Change-Point Analysis with weighted U-Statistics

StatScale ECR Meeting, 14th–16th December

Kata Vuk
Ruhr University Bochum

Joint work with Herold Dehling and Martin Wendler

Mean Change Model

We consider the following model

$$X_i = \mu_i + \xi_i,$$

where

- $(\mu_i)_{i\geq 1}$ is an unknown signal,
- $(\xi_i)_{i\geq 1}$ is a mean zero stationary stochastic process.

Based on observations $X_1, \ldots, X_n$, we want to test the hypothesis

$$H : \mu_1 = \ldots = \mu_n$$

against the alternative

$$A : \mu_1 = \ldots = \mu_{k^*} \neq \mu_{k^*+1} = \ldots = \mu_n, \text{ for some } k^* \in \{1, \ldots, n-1\}.$$

Two-Sample U-Statistics

Consider two samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ and a kernel function $h : \mathbb{R}^2 \to \mathbb{R}$. Then

$$\frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h(X_i, Y_j)$$

is a two-sample U-statistic.

- For change-point problems the two samples are given by $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_n$, where k is the potential change-point.
- Taking suitable functionals of $\sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j),\ 1 \leq k \leq n-1$, a variety of change-point tests can be derived.

Maximum Test Statistic

$$\max_{1 \le k \le n-1} \frac{1}{n^{3/2}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) \right|$$

Specific kernel functions:

- $h(x, y) = y - x$ leads to the CUSUM test statistic
- $h(x, y) = (1_{\{x \le y\}} - \frac{1}{2})$ leads to the Wilcoxon test statistic

CUSUM and Wilcoxon Test Statistic

CUSUM:

$$\max_{1\leq k\leq n-1} \frac{1}{n^{3/2}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_j - X_i) \right| = \max_{1\leq k\leq n-1} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{k} X_i - \frac{k}{n} \sum_{j=1}^{n} X_j \right|$$

Wilcoxon:

$$\max_{1\leq k\leq n-1} \frac{1}{n^{3/2}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \right| = \max_{1\leq k\leq n-1} \frac{1}{n^{3/2}} \left| \sum_{i=1}^{k} \mathrm{rank}(X_i) - \frac{k}{n} \sum_{i=1}^{n} \mathrm{rank}(X_i) \right|$$

- Both tests compare the first part of the sample to the average. The Wilcoxon test involves the rank of the data, whereas the CUSUM involves their values

Asymptotic Distribution under the Hypothesis

$$\max_{1\leq k<n} \frac{1}{n^{3/2}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) \right|$$

converges in distribution to the supremum of a standard Brownian bridge process.

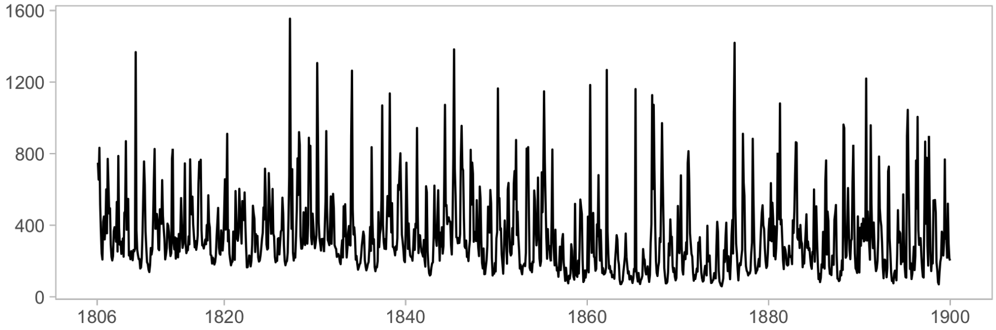
- I.i.d. data: M. Csörgő, L. Horváth (1988). Invariance Principles for Change- point Problems.
- Short range dependent data: H. Dehling, R. Fried, I. Garcia, M. Wendler (2015). Change-Point Detection Under Dependence Based on Two-Sample U-Statistics.

Different limit distribution for

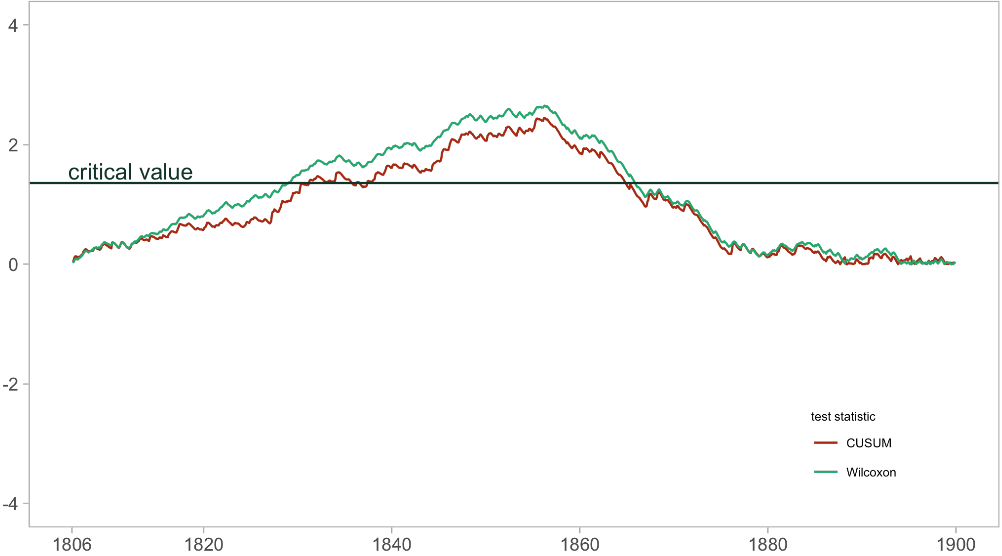
- Long range dependent data: H. Dehling, E. Rooch, M. Taqqu (2017). Two-Sample U-Statistic Processes for Long-Range Dependent Data.

Data Example: Elbe River

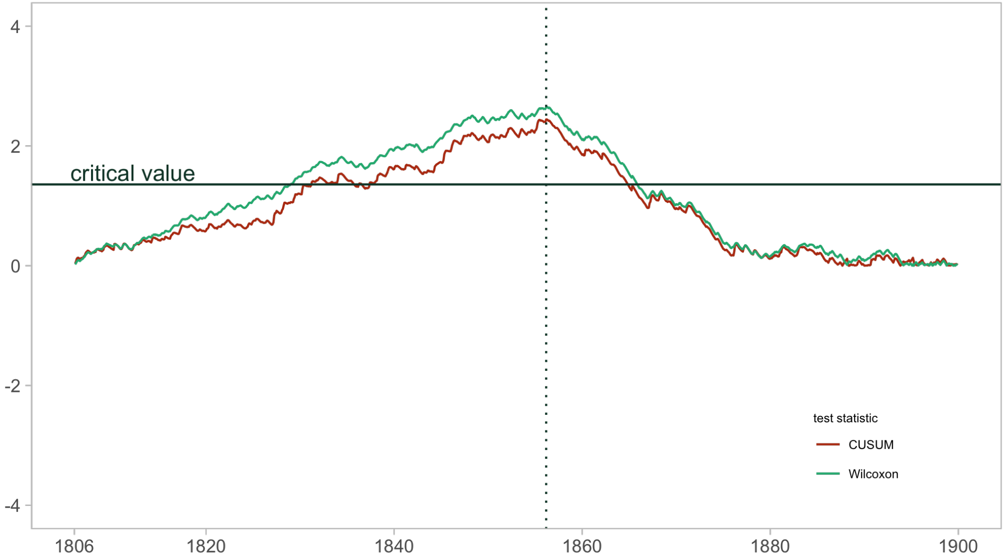
- Monthly average discharge of the river Elbe in Dresden, Germany
- Considered years: 1806 to 1900



Elbe River Data: Values of the Test Statistics

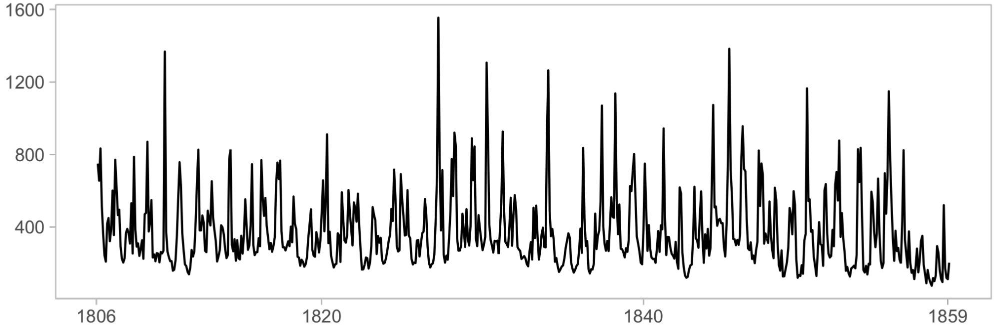


Elbe River Data: Detected Change-Point

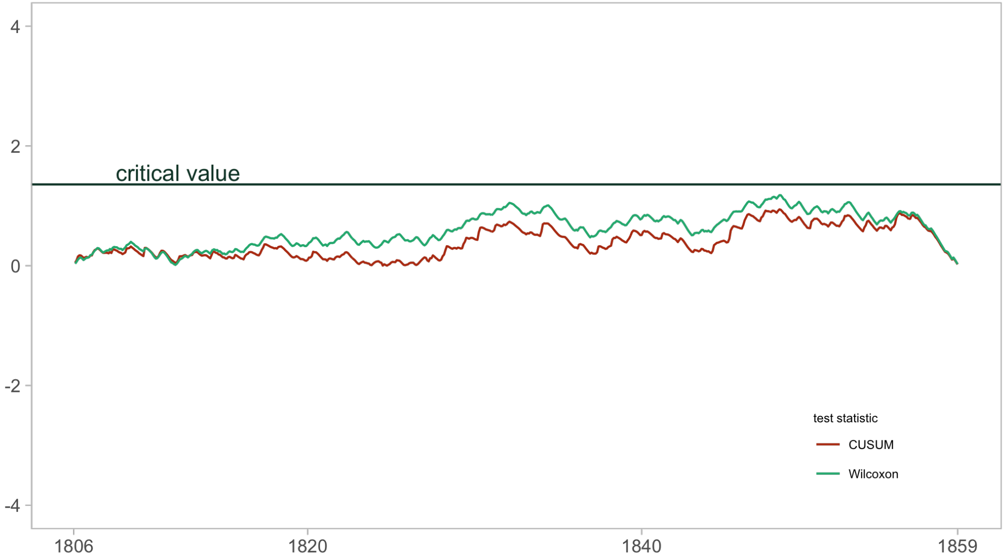


Elbe River: A shorter time period

- Considered years: 1806 to 1859 (shortly after the change-point)



Problem: Test statistics don't detect the change-point



Solution: Weighted Test Statistics

Unweighted $\to$ Weighted Test Statistics

Consider $\sum_{i=1}^{k}\sum_{j=k+1}^{n}(X_j - X_i)$ with $X_1, \ldots, X_n$ i.i.d. and $\mathrm{Var}(X_1) = \sigma^2$.

Standardizing $\sum_{i=1}^{k}\sum_{j=k+1}^{n}(X_j - X_i)$ leads to the process

$$\frac{1}{\sqrt{k(n-k)n}\sigma}\sum_{i=1}^{k}\sum_{j=k+1}^{n}(X_j - X_i)$$

Taking the maximum and replacing $X_j - X_i$ by $h(X_i, X_j)$ leads to the **weighted test statistic**

$$\max_{1\le k\le n-1}\frac{1}{\sqrt{k(n-k)n}\sigma}\sum_{i=1}^{k}\sum_{j=k+1}^{n}h(X_i, X_j).$$

Results under the Hypothesis

Limit distribution of the weighted test statistic

$$\max_{1\leq k\leq n-1} \frac{1}{\sqrt{k(n-k)n}\sigma} \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j)$$

- for i.i.d. data: M. Csörgő, L. Horváth (1988). Invariance Principles for Change- point Problems.

Our contribution:

- For short range dependent data: H. Dehling, K. Vuk, M. Wendler (2022). Change-Point Detection Based on Weighted Two-Sample U-Statistics.

Result: Asymptotic Distribution under the Hypothesis

Theorem (Dehling, V., Wendler, 2022)

Let $(X_i)_{i\geq 1}$ *be an* α*-mixing, strictly stationary process and let* $h(x,y)$ *be a bounded anti-symmetric kernel. Then, under some technical assumptions, it holds under the hypothesis*

$$\frac{\sqrt{2\log\log n}}{\sigma}\max_{1\leq k<n}\frac{1}{\sqrt{k(n-k)n}}\left|\sum_{i=1}^{k}\sum_{j=k+1}^{n}h(X_i,X_j)\right|-\beta_n\xrightarrow{\mathcal{D}}G,\ \text{as } n\to\infty,$$

where G *denotes a Gumbel extreme value distribution, i.e.* $P(G\leq x)=\exp(-2\exp(-x))$, *and where* $\beta_n=2\log\log n+\frac{1}{2}\log\log\log n-\frac{1}{2}\log\pi$. *The variance is given by*

$$\sigma^2=\operatorname{Var}(h_1(X_1))+2\sum_{k=2}^{\infty}\operatorname{Cov}(h_1(X_1),h_1(X_k))$$

where $h_1(x)=\operatorname{E}h(x,Y)-\operatorname{E}h(X,Y)$ *and* X,Y *are independent with the same distribution as* X_1.

Idea of Proof: Hoeffding's Decomposition

- Important technical tool in the study of U-statistics.

If $E|h(X,Y)| < \infty$ for two independent random variables X and Y we have

$$h(x,y) = \theta + h_1(x) + h_2(y) + \Psi(x,y),$$

where

$$\begin{aligned}
\theta &= Eh(X,Y) \\
h_1(x) &= Eh(x,Y) - \theta \\
h_2(y) &= Eh(X,y) - \theta \\
\Psi(x,y) &= h(x,y) - h_1(x) - h_2(y) - \theta.
\end{aligned}$$

Note that $\theta = 0$ for anti-symmetric kernels (i.e. $h(x,y) = -h(y,x)$).

Hoeffding's Decomposition applied to the Test Statistic

We apply Hoeffding's decomposition to the test statistic and obtain

$$\max_{1\le k<n}\sqrt{\frac{\log\log n}{k(n-k)n}}\left|\sum_{i=1}^{k}\sum_{j=k+1}^{n}h(X_i,X_j)\right|$$

$$=\max_{1\le k<n}\sqrt{\frac{\log\log n}{k(n-k)n}}\left|\underbrace{(n-k)\sum_{i=1}^{k}h_1(X_i)+k\sum_{i=k+1}^{n}h_2(X_i)}_{\text{linear part}}+\underbrace{\sum_{i=1}^{k}\sum_{j=k+1}^{n}\Psi(X_i,X_j)}_{\text{degenerate part}}\right|.$$

- The degenerate part is asymptotically negligible.
- The linear part determines the asymptotic behavior.

Result: Consistency under the Alternative

Theorem (Dehling, V., Wendler, 2022)

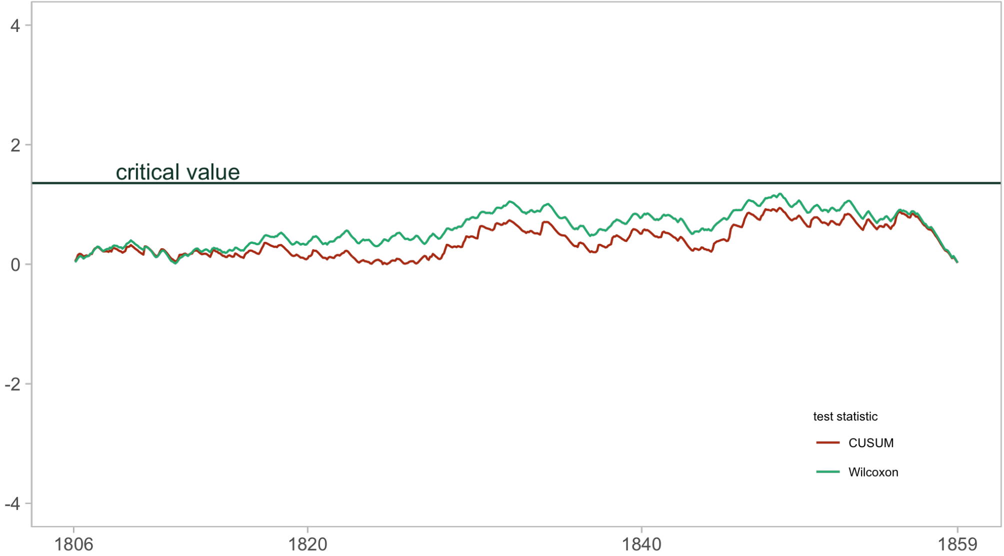
Assume that A holds and let h be a degenerate kernel. Define $\Delta_n := \mu_{k_n^*+1} - \mu_{k_n^*}$. *Under some technical assumptions, the following holds: If*

$$\sqrt{\frac{k_n^*(n-k_n^*)}{n \log\log n}}\,|\Delta_n| \longrightarrow \infty,$$

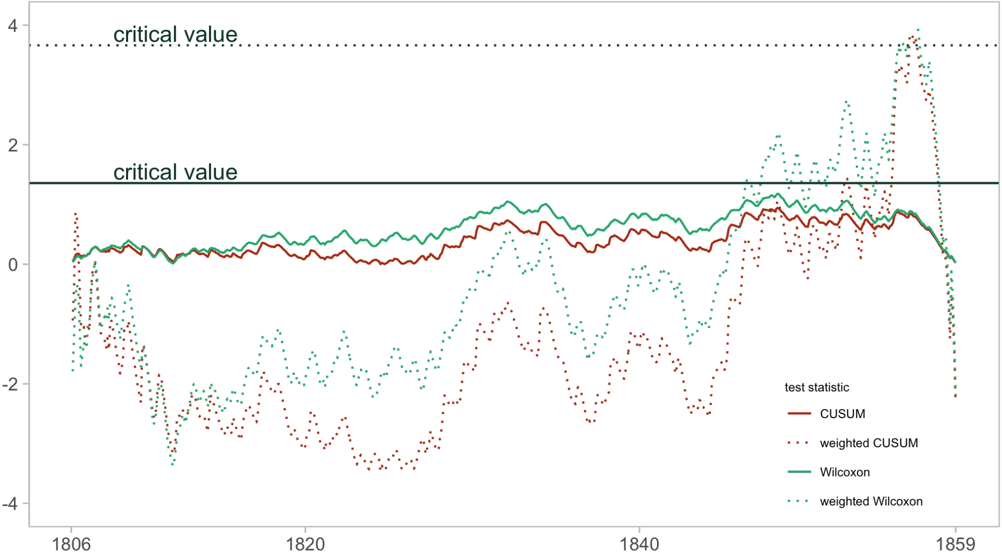
then

$$(\log\log n)^{-1/2} \max_{1\le k<n} \frac{1}{\sqrt{k(n-k)n}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) \right| \xrightarrow{\mathcal{P}} \infty.$$

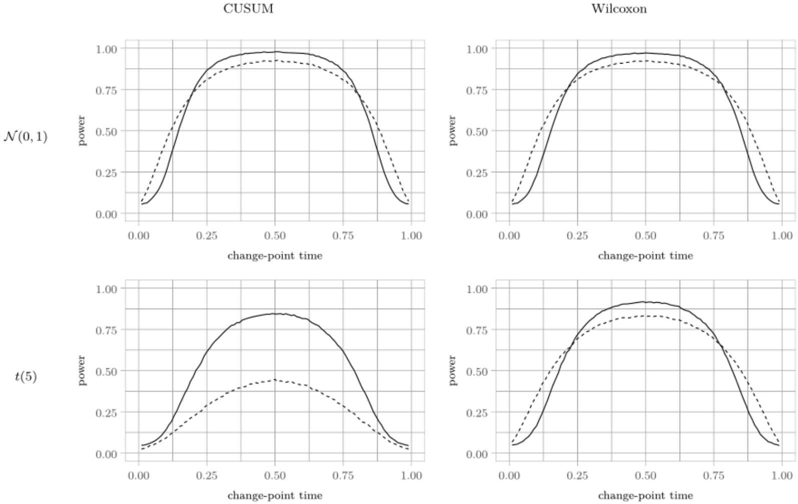
Back to the Elbe River Data Example



Weighted Test Statistics detect the Change-Point



Simulation Study: Power Comparison



Size-corrected power of the — unweighted and - - - weighted test statistics
800 i.i.d. observations; 20000 runs; $k_n^* = [\tau^* n]$; $\Delta = 0.3$

Flexible Weight Function

We can write

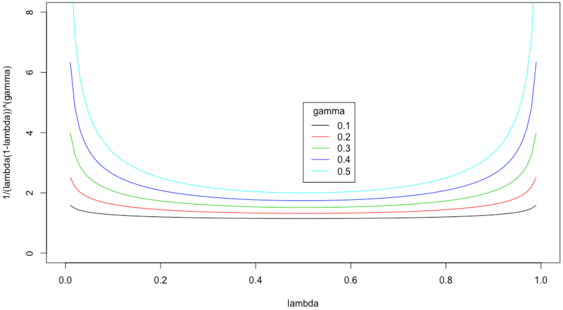
$$\frac{1}{\sqrt{k(n-k)n}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j)$$

$$= \frac{1}{n^{3/2}} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^{1/2}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j).$$

General weight functions of the form
$w(\lambda) = (\lambda(1-\lambda))^{\gamma}$, $\lambda \in (0,1)$, yield

$$\frac{1}{n^{3/2}} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j).$$

$$w(\lambda) = \frac{1}{(\lambda(1-\lambda))^{\gamma}}, \lambda \in (0,1)$$



Two Types of Alternatives

Model: $X_i = \mu_i + \xi_i$

- $(\xi_i)_{i\geq 1}$ is a mean zero stationary stochastic process
- $(\mu_i)_{i\geq 1}$ is an unknown signal

Testing problem:

$$H : \mu_1 = \ldots = \mu_n \text{ vs. } A : \mu_1 = \ldots = \mu_{k_n^*} \neq \mu_{k_n^*+1} = \ldots = \mu_n, \text{ for some } k_n^* \in \{1, \ldots, n-1\}.$$

Alternative 1: A with $\boldsymbol{k_n^* = [\tau^* n]}$, $\tau^* \in (0,1)$, and $\boldsymbol{\Delta_n = \mu_{k_n^*+1} - \mu_{k_n^*} = \frac{c}{\sqrt{n}}}$, where c is a constant.

Alternative 2: A with $\boldsymbol{k_n^* \approx c n^{\kappa}}$, where $\kappa = \frac{1-2\gamma}{2(1-\gamma)}$ and $\boldsymbol{\Delta_n \equiv \Delta}$.

Result: Asymptotic Distribution under Alternative 1

Alternative 1: $k_n^* = [\tau^* n]$, $\tau^* \in (0,1)$, and $\Delta_n = \mu_{k_n^*+1} - \mu_{k_n^*} = \frac{c}{\sqrt{n}}$

Theorem (Dehling, V., Wendler, 2022+)

Assume that ξ_1 has bounded density and that g is an odd function with $g(\xi_2 - \xi_1)$ having finite second moments. Moreover, assume that $\mathrm{Var}(h_1(\xi_1)) \to 0$ and that $c_g = \lim_{n\to\infty} \sqrt{n} u(\Delta_n)$ exists. Under Alternative 1, for $0 \le \gamma < \frac{1}{2}$ and as $n \to \infty$,

$$\max_{1\le k\le n} \frac{1}{n^{3/2}} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} g(X_j - X_i) \xrightarrow{\mathcal{D}} \sup_{0\le\lambda\le 1} \frac{1}{(\lambda(1-\lambda))^{\gamma}} [\sigma W^{(0)}(\lambda) + c_g \phi_{\tau^*}(\lambda)],$$

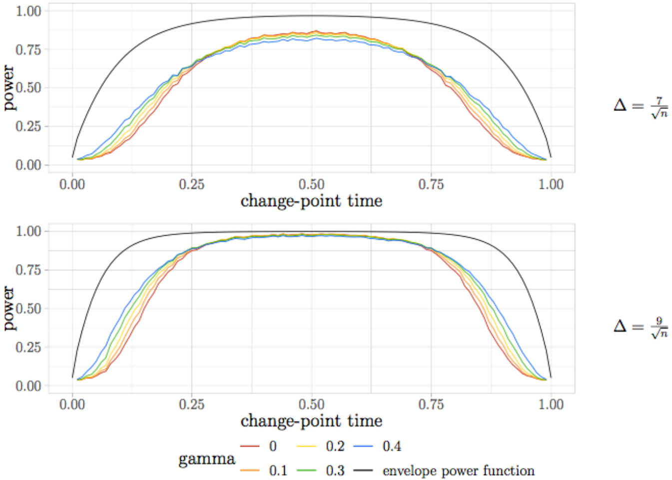
where $W^{(0)}(\lambda)$ is a Brownian bridge process, $\sigma^2 = \mathrm{E}(g_1^2(\xi_1)) > 0$ and

$$g_1(\xi_1) = \mathrm{E}[g(\xi - \xi_1)] - \mathrm{E}(g(\xi - \eta)),$$
$$u(\Delta_n) = \mathrm{E}[g(\xi - \eta + \Delta_n) - g(\xi - \eta)],$$
$$h_1(\xi_1) = \mathrm{E}[g(\xi - \xi_1 + \Delta_n) - g(\xi - \xi_1)] - u(\Delta_n),$$

$$\phi_{\tau^*}(\lambda) = \begin{cases} \lambda(1-\tau^*) & \text{for } \lambda \le \tau^* \\ \tau^*(1-\lambda) & \text{for } \lambda \ge \tau^*. \end{cases}$$

where ξ and η are independent with the same distribution as ξ_1,

Power Comparison under Alternative 1



Size-corrected power. SImulations: 1000 i.i.d. standard normally distributed observations; 5000 runs

Power Comparison under Alternative 1

- Overall-power compared to the envelope power for different values of the parameter γ and different shift heights Δ_n.

Δ_n \ γ	0	0.1	0.2	0.3	0.4
$\frac{5}{\sqrt{n}}$	72.30%	72.71%	74.13%	74.75%	71.53%
$\frac{7}{\sqrt{n}}$	78.97%	79.86%	81.45%	82.65%	81.22%
$\frac{9}{\sqrt{n}}$	83.98%	85.14%	86.87%	88.52%	89.22%

The simulations are based on $n = 1000$ independent, standard normally distributed observations.

Result: Asymptotic Distribution under Alternative 2

Alternative 2: $k_n^* \approx cn^\kappa$, where $\kappa = \frac{1-2\gamma}{2(1-\gamma)}$ and $\Delta_n \equiv \Delta$

Theorem (Dehling, V., Wendler, 2022+)

Assume that $g(\xi_2 - \xi_1)$ has finite second moments. Moreover, assume that $\mathrm{Var}(h_1(\xi_1)) < \infty$. Then, under Alternative 2 and as $n \to \infty$, we have for $\gamma = 0$

$$\max_{1\le k\le n} \frac{1}{n^{3/2}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} g(X_j - X_i) \right| \xrightarrow{\mathcal{D}} \sup_{0\le\lambda\le1} \left| \sigma W^{(0)}(\lambda) + c(1-\lambda)u(\Delta) \right|,$$

and for $0 < \gamma < 1/2$

$$\max_{1\le k\le n} \frac{1}{n^{3/2}} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} g(X_j - X_i) \right| \xrightarrow{\mathcal{D}} \max\left\{ c^{1-\gamma}u(\Delta), \sup_{0\le\lambda\le1} \frac{\sigma}{(\lambda(1-\lambda))^\gamma} \left| W^{(0)}(\lambda) \right| \right\},$$

where σ and $u(\Delta)$ and $h_1(\xi_1)$ are defined as in the theorem under Alternative 1.

Short Summary and Future Work

Test statistic:

$$\frac{1}{n^{3/2}} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j).$$

- Asymptotic distribution under the hypothesis and consistency under the alternative for $\gamma = 1/2$ for short range dependent data
- Asymptotic distribution under the two types of alternative for $0 \leq \gamma < 1/2$ and $h(x,y) = g(y-x)$
- Simulation study: Weighted test statistics have better power when we have early or late changes

Future work:

- Asymptotic theory under the alternative for dependent data
- Gradual changes
- Multiple changes