}{a^A} \vee n^{\frac{2p_k^*}{2p_k^*+1}}\left(\frac{\sigma^2\log(n)}{\Delta_{p_k^*}^2}\right)^{\frac{1}{2p_k^*+1}}\right)\middle| 1 \leqslant k \leqslant N\right\}$$

Here C_1, C_2 *satisfy* $C_1 > 2(p+1)C_2$

Robust Inference using Confidence Sets

- We present a procedure for inference in the presence of **sign independent** $(\text{sign}(\zeta_t) \perp\!\!\!\perp \text{sign}(\zeta_s)\,; t \neq s)$ and **sign symmetric** $(\mathbb{P}(\zeta_t > 0) = \mathbb{P}(\zeta_t < 0))$ contaminating noise
- **Advantage # 1:** we make practically no assumptions on the distribution of the noise but still provide near optimal detection guarantees
- **Advantage # 2:** we consider the most general piecewise polynomial signal setting, whereas competing algorithms consider only signals piecewise constant

Robust Inference using Confidence Sets

- We present a procedure for inference in the presence of **sign independent** $(\text{sign}(\zeta_t) \perp\!\!\!\perp \text{sign}(\zeta_s)\,;\, t \neq s)$ and **sign symmetric** $(\mathbb{P}(\zeta_t > 0) = \mathbb{P}(\zeta_t < 0))$ contaminating noise
- **Advantage # 1:** we make practically no assumptions on the distribution of the noise but still provide near optimal detection guarantees
- **Advantage # 2:** we consider the most general piecewise polynomial signal setting, whereas competing algorithms consider only signals piecewise constant

Robust Inference using Confidence Sets

- We present a procedure for inference in the presence of **sign independent** $(\text{sign}\,(\zeta_t) \perp\!\!\!\perp \text{sign}\,(\zeta_s)\,; t \neq s)$ and **sign symmetric** $(\mathbb{P}\,(\zeta_t > 0) = \mathbb{P}\,(\zeta_t < 0))$ contaminating noise
- **Advantage # 1:** we make practically no assumptions on the distribution of the noise but still provide near optimal detection guarantees
- **Advantage # 2:** we consider the most general piecewise polynomial signal setting, whereas competing algorithms consider only signals piecewise constant

Local Tests

Computationally Infeasible Test

- Start with the following test motivate by Fryzlewicz (2021)

$$T_{s:e} = \mathbf{1}\left\{ \min_{\hat{f}\in\mathscr{P}(p)} \max_{s\leqslant i\leqslant j\leqslant e} \frac{1}{\sqrt{j-i+1}} \left| \sum_{t=i}^{j} \operatorname{sign}\left(Y_t - \hat{f}\,(t/n) \right) \right| > \lambda \right\}$$

- Easy to bound the FWE since under the local null we have

$$T_{s:e} \leqslant \mathbf{1}\left\{ \max_{1\leqslant i\leqslant j\leqslant n} \frac{1}{\sqrt{j-i+1}} \left| \sum_{t=i}^{j} \operatorname{sign}(\zeta_t) \right| > \lambda \right\}$$

- Naïvely each local test has time complexity $\mathcal{O}\left((e-s)^{p+3}\right)$

Computationally Infeasible Test

- Start with the following test motivate by Fryzlewicz (2021)

$$T_{s:e} = \mathbf{1}\left\{\min_{\hat{f}\in\mathscr{P}(p)} \max_{s\leqslant i\leqslant j\leqslant e} \frac{1}{\sqrt{j-i+1}}\left|\sum_{t=i}^{j} \operatorname{sign}\left(Y_t - \hat{f}\left(t/n\right)\right)\right| > \lambda\right\}$$

- Easy to bound the FWE since under the local null we have

$$T_{s:e} \leqslant \mathbf{1}\left\{\max_{1\leqslant i\leqslant j\leqslant n} \frac{1}{\sqrt{j-i+1}}\left|\sum_{t=i}^{j} \operatorname{sign}\left(\zeta_t\right)\right| > \lambda\right\}$$

- Naïvely each local test has time complexity $\mathcal{O}\left((e-s)^{p+3}\right)$

Computationally Infeasible Test

- Start with the following test motivate by Fryzlewicz (2021)

$$T_{s:e} = \mathbf{1}\left\{ \min_{\hat{f}\in\mathscr{P}(p)} \max_{s\leqslant i\leqslant j\leqslant e} \frac{1}{\sqrt{j-i+1}} \left| \sum_{t=i}^{j} \operatorname{sign}\left(Y_t - \hat{f}\,(t/n) \right) \right| > \lambda \right\}$$

- Easy to bound the FWE since under the local null we have

$$T_{s:e} \leqslant \mathbf{1}\left\{ \max_{1\leqslant i\leqslant j\leqslant n} \frac{1}{\sqrt{j-i+1}} \left| \sum_{t=i}^{j} \operatorname{sign}\left(\zeta_t\right) \right| > \lambda \right\}$$

- Naïvely each local test has time complexity $\mathcal{O}\left((e-s)^{p+3}\right)$

Theorem (sup on the complete grid, Kabluchko & Wang 2014)

Introduce the complete grid $\mathcal{G}_n^c = \left\{(s,e) \in \mathbb{N}^2 \mid 1 \leqslant s < e \leqslant n\right\}$ *and define sup of (standardised) partial sums of noise signs as follows*

$$M_n = \max_{(s,e) \in \mathcal{G}_n^c} \left\{ \frac{1}{\sqrt{e-s+1}} \sum_{t=s}^{e} sign\left(\zeta_t\right) \right\}$$

Then putting $a_n = \sqrt{2 \log\left(n \log^{-\frac{1}{2}} \log(n)\right)}$ *there is a constant* H *such that for fixed* $\tau \in \mathbb{R}$ *the following holds*

$$\mathbb{P}\left(\left(M_n - a_n\right) a_n \leqslant \tau\right) = exp\left(-He^{-\tau}\right) + o(1)$$

Towards a Computationally Feasible Test

- For arbitrary f put $\zeta_t^f = Y_t - f\,(t/n)$ and introduce the test

$$\psi_{s:e}^{f} = \mathbf{1}\left\{\max_{s\leqslant i\leqslant j\leqslant e} \frac{1}{\sqrt{j-i+1}}\left|\sum_{t=i}^{j}\mathrm{sign}\left(\hat{\zeta}_t^f\right)\right| > \lambda\right\}$$

- Invert $\{\psi_{s:e}\}$ forming a local confidence set for $f_\circ$

$$\mathcal{C}_{s:e}\,(Y,\lambda) = \left\{f : \psi_{s:e}^{f} \neq 1\right\}$$

- By duality of confidence sets and statistical tests use

$$T_{s:e} = \mathbf{1}\left\{\mathcal{C}_{s:e}\,(Y,\lambda) \cap \mathscr{P}\,(p) = \emptyset\right\}$$

Towards a Computationally Feasible Test

- For arbitrary f put $\zeta_t^f = Y_t - f(t/n)$ and introduce the test

$$\psi_{s:e}^f = \mathbf{1}\left\{ \max_{s \leqslant i \leqslant j \leqslant e} \frac{1}{\sqrt{j-i+1}} \left| \sum_{t=i}^{j} \operatorname{sign}\left(\hat{\zeta}_t^f\right) \right| > \lambda \right\}$$

- Invert $\{\psi_{s:e}\}$ forming a local confidence set for $f_\circ$

$$\mathcal{C}_{s:e}(Y, \lambda) = \left\{ f : \psi_{s:e}^f \neq 1 \right\}$$

- By duality of confidence sets and statistical tests use

$$T_{s:e} = \mathbf{1}\left\{ \mathcal{C}_{s:e}(Y, \lambda) \cap \mathscr{P}(p) = \emptyset \right\}$$

Towards a Computationally Feasible Test

- For arbitrary f put $\zeta_t^f = Y_t - f\left(t/n\right)$ and introduce the test

$$\psi_{s:e}^f = \mathbf{1}\left\{\max_{s\leqslant i\leqslant j\leqslant e} \frac{1}{\sqrt{j-i+1}}\left|\sum_{t=i}^{j} \mathrm{sign}\left(\hat{\zeta}_t^f\right)\right| > \lambda\right\}$$

- Invert $\{\psi_{s:e}\}$ forming a local confidence set for $f_\circ$

$$\mathcal{C}_{s:e}\left(Y,\lambda\right) = \left\{f : \psi_{s:e}^f \neq 1\right\}$$

- By duality of confidence sets and statistical tests use

$$T_{s:e} = \mathbf{1}\left\{\mathcal{C}_{s:e}\left(Y,\lambda\right) \cap \mathscr{P}\left(p\right) = \varnothing\right\}$$

Approximation Through Confidence Sets

- Restrict attention to **monotone functions** and compute point-wise upper and lower bounds on the set

$$L_k^{\uparrow} = \inf \left\{ f\left(k/n\right) \mid f \in \mathcal{C}_{s:e}\left(Y, \lambda\right) \cap \mathcal{F}_{\uparrow} \right\}$$

- Can be thought of as performing two sided tests at all scales and locations on the empirical residuals $\hat{\zeta}_s^f, \ldots, \hat{\zeta}_e^f$

$$L_k^{\uparrow} = \inf \left\{ f\left(k/n\right) \mid f \in \mathcal{F}_{\uparrow}, \max_{s \leqslant i \leqslant j \leqslant e} \frac{1}{\sqrt{j-i+1}} \left| \sum_{t=i}^{j} \operatorname{sign}\left(\zeta_t^f\right) \right| \leqslant \lambda \right\}$$

- Only one tail is important; Dümbgen & Johns (2004) introduce the easily computable lower (upper) bound $\check{L}^{\uparrow}$ for which $\check{L}_k^{\downarrow} \leqslant L_k^{\uparrow}$ for all k

$$\check{L}_k^{\uparrow} = \min \left\{ \bar{f} \in \{-\infty, Y_s, Y_{s+1}, \ldots, Y_k\} \mid \bar{f} \geqslant \check{L}_{k-1}^{\uparrow}, \max_{s \leqslant i \leqslant j \leqslant k} \frac{1}{\sqrt{j-i+1}} \sum_{t=i}^{j} \operatorname{sign}(Y_t - \bar{f}) \leqslant \lambda \right\}$$

Approximation Through Confidence Sets

- Restrict attention to **monotone functions** and compute point-wise upper and lower bounds on the set

$$L_k^{\uparrow} = \inf \left\{ f\left(k/n\right) \mid f \in \mathcal{C}_{s:e}\left(Y, \lambda\right) \cap \mathcal{F}_{\uparrow} \right\}$$

- Can be thought of as performing two sided tests at all scales and locations on the empirical residuals $\hat{\zeta}_s^f, \ldots, \hat{\zeta}_e^f$

$$L_k^{\uparrow} = \inf \left\{ f\left(k/n\right) \mid f \in \mathcal{F}_{\uparrow}, \max_{s \leqslant i \leqslant j \leqslant e} \frac{1}{\sqrt{j-i+1}} \left| \sum_{t=i}^{j} \operatorname{sign}\left(\zeta_t^f\right) \right| \leqslant \lambda \right\}$$

- Only one tail is important; Dümbgen & Johns (2004) introduce the easily computable lower (upper) bound $\check{L}^{\uparrow}$ for which $\check{L}_k^{\downarrow} \leqslant L_k^{\uparrow}$ for all k

$$\check{L}_k^{\uparrow} = \min \left\{ \bar{f} \in \{-\infty, Y_s, Y_{s+1}, \ldots, Y_k\} \mid \bar{f} \geqslant \check{L}_{k-1}^{\uparrow}, \max_{s \leqslant i \leqslant j \leqslant k} \frac{1}{\sqrt{j-i+1}} \sum_{t=i}^{j} \operatorname{sign}(Y_t - \bar{f}) \leqslant \lambda \right\}$$

Approximation Through Confidence Sets

- Restrict attention to **monotone functions** and compute point-wise upper and lower bounds on the set

$$L_k^{\uparrow} = \inf \left\{ f\left(k/n\right) \mid f \in \mathcal{C}_{s:e}\left(Y, \lambda\right) \cap \mathcal{F}_{\uparrow} \right\}$$

- Can be thought of as performing two sided tests at all scales and locations on the empirical residuals $\hat{\zeta}_s^f, \ldots, \hat{\zeta}_e^f$

$$L_k^{\uparrow} = \inf \left\{ f\left(k/n\right) \mid f \in \mathcal{F}_{\uparrow}, \max_{s \leqslant i \leqslant j \leqslant e} \frac{1}{\sqrt{j-i+1}} \left| \sum_{t=i}^{j} \operatorname{sign}\left(\zeta_t^f\right) \right| \leqslant \lambda \right\}$$

- Only one tail is important; Dümbgen & Johns (2004) introduce the easily computable lower (upper) bound $\check{L}^{\uparrow}$ for which $\check{L}_k^{\downarrow} \leqslant L_k^{\uparrow}$ for all k

$$\check{L}_k^{\uparrow} = \min \left\{ \bar{f} \in \{-\infty, Y_s, Y_{s+1}, \ldots, Y_k\} \mid \bar{f} \geqslant \check{L}_{k-1}^{\uparrow}, \max_{s \leqslant i \leqslant j \leqslant k} \frac{1}{\sqrt{j-i+1}} \sum_{t=i}^{j} \operatorname{sign}(Y_t - \bar{f}) \leqslant \lambda \right\}$$

Approximation Through Confidence Sets

- Monotone bounds sufficient when $p \leqslant 1$ and when $p > 1$ we can **adaptively switch monotonicity** of the bounds
- Put $\mathcal{R}^{\uparrow}_{s:e}$ for the region in $[0,1] \times \mathbb{R}$ enclosed by upper and lower bounds, approximate $T_{s:e}$ with

$$T'_{s:e} = \mathbf{1}\left\{\left(\mathcal{R}^{\uparrow}_{s:e} \cap \mathtt{poly}\,(p)\right) = \varnothing\right\} \wedge \mathbf{1}\left\{\left(\mathcal{R}^{\downarrow}_{s:e} \cap \mathtt{poly}\,(p)\right) = \varnothing\right\}$$

- Approximation has time complexity $\mathcal{O}\left((e-s)^2 + \mathrm{cost}_p\,(e-s)\right)$

Approximation Through Confidence Sets

- Monotone bounds sufficient when $p \leqslant 1$ and when $p > 1$ we can **adaptively switch monotonicity** of the bounds
- Put $\mathcal{R}^{\uparrow}_{s:e}$ for the region in $[0,1] \times \mathbb{R}$ enclosed by upper and lower bounds, approximate $T_{s:e}$ with

$$T'_{s:e} = \mathbf{1}\left\{\left(\mathcal{R}^{\uparrow}_{s:e} \cap \texttt{poly}\left(p\right)\right) = \varnothing\right\} \wedge \mathbf{1}\left\{\left(\mathcal{R}^{\downarrow}_{s:e} \cap \texttt{poly}\left(p\right)\right) = \varnothing\right\}$$

- Approximation has time complexity $\mathcal{O}\left(\left(e-s\right)^2 + \text{cost}_p\left(e-s\right)\right)$

Approximation Through Confidence Sets

- Monotone bounds sufficient when $p \leqslant 1$ and when $p > 1$ we can **adaptively switch monotonicity** of the bounds
- Put $\mathcal{R}^{\uparrow}_{s:e}$ for the region in $[0,1] \times \mathbb{R}$ enclosed by upper and lower bounds, approximate $T_{s:e}$ with

$$T'_{s:e} = \mathbf{1}\left\{\left(\mathcal{R}^{\uparrow}_{s:e} \cap \texttt{poly}\,(p)\right) = \varnothing\right\} \wedge \mathbf{1}\left\{\left(\mathcal{R}^{\downarrow}_{s:e} \cap \texttt{poly}\,(p)\right) = \varnothing\right\}$$

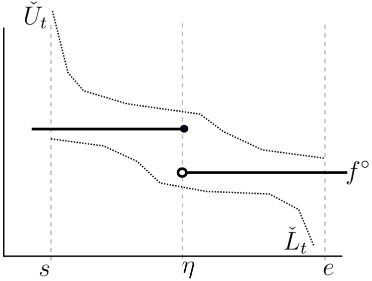
- Approximation has time complexity $\mathcal{O}\left(\left(e-s\right)^2 + \mathrm{cost}_p\left(e-s\right)\right)$

In Practice: piecewise constant signal

Data does not locally agree with $f_\circ$ being a degree zero polynomial if

$$\min_{s \leqslant t \leqslant e} \check{U}_t < \max_{s \leqslant t \leqslant e} \check{L}_t$$

It can be seen that $\text{cost}_0\left(e-s\right) = \mathcal{O}\left(1\right)$

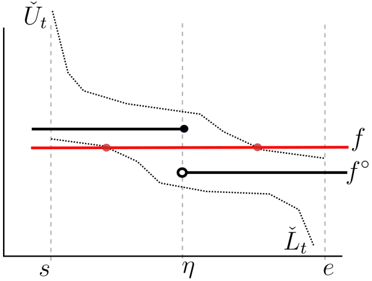


In Practice: piecewise constant signal

Data does not locally agree with $f_\circ$ being a degree zero polynomial if

$$\min_{s\leqslant t\leqslant e} \check{U}_t < \max_{s\leqslant t\leqslant e} \check{L}_t$$

It can be seen that $\mathrm{cost}_0\left(e-s\right) = \mathcal{O}\left(1\right)$

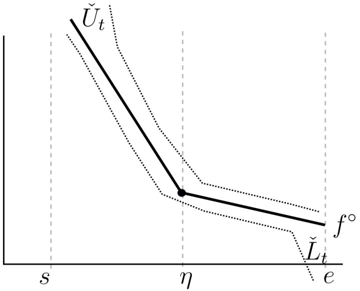


In Practice: piecewise constant signal

Data does not locally agree with $f_\circ$ being a degree one polynomial if

$$\text{vol}\left(\texttt{Hull}\left(\check{U}_{s:e}\right) \cap \texttt{Hull}\left(\check{L}_{s:e}\right)\right) > 0$$

It can be seen (Chazelle & Dobkin 1987) that $\text{cost}_1\left(e-s\right) = \mathcal{O}\left(\log\left(e-s\right)\right)$

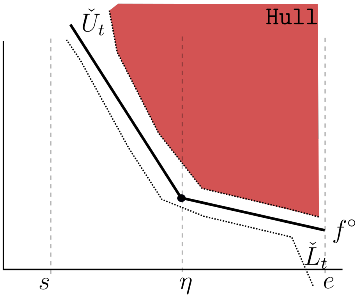


In Practice: piecewise constant signal

Data does not locally agree with $f_\circ$ being a degree one polynomial if

$$\mathrm{vol}\left(\mathtt{Hull}\left(\check{U}_{s:e}\right) \cap \mathtt{Hull}\left(\check{L}_{s:e}\right)\right) > 0$$

It can be seen (Chazelle & Dobkin 1987) that $\mathrm{cost}_1\left(e-s\right) = \mathcal{O}\left(\log\left(e-s\right)\right)$

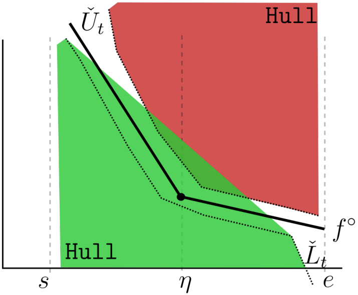


In Practice: piecewise constant signal

Data does not locally agree with $f_\circ$ being a degree one polynomial if

$$\text{vol}\left(\texttt{Hull}\left(\check{U}_{s:e}\right) \cap \texttt{Hull}\left(\check{L}_{s:e}\right)\right) > 0$$

It can be seen (Chazelle & Dobkin 1987) that $\text{cost}_1(e-s) = \mathcal{O}(\log(e-s))$



Performance of the Algorithm

Characterising the Change Point Problem

- The **changes in derivative** are given by $\Delta_{j,k} = \alpha_{j,k} - \beta_{j,k}$ and between change points the **signal** is parametrised as

$$f_\circ(t/n) = \begin{cases} \sum_{j=0}^{p} \alpha_{j,k} \left(t/n - \eta_k/n\right)^j & \eta_{k-1} < t \leqslant \eta_k \\ \sum_{j=0}^{p} \beta_{j,k} \left(t/n - \eta_k/n\right)^j & \eta_k < t < \eta_{k+1} \end{cases}$$

- The **most prominent** change in derivative at the k-th change point is uniquely defined as

$$p_k^* = \underset{0 \leqslant j \leqslant p}{\text{arg-max}} \left\{ \left|\Delta_{j,k}\right| \left(\frac{\delta_k}{n}\right)^j \right\}$$

- Assume that for each $k = 1, \ldots, N$ the following holds

$$\delta_k > C_1 \left(\log(n) \vee n^{\frac{2p_k^*}{2p_k^*+1}} \left(\frac{\log(n)}{\Delta_{p_k^*}^2}\right)^{\frac{1}{2p_k^*+1}} \right)$$

Detection Guarantees

- **(A1) median control:** There is a function $H : [0,1] \to \mathbb{R}^+$ with $H(u) = \infty$ if $u > 1$ such that (i) $\mathbb{P}(\zeta_t \leqslant H(u)) \wedge \mathbb{P}(\zeta_t \geqslant -H(u)) \geqslant (1+u)/2$ and (ii) $\limsup_{u \downarrow 0} H(u)/u < \infty$
- **(A2) derivative control:** for any sequence $(\rho_{k,n})_{n \geqslant 1}$ with $\rho_{k,n} \leqslant (\delta_k/n)$ it holds that $(|\alpha_{j,k}| \vee |\beta_{j,k}|)\, \rho_{k,n}^{j} \leqslant C_f \Delta_{p_k^*} \rho_{k,n}^{p_k^*}$ for each k and all $j \geqslant 1$
- **(A3) extrema control:** let $f_\circ$ have K local extrema, excluding those induced by a change point, at integer locations $0 < \tau_1 < \cdots < \tau_K < n$, it holds that $|\tau_j - \eta_k| > \delta_k$ for each j, k

Detection Guarantees

- **(A1) median control:** There is a function $H : [0,1] \to \mathbb{R}^+$ with $H(u) = \infty$ if $u > 1$ such that (i) $\mathbb{P}(\zeta_t \leqslant H(u)) \wedge \mathbb{P}(\zeta_t \geqslant -H(u)) \geqslant (1+u)/2$ and (ii) $\limsup_{u \downarrow 0} H(u)/u < \infty$
- **(A2) derivative control:** for any sequence $(\rho_{k,n})_{n \geqslant 1}$ with $\rho_{k,n} \leqslant (\delta_k/n)$ it holds that $(|\alpha_{j,k}| \vee |\beta_{j,k}|)\, \rho_{k,n}^{j} \leqslant C_f \Delta_{p_k^*} \rho_{k,n}^{p_k^*}$ for each k and all $j \geqslant 1$
- **(A3) extrema control:** let $f_\circ$ have K local extrema, excluding those induced by a change point, at integer locations $0 < \tau_1 < \cdots < \tau_K < n$, it holds that $|\tau_j - \eta_k| > \delta_k$ for each j, k

Detection Guarantees

- **(A1) median control:** There is a function $H : [0,1] \to \mathbb{R}^+$ with $H(u) = \infty$ if $u > 1$ such that (i) $\mathbb{P}(\zeta_t \leqslant H(u)) \wedge \mathbb{P}(\zeta_t \geqslant -H(u)) \geqslant (1+u)/2$ and (ii) $\limsup_{u \downarrow 0} H(u)/u < \infty$
- **(A2) derivative control:** for any sequence $(\rho_{k,n})_{n \geqslant 1}$ with $\rho_{k,n} \leqslant (\delta_k/n)$ it holds that $(|\alpha_{j,k}| \vee |\beta_{j,k}|)\, \rho_{k,n}^{j} \leqslant C_f \Delta_{p_k^*} \rho_{k,n}^{p_k^*}$ for each k and all $j \geqslant 1$
- **(A3) extrema control:** let $f_\circ$ have K local extrema, excluding those induced by a change point, at integer locations $0 < \tau_1 < \cdots < \tau_K < n$, it holds that $|\tau_j - \eta_k| > \delta_k$ for each j, k

Theorem (change point detection guarantees)

Let $\left\{\hat{I}_1, \dots, \hat{I}_{\hat{N}}\right\}$ *be intervals returned by the generic algorithm with local tests* $\mathcal{T}_{\mathcal{G}_n^c}$ *on the complete grid* $\mathcal{G}_n^c$ *using threshold* λ *defined below for some* $\alpha \in (0,1)$

$$\lambda = \sqrt{2\log\left(n\log^{-\frac{1}{2}} n\right)} + \log\left(\frac{1}{\log\left(\frac{1}{1-\alpha}\right)/2H}\right) \Big/ \sqrt{2\log\left(n\log^{-\frac{1}{2}} n\right)}$$

Then with probability $1-\alpha+o(1)$ *the following events occurs simultaneously*

$$E_1^* = \left\{\hat{N} = N\right\}$$

$$E_2^* = \left\{\forall k = 1, \dots, N \ \hat{I}_k \cap \Theta = \eta_k\right\}$$

$$E_3^* = \left\{\left|\hat{I}_k\right| \leqslant C_2\left(\log(n) \vee n^{\frac{2p_k^*}{2p_k^*+1}}\left(\frac{\log(n)}{\Delta_{p_k^*}^2}\right)^{\frac{1}{2p_k^*+1}}\right) \middle| 1 \leqslant k \leqslant N\right\}$$

Here C_1, C_2 *satisfy* $C_1 > 2(p+1)C_2$

Simulation Studies

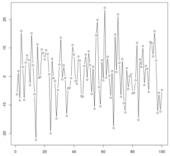
Comparison with Existing Procedures

- Robust algorithms: **RNSP** (Fryzlewicz 2021), **NSP-SN** (Fryzlewicz 2020), **MQS** Jula Vanegas et al. (2021), and **HSMUCE** Pein et al. (2017)
- Non-robust algorithms: **NSP** (Fryzlewicz 2020), **SMUCE** (Frick et al. 2014), **MOSUM** (Cho & Kirch 2022)

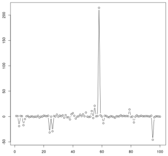
Comparison with Existing Procedures

- Robust algorithms: **RNSP** (Fryzlewicz 2021), **NSP-SN** (Fryzlewicz 2020), **MQS** Jula Vanegas et al. (2021), and **HSMUCE** Pein et al. (2017)
- Non-robust algorithms: **NSP** (Fryzlewicz 2020), **SMUCE** (Frick et al. 2014), **MOSUM** (Cho & Kirch 2022)

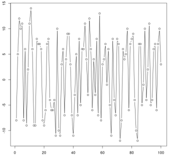
Coverage: robust algorithms



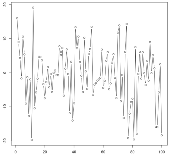
(a) Gaussian



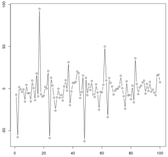
(b) Cauchy



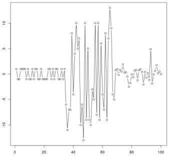
(c) Sym. Poisson



(d) GARCH



(e) TV-Variance



(f) Mix

Coverage: robust algorithms

Table 1: number of times no intervals of significance are returned after applying each method to a vector of pure noise 100 times, and select tuning parameters so that each method outputs intervals with nominal 90% coverage

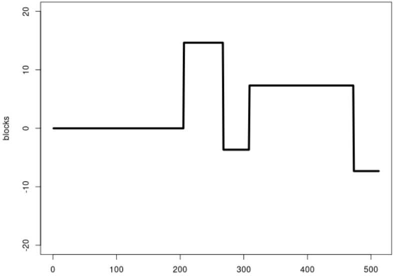
	Our method	NSP-SN	RNSP	H-SMUCE	MQS
Gaussian	99	100	99	100	63
Cauchy	100	61	100	100	66
Sym. Poisson	97	100	99	0	65
GARCH	98	100	99	99	69
TV-Variance	98	100	99	100	65
Mix	94	90	98	0	48

Coverage: non-robust algorithms

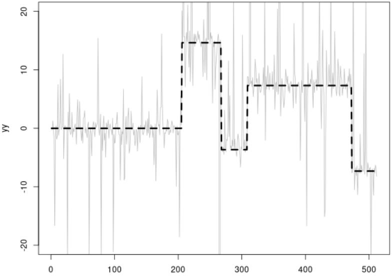
Table 2: proportion of times, over 500 replications, one or more interval of significance are spuriously returned when each algorithm is applied to a vector of standard Gaussian noise having length $n = 500$

	degree 0			degree 1			degree 2		
Significance level (α)	0.100	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
Our method ($a = 2, A = 0$)	0.127	0.060	0.007	0.072	0.031	0.010	0.092	0.044	0.009
Our method ($a = \sqrt{2}, A = 0$)	0.100	0.055	0.006	0.074	0.044	0.006	0.062	0.031	0.008
NSP (asymptotic)	0.019	0.007	0.000	0.010	0.004	0.001	0.010	0.002	0.000
NSP (simulated)	0.094	0.047	0.013	0.081	0.050	0.010	0.102	0.045	0.015
MOSUM (uniscale)	0.021	0.021	0.021	–	–	–	–	–	–
MOSUM (multiscale)	0.106	0.106	0.106	–	–	–	–	–	–
SMUCE	0.023	0.008	0.001	–	–	–	–	–	–

Detection Power: robust algorithms



(a) the `blocks` signal



(b) . . . contaminated with Cauchy noise

Detection Power: robust algorithms

Table 3: length of intervals returned (length), number intervals containing at least one change point (no. genuine), proportion of intervals containing at least one change point (prop. genuine), and whether all intervals contain at least once change point location (coverage) when non-robust algorithms are applied to the blocks signal with different noise types over 100 replications

		Gaussian	Cauchy	Sym. Poisson	GARCH	TV-Variance	Mix
Our method	length	63.58	30.72	52.69	47.16	76.01	44.34
	no. genuine	2.98	4	3.11	3.67	2.44	3.94
	prop. genuine	1	1	1	1	1	0.99
	coverage	1	1	0.99	1	1	0.95
SNP-SN	length	119.13	127.61	79.02	97.03	186.36	31.22
	no. genuine	1.88	1.61	2.51	2.32	1.05	2.97
	prop. genuine	1	0.95	0.86	1	0.9	0.41
	coverage	1	0.92	0.65	1	0.9	0.05
RNSP	length	71.71	33.07	57.47	50.22	79.34	47.37
	no. genuine	2.78	4	3.04	3.53	2.23	3.92
	prop. genuine	1	1	1	1	1	0.99
	coverage	1	1	1	1	1	0.97
HSMUCE	length	84	27.72	40.13	42.79	146.6	8.39
	no. genuine	1.65	1.94	0.03	1.55	1.51	0
	prop. genuine	0.84	0.97	0.01	0.78	0.95	0
	coverage	0.7	0.95	0.01	0.56	0.92	0
MQS	length	71.23	34.8	70.15	53.69	95.81	44.27
	no. genuine	3.5	3.98	3.31	3.89	2.87	3.91
	prop. genuine	0.96	0.98	0.96	0.95	0.95	0.95
	coverage	0.84	0.91	0.85	0.81	0.83	0.82

Detection Power: non-robust algorithms

Table 4: length of intervals returned (length), number intervals containing at least one change point (genuine), and whether all intervals contain at least once change point location (coverage) when non-robust algorithms are applied to three test signal contaminated with Gaussian noise over 500 replications

	blocks ($\sigma = 15$)			waves ($\sigma = 12$)			poly-mix ($\sigma = 1$)		
	length	genuine	coverage	length	genuine	coverage	length	genuine	coverage
Our method ($a = 2, A = 0$)	58.489	2.218	0.962	241.083	1.858	0.976	130.236	2.116	0.972
Our method ($a = \sqrt{2}, A = 0$)	58.126	2.250	0.964	215.339	1.692	0.972	102.041	2.100	0.986
NSP (asymptotic)	122.020	1.420	0.998	254.595	1.444	1.000	116.417	2.006	1.000
SNP (simulated)	97.704	1.954	0.994	213.218	1.718	0.996	98.628	2.116	0.992
MOSUM (uniscale)	4.762	0.362	0.770	–	–	–	–	–	–
MOSUM (multiscale)	21.369	2.400	0.620	–	–	–	–	–	–
SMUCE	72.915	1.952	0.594	–	–	–	–	–	–

Real Data Examples

Ozone Concentration in Los Angeles

- We analyse daily Ozone concentration (maximum of one hour averages) in the Los Angeles basin during 1976
- We estimate change points using popular algorithms for recovering piecewise linear signals: **NOT** (Baranowski et al. 2019) **ID** (Anastasiou & Fryzlewicz 2019), **SC** (Bai & Perron 2003), **FKS** (Spiriti et al. 2013), **MARS** (Friedman 1991), **TF** (Tibshirani et al. 2014), and **CPOP** (Maidstone et al. 2017)

Ozone Concentration in Los Angeles

- We analyse daily Ozone concentration (maximum of one hour averages) in the Los Angeles basin during 1976
- We estimate change points using popular algorithms for recovering piecewise linear signals: **NOT** (Baranowski et al. 2019) **ID** (Anastasiou & Fryzlewicz 2019), **SC** (Bai & Perron 2003), **FKS** (Spiriti et al. 2013), **MARS** (Friedman 1991), **TF** (Tibshirani et al. 2014), and **CPOP** (Maidstone et al. 2017)

Ozone Concentration in Los Angeles

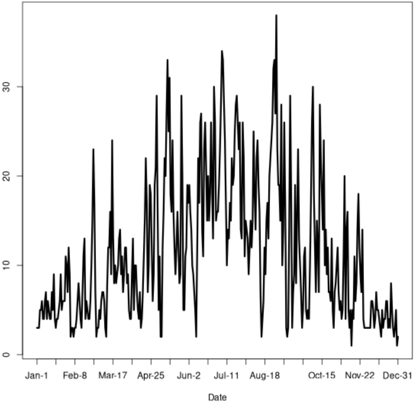


Figure 6: daily Ozone concentrations; it is well documented that Ozone concentrations in the Northern hemisphere follow a pronounced yearly cycle with the maximum occurring towards the middle of the year (Monks 2000)

Ozone Concentration in Los Angeles

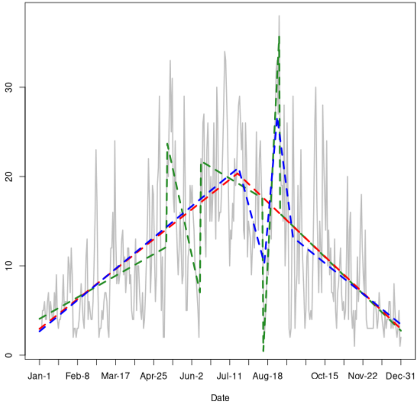


Figure 6: estimated signals using NOT-cont (- - -), NOT (- - -), and ID (- - -)

Ozone Concentration in Los Angeles

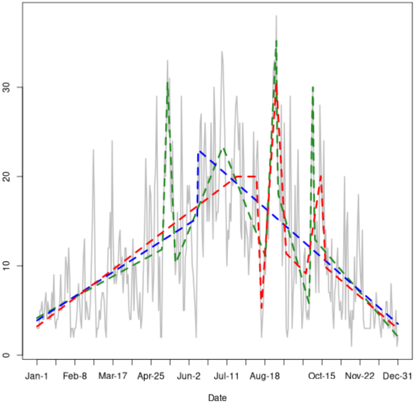


Figure 6: estimated signals using MARS (- - -), FKS (- - -), and SC (- - -)

Ozone Concentration in Los Angeles

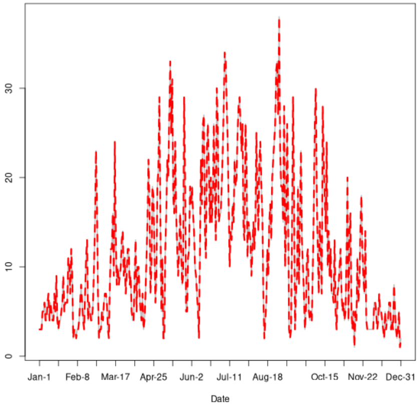


Figure 6: estimated signals using TF (- - -) with penalty term chosen via cross-validation using `cv.trendfilter` function in the `genlasso` package

Ozone Concentration in Los Angeles

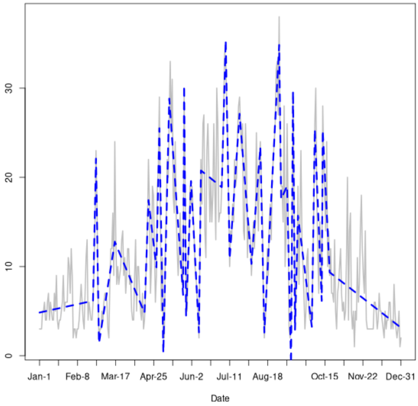


Figure 6: estimated signals using CPOP (- - -) with default $2\log(n)$ penalty and noise scale estimated using MAD

Ozone Concentration in Los Angeles

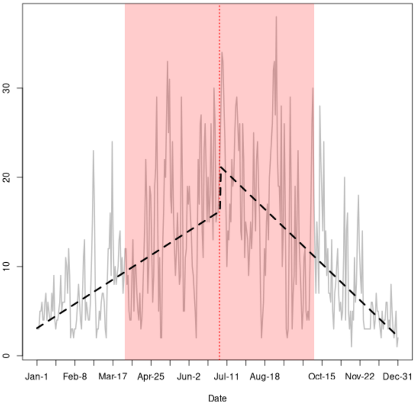


Figure 6: estimated 90% interval of significance obtained with our robust algorithm together with signal fitted to the left and right of the interval's midpoint

Thank you!