Robust Inference for Change Points in Piecewise Polynomials of General Degrees

Shakeel Gavioli-Akilagun   Piotr Fryzlewicz

London School of Economics
Department of Statistics

Problem Statement

We observe data $\mathrm{y}_{1:n} = (y_1, \ldots, y_n)'$ generated by the following ‘signal + noise’ model

$$y_t = f_t^\circ + \zeta_t, \qquad t = 1, \ldots, n$$

- ζ's sign symmetric and independent, otherwise arbitrary
- f° piecewise polynomial with known fixed degree
- **Goal** inferential statements about unknown change point locations

Our Method: piecewise-constant signal

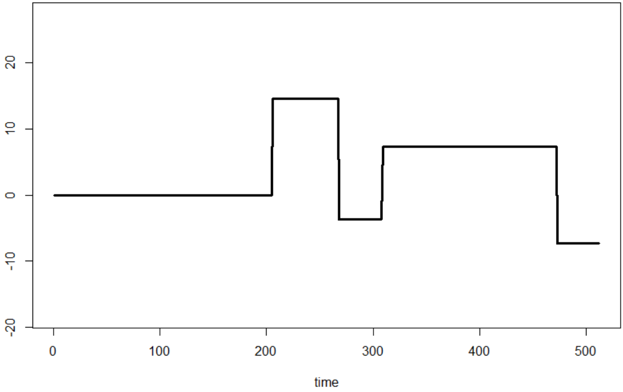


Figure 1: first 512 values of the `blocks` signal

Our Method: piecewise-constant signal

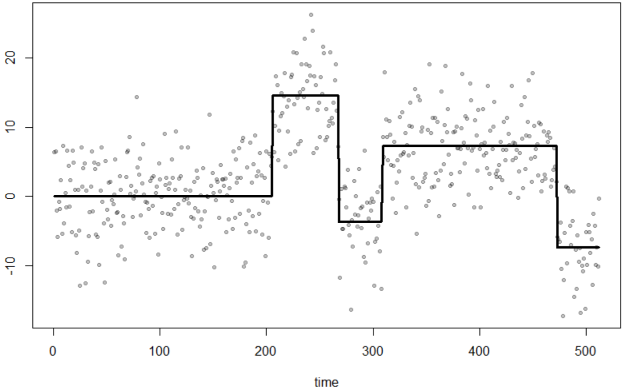


Figure 1: first 512 values of the `blocks` signal with additive noise

Our Method: piecewise-constant signal

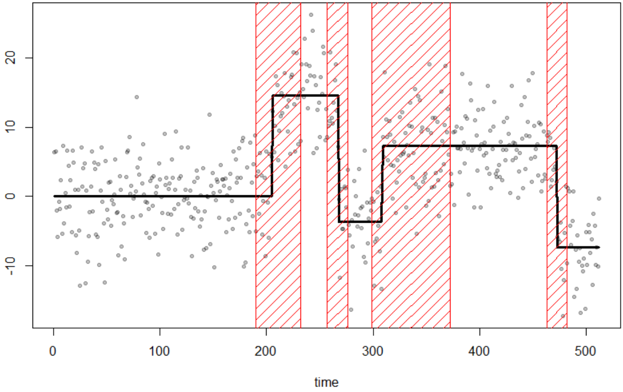


Figure 1: intervals of significance returned by our procedure

Our Method: piecewise-linear signal

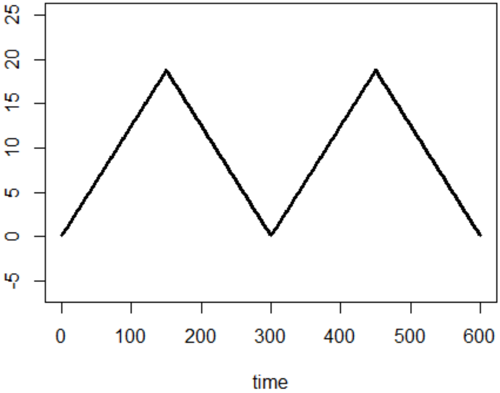


Figure 2: first 600 values of the `waves` signal

Our Method: piecewise-linear signal

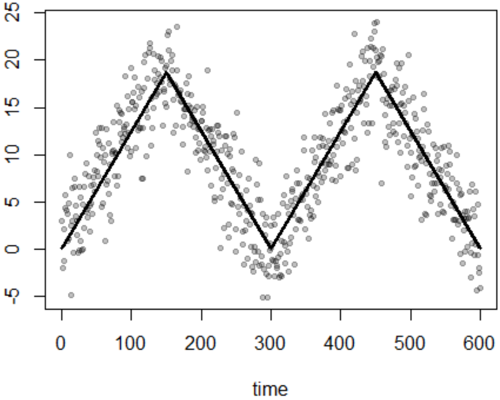


Figure 2: first 600 values of the `waves` signal with additive noise

Our Method: piecewise-linear signal

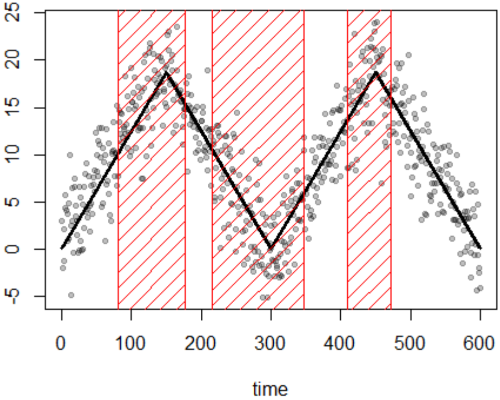


Figure 2: intervals of significance returned by our procedure

Background

Current approaches to change point inference

- **Post selection inference:** Jewell et al. (2019), Hyun et al. (2021), Valiollahi Mehrizi (2021)
- **Simultaneous inference and selection:** Frick et al. (2014), Pein et al. (2015), Jula Vanegas et al. (2021)
- **Post inference selection:** Fryzlewicz (2020), Fryzlewicz (2021)

Our motivations was to...

- Generalize **beyond piecewise constant** signals
- Allow for **arbitrary error distributions**
- Produce a fast - better than $\mathcal{O}(n^2)$ - algorithm

Background

Current approaches to change point inference

- **Post selection inference:** Jewell et al. (2019), Hyun et al. (2021), Valiollahi Mehrizi (2021)
- **Simultaneous inference and selection:** Frick et al. (2014), Pein et al. (2015), Jula Vanegas et al. (2021)
- **Post inference selection:** Fryzlewicz (2020), Fryzlewicz (2021)

Our motivations was to...

- Generalize **beyond piecewise constant** signals
- Allow for **arbitrary error distributions**
- Produce a fast - better than $\mathcal{O}(n^2)$ - algorithm

Recipe for Post Inference Selection (Fryzlewicz 2020)

1. Take a collection of local test $T_{s:e} : (y_s, \ldots, y_e) \mapsto \{0, 1\}$ for

$$\begin{aligned} \mathcal{H}_0^{s:e} &: \text{ on } [s, e] \quad f^\circ \in \mathfrak{M} \\ \mathcal{H}_1^{s:e} &: \text{ on } [s, e] \quad f^\circ \notin \mathfrak{M} \end{aligned}$$

2. Restrict attention to tests with strong false discovery control. For any collection of null sub-intervals $\mathcal{E}$

$$\mathbb{P}\left(\bigcup_{(s,e)\in\mathcal{E}} \{T_{s:e} = 1\} \right) \leqslant \alpha$$

3. Apply tests to a sorted grid of sub-intervals; every time a non-null interval is detected explore sub interval for narrowest region of significance then recur to the left / right of detected interval

Recipe for Post Inference Selection (Fryzlewicz 2020)

1. Take a collection of local test $T_{s:e} : (y_s, \ldots, y_e) \mapsto \{0, 1\}$ for

$$\begin{aligned} \mathcal{H}_0^{s:e} :\ \text{on } [s, e] \quad & f^\circ \in \mathfrak{M} \\ \mathcal{H}_1^{s:e} :\ \text{on } [s, e] \quad & f^\circ \notin \mathfrak{M} \end{aligned}$$

2. Restrict attention to tests with strong false discovery control. For any collection of null sub-intervals $\mathcal{E}$

$$\mathbb{P}\left(\bigcup_{(s,e)\in\mathcal{E}} \{T_{s:e} = 1\} \right) \leqslant \alpha$$

3. Apply tests to a sorted grid of sub-intervals; every time a non-null interval is detected explore sub interval for narrowest region of significance then recur to the left / right of detected interval

Recipe for Post Inference Selection (Fryzlewicz 2020)

1. Take a collection of local test $T_{s:e} : (y_s, \ldots, y_e) \mapsto \{0, 1\}$ for

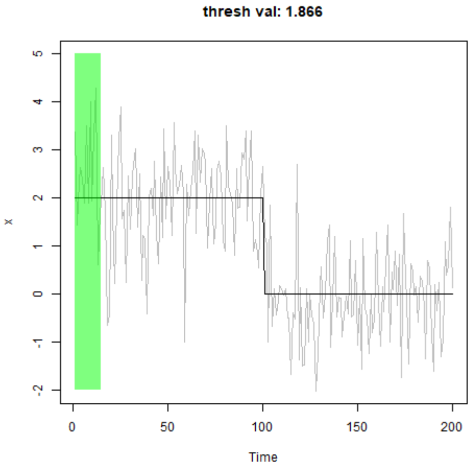
$$\begin{aligned} \mathcal{H}_0^{s:e} &: \text{ on } [s, e] \quad f^\circ \in \mathfrak{M} \\ \mathcal{H}_1^{s:e} &: \text{ on } [s, e] \quad f^\circ \notin \mathfrak{M} \end{aligned}$$

2. Restrict attention to tests with strong false discovery control. For any collection of null sub-intervals $\mathcal{E}$

$$\mathbb{P}\left(\bigcup_{(s,e)\in\mathcal{E}} \{T_{s:e} = 1\} \right) \leqslant \alpha$$

3. Apply tests to a sorted grid of sub-intervals; every time a non-null interval is detected explore sub interval for narrowest region of significance then recur to the left / right of detected interval

Recipe for Post Inference Selection (Fryzlewicz 2020)



New Generic Recipe for Inference on Piecewise Polynomials

1. Given a candidate f define $\hat{\zeta}_t^f = y_t - f_t$; introduce a second local test for whether empirical residuals "look like noise" at some level α

$$\psi_{s:e}^{f,\alpha} : \left(\hat{\zeta}_s^f, \ldots, \hat{\zeta}_e^f\right) \mapsto \{0, 1\}$$

2. Then a local $(1-\alpha)$ level confidence set for f° can be obtained by inverting ψ on any sub-interval as follows

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

3. Finally, construct a local test by intersecting the confidence set with polynomials of degree p

$$T_{s:e} = 1\left\{\mathcal{C}\left(y_{s:e}, \alpha\right) \cap \texttt{poly}(p) \neq \varnothing\right\}$$

New Generic Recipe for Inference on Piecewise Polynomials

1. Given a candidate f define $\hat{\zeta}_t^f = y_t - f_t$; introduce a second local test for whether empirical residuals "look like noise" at some level α

$$\psi_{s:e}^{f,\alpha} : \left(\hat{\zeta}_s^f, \ldots, \hat{\zeta}_e^f\right) \mapsto \{0, 1\}$$

2. Then a local $(1 - \alpha)$ level confidence set for f° can be obtained by inverting ψ on any sub-interval as follows

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

3. Finally, construct a local test by intersecting the confidence set with polynomials of degree p

$$T_{s:e} = 1\left\{\mathcal{C}\left(y_{s:e}, \alpha\right) \cap \texttt{poly}(p) \neq \emptyset\right\}$$

New Generic Recipe for Inference on Piecewise Polynomials

1. Given a candidate f define $\hat{\zeta}_t^f = y_t - f_t$; introduce a second local test for whether empirical residuals "look like noise" at some level α

$$\psi_{s:e}^{f,\alpha} : \left(\hat{\zeta}_s^f, \ldots, \hat{\zeta}_e^f\right) \mapsto \{0, 1\}$$

2. Then a local $(1 - \alpha)$ level confidence set for f° can be obtained by inverting ψ on any sub-interval as follows

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

3. Finally, construct a local test by intersecting the confidence set with polynomials of degree p

$$T_{s:e} = 1\left\{\mathcal{C}\left(y_{s:e}, \alpha\right) \cap \texttt{poly}(p) \neq \varnothing\right\}$$

The Runs Test for Pure Noise

We may use the maximum run of the empirical residuals on a sub-interval to decide whether f represents a good local fit

$$\psi_{s:e}^{f} = 1\left\{\mathrm{max\_run}(\mathrm{y}_{\mathrm{s:e}} - \mathrm{f}_{\mathrm{s:e}}) > \lambda\right\}$$

*The maximum run of either sign is given by

```
0.24 -0.73 0.79 -1.57 -0.25 -1.36 0.60 0.92 0.00 1.13 0.17
```

The Runs Test for Pure Noise

We may use the maximum run of the empirical residuals on a sub-interval to decide whether f represents a good local fit

$$\psi^{f}_{s:e} = 1\left\{\mathrm{max\_run}(\mathrm{y}_{\mathrm{s:e}} - \mathrm{f}_{\mathrm{s:e}}) > \lambda\right\}$$

*The maximum run of either sign is given by

`0.24 -0.73 0.79 -1.57 -0.25 -1.36 0.60 0.92 0.00 1.13 0.17`

The Runs Test for Pure Noise

We may use the maximum run of the empirical residuals on a sub-interval to decide whether f represents a good local fit

$$\psi^{f}_{s:e} = 1\left\{\mathrm{max\_run}(\mathrm{y}_{\mathrm{s:e}} - \mathrm{f}_{\mathrm{s:e}}) > \lambda\right\}$$

*The maximum run of either sign is given by

$$\max\left\{\omega \in \mathbb{N} : \omega = \left(\max_{\substack{s\leqslant j\leqslant k\leqslant e \\ k-j+1=\omega}} \sum_{i=j}^{k} 1_{\{\hat{\zeta}^f_i \geqslant 0\}} \bigwedge \max_{\substack{s\leqslant j\leqslant k\leqslant e \\ k-j+1=M}} \sum_{i=j}^{k} 1_{\{\hat{\zeta}^f_i < 0\}}\right)\right\}$$

Inverting the Runs Test

- No need compute the entire confidence set! Enough to compute the point-wise upper and lower bounds then check if a polynomial of the desired degree may pass between them

$$
\begin{aligned}
l_t &= \inf \left\{ f_t : f \in C\left(\mathrm{y}_{\mathrm{s:e}}, \alpha\right) \right\} \\
u_t &= \sup \left\{ f_t : f \in C\left(\mathrm{y}_{\mathrm{s:e}}, \alpha\right) \right\}
\end{aligned}
$$

- From **Davies (1995), Davies & Kovac (2001)** we have that

$$
u_t = \begin{cases} \min \left\{ u_{t-1}, \max \left\{ y_i : t - \lambda \leqslant i \leqslant t \right\} \right\} & f^\circ \text{ non-increasing} \\ \min \left\{ u_{t+1}, \max \left\{ y_i : t \leqslant i \leqslant t + \lambda \right\} \right\} & f^\circ \text{ non-decreasing} \end{cases}
$$

$$
l_t = \begin{cases} \max \left\{ l_{t+1}, \min \left\{ y_i : t \leqslant i \leqslant t + \lambda \right\} \right\} & f^\circ \text{ non-increasing} \\ \max \left\{ l_{t-1}, \min \left\{ y_i : t - \lambda \leqslant i \leqslant t \right\} \right\} & f^\circ \text{ non-decreasing} \end{cases}
$$

Inverting the Runs Test

- No need compute the entire confidence set! Enough to compute the point-wise upper and lower bounds then check if a polynomial of the desired degree may pass between them

$$
\begin{aligned}
l_t &= \inf \left\{ f_t : f \in C\left(\mathrm{y}_{\mathrm{s:e}}, \alpha\right)\right\} \\
u_t &= \sup \left\{ f_t : f \in C\left(\mathrm{y}_{\mathrm{s:e}}, \alpha\right)\right\}
\end{aligned}
$$

- From **Davies (1995), Davies & Kovac (2001)** we have that

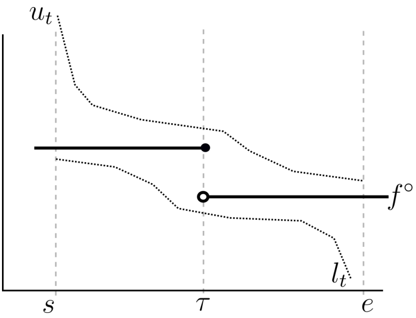
$$
u_t = \begin{cases} \min\left\{u_{t-1}, \max\left\{y_i : t-\lambda \leqslant i \leqslant t\right\}\right\} & f^\circ \text{ non-increasing} \\ \min\left\{u_{t+1}, \max\left\{y_i : t \leqslant i \leqslant t+\lambda\right\}\right\} & f^\circ \text{ non-decreasing} \end{cases}
$$
$$
l_t = \begin{cases} \max\left\{l_{t+1}, \min\left\{y_i : t \leqslant i \leqslant t+\lambda\right\}\right\} & f^\circ \text{ non-increasing} \\ \max\left\{l_{t-1}, \min\left\{y_i : t-\lambda \leqslant i \leqslant t\right\}\right\} & f^\circ \text{ non-decreasing} \end{cases}
$$

Practical Testing in the Piecewise Constant Setting

Data locally agrees with f° being a degree zero polynomial if

$$\min_{s \leqslant t \leqslant e} u_t > \max_{s \leqslant t \leqslant e} l_t$$

Can be tested in $\mathcal{O}(1)$ time

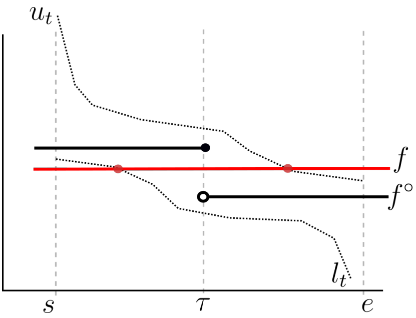


Practical Testing in the Piecewise Constant Setting

Data locally agrees with f° being a degree zero polynomial if

$$\min_{s \leqslant t \leqslant e} u_t > \max_{s \leqslant t \leqslant e} l_t$$

Can be tested in $\mathcal{O}(1)$ time

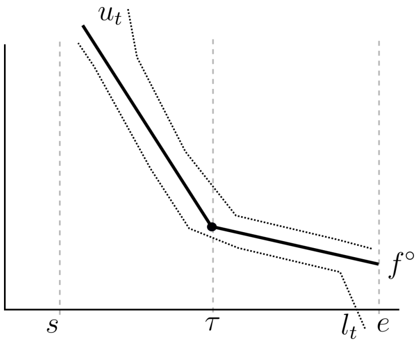


Practical Testing in the Piecewise Linear Setting

Data locally agrees with f° being a polynomial of degree one if

$$\texttt{Hull}\left\{(s, u_s), \ldots, (e, u_e)\right\} \cap \texttt{Hull}\left\{(s, l_s), \ldots, (e, l_e)\right\} = \varnothing$$

Can be tested in $\mathcal{O}(\log n)$ time (Chazelle & Dobkin 1987)

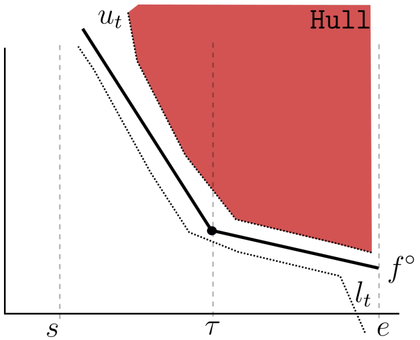


Practical Testing in the Piecewise Linear Setting

Data locally agrees with f° being a polynomial of degree one if

$$\texttt{Hull}\{(s,u_s),\ldots,(e,u_e)\} \cap \texttt{Hull}\{(s,l_s),\ldots,(e,l_e)\} = \varnothing$$

Can be tested in $\mathcal{O}(\log n)$ time (Chazelle & Dobkin 1987)

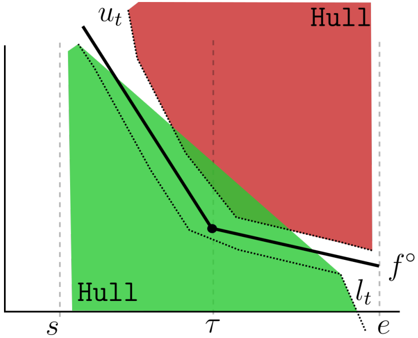


Practical Testing in the Piecewise Linear Setting

Data locally agrees with f° being a polynomial of degree one if

$$\texttt{Hull}\left\{(s, u_s), \ldots, (e, u_e)\right\} \cap \texttt{Hull}\left\{(s, l_s), \ldots, (e, l_e)\right\} = \varnothing$$

Can be tested in $\mathcal{O}(\log n)$ time (Chazelle & Dobkin 1987)



Generic Piecewise Polynomial Setting

Let $f_0, \ldots, f_N$ be polynomials in t of fixed and known degrees and with $f_j \neq f_{j-1}$. Let $\Theta = (\tau_1, \ldots, \tau_N)$ be the set of change point locations, then the signal component is

$$f_t^\circ = \sum_{j=0}^{N} f_{t,j}^\circ \mathbb{1}_{\{t \in [\tau_{j-1}, \tau_j)\}}$$

(A1) $\mathbb{P}(\zeta_t > 0) = \mathbb{P}(\zeta_t \leqslant 0) = \frac{1}{2}$ for all $t = 1, 2, \ldots$

(A2) $\mathbb{P}(\zeta_t > 0 \mid \zeta_s) = \mathbb{P}(\zeta_t > 0)$ for all $t \neq s$

Exact Coverage Guarantees

Let A1 - A2 hold and $\left\{\hat{I}, \ldots, \hat{I}_{\hat{N}}\right\}$ be the set of intervals returned by our algorithm in any generic piecewise polynomial setting with a threshold $\lambda = \lambda_n$ chosen as shown below

$$\lambda_n = \log_2 (n-1) + \log_2 \left(\frac{1}{\log\left(\frac{1}{1-\alpha}\right)} \right)$$

Then the following guarantee holds

$$\mathbb{P}\left(\bigcup_{j=1}^{\hat{N}} \left\{ \hat{I}_j \cap \Theta = \varnothing \right\} \right) \leqslant \alpha + \frac{0.3}{n-2}$$

The Piecewise Constant Setting

For some constants $\theta_0, \ldots, \theta_N$ with $\theta_j \neq \theta_{j-1}$ the signal component is

$$f_t^\circ = \sum_{j=0}^{N} \theta_j 1_{\{t \in [\tau_{j-1}, \tau_j)\}}$$

(A3) Minimum change point spacing is of the order $\mathcal{O}(n)$

(A4) There is a random variable ξ such that $\mathbb{P}(|\zeta_t| > x) \leqslant \mathbb{P}(|\xi| > x)$ for any $x > 0$ and for all $t = 1, 2, \ldots$

(A5) $\exists \beta \in (0, \frac{1}{2})$ such that if $x \in [0, 1]$ then $\mathbb{P}(|\xi| > x) \leqslant \frac{1}{2} - \beta x$

Interval Sub-Sampling (Fryzlewicz 2020)

1. Chooses an integer M which determines how many local test may be applied. Define an integer

$$k^* = \left\lfloor \left(1 + \sqrt{1 + 8M}\right) /2 \right\rfloor$$

2. Define a shift term

$$\delta_M = \left\lfloor \frac{n-1}{k^* - 1} \right\rfloor$$

3. Draw all sub-intervals in $\{1, \ldots, k^*\}$ and re-map to $\{1, \ldots, n\}$ according to the rule

$$(s, e) \mapsto ((s-1) \times \delta_M + 1, (e-1) \times \delta_M + 1)$$

Interval Sub-Sampling (Fryzlewicz 2020)

1. Chooses an integer M which determines how many local test may be applied. Define an integer

$$k^* = \left\lfloor \left(1 + \sqrt{1 + 8M}\right) / 2 \right\rfloor$$

2. Define a shift term

$$\delta_M = \left\lfloor \frac{n-1}{k^* - 1} \right\rfloor$$

3. Draw all sub-intervals in $\{1, \ldots, k^*\}$ and re-map to $\{1, \ldots, n\}$ according to the rule

$$(s, e) \mapsto ((s-1) \times \delta_M + 1, (e-1) \times \delta_M + 1)$$

Interval Sub-Sampling (Fryzlewicz 2020)

1. Chooses an integer M which determines how many local test may be applied. Define an integer

$$k^* = \left\lfloor \left(1 + \sqrt{1 + 8M}\right) / 2 \right\rfloor$$

2. Define a shift term

$$\delta_M = \left\lfloor \frac{n - 1}{k^* - 1} \right\rfloor$$

3. Draw all sub-intervals in $\{1, \ldots, k^*\}$ and re-map to $\{1, \ldots, n\}$ according to the rule

$$(s, e) \mapsto ((s - 1) \times \delta_M + 1, (e - 1) \times \delta_M + 1)$$

Change Point Detection Guarantees

Grant assumption A1 - A5 and let $\left\{\hat{I}_1, \ldots, \hat{I}_{\hat{N}}\right\}$ be the intervals returned by our algorithm in the piecewise constant setting. Define

$$\mathcal{A}_n = \left\{\hat{N} = N, \forall j = 1, \ldots, N \; \exists k \text{ s.t. } \left\{\hat{I}_j \cap \Theta\right\} = \tau_k\right\}$$

If the number of local tests is chosen as $M = M_n$ such that $M_n \to \infty$ as n diverges then writing $\Delta = \min_j \left|\theta_j - \theta_{j-1}\right|$ the following guarantee holds

$$\mathbb{P}\left(\mathcal{A}_n\right) \geqslant 1 - \left(\alpha + \frac{0.3}{n-2} + C_N \exp\left(-C_{M,\beta,\alpha} \times n^{1-\log_2\left(\frac{2}{1+\beta\Delta}\right)}\right)\right)$$

w.p.a.-$(1-\alpha)$ our algorithm detects all change points and isolates them to their own interval as long as the minimum jump size satisfies

$$\Delta > \left(\beta \log_2 n\right)^{-1}$$

Change Point Detection Guarantees

Grant assumption A1 - A5 and let $\left\{\hat{I}_1, \ldots, \hat{I}_{\hat{N}}\right\}$ be the intervals returned by our algorithm in the piecewise constant setting. Define

$$\mathcal{A}_n = \left\{\hat{N} = N, \forall j = 1, \ldots, N \; \exists k \text{ s.t. } \left\{\hat{I}_j \cap \Theta\right\} = \tau_k\right\}$$

If the number of local tests is chosen as $M = M_n$ such that $M_n \to \infty$ as n diverges then writing $\Delta = \min_j |\theta_j - \theta_{j-1}|$ the following guarantee holds

$$\mathbb{P}\left(\mathcal{A}_n\right) \geqslant 1 - \left(\alpha + \frac{0.3}{n-2} + C_N \exp\left(-C_{M,\beta,\alpha} \times n^{1-\log_2\left(\frac{2}{1+\beta\Delta}\right)}\right)\right)$$

w.p.a.-$(1-\alpha)$ our algorithm detects all change points and isolates them to their own interval as long as the minimum jump size satisfies

$$\Delta > \left(\beta \log_2 n\right)^{-1}$$

Alternative Approaches from the Literature

- **Fryzlewicz (2021)** [MR] obtains simultaneous confidence intervals for $\Theta = (\tau_1, \ldots, \tau_N)$ using local tests based

$$T_{s:e} = 1\left\{\min_{f\in\mathbb{R}} \|\mathrm{sgn}(y_{s:e} - f1_{s:e})\|_{MR} > \lambda\right\}$$

- **Jula Vanegas et al. (2021)** [MQS] produces a confidence set for f° by inverting local tests

$$\psi^f_{s:e} = 1\left\{\sqrt{2\mathcal{L}\left(y_{s:e}, f\right)} - \sqrt{2\log\left(\frac{en}{e-s+1}\right)} > q_n\left(\alpha\right)\right\}$$

Execution Speed

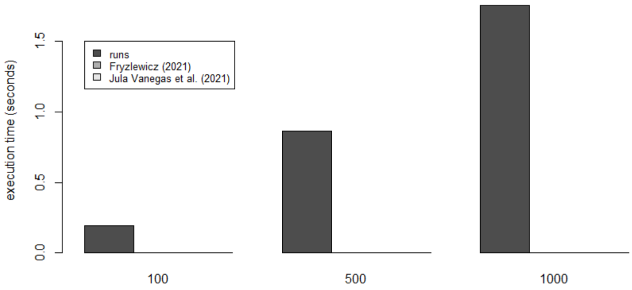


Figure 3: execution times on Gaussian noise, using 2.59 GHz Intel Core i7 CPU

Execution Speed

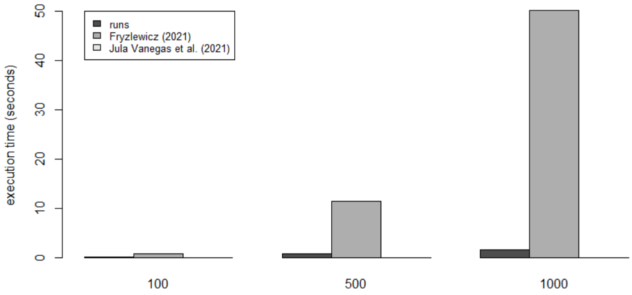


Figure 3: execution times on Gaussian noise, using 2.59 GHz Intel Core i7 CPU

Execution Speed

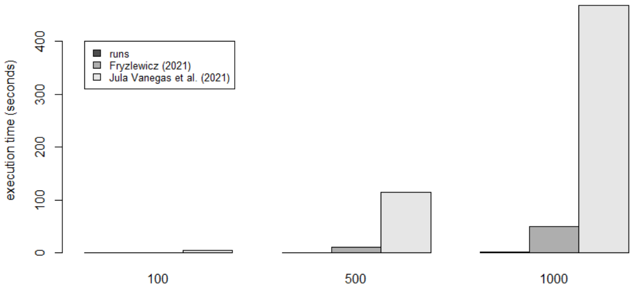


Figure 3: execution times on Gaussian noise, using 2.59 GHz Intel Core i7 CPU

Execution Speed

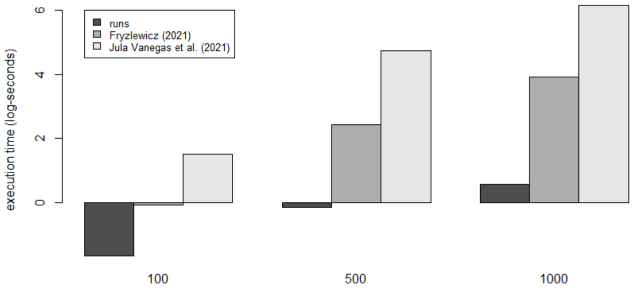
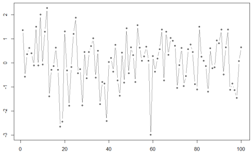
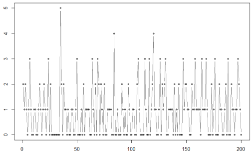


Figure 3: execution times on Gaussian noise, using 2.59 GHz Intel Core i7 CPU

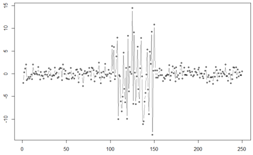
Coverage Guarantees: pure noise test-beds



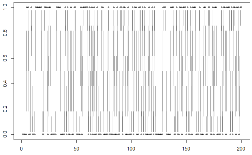
(a) Gaussian



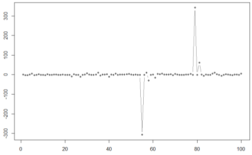
(b) Poisson



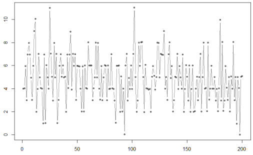
(c) Gaussian (hetero)



(d) Symmetric Bernoulli



(e) Cauchy

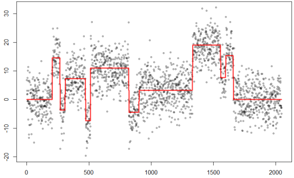


(f) Poisson + Gaussian

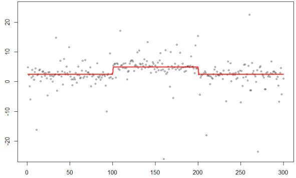
Number of times, out of 100 simulated sample paths, that a particular method indicates no intervals of significance:

	runs	MR	MQS
Plain Gauss	99	100	57
Plain Poisson	100	99	90
Heterogeneous Gauss	100	100	61
Symmetric Bernoulli	99	92	81
Plain Cauchy	99	100	59
Mix	96	100	68

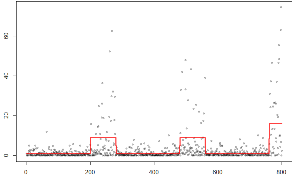
Detection Power: signal + noise test-beds



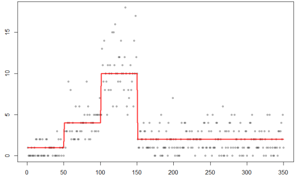
(a) Blocks



(b) Cauchy



(c) Bursts



(d) Poisson

Detection Power: simulation results

Average performance on 100 simulated sample paths of the test beds for each change point inference method:

	Blocks			**Cauchy**		
	runs	MR	MQS	runs	MR	MQS
Spurious	0.00	0.00	0.38	0.00	0.00	0.21
Prop. Genuine	1.00	1.00	0.96	1.00	1.00	0.91
No. Genuine	8.70	9.88	10.79	1.36	2.00	1.96
Avg. Length	48.06	46.44	44.86	75.01	60.08	51.90

Detection Power: simulation results

Average performance on 100 simulated sample paths of the test beds for each change point inference method:

	Bursts			**Poisson**		
	runs	MR	MQS	runs	MR	MQS
Spurious	0.02	0.00	0.13	0.00	0.00	0.07
Prop. Genuine	0.99	1.00	0.97	1.00	1.00	0.98
No. Genuine	1.25	3.12	4.16	2.27	2.92	2.86
Avg. Length	114.29	100.68	114.21	35.80	36.41	31.88

Real Data Analysis: air quality and COVID lock-downs

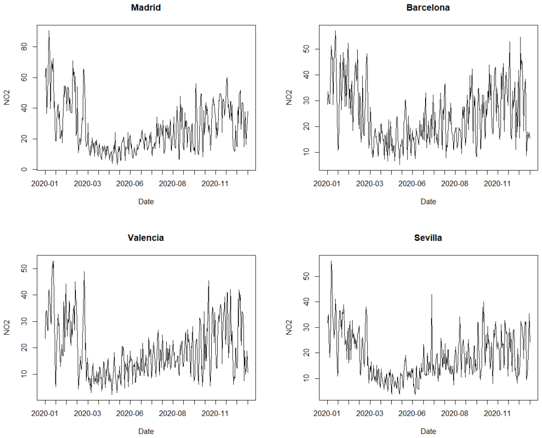


Figure 6: daily NO_2 concentration levels during 2020, de-trended and whitened, for the four largest cities in Spain by population

Real Data Analysis: air quality and COVID lock-downs

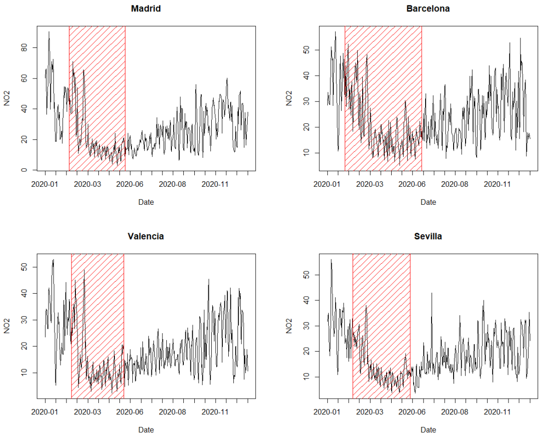


Figure 6: intervals of significance returned by our procedure, testing against f° begin degree 1 polynomial with $\alpha = 0.1$, roughly align with the national state of alarm (March 15) and initial lock-down easing (May 2)

Thank you!