Multiscale detection of practically significant changes in a gradually varying time series

Patrick Bastian, Holger Dette

Ruhr-Universität Bochum Fakultät für Mathematik 44780 Bochum, Germany

Abstract

In many change point problems it is reasonable to assume that compared to a benchmark at a given time point t_0 the properties of the observed stochastic process change gradually over time for $t > t_0$. Often, these gradual changes are not of interest as long as they are small (nonrelevant), but one is interested in the question if the deviations are practically significant in the sense that the deviation of the process compared to the time t_0 (measured by an appropriate metric) exceeds a given threshold, which is of practical significance (relevant change).

In this paper we develop novel and powerful change point analysis for detecting such deviations in a sequence of gradually varying means, which is compared with the average mean from a previous time period. Current approaches to this problem suffer from low power, rely on the selection of smoothing parameters and require a rather regular (smooth) development for the means. We develop a multiscale procedure that alleviates all these issues, validate it theoretically and demonstrate its good finite sample performance on both synthetic and real data.

1 Introduction

Change point analysis is an ubiquitous topic in mathematical statistics with numerous applications in diverse areas such as economics, climatology and linguistics. In this paper we are concerned with the detection of (gradual) changes in the univariate time series

$$X_i = \mu(i/n) + \epsilon_i , \quad i = 1, \dots, n, \tag{1.1}$$

where $(\epsilon_i)_{1 \leq i \leq n}$ is a centered error process and $\mu : [0, 1] \to \mathbb{R}$ is an unknown (mean) function. Starting with the seminal work of (Page, 1955), a large part of the literature considers the case where the function μ is piecewise constant with at most one change and we refer to reviews of (Aue and Horváth, 2013), (Aue and Kirch, 2023) and to the recent textbook of (Horváth and Rice, 2024) and the reference therein. More recently the problem of detecting multiple changes has found considerable interest as well. Among many others, we mention (Fryzlewicz and Rao, 2013), (Frick et al., 2014), (Baranowski et al., 2019), (Dette et al., 2020), who considered one-dimensional data and to (Chen et al., 2022; Li et al., 2023; Madrid Padilla et al., 2022) for some recent results in the high-dimensional and multivariate case. A good review of the current state of the art on data segmentation of a piecewise constant signal can be found in the recent papers of (Truong et al., 2020) and (Cho and Kirch, 2024). A common aspect of the methodology in most of these references consists in the fact that it is based on the construction of testing procedures for the hypothesis

$$H_0: \mu(t) = c \quad \forall t \in [0, 1],$$

where c is an unknown constant, and the different procedures address different forms of the piecewise constant function μ under the alternative. Here a large part of the literature has its focus on piecewise stationary alternatives.

While there are many applications where this can be well justified, at least approximately (see for instance Aston and Kirch, 2012; Hotz et al., 2013, for some examples), there exist also many situations where it is not reasonable to assume that the mean function μ in model (1.1) is piecewise constant over the full period because it is continuously smoothly between potential jumps points. Examples include climate data (Karl et al., 1995), where it is continuously varying with no jumps, financial data (Vogt and Dette, 2015) and medical data (Gao et al., 2018). In this context, (Müller, 1992), (Gijbels et al., 2007) and (Gao et al., 2008) among others, considered model (1.1) with a gradually varying mean function and sudden jumps and developed change point analysis for determining the locations of these jumps. (Vogt and Dette, 2015) considered the problem of detecting gradual changes in a more general context. For the special case of the mean they assumed that μ is constant for some time, then slowly starts to change (with no jumps), and developed a fully nonparametric method to estimate such a smooth change point.

The present paper differs from these works. Although we consider a model of the form (1.1) with smooth regression function and potential jumps, we are not interested in the locations of the jump points or in the first time point where the mean functions starts to vary. Instead we define a change differently as a practical significant deviation of μ in the interval [0, 1] from a given benchmark, say $\mu_0^{t_0}$. More precisely, for a constant $\mu_0^{t_0}$ and a threshold $\Delta > 0$

we are interested in the hypotheses

$$H_0(\Delta) : \sup_{t \in [t_0, 1]} |\mu(t) - \mu_0^{t_0}| \le \Delta \quad vs \quad H_1(\Delta) : \sup_{t \in [t_0, 1]} |\mu(t) - \mu_0^{t_0}| > \Delta,$$
(1.2)

where $t_0 \in [0, 1)$ is a given point defining the time interval of interest. A prominent example for the consideration of these hypotheses is the analysis of global mean temperature anomalies, where one is interested in a significant deviation of the current temperatures from a reference value $\mu_0^{t_0}$ at time t_0 (such as the average temperature before time t_0), and an important problem is to investigate if these deviations exceed Δ after the time t_0 , such as 1.5 degrees Celsius as postulated in the Paris agreement. Hypotheses of the form (1.2) are also considered in quality control (often in an online framework), where one is interested in the "stability" of a given process. This means that the sequence of means stays within a predefined range (as specified by the null hypothesis in (1.2)). While in change-point analysis the focus is often on testing for the presence of a change and on estimating the time at which a change occurs once it has been detected, quality control typically has its focus more on detecting such a change as quickly as possible after it occurs (see, for example, Woodall and Montgomery, 1999).

Despite their importance, a test for the hypotheses in (1.2) has only been recently developed by (Bücher et al., 2021), who used local linear regression techniques to estimate the quantity $\sup_{t \in [t_0,1]} |\mu(t) - \mu_0^{t_0}|$. They proposed to reject the null hypothesis for large values of the estimate, where critical values are either obtained by asymptotic theory (which shows that a properly scaled version of the estimate converges weakly to a Gumbel type distribution) or by resampling based on a Gaussian approximation. As a consequence, the resulting procedure suffers from several deficits First, it is conservative in finite samples, particularly for small sample sizes. Second, the mean function μ in model (1.1) has to be twice differentiable and the difference $\mu(t) - \mu_0^{t_0}$ has to satisfy some convexity properties, making the method unreliable for less smooth or discontinuous functions. Finally, the procedure proposed in (Bücher et al., 2021) relies on a bandwidth parameter that also prevents detection of local alternatives at the standard parametric rate $n^{-1/2}$.

Our contribution in this paper is a novel multiscale test for the hypotheses (1.2) that does not suffer from the aforementioned problems. More precisely, we compare (an estimate of) the benchmark with local means calculated over different scales and establish the weak convergence of this statistic (see Theorem 2.2). The limit is not distribution free and depends on sequences of "extremal sets", which can only be defined implicitly due to certain properties of the Brownian motion. Based on this result, we propose first a testing procedure that relies on a distributional upper bound of the limiting distribution and only requires estimation of a long-run variance. By construction, the resulting test is conservative, but it it can detect local alternatives at a parametric rate and already outperforms the existing methodology in our finite sample study. By estimating the extremal sets we are able to construct a more elaborate procedure which achieves the nominal level asymptotically. The resulting test can detect local alternatives at a parametric rate as well and yields a further substantial improvement of the finite sample performance.

Summarizing, both novel multiscale tests for practically relevant changes in the gradually changing mean function can detect local alternatives converging to the null at a faster rate than the currently available procedure. In contrast to this test, they are applicable for piecewise smooth mean functions without further constraints on their geometry and do not require the choice of smoothing parameters.

2 Multiscale detection of relevant changes

In the following we consider the location scale model

$$X_i = \mu(i/n) + \epsilon_i \quad i = 1, ..., n$$
 (2.1)

where $(\epsilon_i)_{i \in \mathbb{N}}$ is a stationary centered process and μ is a bounded and piecewise Lipschitzcontinuous function. We are interested in detecting significant deviations of the function μ on the interval $[t_0, 1]$ from its long-term average in the past

$$\mu_0^{t_0} := \frac{1}{t_0} \int_0^{t_0} \mu(x) dx \; ,$$

where $0 < t_0 < 1$ is some predefined point in time. We address this problem by testing the hypotheses

$$H_0(\Delta): d_{\infty} \le \Delta \quad vs \quad H_1(\Delta): d_{\infty} > \Delta ,$$

$$(2.2)$$

where

$$d_{\infty} = \sup_{t \in [t_0, 1]} |\mu(t) - \mu_0^{t_0}|$$
(2.3)

denotes the maximum (absolute) deviation of the function μ from $\mu_0^{t_0}$ over the interval $[t_0, 1]$ and $\Delta > 0$ is a given threshold.

A typical application for this benchmark is encountered in the analysis of temperature data where deviations from a pre-industrial average temperature are of interest. In this case the choice of threshold is also quite clear. For example, if we are interested in the global mean temperature anomalies, a reasonable choice is $\Delta = 1.5$ degrees Celsius corresponding to the Paris Agreement adopted at the UN Climate Change Conference(COP21) in Paris, 2015. In other circumstances the choice of Δ might not be so obvious and has to be carefully discussed for each application. We defer further discussion of this issue to Remark 2.7, where we also propose a data based choice of the threshold Δ .

In the remainder of this section we introduce the necessary concepts and assumptions to define and theoretically analyze a multiscale test statistic for the testing problem (2.2).

Assumption 2.1. The random variables in model (2.1) form a triangular array of real valued random variables where $(\epsilon_i)_{i\in\mathbb{Z}}$ is a mean zero stationary sequence with existing long run variance $\sigma^2 = \sum_{i\in\mathbb{Z}} \mathbb{E}[\epsilon_0\epsilon_i]$

(A1) For some $0 there exists a standard Brownian motion B, such that for all <math>k \in \mathbb{N}$

$$\left|\sum_{i=1}^{k} \epsilon_{i} - \sigma B(k)\right| \le Ck^{1/2-p}$$

almost surely for some constant C > 0.

(A2) The function $\mu: [0,1] \to \mathbb{R}$ is piecewise Lipschitz continuous with finitely many jumps.

Assumption (A1) is a high level assumption that is standard in the literature and is satisfied by a large class of weakly dependent time series (see, for instance (Dehling, 1983) for mixing, (Wu, 2005) for physically dependent and (Berkes et al., 2011) for L^p -m-approximable processes). Assumption (A2) is a weak regularity assumption on the function μ , that can in principle be weakened to Hölder continuity with some additional technical effort.

We now introduce the test statistic, and to that end denote for j < k by

$$\hat{\mu}_j^k = \frac{1}{k-j} \sum_{i=j+1}^k X_i$$

the (local) mean of the observations X_{j+1}, \ldots, X_k . Note that

$$\mathbb{E}\left[\hat{\mu}_{j}^{k}\right] = \frac{1}{k-j} \sum_{i=j+1}^{k} \mu(i/n) \simeq \frac{n}{k-j} \int_{j/n}^{k/n} \mu(t) dt$$

and therefore $\hat{\mu}_0^{\lfloor nt_0 \rfloor} - \hat{\mu}_j^k$ (approximately) compares the integral of the function μ over the interval $[0, \lfloor nt_0 \rfloor/n]$ with the "local" integral over the interval [j/n, k/n], which is approximately given by $\frac{n}{k-j} \int_{j/n}^{k/n} \mu(t) dt \approx \mu(\frac{k+j}{2n})$ if k-j is small. We now consider these differences

on different scales and define for a sequence $(c_n)_{n\in\mathbb{N}}$ of natural numbers such that $c_n\to\infty$ and

$$\frac{n^{1-2p}}{c_n} = o(1), \tag{2.4}$$

the test statistic

$$\hat{T}_{n,\Delta} = \sup_{\substack{c_n \leq c \\ c \leq n - \lfloor nt_0 \rfloor \\ c \in \mathbb{N}}} \sup_{\substack{|k-j|=c \\ k>j \geq \lfloor nt_0 \rfloor}} \left(\sqrt{c} \left| \hat{\mu}_0^{\lfloor nt_0 \rfloor} - \hat{\mu}_j^k \right| - \sqrt{2 \log\left(\frac{ne}{c}\right)} - \sqrt{c}\Delta \right).$$
(2.5)

By the discussion of the previous paragraph $\hat{\mu}_j^k$ estimates $\mu(\frac{k+j}{2n})$ for smaller scales, which ensures that relatively short excursions of the function $t \to |\mu(t) - \mu_0^{t_0}|$ above Δ are detected. For larger scales the statistic (2.5) is able to take advantage of longer excursions of the function $t \to |\mu(t) - \mu_0^{t_0}|$ above the threshold Δ , thereby increasing the power of the test substantially. The additive factor

$$\Gamma_n(c) := \sqrt{2\log\left(\frac{ne}{c}\right)}$$

equalizes the magnitude of the different scales which would otherwise be dominated by the small scales. We also note that the scaling factor \sqrt{c} depends on n and is of larger and smaller order than $n^{1/2-p}$ and $n^{1/2}$, which leads to a non-trivial asymptotic distribution of the statistic $\hat{T}_{n,\Delta}$ in the case $d_{\infty} = \Delta$, which we call *boundary* of the hypotheses. Our first main result provides such a weak convergence result for the statistic

$$\hat{T}_n = \sup_{\substack{c_n \le c \le n - \lfloor nt_0 \rfloor \\ c \in \mathbb{N}}} \sup_{\substack{|k-j|=c \\ k>j \ge \lfloor nt_0 \rfloor}} \left(\sqrt{c} \left| \hat{\mu}_0^{\lfloor nt_0 \rfloor} - \hat{\mu}_j^k \right| - \Gamma_n(c) - d_\infty \right),$$

which reduces to the statistic $\hat{T}_{n,\Delta}$ in (2.5), if the centering term d_{∞} defined in (2.3) is replaced by the threshold Δ .

Theorem 2.2. Grant assumptions (A1) and (A2). We then have

$$\hat{T}_n \xrightarrow{d} T_{d_{\infty}}$$
 (2.6)

where

$$T_{d_{\infty}} := \sigma \lim_{\epsilon \downarrow 0} \sup_{(s,t) \in \mathcal{A}_{\epsilon,d_{\infty}}} \Big\{ \mathfrak{s}(s,t) \Big(\sqrt{t-s} \frac{B(t_0)}{t_0} - \frac{B(t) - B(s)}{\sqrt{t-s}} \Big) - \Gamma(t-s) \Big\}, \qquad (2.7)$$

B denotes a standard Brownian motion and

$$\mathcal{A}_{\epsilon,d_{\infty}} = \left\{ (s,t) \in [t_0,1]^2 \ \middle| \ s < t, \ \left| \mu_0^{t_0} - \mu_s^t \right| \ge d_{\infty} - \epsilon \right\}$$
(2.8)
$$\mu_s^t = \frac{1}{t-s} \int_s^t \mu(x) dx$$

$$\mathfrak{s}(s,t) = sgn(\left(\mu_0^{t_0} - \mu_s^t\right))$$

$$\Gamma(t-s) = \sqrt{2\log\left(\frac{e}{t-s}\right)} .$$
(2.9)

Moreover, the distribution of the random variable $T_{d_{\infty}}$ is continuous. In particular,

(1) if
$$d_{\infty} = \Delta$$
, we have $\hat{T}_{n,\Delta} \xrightarrow{d} T_{\Delta}$,
(2) If $d_{\infty} < \Delta$ or $d_{\infty} > \Delta$ we have $\hat{T}_{n,\Delta} \xrightarrow{\mathbb{P}} -\infty$ or $\hat{T}_{n,\Delta} \xrightarrow{\mathbb{P}} \infty$, respectively.

Note that the limit distribution in Theorem 2.2 is not distribution free even if the long run variance of the error process $(\epsilon_i)_{i\in\mathbb{N}}$ would be known. In fact this distribution depends in a rather delicate way on the regression function μ which appears in the definition $T_{d_{\infty}}$ through the set $\mathcal{A}_{\varepsilon,d_{\infty}}$. This is contrast to many other multiscale tests proposed in the literature (see, for example, Dümbgen and Spokoiny, 2001; Dümbgen and Walther, 2008; Schmidt-Hieber et al., 2013; Dette et al., 2020). The difference can be explained by the fact that these and - to the best of our knowledge - all other papers on multiscale testing do not consider relevant hypotheses of the form (1.2) and (2.3) with $\Delta > 0$. In fact, transferring the hypotheses considered in the multiscale testing literature so far to the situation considered in this paper yields the testing problem for the "classical" hypotheses

$$H_0: d_\infty = 0 \quad vs \quad H_1: d_\infty > 0 ,$$

which corresponds to the choice $\Delta = 0$ in (2.2). It follows from the arguments given in the proof of Theorem 2.2 that in the case $d_{\infty} = 0$

$$\hat{T}_n \xrightarrow{d} \sigma M$$

where the random variable M is defined by

$$M = \sup_{t_0 \le s < t \le 1} \left| \sqrt{t - s} \frac{B(t_0)}{t_0} - \frac{B(t) - B(s)}{\sqrt{t - s}} \right| - \Gamma(t - s).$$
(2.10)

As the quantiles of the distribution of M can be obtained by simulation, we can already use this result for the construction of a valid (conservative) inference procedure for the hypotheses (2.2) employing the upper (distributional) bound

$$\mathbb{P}(T_{\Delta} > q) \le \mathbb{P}(\sigma M > q) \tag{2.11}$$

and estimating the long run variance σ^2 . To be specific, following (Wu and Zhao, 2007) we define

$$\hat{\sigma}^2 = \frac{1}{\lfloor n/m \rfloor - 1} \sum_{j=1}^{\lfloor n/m \rfloor - 1} \frac{\left(\hat{\mu}_{(j-1)m}^{jm} - \hat{\mu}_{jm}^{(j+1)m}\right)^2}{2m}$$
(2.12)

as an estimator of the long run variance, where the parameter $m \in \mathbb{N}$ converges to ∞ as $n \to \infty$ and is is proportional to $n^{1/3}$. The null hypothesis is then rejected, whenever

$$T_{n,\Delta} \ge \hat{\sigma} q_{1-\alpha} , \qquad (2.13)$$

where $q_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the distribution of M. We will show in the Section 5.4 of the appendix that (under the assumptions made in this paper) the estimator $\hat{\sigma}^2$ is consistent for σ^2 , that is,

$$\hat{\sigma}^2 = \sigma^2 + O_{\mathbb{P}}(n^{-1/3}) ,$$

which yields the following result.

Theorem 2.3. Under assumptions (A1) and (A2) the test defined by (2.13) is consistent and has asymptotic level α .

We continue investigating the asymptotic power properties of the test (2.13) by considering a class of local alternatives of the form

$$\mu_n(t) - \mu_0^{t_0} = \Delta + \beta_n h(t) , \qquad (2.14)$$

where h is some non-negative and Lipschitz continuous function.

Theorem 2.4. Let assumptions (A1) and (A2) and be satisfied and consider local alternatives of the form (2.14) with $\beta_n = n^{-1/2}$, then

$$\hat{T}_{n,\Delta} \xrightarrow{d} \sigma \sup_{t_0 \le s < t \le 1} \left(\sqrt{t-s} \frac{B(t_0)}{t_0} - \frac{B(t)-B(s)}{\sqrt{t-s}} - \Gamma(t-s) + \frac{1}{\sqrt{t-s}} \int_s^t h(x) dx \right)$$

We collect some observations that follow from this result in the following remark.

Remark 2.5.

(1) By Theorem 2.4 the test (2.13) can detect local alternatives converging to the null hypothesis at a parametric rate $n^{-1/2}$. This establishes a substantial improvement over the results obtained in (Bücher et al., 2021), where the nonparametric rate

$$\beta_n \simeq \left(\sqrt{nh_n \log(h_n)}\right)^{-1}$$

is required to obtain non-trivial power. Here h_n is the bandwidth used for the local linear estimator of the regression function μ .

(2) It is clear from the inequality (2.11) that the test (2.13) is in general conservative even in the case that $|\mu(t) - \mu_0^{t_0}| = \Delta$ for all $t \in [t_0, 1]$. Comparing the definitions of the random variables T_{Δ} and M in (2.7) and (2.10), respectively, the difference $\mathbb{P}(T_{\Delta} > q) - \mathbb{P}(\sigma M > q)$ will be large for those models where $d_{\infty} = \Delta$ and where at the same time the set $\{s \in [t_0, 1] \mid |\mu_0^{t_0} - \mu(s)| < \Delta\}$ is "large". This indicates the need for a test procedure that is able to take into account the structure of the sets $\mathcal{A}_{\varepsilon,d_{\infty}}$ appearing in the definition of the random variable $T_{d_{\infty}}$ in (2.6).

To alleviate the issue raised in the second part of the previous remark we will develop an alternative test which uses quantiles from a distribution which approximates the distribution of the random variable T_{Δ} more directly. Obviously, such an approach has to take the estimation of the sets $\mathcal{A}_{\epsilon,d_{\infty}}$ in (2.8) into account. For this purpose we define for a scale parameter $c \in \mathbb{N}$

$$\hat{\mathcal{E}}_c = \left\{ (j,k) \in \{ \lfloor nt_0 \rfloor, \dots, n\}^2 \ \Big| \ k - j = c, \ \left| \hat{\mu}_0^{\lfloor nt_0 \rfloor} - \hat{\mu}_j^k \right| \ge \hat{d}_{\infty,c} - \hat{\sigma} \log(n) / \sqrt{c} \right\}$$

as an estimator of the extremal set at scale c, where

$$\hat{d}_{\infty,c} = \max_{k-j=c, \lfloor nt_0 \rfloor \le j < k \le n} |\hat{\mu}_0^{\lfloor nt_0 \rfloor} - \hat{\mu}_j^k|.$$

Note that the sequence of sets

$$\hat{\mathcal{A}}_n = igcup_{c_n \le c \le n - \lfloor nt_0 \rfloor} \hat{\mathcal{E}}_c$$

may heuristically be interpreted as a sequence of estimators for the sets $\mathcal{A}_{\epsilon,d_{\infty}}$ for a suitable sequence of $\epsilon \downarrow 0$. Next we introduce

$$\hat{\mathfrak{s}}(j/n,k/n) = \operatorname{sgn}((\hat{\mu}_0^{\lfloor nt_0 \rfloor} - \hat{\mu}_j^k))$$

as an estimator of the sign $\mathfrak{s}(j/n, k/n)$ in (2.9) and consider the statistic

$$\hat{T}_n^* = \hat{\sigma} \sup_{\substack{c_n \le c \le n - \lfloor nt_0 \rfloor \\ c \in \mathbb{N}}} \sup_{(j,k) \in \hat{\mathcal{E}}_c} \hat{\mathfrak{s}}(j/n,k/n) \left(\sqrt{c} \frac{B(\lfloor nt_0 \rfloor)}{\lfloor nt_0 \rfloor} - \frac{B(k) - B(j)}{\sqrt{c}}\right) - \Gamma_n(c)$$

where $\hat{\sigma}^2$ is the estimator of the long run variance defined in (2.12) and suprema over empty sets are defined as 0. We denote the $(1 - \alpha)$ -quantile of the distribution of \hat{T}_n^* by $q_{1-\alpha}^*$, which can easily be simulated. Our second test for a practically relevant deviation from deviation from the average $\mu_0^{t_0}$ rejects the null hypothesis in (2.2), whenever

$$\hat{T}_{n,\Delta} \ge q_{1-\alpha}^* \tag{2.15}$$

and the following result shows that this decision rule defines a consistent asymptotic level α , which can detect local alternatives converging to null hypothesis at a parametric rate.

Theorem 2.6. Under assumptions (A1) and (A2) the test (2.15) is consistent and has asymptotic level α . More precisely,

$$\mathbb{P}(\hat{T}_{n,\Delta} \ge q_{1-\alpha}^*) \to \begin{cases} 0 & d_{\infty} < \Delta \\ \alpha & d_{\infty} = \Delta \\ 1 & d_{\infty} > \Delta \end{cases}$$

Remark 2.7. An important question from a practical point of view is the choice of the threshold $\Delta > 0$, which has to be carefully discussed for each specific application. Essentially, this boils down to the important question when a deviation from the reference value $\mu_0^{t_0}$ is practically significant, which is related to the specification of the effect size (see Cohen, 1988). While in many situations, such as in the climate data example mentioned before, this specification is quite obvious, there are other applications where this choice might be less clear. However, for such cases it is possible to determine a threshold from the data which can serve as measure of evidence for a deviation of μ from the long term average $\mu_0^{t_0}$ with a controlled type I error α .

To be precise, note that the hypotheses $H_0(\Delta_1)$ and $H_0(\Delta_2)$ in (2.2) are nested for $\Delta_1 < \Delta_2$ and that the test statistic (2.5) is monotone in Δ . As the quantile $q_{1-\alpha}^*$ does not depend on Δ , rejecting $H_0(\Delta)$ for $\Delta = \Delta_1$ also implies rejecting $H_0(\Delta)$ for all $\Delta < \Delta_1$. The sequential rejection principle then yields that we may simultaneously test the hypotheses (2.2) for different choices of $\Delta \ge 0$ until we find the minimum value, say $\hat{\Delta}_{\alpha}$, for which $H_0(\Delta)$ is not rejected, that is

$$\hat{\Delta}_{\alpha} := \min\left\{\Delta \ge 0 \,|\, \hat{T}_{n,\Delta} \le q_{1-\alpha}^*\right\} \,. \tag{2.16}$$

Consequently, one may postpone the selection of Δ until one has seen the data. The same arguments of course also hold for the more conservative procedure defined by (2.13).

3 Estimating the time of the first relevant deviation

If a relevant deviation from a benchmark has been detected, it is of interest to determine the first time where this deviation occurs, that is

$$t^* = \min\left\{t \in [t_0, 1] \middle| \mid \mu(t) - \mu_0^{t_0} \mid \ge \Delta\right\}.$$
(3.1)

A natural estimator for t^* is the first time k where at least one estimated difference $\hat{\mu}_0^{\lfloor nt_0 \rfloor} - \hat{\mu}_j^k$ exceeds approximately Δ , and therefore we define

$$\hat{t} = \min\left\{k \ge \lfloor nt_0 \rfloor + c_n \mid \exists j \in \{\lfloor nt_0 \rfloor, \dots, k - c_n\} \text{ such that}$$

$$|\hat{\mu}_0^{\lfloor nt_0 \rfloor} - \hat{\mu}_j^k| \ge \Delta - \frac{\hat{\sigma} \log(n)}{\sqrt{k - j}}\right\}$$

$$(3.2)$$

where the constant c_n satisfies (2.4) (here we define the minimum over an empty set as ∞). In the following discussion we investigate the theoretical performance of this estimator, distinguishing the case where t^* is a point of continuity of $|\mu(t) - \mu_0^{t_0}|$ and where it is not. For this purpose, we introduce the function

$$t \to d(t) = |\mu(t) - \mu_0^{t_0}|$$

and consider first the smooth setting. Intuitively, a change at a point of continuity will be harder to detect if the function μ is very flat at t^* . To quantify this property, we assume that there exists constants $\kappa, c_{\kappa} > 0$ such that

$$\lim_{t\uparrow t^*} \frac{|d(t^*) - d(t)|}{(t^* - t)^{\kappa}} \to c_{\kappa} .$$
(3.3)

To obtain an explicit convergence rate we also need an assumption to ensure that the function d does not behave too irregularly close to the point t^* .

(A3) For some constant $\gamma \in (0, t^*)$ the function $t \to \operatorname{sign}(d(t))d(t)$ is increasing on the set

$$U_{\gamma}(t^*) = \left\{ t \in [t_0, t^*] \mid t > t^* - \gamma \right\}$$

Theorem 3.1. Let Assumptions (A1) - (A3) be satisified.

(a) If $t^* \in (t_0, 1]$, condition (3.3) holds and $c_n^{\kappa+1/2} \lesssim n^{\kappa} \sqrt{\log(n)}$, then

$$\hat{t} = t^* + O_{\mathbb{P}} \left(\frac{\log(n)}{\sqrt{c_n}} \right)^{1/\kappa} \,.$$

(b) If $t^* = \infty$ we have $\mathbb{P}(\hat{t} = \infty) = 1 - o(1)$.

Next we discuss the case, where there is a jump at the point t^* and assume for some $\epsilon > 0$ that

$$|\mu(t) - \mu_0^{t_0}| \begin{cases} <\Delta - \epsilon & \text{if } t < t^* \\ \ge \Delta & \text{if } t = t^* \\ \ge \Delta + O(t - t^*) & \text{if } \tilde{t} > t > t^* \end{cases}$$
(3.4)

where $\tilde{t} > t^*$ is the smallest point with a jump of the function μ at \tilde{t} (if there are no jumps for $t > t^*$ we set $\tilde{t} = 1$).

Theorem 3.2. Let Assumptions (A1) - (A3) be satisfied and let c_n satisfy $c_n^{3/2} \leq n\sqrt{\log(n)}$. Then

(a) If $t^* \in (t_0, 1]$ is a jump discontinuity satisfying (3.4), then

$$\hat{t} = t^* + O_{\mathbb{P}}\left(\frac{c_n}{n}\right).$$

(b) If $t^* = \infty$, we have $\mathbb{P}(\hat{t} = \infty) = 1 - o(1)$.

Comparing the Theorem 3.1 and 3.2, it is readily apparent that detecting smooth changes profits from a large c_n (i.e. the mean is only estimated over longer intervals) while abrupt changes are easier detected if c_n is chosen small (i.e. the mean is also estimated over shorter intervals).

4 Finite sample properties

In this section we investigate the finite sample properties of the proposed methodology by means of a simulation study and illustrate its application by analyzing a real data example.



Figure 1: Plot of the regression function $\mu_a(x)$ in (4.1) for a = 2. The dotted line is given by $\mu_0^{1/4} = 4 \int_0^{1/4} \mu_2(s) ds$.

For the sake of comparison we consider the same scenarios as investigated in (Bücher et al., 2021).

4.1 Synthetic Data

We choose $\Delta = 1$ and the mean function

$$\mu_a(x) = 10 + 1/2\sin(8\pi x) + a\left(x - \frac{1}{4}\right)^2 \mathbb{1}\left\{x > \frac{1}{4}\right\},\tag{4.1}$$

which is displayed in Figure 1 for a = 2. We consider various choices of the parameter a where we choose $t_0 = 1/4$ so that the hypotheses are given by

$$H_0(1): d_\infty \le 1 \quad vs \quad H_1(1): d_\infty > 1$$
 (4.2)

where

$$d_{\infty} = \sup_{t \in [1/4,1]} \left| \mu_a(t) - 4 \int_0^{1/4} \mu_a(s) ds \right|$$

Note that $d_{\infty} = 1$ (boundary of the hypotheses) for $a = \frac{128}{81}$ and that $d_{\infty} > 1$ (alternative) and $d_{\infty} < 1$ (interior of the null hypothesis), whenever $a > \frac{128}{81}$ and $a < \frac{128}{81}$, respectively.

For the error processes $(\epsilon_i)_{i\in\mathbb{Z}}$ in model (2.1) we investigate the processes

(IID)
$$\epsilon_i = \frac{1}{2}\eta_i,$$

(MA) $\epsilon_i = \frac{1}{\sqrt{5}}(\eta_i + \frac{1}{2}\eta_{i-1})$
(AR) $\epsilon_i = \frac{\sqrt{3}}{4}(\eta_i + \frac{1}{2}\epsilon_{i-1}),$ (4.3)

where $(\eta_i)_{i\in\mathbb{Z}}$ is an i.i.d. sequence of standard normally distributed random variables. In particular, we have $\operatorname{Var}(\epsilon_i) = \frac{1}{4}$ for all error processes under consideration. We will compare the novel testing procedures (2.13) and (2.15) proposed in this paper with the most powerful test from (Bücher et al., 2021) which is given in equation (4.6) therein. Throughout this section we generically choose m = 5 (tuning parameter for the long run variance estimation) and $c_n = 20$ (lower bound for scales in the multi-scale statistic (2.5)). The results are fairly stable under perturbation of these parameters as long as they are not chosen too small. The empirical rejection rates are calculated by 1000 simulation runs. For the test (2.13) we calculated the quantiles of the distribution of M by 1000 samples from a Brownian motion sampled on a grid with width 0.001. For the test (2.15) we used 200 samples to calculate the quantile $q_{1-\alpha^*}$ in (2.13) for each of the 1000 simulation runs.

The empirical rejection probabilities of all three tests are recorded in Table 1 and confirm the asymptotic theory. Regarding the interpretation of the empirical findings we note that the null hypothesis in (4.2) is true, whenever the parameter a satisfies $a \leq 128/81 \simeq 1.58$ and that we expect an increasing number of rejections for larger values of a, which yield increasing values for the difference $d_{\infty} - \Delta$. Note that all three tests are conservative in the sense that the empirical size is smaller than 5% at the boundary of the hypotheses (2.2) defined by $d_{\infty} - \Delta = 0$ (in boldface). However, it is worthwhile to mention that the test (2.15) provides a better approximation of the nominal level than its competitors.

μ_a	test	(2.13)		(2.15)			Bücher et al. (2021)			
a	$d_{\infty} - \Delta$	200	500	1000	200	500	1000	200	500	1000
Panel A: iid errors										
1.5	-0.03	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.58	0.00	0.0	0.1	0.2	0.2	1.5	0.2	0.0	0.0	0.2
2.0	0.13	0.6	2.5	3.2	5.5	14.1	15.4	0.0	3.3	23.1
2.5	0.29	10.4	42.6	50.3	42.1	73.6	80.2	0.0	29.9	97.8
3.0	0.45	54.3	93.8	99.7	86.6	99.4	100	0.2	57.3	100
Panel B: MA errors										
1.5	-0.03	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0
1.58	0.0	0.1	0.6	0.1	0.7	2.7	2.4	0.0	0.0	0.3
2.0	0.13	1.2	5.1	4.3	6.3	14.9	16.1	0.0	3.7	18.7
2.5	0.29	10.3	31.0	47.1	26.7	56.8	75.6	0.2	27.0	87.9
3.0	0.45	44.9	81.4	96.6	69.7	94.5	99.7	0.5	52.8	99.7
Panel C: AR errors										
1.5	-0.03	0.7	1.1	1.1	0.0	0.0	0.0	0.1	0.5	1.0
1.58	0.00	0.8	1.1	2.0	2.6	5.9	6.7	0.1	1.4	1.5
2.0	0.13	3.1	9.4	10.7	7.8	22.6	27.0	0.0	7.8	23.1
2.5	0.29	15.6	36.1	51.1	29.6	60.6	75.1	0.4	27.3	77.7
3.0	0.45	43.5	81.2	95.9	67.3	91.7	99.5	1.1	53.9	98.4

Table 1: Empirical rejection rates of the tests (2.13) and (2.15) and the test proposed in equation (4.6) of (Bücher et al., 2021) for the hypotheses (4.2). Different values for the parameter a in the mean function (4.1), error processes, and sample sizes n = 200, 500, 1000 are considered.

A comparison of the power properties of the different procedures shows that the conservative multiscale test (2.13) outperforms the test in (Bücher et al., 2021) for moderates samples sizes (n = 200, 500), but the last-named test yields larger rejection probabilities if the sample size n = 1000, in particular if the deviation is $d_{\infty} - \Delta = 2$. On the other hand the multi-scale test (2.15) yields an even larger power for sample size n = 200, 500, and for n = 1000 it shows a similar performance as the procedure in (Bücher et al., 2021).

Finally, we present a brief simulation study to investigate the performance of the estimator (3.2) for the first time of a relevant deviation as defined in (3.1), where we use m = 5 and $c_n = 20 + n^{1/2}$. We consider the mean function (4.1) with a = 2, thus the time of a first relevant deviation is $t^* = 0.791$. The error process is given by (4.3) and the sample size n = 500. In Figure 2 we display histograms of the estimator (3.2) and the estimator proposed in equation (5.2) of (Bücher et al., 2021) based on 10.000 simulation runs. We observe that the estimator introduced in (Bücher et al., 2021) does not detect a relevant deviation in more than 90% of all cases for many of the considered settings while our estimator detects

such a deviation almost always. However, due to the noise of the error process there exist also cases where \hat{t} underestimates the true point t^* and delivers an estimate for the local maximum of the function μ at t = 0.57 which is the closest local peak to the time t = 0.8 (here the deviation is $|\mu(0.57) - \mu_0^{1/4}| \simeq 0.695 < 1$.

We therefore refrain from a direct comparison of the tow estimators and display in Table 2 the empirical bias, standard deviation and the detection rate of the estimator (3.2) proposed in this paper.



Figure 2: Histograms for the estimator \hat{t} defined in (3.2) (left) and the estimator proposed in equation (5.2) of (Bücher et al., 2021) (right).

Note that all choices of the parameter a in Table 2 correspond to a violation of the null hypotheses. Therefore, a "good" estimator of the time point of the first relevant deviation should be finite in as many cases as possible. We observe that the estimator (3.2) proposed in this paper is finite almost always for all choices of a even for the sample size n = 200. The bias and standard deviation of the estimator first increase when the sample size increases to 500. This is due to the method sometimes detecting a change at the second highest peak of the the curve (see Figure (2)) when the sample size becomes larger. This happens more rarely when the sample size grows further, as is reflected by a lower bias for n = 1000 compared to n = 500. The estimator generally performs better for less dependent data but nonetheless performs well across all settings.

		$\hat{E}[1(t^* = \infty)]$							
$n \setminus a$	2.0	2.5	3.0	2.0	2.5	3.0			
Panel A: iid errors									
200	0.864(0.029)	0.839(0.036)	0.812(0.052)	0.031	0.002	0.000			
500	0.809(0.044)	$0.780\ (0.069)$	$0.738\ (0.090)$	0.001	0.000	0.000			
1000	0.803(0.023)	$0.783\ (0.046)$	$0.749\ (0.076)$	0.000	0.000	0.000			
Panel B: MA errors									
200	$0.854\ (0.054)$	$0.826\ (0.063)$	0.794(0.074)	0.057	0.029	0.000			
500	0.788(0.078)	$0.754\ (0.093)$	$0.711 \ (0.101)$	0.001	0.000	0.000			
1000	$0.789\ (0.057)$	$0.757\ (0.080)$	$0.715\ (0.095)$	0.000	0.000	0.000			
Panel C: AR errors									
200	0.835(0.084)	0.810(0.087)	0.778(0.090)	0.094	0.011	0.001			
500	0.764(0.108)	0.727(0.116)	0.693(0.111)	0.009	0.001	0.000			
1000	0.767(0.087)	$0.731 \ (0.099)$	0.692(0.104)	0.002	0.000	0.000			
t^*	0.79	0.78	0.77						

Table 2: Empirical bias, standard deviation and detection rate of the estimator \hat{t} defined in (3.2). Central part: empirical mean and standard deviation (in brackets) of \hat{t} , conditional on $t^* \neq \infty$. Right part: proportion of cases for which $\hat{t} = \infty$. Last line: true change point t^* .

4.2 Real Data Application

We consider the mean of daily minimal temperatures (in degree Celsius) over the month of July for different weather stations in Australia. The data set is available via the R package fChange (see Sonmez et al., 2025) on Github and the sample size varies between 100 and 150, depending on the weather station. For each weather station we test the hypotheses (2.2) for different thresholds $\Delta \in \{0.5, 1, 1.5\}$, where t_0 is chosen such that the years $1, ..., nt_0$ correspond to the time frame until the year 1950. In Table 3 we record the *p*-values of the multi-scale test (2.15) and the test proposed in equation (4.6) of (Bücher et al., 2021). For the test (2.15) these *p*-values are calculated by 1000 bootstrap repetitions, while the parameters *m* and c_n are chosen as in the previous section, that is $c_n = 20$ and m = 5.

Δ	0.5	1.0	1.5	0.5	1.0	1.5	
test	Büche	er et al.	(2021)	(2.15)			
Boulia/p-value	29.0	73.1	98.0	0.0	0.8	37.7	
Boulia/year	-	-	-	1950	1956	-	
Cape Otway/p-value	11.4	98.0	100.0	7.4	78.8	99.9	
Cape Otway/year	-	-	-	-	-	-	
Gayndah/p-value	0.2	1.4	3.2	0.0	0.0	0.3	
Gayndah/year	1952	1968	1974	1950	1950	1984	
Gunnedah/p-value	0.8	2.7	10.3	0.0	0.0	1.4	
Gunnedah/year	1952	1955	-	1962	1973	1984	
Hobart/p-value	94.9	100.0	100.0	34.1	95.5	100.0	
Hobart/year	-	-	-	-	-	-	
Melbourne/p-value	0.0	1.1	27.3	0.0	0.0	3.7	
Melbourne/year	1968	1976	-	1950	1972	1990	
Robe/p-value	3.3	44.9	98.2	6.1	68.3	99.8	
Robe/year	1953	-	-	-	-	-	
Sydney/p-value	41.1	98.1	100	0.05	0.14	89.9	
Sydney/year	-	-	-	1950	-	-	

Table 3: p-values and estimates for t^* of the test in Bücher et al. (2021) (left part) and the bootstrap test (2.15) proposed in this paper (right part) for the hypotheses (2.2) for various values of the threshold Δ .

Except for the Robe weather station the p values of the multi-scale test test (2.15) are either similar or substantially smaller than the p values obtained by the test in (Bücher et al., 2021). In particular the new test detects changes in Boulia, Melbourne and Sydney that the procedure from (Bücher et al., 2021) was not able to identify. We also observe that in general the new estimator \hat{t} proposed in this paper generally dates deviations earlier than its counterpart from (Bücher et al., 2021). The only exception is the station Robe, where the test from (Bücher et al., 2021) detects a difference of at least 0.5 degrees Celsius that our method does not detect at significance level $\alpha = 0.05$. However, a more precise look at the minimum value $\hat{\Delta}_{\alpha}$, for which $H_0(\Delta)$ is not rejected at a controlled type I error α (see equation (2.16) and equation (5.2) in (Bücher et al., 2021)) shows that the difference between the two tests are small: the method proposed in (Bücher et al., 2021) gives $\hat{\Delta}_{0.05} = 0.53$, while the estimator (2.16) yields $\hat{\Delta}_{0.05} = 0.49$.

These values are taken from Table 4, which displays the values $\hat{\Delta}_{0.05}$ for both methods (here we have recalculated the results of the test of (Bücher et al., 2021)). The results further confirm the previous findings. Except for the weather station in Robe the test (2.15) always detects larger differences than the test proposed in (Bücher et al., 2021). In particular we are able to detect changes in Boulia and Hobart where the method (Bücher et al., 2021) does not detect any relevant deviation. While at Hobart the difference in the value $\hat{\delta}_{0.05}$ is small, it is larger than 1 degree Celsius at Boulia.

$\hat{\Delta}_{0.05}$	Boulia	Cape Otway	Gayndah	Gunnedah	Hobart	Melbourne	Robe	Sydney
BüDH	0.00	0.38	1.69	1.22	0.00	1.17	0.53	0.13
BaD	1.16	0.45	1.74	1.64	0.17	1.52	0.49	0.87

Table 4: The minimum value $\hat{\Delta}_{0.05}$, for which $H_0(\Delta)$ is not rejected at a controlled type I error of 5% (see equation (2.16) and equation (5.2) in Bücher et al. (2021)).

Acknowldgements: This research was partially funded in the course of TRR 391 Spatiotemporal Statistics for the Transition of Energy and Transport (520388526) by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

5 Proofs

Throughout this section we use the notation $k_0 = \lfloor nt_0 \rfloor$ and define for any constant a > 0 the sets

$$\mathcal{E}_{c}(a) = \left\{ (j,k) \in \{k_{0},...,n\}^{2} \middle| k - j = c, \left| \mu_{0}^{t_{0}} - \mu_{j/n}^{k/n} \right| \ge d_{\infty} - a \log(n) / \sqrt{c} \right\}$$

$$\mathcal{E}_{c}^{co}(a) = \left\{ (j,k) \in \{k_{0},...,n\}^{2} \middle| k - j = c \right\} \setminus \mathcal{E}_{c}(a) .$$
(5.1)

We will generally suppress the dependence on a in the notation except for the subsection about the bootstrap procedure. In the following we shall assume that $\sigma = 1$, the general case follows by simple rescaling.

5.1 Proof of Theorem 2.2

Define

$$\bar{\mu}_{j}^{k} = \frac{1}{k-j} \sum_{i=j+1}^{k} \mu(i/n)$$

and denote by $\check{B}(s) = n^{-1/2}B(ns)$ a rescaled version of the Brownian motion *B*. We will start stating four auxiliary results, which will be used in the proof of Proof of Theorem 2.2 and of other results. The proofs are given at the end of this section.

Lemma 5.1. Grant assumption (A2). It then holds that

$$\sup_{1 \le j < k \le n} (k - j)(\bar{\mu}_j^k - \mu_{j/n}^{k/n}) = O(1)$$

Lemma 5.2. If $d_{\infty} > 0$ and assumption (A1) and (A2) are satisfied, we have with high probability that

$$\hat{T}_n \le \sup_{\substack{c_n \le c \le n-k_0 \ (j,k) \in \mathcal{E}_c}} \sup_{(j,k) \in \mathcal{E}_c} \mathfrak{s}(j/n,k/n) \left(\sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}}\right) - \Gamma_n(c)\right)$$

where B is a Brownian motion with variance σ^2 and suprema over empty sets are defined as $as -\infty$.

Lemma 5.3. If $d_{\infty} > 0$ and Assumption (A1) and (A2) are satisfied, we have

$$\sup_{\substack{c_n \leq c \leq n-k_0 \ (j,k) \in \mathcal{E}_c \\ c \in \mathbb{N}}} \sup_{\substack{s(j) \in \mathcal{A}_{\epsilon,d_{\infty}}}} \mathfrak{s}(s,t) \Big(\sqrt{t-s} \frac{\check{B}(k_0)}{k_0} - \frac{\check{B}(k) - \check{B}(j)}{\sqrt{c}} \Big) - \Gamma_n(c)$$

$$\leq \lim_{\epsilon \downarrow 0} \sup_{(s,t) \in \mathcal{A}_{\epsilon,d_{\infty}}} \mathfrak{s}(s,t) \Big(\sqrt{t-s} \frac{\check{B}(t_0)}{t_0} - \frac{\check{B}(t) - \check{B}(s)}{\sqrt{t-s}} \Big) - \Gamma(t-s) + o_{a.s.}(1),$$

where

$$\mathcal{A}_{\epsilon,d_{\infty}} = \left\{ (s,t) \in [t_0,1]^2 \left| s < t, \left| \mu_0^{t_0} - \mu_s^t \right| \ge d_{\infty} - \epsilon \right\} \right\}$$

Lemma 5.4. If $d_{\infty} > 0$ and Assumption (A1) and (A2) are satisfied, we have

$$\begin{split} \lim_{\epsilon \downarrow 0} \sup_{(s,t) \in \mathcal{A}_{\epsilon,d_{\infty}}} \mathfrak{s}(s,t) \Big(\sqrt{t-s} \frac{\mathring{B}(t_0)}{t_0} - \frac{\mathring{B}(t) - \mathring{B}(s)}{\sqrt{t-s}} \Big) - \Gamma(t-s) \\ \leq \sup_{\substack{c_n \leq c \leq n-k_0 \\ c \in \mathbb{N}}} \sup_{(j,k) \in \mathcal{E}_c} \mathfrak{s}(j/n,k/n) \Big(\sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} - \sqrt{c} \Big(\mu_0^{t_0} - \mu_{j/n}^{k/n} \Big) \Big) \\ - \Gamma_n(c) - \sqrt{c} d_{\infty} \Big) + o_{\mathbb{P}}(1) \end{split}$$

Proof of Theorem 2.2

The weak convergence of the statistic \hat{T}_n in (2.6) follows directly from Lemma 5.2 - 5.4. Regarding the continuity of the distribution of $T_{d_{\infty}}$ we note that $T_{d_{\infty}}$ is a limit of convex functions of a Gaussian process and therefore a convex function of a Gaussian process itself. The continuity then follows by Theorem 4.4.1 from (Bogachev, 2015).

The statements (1) and (2) regarding the asymptotic properties of the test statistic $T_{n,\Delta}$ are a direct consequence of (2.6) if $d_{\infty} \leq \Delta$. In the case $d_{\infty} > \Delta$ we note that piecewise Lipschitz continuity of μ yields that there exists a sequence of j_n, k_n with $k_n - j_n \simeq n$ such that for some $\rho > 0$ we have $|\mu(i/n) - \mu_0^{t_0}| \geq \Delta + \rho$ for all $j_n \leq i \leq k_n$. Consequently, using Lemma 5.1, we obtain

$$\left|\frac{1}{k_0}\sum_{i=1}^{k_0}\mu(i/n) - \frac{1}{c}\sum_{i=j_n+1}^{k_n}\mu(i/n)\right| \ge \Delta + \rho - O(n^{-1})$$

which yields $\hat{T}_{n,\Delta} \to \infty$ by an application of the triangle inequality.

5.1.1 Proof of Lemma 5.1 - 5.4

Proof of Lemma 5.1. Let us first assume that μ has no discontinuities, then

$$\begin{split} \bar{\mu}_{j}^{k} - \mu_{j/n}^{k/n} &= \frac{1}{k-j} \sum_{i=j+1}^{k} \left(\mu(i/n) - n \int_{(i-1)/n}^{i/n} \mu(t) dt \right) \\ &= \frac{1}{k-j} \sum_{i=j+1}^{k} n \int_{(i-1)/n}^{i/n} \mu(t) - \mu(i/n) dt \\ &\lesssim \frac{1}{k-j} \sum_{i=j+1}^{k} n^{-1} = n^{-1} \end{split}$$

The general case follows by splitting up the integrals containing the discontinuities, leading to finitely many additional terms in the sum that can be bounded only by a constant instead of n^{-1} .

Proof of Lemma 5.2. By Assumption (A1) and condition (2.4) we have

$$\sup_{n \ge k-j \ge c_n} \sqrt{k-j} |\hat{\mu}_0^{k_0} - \hat{\mu}_j^k| = \sup_{|k-j| \ge c_n} \frac{|B(k) - B(j)|}{\sqrt{k-j}} + \frac{O(n^{1/2-p})}{\sqrt{k-j}}$$
$$= \sup_{|k-j| \ge c_n} \frac{|B(k) - B(j)|}{\sqrt{k-j}} + o_{\mathbb{P}}(1) .$$

Using this and Lemma 5.1 we therefore obtain that

$$\begin{split} \hat{T}_n &= \sup_{\substack{c_n \leq c \leq n-k_0 \\ c \in \mathbb{N}}} \sup_{\substack{|k-j| = c \\ k > j \geq k_0}} \left(\left| \sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} - \sqrt{c} \left(\frac{1}{k_0} \sum_{i=1}^{k_0} \mu(i/n) - \frac{1}{c} \sum_{i=j+1}^k \mu(i/n) \right) \right| \\ &- \Gamma_n(c) - \sqrt{c} d_\infty \right) + o_{\mathbb{P}}(1) \\ &= \sup_{\substack{c_n \leq c \leq n-k_0 \\ c \in \mathbb{N}}} \sup_{\substack{|k-j| = c \\ k > j \geq k_0}} \left(\left| \sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} - \sqrt{c} \left(\mu_0^{t_0} - \mu_{j/n}^{k/n} \right) \right| \right. \\ &- \Gamma_n(c) - \sqrt{c} d_\infty \right) + o_{\mathbb{P}}(1) \end{split}$$

Now note that it follows from the discussion in Section 2.2 of (Frick et al., 2014) that the random variable M defined in (2.10) is finite with probability 1, which implies

$$\sup_{\substack{c_n \le c \le n-k_0 \ k>j \ge k_0 \\ k>j \ge k_0}} \sup_{\substack{k-j|=c \\ k>j \ge k_0}} \left| \sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} \right| - \Gamma_n(c) = O_{\mathbb{P}}(1).$$

This yields

$$\sup_{\substack{c_n \leq c \leq n-k_0 \ (j,k) \in \mathcal{E}_c^{co} \\ c \in \mathbb{N}}} \sup_{\substack{c_n \leq c \leq n-k_0 \ (j,k) \in \mathcal{E}_c^{co} \\ c \in \mathbb{N}}} \left(\left| \sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} - \sqrt{c} \left(\mu_0^{t_0} - \mu_{j/n}^{k/n} \right) \right| - \Gamma_n(c) - \sqrt{c} d_\infty \right)$$

$$\lesssim -\log(n)$$

with high probability by the definition of the set \mathcal{E}_c^{co} in (5.1). Therefore,

$$\begin{split} \sup_{\substack{c_n \leq c \leq n-k_0 \ |k-j| = c \\ c \in \mathbb{N}}} \sup_{\substack{k>j \geq k_0}} \left(\left| \sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} - \sqrt{c} \left(\mu_0^{t_0} - \mu_{j/n}^{k/n} \right) \right| \right) \\ &- \Gamma_n(c) - \sqrt{c} d_\infty \right) \\ = \sup_{\substack{c_n \leq c \leq n-k_0 \ (j,k) \in \mathcal{E}_c}} \sup_{\substack{(j,k) \in \mathcal{E}_c}} \left(\left| \sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} - \sqrt{c} \left(\mu_0^{t_0} - \mu_{j/n}^{k/n} \right) \right| \right) \\ &- \Gamma_n(c) - \sqrt{c} d_\infty \right) + o_{\mathbb{P}}(1) \\ = \sup_{\substack{c_n \leq c \leq n-k_0 \ (j,k) \in \mathcal{E}_c}} \sup_{\substack{(j,k) \in \mathcal{E}_c}} \mathfrak{s}(j/n, k/n) \left(\sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} - \sqrt{c} \left(\mu_0^{t_0} - \mu_{j/n}^{k/n} \right) \right) \\ &- \Gamma_n(c) - \sqrt{c} d_\infty \right) + o_{\mathbb{P}}(1) \\ \leq \sup_{\substack{c_n \leq c \leq n-k_0 \ (j,k) \in \mathcal{E}_c}} \sup_{\substack{(j/n, k/n) \ (\sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} - \sqrt{c} \left(\mu_0^{t_0} - \mu_{j/n}^{k/n} \right) \right)} \\ \leq \sup_{\substack{c_n \leq c \leq n-k_0 \ (j,k) \in \mathcal{E}_c}} \sup_{\substack{(j/n, k/n) \ (\sqrt{c} \frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{c}} - \Gamma_n(c) - \Gamma_n(c) \right)} \\ \end{cases}$$

with high probability, which yields the desired statement.

Proof of Lemma 5.3. Existence of the limit with respect to ϵ follows because the quantity is, as a function on the probability space, pointwise monotonically non-increasing in ϵ and bounded because the random variable M is finite almost surely. The asymptotic inequality follows because $\bigcup_{c_n \leq c \leq n-k_0} \mathcal{E}_c$ is, for any $\epsilon > 0$, eventually a subset of $\mathcal{A}_{\epsilon,d_{\infty}}$.

Proof of Lemma 5.4. The equality has been established in the proof of Lemma 5.2 already. For the upper bound we proceed in three steps:

(I) For any sequence with $b_n = o(n^{-1/2})$ we have

$$\sup_{\substack{(j,k)\in \bigcup_{c_n\leq c\leq n-k_0}\mathcal{E}_c}} \mathfrak{s}(j/n,k/n) \Big(\sqrt{k-j}\frac{B(k_0)}{k_0} - \frac{B(k) - B(j)}{\sqrt{k-j}} - \sqrt{k-j}\Big(\mu_0^{t_0} - \mu_{j/n}^{k/n}\Big)\Big) - \Gamma_n(k-j) - \sqrt{k-j}d_\infty\Big)$$

$$\geq \sup_{\substack{(s,t)\in A_{b_n}\cap\{k_0/n,\dots,1\}^2\\|s-t|\geq c_n/n}} \mathfrak{s}(s,t) \Big(\sqrt{t-s}\frac{\check{B}(t_0)}{t_0} - \frac{\check{B}(t) - \check{B}(s)}{\sqrt{t-s}} - \sqrt{(t-s)n}\Big(\mu_0^{t_0} - \mu_s^t\Big)\Big) - \Gamma(t-s) - \sqrt{(t-s)n}d_\infty\Big)$$

by definition of the involved sets and of \mathring{B} .

(II) By the definition of A_{b_n} we then obtain

$$\sup_{\substack{(s,t)\in A_{b_n}\cap\{k_0/n,\dots,1\}^2\\|s-t|\ge c_n/n}} \mathfrak{s}(s,t) \left(\sqrt{t-s}\frac{\check{B}(t_0)}{t_0} - \frac{\check{B}(t) - \check{B}(s)}{\sqrt{t-s}} - \sqrt{(t-s)n} \left(\mu_0^{t_0} - \mu_s^t\right)\right)$$
$$-\Gamma(t-s) - \sqrt{(t-s)n}d_{\infty}\right)$$
$$= \sup_{\substack{(s,t)\in A_{b_n}\cap\{k_0/n,\dots,1\}^2\\|s-t|\ge c_n/n}} \left(\mathfrak{s}(s,t) \left(\sqrt{c}\frac{\check{B}(t_0)}{t_0} - \frac{\check{B}(t) - \check{B}(s)}{\sqrt{t-s}}\right) - \Gamma(t-s)\right) + o(1)$$

(III) Using similar arguments as in the proof of Theorem 2.1 in (Dümbgen, 2002) (use Theorem 7.1 and Lemma 7.2 in (Dümbgen and Walther, 2008) with $\beta(x) = \mathbb{1}\{x \in [0, 1]\}$ instead of Proposition 7.1 from (Dümbgen, 2002)) we obtain

$$\sup_{\substack{(s,t)\in A_{b_n}\cap\{k_0/n,\dots,1\}^2\\|s-t|\ge c_n/n}} \left(\mathfrak{s}(j/n,k/n)\left(\sqrt{c}\frac{\check{B}(k_0)}{k_0} - \frac{\check{B}(k) - \check{B}(j)}{\sqrt{c}}\right) - \Gamma_n(c)\right) \quad (5.2)$$

$$= \sup_{(s,t)\in A_{b_n}} \left(\mathfrak{s}(j/n,k/n)\left(\sqrt{c}\frac{\check{B}(k_0)}{k_0} - \frac{\check{B}(k) - \check{B}(j)}{\sqrt{c}}\right) - \Gamma_n(c)\right) + o_{\mathbb{P}}(1)$$

To be precise we proceed as in the proof of Theorem 2.1 in (Dümbgen, 2002) to obtain, for

any $\delta_1 > 0$, a set A_1 with probability $1 - \delta_1$ on which there exists $\delta_2 > 0$ such that

$$\sup_{\substack{(s,t)\in A_{b_n}\cap\{k_0/n,\dots,1\}^2\\|s-t|\ge c_n/n}} \left(\mathfrak{s}(j/n,k/n)\left(\sqrt{c}\frac{\check{B}(k_0)}{k_0} - \frac{\check{B}(k) - \check{B}(j)}{\sqrt{c}}\right) - \Gamma_n(c)\right)$$
$$= \sup_{\substack{(s,t)\in A_{b_n}\cap\{k_0/n,\dots,1\}^2\\|s-t|\ge \delta_2}} \left(\mathfrak{s}(j/n,k/n)\left(\sqrt{c}\frac{\check{B}(k_0)}{k_0} - \frac{\check{B}(k) - \check{B}(j)}{\sqrt{c}}\right) - \Gamma_n(c)\right)$$

holds. Exactly the same arguments also yield a set A_2 with probability $1 - \delta_1$ on which

$$\sup_{\substack{(s,t)\in A_{b_n}}} \left(\mathfrak{s}(j/n,k/n)\left(\sqrt{c}\frac{\check{B}(k_0)}{k_0} - \frac{\check{B}(k) - \check{B}(j)}{\sqrt{c}}\right) - \Gamma_n(c)\right)$$
$$= \sup_{\substack{(s,t)\in A_{b_n}\\|s-t|\geq \delta_2}} \left(\mathfrak{s}(j/n,k/n)\left(\sqrt{c}\frac{\check{B}(k_0)}{k_0} - \frac{\check{B}(k) - \check{B}(j)}{\sqrt{c}}\right) - \Gamma_n(c)\right)$$

holds. Equation (5.2) then follows by a standard argument involving the uniform continuity of the process $(s,t) \rightarrow \frac{B(t)-B(s)}{\sqrt{t-s}}$ on the set $\{(s,t)||s-t| \ge \delta_2\}$.

The Lemma then by the fact that

$$\sup_{(s,t)\in\mathcal{A}_{\epsilon,d_{\infty}}}\mathfrak{s}(s,t)\Big(\sqrt{t-s}\frac{\mathring{B}(t_{0})}{t_{0}}-\frac{\mathring{B}(t)-\mathring{B}(s)}{\sqrt{t-s}}\Big)-\Gamma(t-s)$$

is a decreasing function of ϵ .

5.2 Proof of Theorem 2.4

The proof follows by a straightforward modification of Lemmas 5.2 to 5.4. The key difference is that the quantity

$$\sqrt{k-j} \left(\left(\mu_0^{t_0} - \mu_{j/n}^{k/n} \right) - \Delta \right)$$

is not upper bounded by (or in some cases converges to) 0 anymore. Instead on uses the expansion

$$\left(\mu_0^{t_0} - \mu_{j/n}^{k/n}\right) - \Delta = \Delta + \frac{\beta_n}{k/n - j/n} \int_{j/n}^{k/n} h(x) dx - \Delta$$
$$= \frac{\beta_n}{k/n - j/n} \int_{j/n}^{k/n} h(x) dx$$

in the last inequality in the proof of Lemma 5.2 and in step (II) of the proof of Lemma 5.4. $\hfill \Box$

5.3 Proof of Theorem 2.6

We define the quantity

$$\hat{\mathcal{E}}_{c}(a) = \left\{ (j,k) \in \{k_{0},...,n\}^{2} \middle| k-j = c, \left| \hat{\mu}_{0}^{k_{0}} - \hat{\mu}_{j}^{k} \right| \ge \Delta - a \log(n) / \sqrt{c} \right\}.$$

and prove at the end of this section the following auxiliary result.

Lemma 5.5. Let $\Delta = d_{\infty}$. For any fixed a > 0 we have for n large enough that

$$\mathcal{E}_c(a-\epsilon) \subset \hat{\mathcal{E}}_c(a) \subset \mathcal{E}_c(a+\epsilon)$$

for all $c \geq c_n$.

Proof of Theorem 2.6. Let us first consider the case $d_{\infty} = \Delta$. By the consistency of the long run variance estimate $\hat{\sigma}^2$ (see Section 5.4 for a proof) we have that

$$\hat{\mathcal{E}}_c(2\sigma) \subset \hat{\mathcal{E}}_c \subset \hat{\mathcal{E}}_c(\sigma/2)$$

with high probability. Lemmas 5.5 and 5.3 then yield the desired statement if we can show that $\hat{\mathfrak{s}}(j/n, k/n) = \mathfrak{s}(j/n, k/n)$ holds with high probability uniformly over k, j with $(j/n, k/n) \in \mathcal{A}_{\epsilon, d_{\infty}}$ for some $\epsilon < \Delta$. This is an easy consequence of Lemma 5.1 and equation (5.3).

For the other cases we note that the statistic \hat{T}_n^* is stochastically bounded because the random variable M defined in (2.10) is almost surely finite. As $\hat{T}_{n,\Delta} \to -\infty$ (∞) if $d_{\infty} < \Delta$ (> Δ) the assertion follows.

Proof of Lemma 5.5. By Assumption (A1) and a union bound it follows that the inequality

$$|\bar{\mu}_{j}^{k} - \hat{\mu}_{j}^{k}| = \left| (k-j)^{-1} \sum_{i=j+1}^{k} \epsilon_{i} \right|$$

$$\lesssim \left| (k-j)^{-1} (B(k) - B(j)) \right| + n^{1/2-p} / (k-j)$$

$$\lesssim \frac{\sqrt{\log(n)}}{\sqrt{k-j}} + n^{1/2-p} / (k-j)$$
(5.3)

holds uniformly with respect to $1 \leq k-j \leq n$ with high probability. As a consequence we have

$$\mathbb{P}\left(|\bar{\mu}_{j}^{k} - \hat{\mu}_{j}^{k}| < \epsilon \log(n)/\sqrt{k-j}, c_{n} \leq k-j \leq n\right)$$

$$\geq \mathbb{P}\left(\frac{\sqrt{\log(n)}}{\sqrt{k-j}} + \frac{n^{1/2-p}}{(k-j)} < \epsilon \log(n)/\sqrt{k-j}, c_{n} \leq k-j \leq n\right).$$
(5.4)

Now condition (2.4) implies uniformly with resepcet to $c_n \leq k - j \leq n$ that

$$\frac{n^{1/2-p}}{k-j} \le \frac{n^{1/2-p}}{\sqrt{c_n}\sqrt{k-j}} \lesssim \frac{o(1)}{\sqrt{k-j}}$$

which then implies that the probability in (5.4) is of order 1 - o(1). This yields the desired set inclusions as they are true whenever

$$|\bar{\mu}_j^k - \hat{\mu}_j^k| < \epsilon \log(n) / \sqrt{k - j} , \quad c_n \le k - j \le n .$$

5.4 Consistency of the long run variance estimate

Lemma 5.6. Under assumption (A1) and (A2) we have

$$\hat{\sigma}^2 = \sigma^2 + O_{\mathbb{P}}(n^{-1/3}) \ .$$

Proof. By assumption (A1) we have that

$$\hat{\sigma}^2 = \frac{1}{\lfloor n/m \rfloor - 1} \sum_{i=1}^{\lfloor n/m \rfloor - 1} \frac{\left(2B(im) - B((i-1)m) - B((i+1)m) + A_i\right)^2}{2m} + o_{\mathbb{P}}(n^{-1/3})$$

where

$$A_i := \bar{\mu}_{(i-1)m}^{im} - \bar{\mu}_{im}^{(i+1)m}$$
.

By Lemma 5.1 we have

$$|A_i| = \frac{n}{m} \left| \mu_{(i-1)m/n}^{im/n} - \mu_{im/n}^{(i+1)m/n} \right| + O(m^{-1})$$

\$\le O(1)\$

where the inequality follows by the Lipschitz continuity of μ_s^t in s and t. Standard arguments then yield

$$\hat{\sigma}^2 = \frac{1}{\lfloor n/m \rfloor - 1} \sum_{i=1}^{\lfloor n/m \rfloor - 1} \frac{\left(2B(im) - B((i-1)m) - B((i+1)m)\right)^2}{2m} + O_{\mathbb{P}}(n^{-1/3})$$

which in turn yields the desired statement by noting that $Z_{im} = B(im) - B((i-1)m)$ is a triangular array of independent $\mathcal{N}(0, \sigma^2)$ variables.

5.5 Proof of Theorem 3.1 and 3.2

Proof of Theorem 3.1 We only consider the case $t^* < \infty$, the case $t^* = \infty$ follows by easier and analogous arguments. We give an upper and a lower bound which establish the desired result upon combining them.

Upper bound: We first note that it follows by Assumption (A1), condition (2.4) and the fact that the random variable M is almost surely finite that

$$\sup_{c_n \le |k-j| \le n} \left| |\hat{\mu}_0^{k_0} - \hat{\mu}_j^k| - |\bar{\mu}_0^{k_0} - \bar{\mu}_j^k| \right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log(n)}{c_n}}\right).$$
(5.5)

We also note that the identity

$$\mu(t^*) - \mu_0^{t_0} = \Delta$$

implies

$$\bar{\mu}_{j_n}^{k_n} - \mu_0^{t_0} \ge \frac{1}{k_n - j_n} \sum_{i=j_n+1}^{k_n} (\Delta - C(t - i/n)^\kappa) = \Delta - \frac{C}{k_n - j_n} \sum_{i=j_n+1}^{k_n} (t - i/n)^\kappa$$

for $k_n = \lfloor nt \rfloor$ and $j_n = k_n - c_n$, where C is some constant that only depends on μ and c_{κ} . We thus obtain

$$\bar{\mu}_{j_n}^{k_n} - \mu_0^{t_0} \gtrsim \Delta - (c_n/n)^{\kappa} ,$$
(5.6)

and a similar argument is valid when $\mu(t) - \mu_0^{t_0} = -\Delta$. Combining (5.5) and (5.6) therefore yields

$$|\hat{\mu}_0^{k_0} - \hat{\mu}_{j_n}^{k_n}| \gtrsim \Delta - (c_n/n)^{\kappa} - O_{\mathbb{P}}\left(\sqrt{\frac{\log(n)}{c_n}}\right) , \qquad (5.7)$$

which implies that $\hat{t} \leq t^*$ holds with high probability.

Lower bound: We give the argument for $\mu(t^*) - \mu_0^{t_0} = \Delta$, the other case follows analogously. By (3.3) we know that

$$\mu(t^* - x) = \Delta - c_{\kappa} x^{\kappa} + o(x^{\kappa}) \; .$$

Thereby, choosing $x = \left(\frac{3\sigma \log(n)}{c_{\kappa}\sqrt{c_n}}\right)^{1/\kappa}$, we have

$$\mu(t^* - x) - \mu_0^{t_0} = \Delta - 3\sigma \log(n) / \sqrt{c_n} + o\left(\sqrt{\log(n)/c_n}\right)$$

Consequently, using the continuity of $\mu(t) - \mu_0^{t_0}$ and Assumption (A3) we obtain

$$\max_{t \in [t_0, t^* - x]} |\mu(t^* - x) - \mu_0^{t_0}| \le \Delta - 2\sigma \log(n) / \sqrt{c_n}$$

when n is sufficiently large. Using similar arguments as in the derivation of the upper bound we therefore obtain

$$\sup_{k/n \in [t_0, t^* - x]} \sup_{j < k, k - j \ge c_n} |\hat{\mu}_0^{k_0} - \hat{\mu}_j^k| \le \Delta - 2\sigma \log(n) / \sqrt{c_n} + O_{\mathbb{P}}(\sqrt{\log(n)/c_n}) .$$
(5.8)

Now suppose that there exists $j_n < k_n$, satisfying $k_n - j_n \ge c_n, k_n/n \in [t_0, t^* - x]$, such that

$$|\hat{\mu}_0^{k_0} - \hat{\mu}_{j_n}^{k_n}| \ge \Delta - \frac{\hat{\sigma}\log(n)}{\sqrt{k_n - j_n}}$$

holds. By (5.8) this would imply

$$\Delta - 2\sigma \log(n) / \sqrt{c_n} + O_{\mathbb{P}}(\sqrt{\log(n)/c_n}) \ge \Delta - \frac{\hat{\sigma}\log(n)}{\sqrt{c_n}}$$

which happens only with probability o(1) by Lemma 5.6. Consequently we have that $\hat{t} \ge t^* - x$ with high probability, i.e.

$$\hat{t} \ge t^* - O_{\mathbb{P}} \Big(\frac{\log(n)}{\sqrt{c_n}} \Big)^{1/\kappa}$$

as desired.

5.5.1 Proof of Theorem 3.2

Again we only consider the case $t^* < \infty$ and note that the case $t^* = \infty$ follows by easier and analogous arguments. We give an upper and a lower bound which establish the desired result upon combination.

Lower bound:

Consider any $t < t^*$ and assume WLOG that $\mu(t^*) - \mu_0^{t_0} = \Delta$. Then, by (3.4), we have

$$\bar{\mu}_j^k - \mu_0^{t_0} < \Delta - \epsilon$$

Equation (5.5) then yields that $\hat{t} \ge t^*$ with high probability.

Upper bound:

Assume WLOG (last inquality of (3.4)) that for some $\delta > 0$ we have for any $t^* \le t \le t^* + \delta$ that

$$\mu(t) - \mu_0^{t_0} \ge \Delta + O(t - t^*)$$
.

Consequently, by the same arguments leading to (5.7), we have

$$|\hat{\mu}_{0}^{k_{0}} - \hat{\mu}_{k_{0}}^{k_{0}+c_{n}}| \gtrsim \Delta - c_{n}/n - O_{\mathbb{P}}\left(\sqrt{\frac{\log(n)}{c_{n}}}\right) ,$$

which yields $\hat{t} \leq t^* + c_n/n$.

References

- Aston, J. A. D. and Kirch, C. (2012). Evaluating stationarity via change-point alternatives with applications to fMRI data. *The Annals of Applied Statistics*, 6(4):1906 1948.
- Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series* Analysis, 34(1):1–16.
- Aue, A. and Kirch, C. (2023). The state of cumulative sum sequential changepoint testing 70 years after page. *Biometrika*, 111(2):367–391.
- Baranowski, R., Chen, Y., and Fryzlewicz, P. (2019). Narrowest-Over-Threshold Detection of Multiple Change Points and Change-Point-Like Features. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(3):649–672.
- Berkes, I., Hörmann, S., and Schauer, J. (2011). Split invariance principles for stationary processes. *The Annals of Probability*, 39(6):2441 2473.

- Bogachev, V. (2015). *Gaussian Measures*. Mathematical Surveys and Monographs. American Mathematical Society.
- Bücher, A., Dette, H., and Heinrichs, F. (2021). Are deviations in a gradually varying mean relevant? a testing approach based on sup-norm estimators. *The Annals of Statistics*, 49(6):3583 – 3617.
- Chen, Y., Wang, T., and Samworth, R. J. (2022). High-dimensional, multiscale online changepoint detection. J. R. Stat. Soc. Ser. B. Stat. Methodol., 84(1):234–266.
- Cho, H. and Kirch, C. (2024). Data segmentation algorithms: Univariate mean change and beyond. *Econometrics and Statistics*, 30:76–95.
- Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Dehling, H. (1983). Limit theorems for sums of weakly dependent banach space valued random variables. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 63:393– 432.
- Dette, H., Eckle, T., and Vetter, M. (2020). Multiscale change point detection for dependent data. *Scandinavian Journal of Statistics*, 47(4):1243–1274.
- Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale Testing of Qualitative Hypotheses. The Annals of Statistics, 29(1):124 – 152.
- Dümbgen, L. (2002). Application of local rank tests to nonparametric regression. *Journal* of Nonparametric Statistics, 14(5):511–537.
- Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. The Annals of Statistics, 36(4):1758–1785.
- Frick, K., Munk, A., and Sieling, H. (2014). Multiscale change point inference. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 76(3):495–580.
- Fryzlewicz, P. and Rao, S. (2013). Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76.
- Gao, J., Gijbels, I., and Van Bellegem, S. (2008). Nonparametric simultaneous testing for structural breaks. *Journal of Econometrics*, 143(1):123–142. Specification testing.
- Gao, Z., Shang, Z., Du, P., and Robertson, J. (2018). Variance change point detection under a smoothly-changing mean trend with application to liver procurement. *Journal of the American Statistical Association*, 114:1–29.
- Gijbels, I., Lambert, A., and Qiu, P. (2007). Jump-preserving regression and smoothing using local linear fitting: A compromise. Annals of the Institute of Statistical Mathematics, 59:235–272.

- Horváth, L. and Rice, G. (2024). *Change point analysis for time series*. Springer Series in Statistics. Springer Nature,.
- Hotz, T., Schütte, O. M., Sieling, H., Polupanow, T., Diederichsen, U., Steinem, C., and Munk, A. (2013). Idealizing ion channel recordings by a jump segmentation multiresolution filter. *IEEE Transactions on NanoBioscience*, 12(4):376–386.
- Karl, T., Kniught, R., and Plummer, N. (1995). Trends in high frequency climate variability in the twentieth century. *Nature*, 377:217–220.
- Li, W., Wang, D., and Rinaldo, A. (2023). Divide and conquer dynamic programming: An almost linear time change point detection methodology in high dimensions. In *International Conference on Machine Learning*, pages 20065–20148. PMLR.
- Madrid Padilla, O. H., Yu, Y., Wang, D., and Rinaldo, A. (2022). Optimal nonparametric multivariate change point detection and localization. *IEEE Trans. Inform. Theory*, 68(3):1922–1944.
- Müller, H.-G. (1992). Change-Points in Nonparametric Regression Analysis. The Annals of Statistics, 20(2):737 – 761.
- Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3-4):523-527.
- Schmidt-Hieber, J., Munk, A., and Dümbgen, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *The Annals* of Statistics, 41(3):1299 – 1328.
- Sonmez, O., Aue, A., and Rice, G. (2025). fChange: Change Point Analysis in Functional Data. R package version 0.2.0.
- Truong, C., Oudre, L., and Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, 167:107299.
- Vogt, M. and Dette, H. (2015). Detecting gradual changes in locally stationary processes. The Annals of Statistics, 43(2):713–740.
- Woodall, W. H. and Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31(4):376–386.
- Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. Proceedings of the National Academy of Sciences, 102(40):14150–14154.
- Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(3):391–410.