At the edge of Donsker's Theorem: Asymptotics of multiscale scan statistics

Johann Köhne[1] and Fabian Mies[2]

[1]University of Göttingen
[2]Delft University of Technology

Abstract

Bandwidth selection is a critical aspect of statistical tests based on nonparametric regression or density estimation, and multiscale test procedures address this issue by incorporating all bandwidths simultaneously. Remarkably, this approach avoids a severe multiple testing penalty and instead offers a "free lunch" in statistical inference: the test achieves asymptotically optimal detection power across all scales without prior knowledge of the effect's length scale. However, constructing valid multiscale procedures is highly dependent on the sampling setting and often relies on bounding unknown and difficult-to-identify quantities, such as exponential tail bounds.

In this work, we develop a feasible multiscale test for a broad range of sampling settings via weak convergence arguments, by replacing the additive multiscale penalty with a multiplicative weighting. Specifically, we derive tightness conditions for the functional central limit in Hölder spaces with a critical modulus of continuity, where Donsker's theorem fails to hold. Probabilistically, we discover a novel form of restricted weak convergence that holds only in the tail of the distribution. This new theoretical foundation preserves the optimal detection properties of multiscale tests and extends their applicability to nonstationary nonlinear time series via a tailored bootstrap scheme. Applications to signal discovery, goodness-of-fit testing of regression functions, and multiple changepoint detection are studied in detail. By offering a fresh perspective on multiscale statistics, we aim to facilitate their adaptation to a broader range of statistical problems.

1 Introduction

The fundamental problem of signal discovery, or anomaly detection, may be modeled as

$$Y_t = f_n(t) + \eta_t, \quad t = 1, \ldots, n, \tag{1}$$

for iid centered random variables η_t with variance $\sigma^2 = \mathrm{Var}(\eta_t)$, and a regression function of the form $f_n(t) = \mu_n \mathbb{1}(a_n < t \leq b_n)$. The statistical problem is to detect if a signal is present, that is, to test the null hypothesis $\mathbb{H}_0 : \mu_n = 0$ versus $\mathbb{H}_1 : \mu_n \neq 0$. An established statistical procedure, both in theory and practice (Glaz et al., 2009), is based on the local scan statistic $T_n(I) = |\sum_{t\in I} Y_t|/\sqrt{|I|}$ for any interval $I = (a, b]$ with $0 \leq a < b \leq n$. If $a, b \notin \mathbb{N}$, the sum may be extended by linear interpolation. The null hypothesis is rejected for large values of the global scan statistic $T_n^{\text{SCAN}} = \max_I T_n(I)$. As concisely reviewed by Walther and Perry (2022), the statistic T_n^{SCAN} will be dominated by the small intervals I, and accordingly has an extreme value limit distribution of Gumbel type for standard Gaussian errors $\eta_t \sim \mathcal{N}(0, 1)$ (Sharpnack and Arias-Castro, 2016). In particular, $T_n^{\text{SCAN}} = \sqrt{2 \log n} + o_P(1)$, and thus the signal with size μ_n and length $l_n = b_n - a_n$ can only be detected consistently if $|\mu_n|^2 l_n > 2 \log n$. Due to the dominance of short intervals, this threshold is asymptotically suboptimal for longer signals of length $l_n \propto n$. In fact, Dümbgen and Spokoiny (2001) show that a signal may be consistently detected if $\mu_n^2 l_n > 2 \log(\frac{en}{l_n})$, which is achieved by a multiscale statistic of the form

$$T_n^{\text{DS}} = \max_I \left\{ T_n(I) - \sigma\sqrt{2 \log \frac{en}{|I|}} \right\}_+, \tag{DS}$$

where $x_+ = \max(x, 0)$. The asymptotic distribution of T_n^{DS} with Gaussian errors is accurately described by Dümbgen and Spokoiny (2001) and critical values may be computed. The concept of multiscale test statistics has since found multiple further applications, in particular in changepoint inference (Frick et al., 2014; Fryzlewicz, 2024a), shape inference for a density from direct measurements (Dümbgen and Walther, 2008) or via deconvolution (Schmidt-Hieber et al., 2013), and rank-based tests (Rohde, 2008).

The additive penalty for the multiscale statistic and the corresponding critical values are only valid for Gaussian noise. The latter setting is often considered as prototypical, as methods developed for the the Gaussian case can usually be applied to non-Gaussian data by virtue of the central limit theorem. However, for the multiscale statistic T_n^{DS}, the situation is more involved: The following proposition shows that if the noise has only slightly heavier, but still sub-Gaussian tails, the statistical inference is invalidated.

Proposition 1.1. *Let $\mu_n = 0$, i.e. no signal, and $\eta_t = \epsilon_t Z_t$ for $Z_t \sim \mathcal{N}(0, 1/p)$ and $\epsilon_t \sim Bin(1, p)$ for some $p \in (0, 1)$, such that* $\text{Var}(\eta_t) = 1$. *Then $T_n^{\text{DS}} \to \infty$ in probability.*

This finding is a direct consequence of Proposition 3.2 below. It reveals that for multiscale testing, one can not directly rely on asymptotic arguments to reduce the non-Gaussian case to the Gaussian. As a remedy, it has been suggested to impose a lower bound $|I| > h_n$ on the length of the considered intervals in (DS) which accounts for the speed of convergence of the central limit theorem for $T_n(I)$, and study the statistic $T_{n,h_n}^{\text{DS}} = \max_{|I|\geq h_n}\{T_n(I) - \sigma\sqrt{2 \log(en/|I|)}\}_+$. For sub-exponential errors, Schmidt-Hieber et al. (2013); Frick et al. (2014) impose $h_n \gg \log^3(n)$, and König et al. (2020) require $h_n \gg \log(n)^{12}$. For auto-correlated errors with polynomial tails, Dette et al. (2020) and Khismatullina and Vogt (2020) require $h_n \gg n^q$ for some exponent $q \in (0, 1)$ depending on the order of the tail bound and the decay of temporal dependence. As Proposition 1.1 shows, choosing h_n too small will invalidate the statistical analysis, but this theoretical lower bound is not known in practice. Thus, h_n will necessarily be chosen too big, sacrificing power against short signals. Even in the idealized situation where a sharp and admissible h_n is known, the test lacks power in the regime $l_n \ll h_n$. We show that the loss of power can be expressed in terms of the ratio l_n/h_n.

Proposition 1.2. *Suppose that η_t are iid standard Gaussian, and $l_n \leq h_n$. If $\mu_n^2 l_n(\frac{l_n}{h_n}) \gg \log(\frac{en}{h_n})$, then $T_{n,h_n}^{DS} \to \infty$ and the test is consistent. On the other hand, if $\mu_n^2 l_n(\frac{l_n}{h_n}) = \mathcal{O}(1)$, then $T_{n,h_n}^{DS} = \mathcal{O}_P(1)$ and the test fails to be consistent.*

To summarize, the multiscale statistic T_n^{DS} of Dümbgen and Spokoiny (2001) does not maintain the size for non-Gaussian errors, and the truncated version T_{n,h_n}^{DS} is suboptimal for short signals.

The contribution of this paper is to develop a broadly applicable asymptotic theory for multiscale statistics of non-Gaussian data, giving rise to a feasible and optimal multiscale testing procedure in a wide range of sampling settings. Instead of the additively penalized statistic T_n^{DS}, we suggest to use the multiplicatively weighted statistic

$$T_n^* = \max_I \frac{T_n(I)}{\sqrt{\log e \frac{n}{|I|}}}.$$

We show that T_n^*/σ is asymptotically pivotal under the assumption of sub-Gaussian errors. Importantly, the statistical methodology is agnostic of the exact tail bound and only uses the noise variance, which can be estimated reliably. Our probabilistic results are based on the analysis of edge cases in Donsker's theorem in Hölder-type spaces. Introducing the interpolated partial sum process $S_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} Y_t + \frac{un - \lfloor un \rfloor}{\sqrt{n}} Y_{\lfloor un \rfloor + 1}$, and the modulus of continuity $\rho_2(h) = \sqrt{h \log \frac{e}{h}}, h \in (0,1)$, the statistic T_n^* may be expressed as the Hölder-type seminorm

$$T_n^* = |S_n|_{\rho_2} = \sup_{u,v \in [0,1]} \frac{|S_n(u) - S_n(v)|}{\rho_2(|u-v|)}. \tag{2}$$

Donsker's theorem (Billingsley, 1999) yields weak convergence of $S_n(u)$ towards a Brownian motion $\sigma B(u)$ in $C[0,1]$, whereas the treatment of (2) requires weak convergence in the stronger Hölder-type space C^{ρ_2}. In view of Donsker's theorem in Hölder spaces (Lamperti, 1962; Hamadouche, 1998; Račkauskas and Suquet, 2004b,a; Račkauskas and Wendler, 2020), the modulus ρ_2 presents an intriguing special case: the limiting Brownian motion is stochastically bounded in C^{ρ_2}, but not tight. Thus, Prokhorov's Theorem can not be used to show weak convergence in this space, which indeed does not even hold. Surprisingly, it is possible to show that $P(|S_n|_{\rho_2}/\sigma > t) \to P(|B|_{\rho_2} > t)$, but only for sufficiently large $t > t_0$, and we can provide exact bounds on t_0 in terms of the tails of η_t. The lower bound on t is not merely a deficiency of our proof, but the weak convergence actually does not hold below this threshold; see Proposition 2.5. We term this novel phenomenon *thresholded weak convergence*, and we study its properties in more detail in Section 2. The implication of this probabilistic result for statistical inference is that we may choose critical values based on the pivotal limit distribution of $|B|_{\rho_2}$, and for sufficiently small significance level $\alpha \leq \alpha_0$, these critical values will be asymptotically valid. We stress that α_0 does not vanish as $n \to \infty$, but is a fixed value.

Our new results on thresholded weak convergence are broadly applicable beyond the prototypical signal discovery problem (1), and beyond iid observations. In Section 2, we in particular showcase how to apply our results to nonstationary time series. We also derive a novel sub-Gaussian concentration inequality for nonlinear time series (Theorem 2.6), which might be of independent interest. After discussing the signal discovery problem in detail in Section 3.1, we formulate a general goodness-of-fit test in Section 3.2, and propose novel inferential procedures for multiple changepoint detection in Section 3.3.

The multiscale test statistic T_n^{DS} has been criticized by Walther and Perry (2022) for loosing too much finite sample power against short signals. We address this criticism via variants of test statistics $|S_n|_\rho$ for the signal discovery problem based on alternative moduli ρ such that $\rho(h)/\rho_2(h) \to 1$ as $h \to 0$. Thus, the procedure maintains its optimal asymptotic properties, while being more sensitive to short signals. Simulations presented in Section 3.1 confirm the improved finite sample detection performance for signals of different lengths. An extended simulation study is presented in Section **??**, where we assess the distributional approximation under the null hypothesis and showcase our methodology on a simulated data set as well as observational data of a substation of the Dutch power grid. All mathematical proofs are gathered in the Appendix.

Notation

For two sequences a_n, b_n of real numbers, we denote $a_n \ll b_n$ to mean $a_n/b_n \to 0$ as $n \to \infty$. We denote weak convergence of a random element by $\Rightarrow$, weak convergence of the finite dimensional marginals of a sequence of stochastic processes by $\Rightarrow_{\mathrm{fidi}}$, and thresholded weak convergence of random variables by $\Rightarrow_\tau$ (introduced in Section 2). The space of continuous functions on $[0,1]$ is denoted by $C[0,1]$. A modulus of continuity is a continuously increasing function $\rho : [0,1] \to (0,\infty)$ with $\rho(0) = 0$, and the corresponding Hölder-type seminorm on $C[0,1]$ is $|f|_\rho = \sup_{u,v\in[0,1]} |f(u) - f(v)|/\rho(|u-v|)$, inducing the Hölder-type Banach space $C^\rho = \{f \in C[0,1] : \|f\|_\rho = \|f\|_\infty + |f|_\rho\}$. Of particular relevance are the special cases $\rho_\alpha(h) = \sqrt{h}\log^{\frac{1}{\alpha}}(e/h)$, for $\alpha \in [0,2]$, and $\rho_{2,a}(h) = \sqrt{h(a + \log(e/h))}$. For a given modulus of continuity we denote by C_0^ρ the subspace of functions $f \in C^\rho$ such that

$$\lim_{\delta\to 0} \sup_{0<|u-v|<\delta} \frac{|f(u) - f(v)|}{\rho(|u-v|)} = 0.$$

For two moduli of continuity ρ_1 and ρ_2 we denote by $\rho_1 \gg \rho_2$ that $\rho_2(h)/\rho_1(h) \to 0$ as $h \to 0$. The interpolated partial sum process based on Y_t is denoted as $S_n(u) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor un\rfloor} Y_t + \frac{un-\lfloor un\rfloor}{\sqrt{n}} Y_{\lfloor un\rfloor+1}$, and $\widetilde{S}_n(u)$ denotes the interpolated partial sum process based on the noise variables η_t, that is $\tilde{S}_n(u) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor un\rfloor} \eta_t + \frac{un-\lfloor un\rfloor}{\sqrt{n}}\eta_{\lfloor un\rfloor+1}$.

2 Asymptotic theory

Central to our statistical methodology is a detailed asymptotic treatment of the interpolated partial sum process of the random variables η_t,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor un\rfloor}\eta_t + \frac{un-\lfloor un\rfloor}{\sqrt{n}}\eta_{\lfloor un\rfloor+1}.$$

The process $\widetilde{S}_n$ is a random element in the Polish space $C[0,1]$, and for iid η_t, Donsker's Theorem (Billingsley, 1999) establishes its weak convergence towards a Brownian motion, $\widetilde{S}_n \Rightarrow \sigma B$. The continuous mapping theorem implies that $T(\widetilde{S}_n) \Rightarrow T(\sigma B)$ for any continuous functional $T : C[0,1] \to \mathbb{R}$. In an inferential setting, $T(S_n)$ is a test statistic, and the continuity requirement restricts the class of possible statistics. Since the paths of $u \mapsto \widetilde{S}_n(u)$ are highly regular as piecewise linear functions, one may instead consider weak convergence in the smaller Hölder-type spaces C_0^ρ for modulus of continuity ρ, thus allowing for a broader class of functionals; see Lamperti (1962) for an early treatment of this idea. The boundary case is marked by $\rho_2(h) = \sqrt{h\log(e/h)}$, which is the modulus of continuity of the limiting Brownian motion such that $B \in C^{\rho_2}$, but not $B \in C_0^{\rho_2}$. Thus we can not expect a central limit theorem in $C_0^{\rho_2}$, but at best in C^{ρ_2}. However, as Proposition 2.5 below shows, $\widetilde{S}_n$ does not converge weakly in C^{ρ_2}.

As a way to resolve this situation, we observe that not all continuous functionals $T : C^{\rho_2} \to \mathbb{R}$ are relevant for statistical inference. As outlined in the introduction, we are particularly interested in the functional $T(\widetilde{S}_n) = |\widetilde{S}_n|_{\rho_2}$, i.e. the Hölder-type seminorm in C^{ρ_2}. We show that the distribution of this specific random variable converges on the majority of its support, but not everywhere. To formalize this notion, we introduce the new concept of thresholded weak convergence as follows.

Definition 2.1. A a sequence of real-valued random variables $(X_n)_n$ convergences weakly beyond a threshold to a random variable X, if there exists some $\tau \in \mathbb{R}$ such that

$$P(X_n > t) \to P(X > t) \quad \forall t > \tau,\ t \in \mathcal{C}(X),$$

where $\mathcal{C}(X)$ denotes the points of continuity of the distribution function of X. We denote this by $X_n \Rightarrow_\tau X$. Equivalently, $(X_n \vee \tau) \Rightarrow (X \vee \tau)$.

Our central probabilistic result is the following sufficient criterion for thresholded weak convergence of Hölder-type seminorms.

Theorem 2.2. *Let W_n be a sequence of stochastic processes with paths in $C[0,1]$, and $\rho_0 : [0,1] \to [0,\infty)$ an increasing function with $\rho_0(0) = 0$. Suppose there exist $C > 0$ and for any $t > C$ a $\kappa(t) > 1$ and $K(t) > 0$, such that*

$$P\left(\frac{|W_n(u) - W_n(v)|}{\rho_0(|u-v|)} > t\right) \le K(t)|u-v|^{\kappa(t)}, \qquad u, v \in [0,1]. \tag{T}$$

Assume further that there exist $p > 0$ and $\widetilde{K} > 0$ such that

$$\rho_0(zh) \leq \widetilde{K} z^p \rho_0(h), \qquad h, z \in [0,1]. \tag{R}$$

Then, for any $t > C$,

$$\sup_n P\left(\sup_{u,v \in [0,1], |u-v| \leq h} \frac{|W_n(u) - W_n(v)|}{\rho_0(|u-v|)} > t \right) \longrightarrow 0, \quad \text{as } h \to 0, \tag{3}$$

and for any modulus of continuity ρ such that $\rho(h)/\rho_0(h) \to \infty$ as $h \to 0$, the sequence $\|W_n\|_\rho$ is stochastically bounded. If furthermore there exists a limiting process W such that $W_n \Rightarrow_{\text{fidi}} W$, then

$$\begin{aligned} &W_n \Rightarrow W \text{ in } C_0^\rho, && \text{if } \rho(h)/\rho_0(h) \stackrel{h\to 0}{\longrightarrow} \infty, \\ &|W_n|_\rho \Rightarrow_C |W|_\rho, && \text{if } \rho(h)/\rho_0(h) \stackrel{h\to 0}{\longrightarrow} 1. \end{aligned} \tag{4}$$

The previous result also extends to processes with values in metric spaces, see Theorem A.3 in the Appendix. In the sequel, we will mostly use Theorem 2.2 for $\rho_0(h) = \rho_\alpha(h) = \sqrt{h}|\log(eh)|^{\frac{1}{\alpha}}$, which satisfies (R), see Lemma A.2. In this situation, a sufficient condition for (T) is

$$P\left(|W_n(u) - W_n(v)| > r\, C\sqrt{|u-v|}\right) \leq K \exp(-r^\alpha), \quad u, v \in [0,1],\ r > 0. \tag{T-α}$$

The tightness condition (T-α) can be interpreted as an entropy bound in terms of the sub-Weibull Orlicz norms (Vladimirova et al., 2020). For the special case that W_n is a standard Brownian motion, then (T-α) holds with $C = \sqrt{2}$ and $\alpha = 2$, and (3) matches Levy's result on the modulus of continuity of Brownian motion (Schilling, 2021, Thm. 10.6).

Remark 1. Invariance principles in Hölder spaces have been first studied by Lamperti (1962), for the moduli $\rho(h) = h^\gamma$, who formulates the sufficient condition that for some $a, b > 0$ with $\gamma < \frac{b}{a}$, the moment condition

$$\mathbb{E}|W_n(u) - W_n(v)|^a \leq C|u-v|^{1+b}$$

holds. Via Markov's inequality, this can be transformed into the form (T). For the statistically most relevant case $W_n = \widetilde{S}_n$, Račkauskas and Suquet (2004b) reformulates the sufficient conditions in terms of polynomial tail bounds. Results for dependent data under mixing conditions are due to Hamadouche (2000), and for linear processes in Hilbert spaces by Račkauskas and Suquet (2009). The stronger modulus $\rho_\alpha(h) = \sqrt{h}\log^{\frac{1}{\alpha}}(e/h)$ for $\alpha < 2$ is studied by Račkauskas and Suquet (2004a), who also allow for observations in infinite dimensional Banach spaces. For the latter modulus, they formulate the condition $P(\|Y_t\| > \log^{\frac{1}{\alpha}}(er)) \ll 1/r$ as $r \to \infty$, or equivalently $P(\|Y_t\| > z) \ll \exp(-z^\alpha)$ as $z \to \infty$. As all these studies aim for a functional CLT, they necessarily omit the critical case ρ_2, where classical weak convergence fails. Our result on thresholded weak convergence shows that in the critical case, one can still obtain certain distributional limits which is sufficient for many statistical purposes, as described in Section 3.

Remark 2. Property (3) also implies that $|W_n|_\rho$ and the discretization $\max_{i,j=1,\ldots,n} |W_n(\frac{i}{n}) - W_n(\frac{j}{n})|/\rho(\frac{|i-j|}{n})$ have the same limit distribution as $n \to \infty$, which facilitates numerical implementation.

As a theoretical complement to the definition of thresholded weak convergence, we provide a version of Skorokhod's Representation Theorem.

Theorem 2.3 (Skorokhod representation)**.** *Let $W_n \in C[0,1]$ such that $W_n \Rightarrow_{\text{fidi}} W$, and $|W_n|_{\rho^*} \Rightarrow_T |W|_{\rho^*}$. Then there exists a probability space with random elements W_n and W such that*

$$\|W_n - W\|_\rho \to 0, \qquad \forall \rho \gg \rho^*.$$

We emphasize that Theorem 2.3 yields a single coupling which approximates the limit simultaneously in all Hölder-type spaces C^ρ with $\rho \gg \rho^*$.

We proceed to specify the abstract Theorem 2.2 for the statistically interesting partial sum process $\widetilde{S}_n$. If the random variables η_t are sub-Gaussian, we obtain the following extension of Donsker's Theorem.

Corollary 2.4. *Let η_t be iid centered random variables with unit variance such that $\mathbb{E}\exp(r\eta_t) \leq \exp(\frac{r^2}{C^2})$. Then the interpolated partial sum process $\widetilde{S}_n$ satisfies* (4) *with threshold C, exponent $\alpha = 2$, and the limit process W is a standard Brownian motion.*

What happens in the lower part of the distribution in (4)? The following proposition shows that the thresholding is not just an artifact of our proof, but indeed necessary. In other words, the lower part of the distribution depends on the distribution of S_n and might converge to various different limits depending on the tails of W_n.

Proposition 2.5. *For any $T > 0$, there exist sub-Gaussian iid random variables η_t with $\mathrm{Var}(\eta_t) = 1, \mathbb{E}(\eta_t) = 0$, and $\|\eta_t\|_{\psi_2} < \infty$, such that Theorem 2.2 applies with $\rho_0 = \rho_2$, but $\liminf_{n\to\infty} P(|\widetilde{S}_n|_{\rho_2} \geq T) = 1$.*

An essential statistical benefit of the multiplicatively weighted statistic T_n^* is that its limit theory can be readily extended to other sampling situations. Here, we study a general nonlinear, non-stationary time series given by the very general Bernoulli shift model $\eta_t = G_t(\boldsymbol{\epsilon}_t)$ for $\boldsymbol{\epsilon}_t = (\epsilon_t, \epsilon_{t-1}, \ldots)$ with $\epsilon_i \sim U(0,1)$ iid random seeds. To quantify the temporal dependence, we introduce $\boldsymbol{\epsilon}_{t,h} = (\epsilon_t, \ldots, \tilde{\epsilon}_{t-h}, \ldots)$ for an independent copy $\tilde{\epsilon}_i \sim U(0,1)$. Following the general idea of Wu (2005), we define the physical dependence measure w.r.t. the sub-Gaussian norm $\|X\|_{\psi_2} = \inf\{c : \mathbb{E}\exp(X^2/c^2) \leq 2\}$ as

$$\delta_{\psi_2}(h) = \sup_t \|G_t(\boldsymbol{\epsilon}_t) - G_t(\boldsymbol{\epsilon}_{t,h})\|_{\psi_2}.$$

With this concept, we obtain the following novel sub-Gaussian concentration bound for dependent data, which might be of independent interest.

Theorem 2.6 (Sub-Gaussian concentration with dependence). *There exists a universal K such that for any centered time series η_t of the form $\eta_t = G(\boldsymbol{\epsilon}_t)$,*

$$\left\| \sum_{t=1}^{n} w_t \eta_t \right\|_{\psi_2} \leq K \sqrt{\sum_{t=1}^{n} |w_t|^2} \left(\sum_{j=1}^{\infty} \sqrt{j} \delta_{\psi_2}(j) \right),$$

where $w_t \in \mathbb{R}$ is a sequence of weights.

Remark 3. The physical dependence measure is usually defined in terms of $L_p(P)$ norms instead of sub-Gaussian Orlicz norms, and denoted as $\delta_p(j)$. For the latter, it holds (Liu et al., 2013)

$$\left\| \sum_{t=1}^{n} \eta_t \right\|_{L_p} \leq C\sqrt{n} \sum_{j=0}^{\infty} \delta_p(j).$$

This bound differs from Theorem 2.6 by a factor $\sqrt{j}$ in the series. This difference may be explained as follows: For iid random variables Z_i, we have $\| \sum_{i=1}^{k} Z_i \|_{L_p} \leq C\sqrt{k} \max_i \|Z\|_{L_p}$, and $\| \sum_{i=1}^{k} Z_i \|_{\psi_2} \leq C\sqrt{k} \max_i \|Z_i\|_{\psi_2}$. For the L_p norm, this inequality remains valid if the Z_i are martingale differences (Pinelis, 1994). On the other hand, extending the sub-Gaussian inequality to martingales requires a uniform bound on the sub-Gaussian norm of Z_i conditional on $\mathcal{F}_{i-1}$, that is

$$\|Z_i\|_{\psi_2, \mathcal{F}_{i-1}} := \inf \left\{ c > 0 \,:\, \mathbb{E}\left(\exp(Z_i^2/c) | \mathcal{F}_{i-1} \right) \leq 2 \right\},$$

similar to the Azuma–Hoeffding inequality. This is more restrictive than imposing an upper bound on $\|Z_i\|_{\psi_2}$. In contrast, our novel formulation of the sub-Gaussian physical dependence does not impose this kind of uniformity, and this weaker assumption leads to the extra term $\sqrt{j}$ in the concentration inequalities.

To ensure that the partial sum process of nonstationary random variables has a well-defined limit, we introduce the additional assumption that the noise process is locally stationary. This concept was initially introduced by Dahlhaus (1997), see also Dahlhaus et al. (2019), and consists of rescaling the non-stationarity in $t \in \{1, \ldots, n\}$ to relative times $t/n \in [0, 1]$. Specifically, we suppose that $\eta_t = \eta_{t,n} = G_{t,n}(\boldsymbol{\epsilon}_t)$ forms a triangular array of random variables, and for any $u \in [0, 1]$, there exists a measurable $\tilde{G}_u : \mathbb{R}^\infty \to \mathbb{R}$ such that

$$\int_0^1 \left\| G_{\lfloor un \rfloor, n}(\boldsymbol{\epsilon}_0) - \tilde{G}_u(\boldsymbol{\epsilon}_0) \right\|_{L_2} du \longrightarrow 0. \tag{LS-1}$$

Moreover, we impose bounded variation of the mapping $t \mapsto G_{t,n}$, that is,

$$\sup_n \left(\|G_{1,n}(\boldsymbol{\epsilon}_0)\|_{L_2} + \sum_{t=2}^{n} \|G_{t,n}(\boldsymbol{\epsilon}_0) - G_{t-1,n}(\boldsymbol{\epsilon}_0)\|_{L_2} \right) < \infty. \tag{LS-2}$$

The formulation of local stationarity in terms of the Bernoulli shift model is due to Wu and Zhou (2011) under the condition that $G_{t,n} = \tilde{G}_{t/n}$ and $u \mapsto \tilde{G}_u$ is Lipschitz, and Zhou (2013) extended this to piecewise-Lipschitz with finitely many discontinuities. In prior work (Mies, 2023), we relaxed this regularity requirement to bounded p-variations, with (LS-2) being a special case for $p = 1$. Assumption (LS-1) was introduced in Mies (2024) as a relaxation of the assumption that $G_{\lfloor un \rfloor, n} \to \tilde{G}_u$ uniformly in u. Under the additional assumption

$$\delta_{\psi_2}(h) = \mathcal{O}(h^{-\beta}) \tag{LS-3}$$

for some $\beta > 1$, we find that $\widetilde{S}_n(u) \Rightarrow B(\Sigma(u)) \overset{d}{=} \int_0^u \sigma_\infty(v)\, dB_v$ for a standard Brownian motion B and $\Sigma(u) = \int_0^u \sigma_\infty^2(v)\, dv$, where $\sigma_\infty^2(v) = \sum_{h=-\infty}^{\infty} \operatorname{Cov}(\tilde{G}_v(\boldsymbol{\epsilon}_0), \tilde{G}_v(\boldsymbol{\epsilon}_h))$ is the local long-run variance. If we combine this with the novel concentration inequality of Theorem 2.6, we are able to derive a multiscale central limit theorem.

Corollary 2.7. *Suppose that $\eta_t = G_{t,n}(\boldsymbol{\epsilon}_t)$ satisfies* (LS-1), (LS-2), *and* (LS-3) *for some $\beta > \frac{3}{2}$. Then the interpolated partial sum process $\widetilde{S}_n$ satisfies* (4) *for some threshold C, and exponent $\alpha = 2$. and the limit process is $W(u) = B(\Sigma(u))$.*

Besides the thresholded weak convergence for the critical modulus ρ_2, Corollary 2.7 yields a Hölderian invariance principle for dependent data. Unlike the results of Hamadouche (2000) and Račkauskas and Suquet (2009), we impose ergodicity in the form of the physical dependence measure, instead of mixing conditions or models in terms of linear processes. Moreover, we also allow for nonstationarity.

3 Statistical applications

3.1 Signal discovery

Returning to the signal discovery problem described in the introduction, we consider the setting of iid noise terms η_t which are sub-Gaussian such that $\mathbb{E}\exp(r\eta_t) \leq \exp(\frac{r^2}{C_\eta^2})$, but not necessarily normally distributed. The multiplicatively weighted multiscale test rejects the null hypothesis $\mathbb{H}_0 : \mu_n = 0$ for large values of $T_n^* = \max_I T_n(I)/\sqrt{\log \frac{en}{|I|}} = |S_n|_{\rho_2}$, which converges weakly beyond a threshold to $\sigma |B|_{\rho_2}$. The variance σ^2 may be estimated as $\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{t=2}^n (Y_t - Y_{t-1})^2/2$. Compared to the usual sample variance, this estimator is consistent under the alternative as long as $\mu_n \ll \sqrt{n}$. As a result, we obtain a feasible test which consistently detects a signal at the optimal rate. To choose a critical value for a significance level $\alpha \in (0,1)$, we denote by q_α the $(1-\alpha)$-quantile of $|B|_{\rho_2}$ for a standard Brownian motion B, which is tabulated in the first row of Table 1. Note that $|B|_{\rho_2} \geq \sqrt{2}$ almost surely, such that $q_\alpha > \sqrt{2}$,

Theorem 3.1. *Under $\mathbb{H}_0$, for any α small enough such that $q_\alpha > C_\eta$, the test maintains size α asymptotically:*

$$\limsup_{n\to\infty} P(T_n^* > \widehat{\sigma}_n q_\alpha) \leq \alpha$$

α	10%	5%	1%	0.1%	0.01%
$\rho_{2,0} = \rho_2$	2.384	2.601	3.084	3.695	4.229
$\rho_{2,50}$	0.641	0.661	0.704	0.758	0.808
$\rho_{2,100}$	0.468	0.484	0.516	0.559	0.599
$\rho_{2,500}$	0.216	0.222	0.237	0.256	0.274
$\rho_{2,1000}$	0.153	0.158	0.169	0.182	0.191
Sparse grid $\mathcal{G}_{\mathrm{RW}}$					
$\rho_{2,0} = \rho_2$	2.173	2.370	2.824	3.408	3.928
$\rho_{2,50}$	0.636	0.656	0.700	0.753	0.807
$\rho_{2,100}$	0.466	0.481	0.513	0.556	0.599
$\rho_{2,500}$	0.215	0.222	0.236	0.256	0.273
$\rho_{2,1000}$	0.153	0.158	0.168	0.181	0.193
Sparse grid $\mathcal{G}_{\mathrm{dyadic}}$					
$\rho_{2,0} = \rho_2$	1.907	2.118	2.631	3.316	3.904
$\rho_{2,50}$	0.586	0.606	0.650	0.706	0.761
$\rho_{2,100}$	0.431	0.446	0.479	0.518	0.557
$\rho_{2,500}$	0.199	0.206	0.221	0.239	0.254
$\rho_{2,1000}$	0.141	0.146	0.156	0.170	0.182

Table 1: $(1-\alpha)$-quantiles of $|B|_{\rho_{2,a}}$ evaluated on various grids based on 10^5 simulations, discretizing the Brownian motion via 10^4 grid points.

Under the sequence of alternatives $f_n(t) = \mu_n \mathbb{1}(t \in I_n)$ with $I_n = [a_n, a_n + l_n) \subset [1, n]$, such that $\mu_n^2 l_n \gg \log(\frac{en}{l_n})$ and $\mu_n \ll \sqrt{n}$, the test is consistent:

$$\lim_{n\to\infty} P(T_n^* > \widehat{\sigma}_n q_\alpha) \to 1. \tag{5}$$

To compare this procedure with the multiscale methodology of Dümbgen and Spokoiny (2001), suppose for simplicity that $\sigma = 1$ is known. Our test rejects the null if $T_n(I) \geq \zeta^{\mathrm{MULT}}(|I|) = q_\alpha \sqrt{\log(en/|I|)}$, whereas the test based on T_n^{DS} rejects if $T_n(I) \geq \zeta^{\mathrm{DS}}(|I|) = \sqrt{2}\sqrt{\log(en/|I|)} + c_\alpha$ for a critical value $c_\alpha > 0$. Since $q_\alpha > \sqrt{2}$, the second threshold is sharper for short intervals $|I|$. However, it is only valid for Gaussian errors, while our new threshold rule is robust to non-Gaussian errors. To further illustrate this difference, we can try to construct a threshold rule of the form $\zeta_A^{\mathrm{DS}}(|I|) = A\sqrt{\log(en/|I|)} + c_\alpha$ for some fixed $A > \sqrt{2}$. The following result shows that this is theoretically possible but statistically infeasible, as the value of A will depend on the unknown sub-Gaussian tail bound.

Proposition 3.2. *For any $A > C_\eta$,*

$$\sup_I \left\{ T_n(I) - A\sqrt{\log \tfrac{en}{|I|}} \right\}_+ \Rightarrow \sup_{u \leq v} \left\{ \tfrac{|B(v)-B(u)|}{\sqrt{v-u}} - A\sqrt{\log \tfrac{e}{|u-v|}} \right\}_+ .$$

On the other hand, for any $A \geq 0$, there exist iid sub-Gaussian random variables η_t such that $\sup_I \left\{ T_n(I) - A\sqrt{\log \frac{en}{|I|}} \right\}_+ \to \infty$ *in probability.*

While both threshold rules, ζ_A^{DS} and ζ^{MULT}, are only consistent in the regime $\mu_n^2 l_n \gg \log(n/l_n)$, the former puts slightly more weight on shorter signals. Thus, we expect a different assignment of statistical power to the length scales. To quantify the power against signals of length l_n in finite samples, Walther and Perry (2022) suggested the *realised exponent* $e_n(l_n)$ defined as

$$e_n(l_n) = \frac{\mu_{\min}(l_n)}{\sqrt{2}\rho_2(l_n)},$$

where $\mu_{\min}(l_n) \geq 0$ is the smallest value such that the test with significance level $\alpha = 10\%$ has 80% power against the alternative $f_n(t) = \mu_n \mathbb{1}(t \in I_n)$ for a randomly placed interval $I_n \subset [1, n]$ of length l_n. Table 2 presents the realised exponents for different threshold rules and Gaussian errors for sample size $n = 10^4$. SCAN refers to a constant threshold $\zeta^{\mathrm{SCAN}}(|I|) = d_{\alpha,n}$ as a quantile of the uncorrected scan statistic $\sup_I T_n(I)$, which necessarily depends on n. SAC is an adaptation of the threshold of Sharpnack and Arias-Castro (2016) presented in Walther and Perry (2022), and of the form $\zeta^{\mathrm{SAC}}(|I|) = \sqrt{2\log\left[en(1+\log|I|)^2/|I|\right]} + c_\alpha$, and BlockedScan and BonferroniScan are two further suggestions of Walther and Perry (2022) which we use as benchmark. The comparsion reveals that the threshold ζ^{MULT} (line 3) has less power on shorter length scales. To increase the power of the multiplicatively weighted procedure, we may consider the alternative modulus of continuity $\rho_{2,a}(h) = \sqrt{h(a + \log(e/h))}$ for $a \geq 0$. Since $\rho_2(h)/\rho_{2,a}(h) \to 1$ as $h \to 0$, the theory of Section 2 as well as Theorem 3.1 still apply, and Table 1 presents the corresponding critical values. Table 2 shows that the use of $\rho_{2,a}$ improves the finite sample performance against short signals, being competitive with the DS threshold. Unlike the benchmark thresholds, our proposed threshold scheme is also applicable for non-Gaussian errors thanks to the new asymptotic theory.

3.2 Goodness-of-fit testing

The signal discovery problem can be interpreted as testing for a very specific mean function $f_n(t) = 0$ in (1). More generally, we may perform a goodness-of-fit test for a class of functions $\mathbb{F}_0 \subset \{f : \{1, \ldots, n\} \to \mathbb{R}\}$. A simple example consists of parametric classes $\{f(t) = a + bt \,|\, a, b \in \mathbb{R}\}$ or $\{f(t) = a\sin(rt) \,|\, a \in \mathbb{R}, r > 0\}$. Alternatively, we may also test certain shape constraints, such as monotonicity $\mathbb{F}_{\uparrow} = \{f : f(t) \leq f(t+1)\}$, convexity $\mathbb{F}_{\mathrm{conv}} = \{f(t+1) - f(t) \geq f(t) - f(t-1)\}$, or non-negativity $\mathbb{F}_{\geq 0} = \{f \geq 0\}$, which are also studied by Dümbgen and Spokoiny (2001). Other shape constraints include unimodality (Chatterjee and Lafferty, 2019), or S-shape (Feng et al., 2022). The choice $\mathbb{F}_0 = \{0\}$ recovers the signal discovery problem, and the case of a singleton null $\mathbb{F}_0 = \{f_0\}$ has been treated by Rohde (2008) via a multiscale methodology for symmetric errors. In this Section, we propose a generic testing procedure which is based on our

$\|I_n\|$	1	5	10	15	50	100	500	1000
SCAN	1.41	1.58	1.74	1.85	2.19	2.47	3.48	4.34
DS	1.80	1.79	1.86	1.90	1.92	1.94	2.08	2.25
SAC	1.41	1.62	1.76	1.85	2.04	2.18	2.73	3.18
BlockedScan	1.49	1.67	1.83	1.87	1.91	2.01	2.43	2.80
BonferroniScan	1.60	1.81	1.98	2.03	2.13	2.25	2.75	3.17
$\rho_{2,0}$	3.42	3.42	3.38	3.27	3.15	3.04	2.81	2.58
$\rho_{2,50}$	1.62	1.74	1.79	1.89	2.03	2.19	2.89	3.46
$\rho_{2,100}$	1.58	1.76	1.76	1.90	2.06	2.25	3.03	3.56
$\rho_{2,500}$	1.54	1.69	1.77	1.88	2.05	2.28	3.13	3.69
$\rho_{2,1000}$	1.54	1.72	1.86	1.87	2.07	2.31	3.16	3.72
$\rho_{2,10^6}$	1.54	1.73	1.79	1.84	2.12	2.29	3.09	3.74

Table 2: Realized exponents for Gaussian noise and different threshold rules. Values for SCAN, DS, SAC, BlockedScan, and BonferroniScan, are taken from Walther and Perry (2022). For the thresholds ζ^{MULT} (third row and below), we determine the realized exponents based on 2000 simulations.

novel probabilistic results, and which can be used to perform goodness-of-fit tests for all classes discussed above.

As multiscale test statistic for the composite null hypothesis $\mathbb{H}_0 : f_n \in \mathbb{F}_0$, we suggest

$$T_n^*(\mathbb{F}_0) = \inf_{f\in\mathbb{F}_0} T_n^*(f), \quad \text{where } T_n^*(f) = |S_n^f|_{\rho_2}$$

$$\text{and} \quad S_n^f(u) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor un\rfloor}[Y_t - f(t)] + \frac{un - \lfloor un\rfloor}{\sqrt{n}}[Y_{\lfloor un\rfloor+1} - f(\lfloor un\rfloor + 1)].$$

Under the null, we have $T_n^*(\mathbb{F}_0) \leq T_n^*(f_n) = |\widetilde{S}_n|_{\rho_2}$, which converges weakly beyond a threshold by virtue of the results of Section 2. Thus, we may combine the critical values reported in Table 1 with an estimate of the variance to construct an asymptotically valid multiscale test procedure for the goodness-of-fit problem.

Theorem 3.3. *Suppose η_t are iid sub-Gaussian such that $\mathbb{E}\exp(r\eta_t) \leq \exp(\frac{r^2}{C_\eta^2})$, and $\sum_{t=2}^n |f_n(t) - f_n(t-1)|^2 \ll n$. Under $\mathbb{H}_0$, for any α small enough such that $q_\alpha > C_\eta$,*

$$\limsup_{n\to\infty} P\left(T_n^*(\mathbb{F}_0) > \widehat{\sigma}_n q_\alpha\right) \leq \alpha,$$

for q_α and $\widehat{\sigma}_n^2$ as in Section 3.1.

In addition to a global statement about the structural properties of f_n, one is often also interested in its local shape. That is, we want to infer if f_n is convex or monotone or non-negative on a given interval I. The goodness-of-fit test may also be localized to obtain such insight into the qualitative nature of the regression function. To this end, for any interval $I \subset [1, n]$ and a candidate class $\mathbb{F}_0$, we define the localization

$$\mathbb{F}_0^I = \{f : \{1, \ldots, n\} \to \mathbb{R} \mid \exists \tilde{f} \in \mathbb{F}_0 \text{ such that } \tilde{f}(t) = f(t) \text{ for all } t \in I\} \supset \mathbb{F}_0.$$

Since the localization is a weaker requirement, we have $T_n^*(\mathbb{F}_0^I) \leq T_n^*(\mathbb{F}_0)$ and $\sup_I T_n^*(\mathbb{F}_0^I) = T_n^*(\mathbb{F}_0)$. This means that the critical value $\widehat{\sigma}_n q_\alpha$ allows for testing all local hypotheses simultaneously while controlling the type I error. Inverting these tests yields the class of intervals

$$\mathcal{I}_n = \left\{ I \mid T_n^*(\mathbb{F}_0^I) > \widehat{\sigma}_n \cdot q_\alpha \right\}.$$

In line with Fryzlewicz (2024a), we call any $I \in \mathcal{I}_n$ an interval of significance. For example, if $\mathbb{F}_0$ is the class of increasing functions, then $I \in \mathcal{I}_n$ is interpreted as evidence that the function f_n has a strict local minimum or maximum in I.

Proposition 3.4. *Under the conditions of Theorem 3.3, with asymptotic probability at least $1 - \alpha$, all intervals of significance are true discoveries, that is*

$$\limsup_{n \to \infty} P\left(f_n \notin \mathbb{F}_0^I \textit{ for all } I \in \mathcal{I}_n\right) \geq 1 - \alpha.$$

The asymptotic power of the goodness-of-fit test can be analyzed for special cases upon identifying alternatives of interest. In the next section, we pursue this for the commonly studied problem of multiple changepoint testing, and show that our procedure achieves asymptotically optimal detection performance.

3.3 Multiple changepoint detection

In the multiple changepoint problem (Cho and Kirch, 2021), the regression function f_n is of the form

$$f_n(t) = \delta_0 + \sum_{k=1}^{\kappa} \delta_k \mathbb{1}(t \geq \tau_k). \tag{CP}$$

In words, f_n is a step function with κ jumps at locations $\tau_k \in \{2, \ldots, n\}$ and jump sizes $\delta_k \in \mathbb{R}$. This model is widespread in applications as the changepoints τ_k can clearly be interpreted as points of interest, e.g. to identify regions with interesting copy number variations in DNA (Niu and Zhang, 2012), or changes in measurement techniques or observation locations in climate time series (Reeves et al., 2007). The main statistical objective is to perform inference on (a) the number κ of changes, and (b) the locations τ_k.

A prominent multiscale procedure for the changepoint problem has been proposed by Frick et al. (2014). For any $J \in \mathbb{N}$, they formulate the problem as a goodness-of-fit test for $\mathbb{F}_0^J = \{f \text{ of the form (CP) with } \kappa \leq J\}$, and estimate κ by the smallest number J such that the hypothesis $f_n \in \mathbb{F}_0^J$ is accepted. Confidence statements for τ_k may then be derived by test inversion, including all functions $f \in \mathbb{F}_0^J$ such that the null $f_n = f$ is not rejected. They use a test statistic of the form (DS), and impose a lower bound of order $\log(n)^3$ on the interval length, leading to suboptimal performance for short-lived changes; see Proposition 1.2.

Here, we derive confidence statements following the alternative approach of narrowest significance pursuit, inspired by Fryzlewicz (2024a). The central idea is to introduce the class $\mathbb{F}_{\text{const}} = \{f_n \text{ constant}\}$ of constant regression functions, observing that f has no changepoint in the interval I if and only if $f \in \mathbb{F}_{\text{const}}^I$. As a special case of the goodness-of-fit test of Section 3.2, the localized changepoint test statistic takes the form

$$T_n^*\left(\mathbb{F}_{\text{const}}^I\right) = \inf_{a\in\mathbb{R}} \sup_{[un,vn]\subset I} \frac{|S_n(v) - S_n(u) - [v-u]a|}{\rho_2(v-u)}.$$

As the right hand side is a convex function of a, the minimum is attained and may be determined numerically via bisection. We may then obtain intervals of significance for the change locations τ_k as

$$\mathcal{I}_n^\tau = \left\{ I \,|\, T_n^*\left(\mathbb{F}_{\text{const}}^I\right) > \widehat{\sigma}_n \cdot q_\alpha \right\},$$

where q_α and $\widehat{\sigma}_n$ are as in Section 3.1. Note that although we borrow the terminology of Fryzlewicz (2024a), a similar idea of constructing confidence intervals is also described in (Verzelen et al., 2023, Sec. 5.4). Subsequent works building on this idea include Pilliat et al. (2023); Fryzlewicz (2024b); Gavioli-Akilagun and Fryzlewicz (2025).

Conditions for consistency of the procedure may be formulated in terms of τ_k and δ_k, and our test turns out to achieve the optimal detection rates for this problem derived by Verzelen et al. (2023). In the sequel, all parameters τ_k, δ_k, and κ may depend on n implicitly, and we denote by $\mathcal{D} = \{\tau_1, \ldots, \tau_\kappa\} \subset \{1, \ldots, n\}$ the set of changepoints. Moreover, we introduce the notation $L_k = \min(\tau_k - \tau_{k-1}, \tau_{k+1} - \tau_0)$ for the length of the k-th changepoint, where $\tau_0 = 0$ and $\tau_{\kappa+1} = n+1$. Prior research into statistical lower bounds for this problem (Arias-Castro et al., 2011; Chan and Walther, 2013; Verzelen et al., 2023) has revealed that the changepoint τ_k is only detectable if $\Delta_k^2 L_k \gg \log(n/L_k)$. Thus, for any $z > 0$, we define the set $\mathcal{D}(z) = \{\tau_j \,|\, \delta_j^2 L_j \geq z \cdot \log(\frac{n}{L_j})\} \subset \mathcal{D}$ of detectable changepoints, and for any $\tau \in \mathcal{D}(z)$, denote by $w(\tau, z) = w(\tau_k, z) = \inf\{r \mid 2\delta_k^2 r \geq z \cdot \log(\frac{n}{2r})\} \leq \frac{L_k}{2}$ its detectable locality.

Theorem 3.5. *Under the conditions of Theorem 3.3, with asymptotic probability at least $1-\alpha$, all intervals of significance contain at least one changepoint, that is*

$$\limsup_{n\to\infty} P\left(I \cap \mathcal{D} \neq \emptyset \textit{ for all } I \in \mathcal{I}_n^\tau\right) \geq 1-\alpha.$$

Moreover, detectable changepoints are asymptotically isolated and located:

$$\lim_{z\to\infty} \limsup_{n\to\infty} P\begin{pmatrix} \forall \tau \in \mathcal{D}(z)\, \exists I \in \mathcal{I}_n^\tau \textit{ such that} \\ I \cap \mathcal{D} = \{\tau\} \textit{ and } |I| \leq w(\tau, z) \end{pmatrix} = 1. \tag{6}$$

In comparison, (Fryzlewicz, 2024a, Thm. 4.1) detects the changepoint τ_k if $\delta_k^2 L_k \gg \log(n)$ and localizes it by an interval with length of order $\log(n)/\delta_k^2$. For longer intervals resp. smaller changes, this localization and detection rate is suboptimal as Fryzlewicz (2024a) uses a uniform threshold instead of a multiscale correction. Incorporating the multiscale idea of Dümbgen and Spokoiny (2001), Pilliat et al. (2023) derive the same order $w(\tau_k, z)$ for the length of the interval, but for a different threshold rule which requires knowledge of the sub-Gaussian norm of the errors. Theorem 3.5 shows that our procedure attains the same localization rate, while being statistically feasible since we can specify critical values without knowing the exact sub-Gaussian bound. Moreover, our asymptotic treatment readily allows for other sampling settings, in particular non-stationary and temporally dependent data as studied in the next subsection.

While the detectability condition for a changepoint matches the optimal lower bound of Verzelen et al. (2023), the latter authors show that for Gaussian errors the localization rate can be improved to $\mathcal{O}(1/\delta_k^2)$. It is not clear if this sharper localization is also attainable for non-Gaussian errors.

The class $\mathcal{I}_n^\tau$ of significant intervals is highly redundant, and contains many overly large as well as intersecting intervals. To interpret the statistical findings, it is preferable to report disjoint intervals, specifically many small intervals for high statistical power. This can be achieved by a postprocessing of $\mathcal{I}_n^\tau$. For any class of intervals $\mathcal{I}$, define $\mathrm{NSP}(\mathcal{I})$ as the output of the following *narrowest significance pursuit* routine:

1. Initialize $\mathcal{I}^0 = \mathcal{I}$.
2. Choose $\widehat{I}_k$ as the shortest interval in $\mathcal{I}^{k-1}$, with arbitrary tie breaking.
3. Update $\mathcal{I}^k = \{I \in \mathcal{I}^{k-1} \mid I \cap \widehat{I}_k = \emptyset\}$.
4. Terminate after iteration K if $\mathcal{I}^K = \emptyset$, and return $\mathrm{NSP}(\mathcal{I}) = \{\widehat{I}_1, \ldots, \widehat{I}_K\}$.

This generic formulation is identical to Algorithm 1 of Pilliat et al. (2023).

Proposition 3.6. *Suppose that a class of intervals $\mathcal{I}$, a subset $\widetilde{\mathcal{D}} \subset \mathcal{D} = \{\tau_1, \ldots, \tau_\kappa\}$, and non-negative real numbers $r(\tau) = r(\tau_k) \leq \frac{L_k}{2}$ satisfy*

(i) $I \cap \mathcal{D} \neq \emptyset$ for all $I \in \mathcal{I}$, and

(ii) for any $\tau \in \widetilde{\mathcal{D}}$ there exist $I(\tau) \in \mathcal{I}$ such that $I(\tau) \cap \mathcal{D} = \{\tau\}$, and $|I(\tau)| \leq r(\tau)$.

Then the class $\mathrm{NSP}(\mathcal{I}) = \{\widehat{I}_1, \ldots, \widehat{I}_{|\widetilde{\mathcal{D}}|}\}$ consists of exactly $|\widetilde{\mathcal{D}}|$ disjoint intervals, uniquely localizing all $\tau \in \widetilde{\mathcal{D}}$ in the sense of (ii).

That is, the reduced class $\mathrm{NSP}(\mathcal{I})$ satisfies the same statistical guarantees as the full class $\mathcal{I}$, while consisting of disjoint intervals. In particular, $|\mathrm{NSP}(\mathcal{I}_n^\tau)|$ provides a lower confidence bound on the total number of changepoints.

3.4 Critical values for nonstationary dependent errors

The assumption of iid errors is too simplistic for many applications. Instead, data is often heteroskedastic and temporally dependent, which may be modeled in terms of nonstationary time series. Especially for changepoint inference, the need to account for nonstationarity was initially observed by Zhou (2013), who showed how to adapt the critical values of a standard CUSUM statistic. Subsequent works include Vogt and Dette (2015), Dette et al. (2019), Cui et al. (2021), and Mies (2023). None of these references consider a multiscale threshold, and hence do not achieve simultaneously optimal detection against changes of different lengths. A multiscale procedure for heteroskedastic, independent Gaussian noise was suggested by Pein et al. (2017), however under the restriction that variance and mean change at the same time. In the sequel, we show how our new asymptotic framework allows for extension to nonstationary time series, i.e. to dependent and heteroskedastic data.

If we postulate the model framework $\eta_t = \eta_{t,n} = G_{t,n}(\boldsymbol{\epsilon}_t)$ introduced in Section 2, the distributional limit of the partial sum process $\widetilde{S}_n$ is an inhomogeneous Brownian motion $W(u) = B(\Sigma(u))$, and the limit of the goodness-of-fit statistic $T_n^*(f_n)$ is given by the random variable $|W|_{\rho_2}$, see Corollary 2.7. As a consequence of the nonstationarity, determining critical values requires an estimate for the whole function $\Sigma(u)$, instead of just a single value. To this end, we adapt the estimator of Dette et al. (2020), which is in turn inspired by Wu and Zhao (2007), to the nonstationary case. Specifically, choose a window size b_n such that $1 \ll b_n \ll n$ and define

$$\widehat{\Sigma}_n(u) = \frac{1}{2n} \sum_{t=b_n}^{un-b_n} \left(\frac{1}{\sqrt{b_n}} \sum_{i=0}^{b_n-1} (Y_{t-i} - Y_{t+1+i}) \right)^2,$$

for $u = \frac{i}{n}$, and interpolate linearly in between. As for the iid case, differencing removes the signal f_n, while the block sum is introduced to capture the serial correlation, and partial sums mimic the functional structure of $u \mapsto \Sigma(u)$. The quadratic terms $\widehat{\sigma}_n^2(\frac{t}{n}) = (1/\sqrt{b_n} \sum_{i=0}^{b_n-1} (Y_{t-i} - Y_{t+1+i}))^2$ should be interpreted as a noisy estimate of the local long-run variance $\sigma_\infty^2(t/n)$. Since $\widehat{\Sigma}_n(u) \to \Sigma(u)$ uniformly (Lemma A.4), we have $B(\widehat{\Sigma}_n(u)) \to B(\Sigma(u))$ uniformly. Thus, it would be natural to estimate the $(1-\alpha)$-quantile q_α of $|B(\Sigma(u))|_{\rho_2}$ by the corresponding quantile of $|B(\widehat{\Sigma}_n(u))|_{\rho_2}$. However, this approach does not yield consistent critical values, as the latter random variable diverges. In particular, Levy's Theorem on the modulus of continuity of the Brownian motion shows that $|B(\widehat{\Sigma}_n(u))|_{\rho_2} \geq \sqrt{2} \max_t \widehat{\sigma}_n^2(\frac{t}{n})$, which is stochastically unbounded.

To solve this issue and obtain asymptotically valid critical values, we suggest to use the $(1-\alpha)$-quantile $\widehat{q}_{\alpha,n}$, conditionally on $\widehat{\Sigma}_n$, of the random variable

$$\sup_{|u-v|>c_n} \frac{|B(\widehat{\Sigma}_n(u)) - B(\widehat{\Sigma}_n(v))|}{\rho_2(|u-v|)},$$

for a sequence c_n tending to zero slowly. The lower bound on the interval length does lead to correct critical values because under the null, the very short intervals have a vanishing contribution to the distribution of the statistic, as a consequence of (3). On the other hand, under the alternative, these short intervals might carry relevant information, and hence should be included when computing the test statistic.

Theorem 3.7. *Let $\eta_t = \eta_{t,n}$ be an array of locally stationary time series satisfying* (LS-1)*,* (LS-2)*, and* (LS-3) *for some $\beta > 2$. Suppose moreover that $1 \ll b_n \ll n$, and $\sum_{t=2}^n |f_n(t) - f_n(t-1)|^2 \leq v_n$ for a sequence $v_n \geq 1$, such that*

$$1 \gg c_n \gg \sqrt{\frac{v_n b_n}{n}}.$$

Then there exists a $\tau > 0$, such that for all α small enough such that $q_\alpha > \tau$, we have $\widehat{q}_{\alpha,n} \to q_\alpha$ in probability.

In particular, we may use $\widehat{q}_{\alpha,n}$ as critical value in the procedures of Sections 3.1, 3.2, and 3.3, and maintain the same statistical guarantees.

3.5 Sparse grids for faster evaluation

In practice, the statistic $|S_n|_{\rho_2}$ is evaluated at the grid points $\frac{i}{n}$, leading to a computational cost of $\mathcal{O}(n^2)$. A reduction to $\mathcal{O}(n)$ is possible if we restrict attention to a sparse subset of candidate intervals. To this end, let $\mathcal{G} \subset [0,1]^2$, and consider the sparse multiscale statistic

$$|S_n|_{\rho,\mathcal{G}} = \sup_{(u,v)\in\mathcal{G}} \frac{|S_n(u) - S_n(v)|}{\rho(|u-v|)}.$$

The same arguments as in Theorem 2.2, in particular property (3), show that $|\widetilde{S}_n|_{\rho_2,\mathcal{G}} \Rightarrow_\tau |W|_{\rho_2,\mathcal{G}}$ beyond a threshold τ. In particular, under the conditions of Corollary 2.4 or Corollary 2.7, the limit process is W is Gaussian, and the the threshold τ is the same as for the full grid. We may thus either use critical values based on the limit distribution $|W|_{\rho_2,\mathcal{G}}$, or the more conservative critical values based on quantiles of $|W|_{\rho_2} \geq |W|_{\rho_2,\mathcal{G}}$ which are tabulated in Table 1 for iid noise.

We highlight two specific choices of sparse grids. The first option is the dyadic grid given by

$$\mathcal{G}_{\text{dyadic}} = \bigcup_{m=1}^{\infty} \mathcal{G}_{\text{dyadic},m}, \qquad \mathcal{G}_{\text{dyadic},m} = \bigcup_{l=0}^{\lfloor \log_2 m \rfloor} \{(k2^{-l}, (k+2)\,2^{-l}) \,|\, k = 0, 1, \ldots, 2^l - 2\}.$$

The second option, suggested by Rivera and Walther (2013), is to consider a finer resolution for shorter intervals, specifically $\mathcal{G}_{\text{RW}} = \bigcup_{m=1}^{\infty} \mathcal{G}_{\text{RW},m}$ for

$$\mathcal{G}_{\text{RW},m} = \bigcup_{l=0}^{\lfloor \log_2 m \rfloor} \left\{ \left(\tfrac{k}{6\sqrt{l}} 2^{-l}, \tfrac{j}{6\sqrt{l}} 2^{-l} \right) \,\middle|\, k, j = 0, \ldots, \lfloor 2^l \cdot 6\sqrt{l} \rfloor \text{ such that } 1 \leq \tfrac{|k-j|}{6\sqrt{l}} \leq 2 \right\}.$$

In practice, based on sample size n, we evaluate the statistic on the grid $\mathcal{G}_{\text{dyadic},n}$ resp. $\mathcal{G}_{\text{RW},n}$. Since both sets have a cardinality of order $\mathcal{O}(n)$ resp. $\mathcal{O}(n \log n)$, this leads to a significant computational speedup.

All guarantees on the false discoveries remain valid, including Theorem 3.3, Proposition 3.4, the first claim of Theorem 3.5, and Theorem 3.7. Moreover, the procedure based on the sparse grid achieves the same asymptotic detection performance as the full grid, as $\mathcal{G}_{\text{dyadic}}$, and thus also $\mathcal{G}_{\text{RW}} \supset \mathcal{G}_{\text{dyadic}}$, satisfy condition (7) below, for $K = 3$.

Proposition 3.8. *Suppose that $\mathcal{G}$ is such that for all $0 \leq u < v \leq 1$, there exist $(u', v') \in \mathcal{G}$ such that for some $K \geq 1$,*

$$u' \leq \ u < v \ \leq v' \quad \textit{and} \quad |u' - v'| \leq K|u - v|. \tag{7}$$

Then the sparse test statistic $|S_n|_{\rho_2,\mathcal{G}}$ maintains detection power in the signal discovery problem, i.e. (5)*. Moreover, the sparse procedure isolates and localizes detectable changes in the sense of* (6)*.*

Table 1 presents the quantiles of $|B|_{\rho_2,\mathcal{G}}$ for both sparse grids, where B is a standard Brownian motion. These quantiles serve as critical values for the multiplicatively weighted multiscale tests.

A Proofs

We will repeatedly make use of the following version of the Arzela-Ascoli theorem. For completeness, we include a proof.

Theorem A.1 (Arzela-Ascoli)**.** *Let ρ and ρ' be moduli of continuity such that $\rho' \gg \rho$. Then the embedding $C^\rho \hookrightarrow C_0^{\rho'}$ is compact.*

Proof of Theorem A.1. It suffices to show that every bounded sequence in C^ρ admits a convergent subsequence in $C_0^{\rho'}$. To this end, let $(x_n)_n$ be a bounded sequence in C^ρ, i.e. $\|x_n\|_\rho \leq M$ for some $M > 0$. Note that, by Arzela-Ascoli, x_n admits a convergent subsequence in $C[0,1]$ with respect to the usual sup-norm, since x_n is uniformly continuous, as $|x_n(u) - x_n(v)| \leq M\rho(|u-v|) < \tilde{\varepsilon}$ for all $u, v \in [0,1]$, $|u - v| \leq \delta$, where $\delta > 0$ is chosen such that $\rho(|u - v|) < \tilde{\varepsilon}/M$, $|u - v| \leq \delta$ for arbitrary $\tilde{\varepsilon} > 0$. Without loss of generality, assume that $x_n \to x$ in C[0,1] for some $x \in C[0,1]$. In particular, x_n is Cauchy with respect to $|\cdot(0)|$ and the sup norm in $C[0,1]$. Since $C_0^{\rho'}$ is a Banach space, it suffices to show that x_n is Cauchy with respect to the seminorm $|\cdot|_{\rho'}$. To this end, for any $\Delta \in (0,1)$, introduce the notation

$$|x|_{\rho,\geq\Delta} = \sup_{|u-v|\geq\Delta} |x(u) - x(v)|/\rho(|u - v|),$$
$$|x|_{\rho,\leq\Delta} = \sup_{|u-v|\leq\Delta} |x(u) - x(v)|/\rho(|u - v|).$$

Now let $\varepsilon > 0$ be arbitrary. For any $n, m \in \mathbb{N}$ and $\Delta > 0$ such that $\rho(|u-v|)/\rho'(|u-v|) < \varepsilon/(4M)$ for all $|u-v| \leq \Delta$, we find that

$$\begin{aligned}
|x_n - x_m|_{\rho'} &\leq |x_n - x_m|_{\rho', \geq \Delta} + |x_n - x_m|_{\rho', \leq \Delta} \\
&\leq \frac{\|x_n - x_m\|_{C[0,1]}}{\rho'(\Delta)} + (|x_n|_{\rho, \leq \Delta} + |x_m|_{\rho, \leq \Delta}) \sup_{|u-v| \leq \Delta} \frac{\rho(|u-v|)}{\rho'(|u-v|)} \\
&\leq \frac{\|x_n - x_m\|_{C[0,1]}}{\rho'(\Delta)} + \frac{\varepsilon}{4M} (|x_n|_\rho + |x_m|_\rho) \\
&= \frac{\|x_n - x_m\|_{C[0,1]}}{\rho'(\Delta)} + \frac{\varepsilon}{2}
\end{aligned}$$

Thus, for $n, m \geq N$, where $N \in \mathbb{N}$ is such that $\|x_n - x_m\|_{C[0,1]} \leq \varepsilon \rho'(\Delta)/2$, we obtain $|x_n - x_m|_{\rho'} < \varepsilon$, i.e. x_n is Cauchy with respect to $|\cdot|_{\rho'}$ and thus also with respect to $\|\cdot\|_{\rho'}$. Hence, $C^\rho \hookrightarrow C_0^{\rho'}$ is a compact embedding. □

Proof of Proposition 1.2

Let $I_n \supseteq [a_n, b_n]$ such that $|I_n| \asymp h_n \geq l_n$. Then $\mathbb{E}(T_n(I_n)) = \mu_n l_n / \sqrt{|I_n|}$ and $\mathrm{Var}(T_n(I_n)) = \sigma^2$. If $\mu_n^2 l_n \frac{l_n}{h_n} \gg \log(\frac{n}{h_n})$, then $T_n(I_n) - \sigma\sqrt{2\log(\frac{n}{|I_n|})} \to \infty$ in probability, and thus the test is consistent.

On the other hand, if $\mu_n^2 l_n (\frac{l_n}{h_n}) = \mathcal{O}(1)$, then $\sup_{|I| \geq h_n} |\mathbb{E}(T_n(I))| = \mathcal{O}(1)$, and thus

$$T_{n,h_n}^{\mathrm{DS}} \leq \sup_{|I| \geq h_n} |\mathbb{E}(T_n(I))| + \sup_{|I| \geq h_n} \left\{ T_n(I) - \mathbb{E}(T_n(I)) - \sigma\sqrt{2\log(n/|I|)} \right\}_+ = \mathcal{O}_P(1),$$

as the second term is stochastically bounded (Dümbgen and Spokoiny, 2001).

Proof of Theorem 2.2

We make use of the following lemma, which is adapted from (Schilling, 2021, Lemma 10.4).

Lemma A.2. *Let $\rho_\alpha(h) = \sqrt{h} \log^{1/\alpha}(1/h)$. Then for all $\kappa \in \mathbb{R}$ and $h < 1/2$*

$$\rho_\alpha(2^\kappa h) \leq [|\kappa| + 1]^{\frac{1}{\alpha}} \sqrt{2^\kappa} \rho_\alpha(h),$$

i.e., weights of the form ρ_α satisfy condition (R) *in Theorem 2.2.*

Proof of Lemma A.2. We have for $h > 0$ such that $\log(2) < |\log(h)|$, i.e. $h < 1/2$,

$$\frac{\rho_\alpha^\alpha(2^\kappa h)}{\rho_\alpha^\alpha(h)} = \frac{(2^\kappa)^{\frac{\alpha}{2}} |\log(2^\kappa h)|}{|\log(h)|} \leq \frac{(2^\kappa)^{\frac{\alpha}{2}} (|\log(2^\kappa)| + |\log(h)|)}{|\log(h)|} \leq (2^\kappa)^{\frac{\alpha}{2}} (|\kappa| + 1),$$

and thus $\rho_\alpha(2^\kappa h) \leq \sqrt{2^\kappa} (|\kappa| + 1)^{\frac{1}{\alpha}} \rho_\alpha(h)$. □

Theorem A.3. *Let W_n be a sequence of stochastic processes taking values in a metric space $(\mathcal{M}, d)$, and with paths in $C[0,1]$, and $\rho_0 : [0,1] \to [0,\infty)$ an increasing function with $\rho_0(0) = 0$. Suppose there exist $C > 0$ and for any $t > C$ a $\kappa(t) > 1$ and $K(t) > 0$, such that*

$$P\left(\frac{d(W_n(u), W_n(v))}{\rho_0(|u-v|)} > t\right) \leq K(t)|u-v|^{\kappa(t)}, \qquad u, v \in [0,1]. \tag{T-M}$$

Assume further that there exist $p > 0$ and $\widetilde{K} > 0$ such that

$$\rho_0(zh) \leq \widetilde{K} z^p \rho_0(h), \qquad h, z \in [0,1]. \tag{R}$$

Then, for any $t > C$,

$$\sup_n P\left(\sup_{u,v\in[0,1],|u-v|\leq h} \frac{d(W_n(u), W_n(v))}{\rho_0(|u-v|)} > t\right) \longrightarrow 0, \quad \text{as } h \to 0, \tag{8}$$

Proof of Theorem A.3. We establish (8) analogously to the proof of Levy's continuity theorem (Schilling, 2021, Thm. 10.6). Let $t > C$ be arbitrary and fix some small $\delta > 0$ and define the event A_m and the probability R_m as

$$\begin{aligned}
A_m &= \left\{ \max_{l=1,\ldots,2^{\lfloor m\delta \rfloor}} \max_{j=0,\ldots,2^m-l} \frac{d(W_n(\frac{l+j}{2^m}), W_n(\frac{j}{2^m}))}{\rho_0(\frac{l}{2^m})} \leq t \right\}, \\
R_m = P(A_m^c) &= P\left(\max_{l=1,\ldots,2^{\lfloor m\delta \rfloor}} \max_{j=0,\ldots,2^m-l} \frac{d(W_n(\frac{l+j}{2^m}), W_n(\frac{j}{2^m}))}{\rho_0(\frac{l}{2^m})} > t \right) \\
&\leq \sum_{l=1}^{2^{\lfloor m\delta \rfloor}} \sum_{j=0}^{2^m-l} P\left(d(W_n(\tfrac{l+j}{2^m}), W_n(\tfrac{j}{2^m})) > t\rho_0(\tfrac{l}{2^m})\right) \\
&\leq 2^m \sum_{l=1}^{2^{\lfloor m\delta \rfloor}} \left(\frac{l}{2^m}\right)^{\kappa(t)} \\
&\leq 2^m 2^{\lfloor m\delta \rfloor} \left(\frac{2^{\lfloor m\delta \rfloor}}{2^m}\right)^{\kappa(t)} \approx 2^{m(1+\delta)-m(1-\delta)\kappa(t)},
\end{aligned}$$

which is summable for any $\kappa(t) > 1$, choosing $0 < \delta < [\kappa(t) - 1] \cdot [1 + \kappa(t)]^{-1}$.
Now we apply a chaining argument as follows: Denote $T_k = \{j2^{-k} \mid j = 0, \ldots, 2^k\}$. For any $0 \leq u \leq v \leq 1$ with $|u-v| \leq h$, choose $m \in \mathbb{N}$ such that $2^{-(m+1)(1-\delta)} < h \leq 2^{-m(1-\delta)}$. We can then find $u_m, v_m \in T_m$ such that $u \leq u_m \leq v_m \leq v$, and sequences $u_k, v_k \in T_k$, $k \geq m$, such that

$$\begin{aligned}
u_k \downarrow u, &\qquad |u_k - u_{k+1}| \leq 2^{-k-1}, \\
v_k \uparrow v, &\qquad |v_k - v_{k+1}| \leq 2^{-k-1}.
\end{aligned}$$

This construction yields, on the event $\underline{A}_m = \bigcap_{k=m}^{\infty} A_k$ with probability $P(\underline{A}_m) \to 1$,

$$
\begin{aligned}
&\sup_{|u-v|\le h} d(W_n(u), W_n(v)) \\
&\le \sup_{|u-v|\le h} d(W_n(u_m), W_n(v_m)) + \sum_{k=m}^{\infty} [d(W_n(u_k), W_n(u_{k+1})) + d(W_n(v_k), W_n(v_{k+1}))] \\
&\le t\rho_0(u_m - v_m) + 2t \sum_{k=m+1}^{\infty} \rho_0(2^{-k}) \\
&= t\rho_0(u_m - v_m) + 2t \sum_{k=0}^{\infty} \rho_0(2^{-k} 2^{-(m+1)(1-\delta)} 2^{(m+1)\delta}) \\
&\le t\rho_0(u_m - v_m) + 2t \sum_{k=0}^{\infty} \rho_0(2^{-k} h 2^{(m+1)\delta}) \\
&\le t\rho_0(h) + 2t \sum_{k=m+1}^{\infty} \rho_0(2^{-k(1-\delta)} h) \\
&\overset{*}{\le} t\rho_0(h) + 2t \sum_{k=m+1}^{\infty} \tilde{K} 2^{-k(1-\delta)p} \rho_0(h) \\
&\le t\rho_0(h) \left[1 + 2\epsilon_m\right],
\end{aligned}
$$

for some sequence $\epsilon_m \to 0$. At the step $(*)$, we use (R). Now let $M \in \mathbb{N}$ such that $\Delta \le 2^{-M(1-\delta)}$. Since $\underline{A}_M \subset \underline{A}_m$ for all $m \ge M$, we may conclude that in the event $\underline{A}_M$,

$$
\sup_{|u-v|\le h} d(W_n(u), W_n(v)) \le t\rho_0(h)[1 + \epsilon_M], \qquad \forall h \le \Delta.
$$

Hence, for any $\epsilon > 0$ and Δ small enough such that $2\epsilon_M = 2\epsilon_{M(\Delta)} < \epsilon$,

$$
P\left(\sup_{|u-v|\le\Delta} \frac{d(W_n(u), W_n(v))}{\rho_0(|u-v|)} \;>\; t(1+\epsilon) \right) \le P(\underline{A}_M^c),
$$

which tends to zero as $M \to \infty$, i.e. as $\Delta \to 0$. As $\epsilon > 0$ and $t > C$ are arbitrary, we actually obtain for any $t > C$ that

$$
Q(t, \Delta) := \sup_n P\left(\sup_{|u-v|\le\Delta} \frac{d(W_n(u), W_n(v))}{\rho_0(|u-v|)} \;>\; t \right) \to 0, \qquad \text{as} \quad \Delta \to 0.
$$

This establishes (3). □

Now proceed to the proof of Theorem 2.2. First, observe that (3) is a consequence of (8) for the metric space $(\mathbb{R}, |\cdot|)$.

From (3), we may conclude that $\|W_n\|_{\rho_0}$ is stochastically bounded as follows. For any $N \in \mathbb{N}$ and any $t > 6$,

$$
P\left(\|W_n\|_{\rho_0} > t\right) \le Q(\tfrac{t}{3}, \tfrac{1}{N}) + \sum_{i,j=1}^{N} P\left(|W_n(\tfrac{i}{N}) - W_n(\tfrac{j}{N})| > \tfrac{t}{3}\rho_0(\tfrac{1}{N})\right)
$$

$$\leq Q(2, \tfrac{1}{N}) + N^2 \max_i P\left(|W_n(\tfrac{i}{N})| > \tfrac{t}{6}\rho_0(\tfrac{1}{N})\right).$$

The second term tends to zero as $t \to \infty$, uniformly in n, because the finite dimensional distributions of W_n converge. For any $\epsilon > 0$, we may choose $N = N(\epsilon)$ such that the first term is smaller than $\frac{\epsilon}{2}$, and then $t = t(N(\epsilon), \epsilon) = t(\epsilon)$ big enough such that the second term is smaller than $\frac{\epsilon}{2}$. Thus, $\|W_n\|_{\rho_0}$ is stochastically bounded. By virtue of Theorem A.1, this boundedness also implies tightness in C_0^ρ for any $\rho \ll \rho_0$, and thus establishes the first claim of (4).

Regarding the second claim of (4), i.e. the thresholed weak convergence, observe that for any $\Delta > 0$ and ρ such that $\rho(h)/\rho_0(h) \to 1$ as $h \to 0$,

$$\begin{aligned} &P(\Delta, W_n) \leq P\left(\|W_n\|_\rho > t\right) \leq P(\Delta, W_n) + R(\Delta, W_n), \\ \text{where}\quad &P(\Delta, W_n) = P\left(\sup_{|u-v|\geq\Delta} \frac{|W_n(u) - W_n(v)|}{\rho(|u-v|)} > t\right), \\ &R(\Delta, W_n) = P\left(\sup_{|u-v|\leq\Delta} \frac{|W_n(u) - W_n(v)|}{\rho(|u-v|)} > t\right) \end{aligned}$$

By virtue of the weak convergence $W_n \Rightarrow W$ in $C[0,1]$ established above, we conclude that $P(\Delta, W_n) \to P(\Delta, W)$ for any $\Delta > 0$ as $n \to \infty$. Moreover, (3) yields $\lim_{\Delta\to 0}\sup_n R(\Delta, S_n) = 0$, also for ρ, since for arbitrary $\tilde{\varepsilon} > 0$ and small enough $\tilde{\Delta} > 0$ it holds that

$$R(\tilde{\Delta}, W_n) \leq P\left(\sup_{|u-v|\leq\tilde{\Delta}} \frac{|W_n(u) - W_n(v)|}{\rho_0(|u-v|)} > t(1-\tilde{\varepsilon})\right).$$

Thus the claim holds for all $t > C/(1-\tilde{\varepsilon})$, and since $\tilde{\varepsilon} > 0$ can be chosen arbitrary, the second claim of (4) follows for all $t > C$ as desired. This completes the proof of Theorem 2.2.

Proof of Theorem 2.3

The boundedness in C^{ρ^*} implies tightness in $C[0,1]$ (Theorem A.1), and thus $W_n \Rightarrow W$ in $C[0,1]$. By Skorokhod's representation theorem, upon potentially changing the probability space, $\|W_n - W\|_{C[0,1]} \xrightarrow{\mathbb{P}} 0$. Next, for any $\Delta > 0$ and any stochastic process X, denote the quantities

$$\begin{aligned} \|X\|_{\rho^*,\geq\Delta} &= \sup_{|u-v|\geq\Delta} \frac{|X(u) - X(v)|}{\rho^*(|u-v|)}, \\ \|X\|_{\rho^*,\leq\Delta} &= \sup_{|u-v|\leq\Delta} \frac{|X(u) - X(v)|}{\rho^*(|u-v|)}, \end{aligned}$$

and $\|X\|_{\rho,\geq\Delta}$ and $\|X\|_{\rho,\leq\Delta}$ are to be understood in the same way. Then

$$\|W_n - W\|_\rho \leq \|W_n - W\|_{\rho,\geq\Delta} + \|W_n - W\|_{\rho,\leq\Delta}.$$

For any fixed $\Delta > 0$, the first term tends to zero in probability as $n \to \infty$ because $\|W_n - W\|_{C[0,1]} \xrightarrow{\mathbb{P}} 0$. For the right hand term, note that

$$\begin{aligned}
\|W_n - W\|_{\rho,\leq\Delta} &\leq \|W_n\|_{\rho,\leq\Delta} + \|W\|_{\rho,\leq\Delta} \\
&= \left(\sup_{h\leq\Delta} \frac{\rho^*(h)}{\rho(h)}\right) (\|W_n\|_{\rho^*,\leq\Delta} + \|W\|_{\rho^*,\Delta}) \quad \leq C(\Delta) \left(\|W_n\|_{\rho^*} + \|W\|_{\rho^*}\right),
\end{aligned}$$

where $C(\Delta) \to 0$, as $\Delta \to 0$, since $\rho \gg \rho^*$. Since $\|W_n\|_{\rho^*} + \|W\|_{\rho^*}$ is stochastically bounded, for any $\varepsilon > 0$ and $\delta > 0$, there are $N \in \mathbb{N}$ and $\tilde{\Delta} > 0$ such that

$$P(\|W_n - W\|_{\rho,\leq\Delta} > \varepsilon) < \delta,\ n \geq N,\ 0 < \Delta \leq \tilde{\Delta}.$$

Thus, $\|W_n - W\|_\rho \xrightarrow{\mathbb{P}} 0$ follows for every $\rho \gg \rho^*$.

Proof of Proposition 2.5

For arbitrary $T > 0$, let $\sigma > 0$ and $p \in (0, 1)$ be such that $1/(\sigma\sqrt{p}) > T$. Let ϵ_t be iid symmetric random variables with $P(|\epsilon_t| > r) = \exp(-r^2)$, and $\sigma^2 = \mathrm{Var}(\epsilon_t)$. For any $p \in (0, 1)$, let $\xi_t \overset{iid}{\sim} \mathrm{bin}(1, p)$ and define $\eta_t = \epsilon_t\xi_t/\sqrt{\sigma^2 p}$, such that $\mathrm{Var}(\eta_t) = 1$. Define the interpolated partial sum process

$$S_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nu\rfloor} \eta_t + \sqrt{n}\left[u - \tfrac{\lfloor un\rfloor}{n}\right] \eta_{\lceil un\rceil}.$$

Then $\|S_n\|_{\rho_2} \geq \max_{t=1,\ldots,n} |\eta_t|/\sqrt{1 + \log n}$. Hence,

$$\begin{aligned}
P\left(\|S_n\|_{\rho_\alpha} \leq c\right) &\leq P\left(|\eta_1| \leq c(1 + \log n)^{\frac{1}{2}}\right)^n \\
&= \left(1 - p + p\exp\left(-\left(c\sigma\sqrt{p}\right)^2 (1 + \log n)\right)\right)^n \\
&\leq \left(1 - pn^{-\left(\frac{c}{\sigma\sqrt{p}}\right)^2}\right)^n.
\end{aligned}$$

For $c < 1/\sqrt{\sigma^2 p}$, the latter term tends to zero as $n \to \infty$. This shows that $P(\|S_n\|_{\rho_2} \geq \frac{1}{\sigma\sqrt{p}}) \to 1$, proving Proposition 2.5 by the specific choice of σ and p. .

Proof of Theorem 2.6

This proof is inspired by (Liu et al., 2013, Thm. 1). Let $S_n = \sum_{t=1}^n w_t\eta_t$, and

$$\begin{aligned}
\eta_{t,j} &= \mathbb{E}(\eta_t|\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-j}), \\
S_{n,j} &= \sum_{t=1}^n w_t\eta_{t,j},
\end{aligned}$$

$$Y_{i,j} = \sum_{t=(i-1)j+1}^{(ij)\wedge n} w_t(\eta_{t,j} - \eta_{t,j-1}), \qquad i = 1, \ldots, \lfloor n/j \rfloor + 1.$$

By telescoping, we find that

$$\begin{aligned} \|S_n\|_{\psi_2} &\leq \sum_{j=1}^{\infty} \|S_{n,j} - S_{n,j-1}\|_{\psi_2} \\ &\leq \sum_{j=1}^{\infty} \left\| \sum_{i \text{ is odd}} Y_{i,j} \right\|_{\psi_2} + \sum_{j=1}^{\infty} \left\| \sum_{i \text{ is even}} Y_{i,j} \right\|_{\psi_2}. \end{aligned} \tag{9}$$

Observe that the $Y_{1,j}, Y_{3,j}, \ldots$ are independent by construction, and the same holds for the even indices. Hence, using that $\|\eta_{t,j} - \eta_{t,j-1}\|_{\psi_2} \leq \delta(j)$,

$$\begin{aligned} \left\| \sum_{i \text{ is odd}} Y_{i,j} \right\|_{\psi_2} &\leq \tilde{K} \sqrt{\sum_i \|Y_{i,j}\|_{\psi_2}^2} \\ &\leq \tilde{K} \sqrt{\sum_i \left(\sum_{t=(i-1)j+1}^{(ij)\wedge n} |w_t| \delta_{\psi_2}(j) \right)^2} \\ &\leq \tilde{K} \sqrt{\sum_i \left(\sum_{t=(i-1)j+1}^{(ij)\wedge n} |w_t|^2 \right) (j \delta_{\psi_2}(j)^2)} \\ &= \tilde{K} \sqrt{j} \delta_{\psi_2}(j) \sqrt{\sum_{t=1}^{n} |w_t|^2}. \end{aligned} \tag{10}$$

for some universal constant $\tilde{K} > 0$, stemming from the sub-Gaussian concentration bound. The same bound holds for the even indices i. Combining (9) and (10) yields the result for $K = 2\tilde{K}$.

Proof of Corollaries 2.4 and 2.7

Proof of Corollary 2.4. We just need to verify the tightness condition (T). For arbitrary $0 \leq u < v \leq 1$, we may write $S_n(u) - S_n(v) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \eta_t w_n(t, u, v)$ for weights $w_n(t, u, v) \in [0, 1]$ such that $\sum_{t=1}^{n} w_n(t, u, v)^2 \leq \sum_t w_n(t, u, v) = |v - u|$. Then (T) is a consequence of Hoeffding's inequality. □

Proof of Corollary 2.7. As in Corollary 2.4, we obtain (T) as a consequence of Theorem 2.6. The convergence of the finite dimensional marginals may be obtained, for example, via Theorem 5 of Mies (2024). □

Proof of Theorem 3.1

Under the null, $\widehat{\sigma}_n^2$ is consistent as an average of a 1-dependent sequence. Corollary 2.4 yields $(T_n^* \vee \tau) \Rightarrow (\sigma|B|_{\rho_2} \vee \tau)$ for any $\tau > \sigma C_\eta$. By Slutsky's Lemma and the consistency of $\widehat{\sigma}_n^2$, we find $(\frac{T_n^*}{\widehat{\sigma}_n} \vee \tau) \Rightarrow (|B|_{\rho_2} \vee \tau)$ for any $\tau > C_\eta$, which yields the first claim.

Under the alternative, $\widehat{\sigma}_n^2$ is still consistent because

$$\widehat{\sigma}_n^2 = \frac{1}{2(n-1)} \sum_{t=2}^{n} (\eta_t - \eta_{t-1} + f(t) - f(t-1))^2 = \frac{1}{2n} \sum_{t=2}^{n} (\eta_t - \eta_{t-1})^2 + \mathcal{O}_P(\mu_n^2/n).$$

Moreover, $T_n^* \geq T_n(I_n)/\sqrt{\log \frac{en}{|I_n|}} \to \infty$ in probability. The latter convergence holds because $\mathrm{Var}(\sum_{t \in I_n} Y_t/\sqrt{|I_n|}) = \sigma$ whereas $\mathbb{E}(\sum_{t \in I_n} Y_t/\sqrt{|I_n|}) = \sqrt{|I_n|}\mu_n \to \infty$, which implies $T_n(I_n) \asymp \sqrt{|I_n|}\mu_n \gg \sqrt{\log \frac{en}{|I_n|}}$ by assumption.

Proof of Proposition 3.2

For $A > C_\eta$, Corollary 2.4 and (3) yield for any $t > 0$

$$\lim_{h \downarrow 0} \sup_n P\left(\sup_{|I| \leq h} \{T_n(I) - A\sqrt{\log \tfrac{en}{|I|}}\}_+ > t \right) = 0.$$

Moreover,

$$\begin{aligned}
\sup_{|I|>h} \left\{ T_n(I) - A\sqrt{\log \tfrac{en}{|I|}} \right\}_+ &\underset{n\to\infty}{\Rightarrow} \sup_{\substack{u \leq v \\ |u-v|>h}} \left\{ \frac{|B(v) - B(u)|}{\sqrt{v-u}} - A\sqrt{\log \tfrac{e}{|u-v|}} \right\}_+ \\
&\underset{h\to 0}{\Rightarrow} \sup_{u \leq v} \left\{ \frac{|B(v) - B(u)|}{\sqrt{v-u}} - A\sqrt{\log \tfrac{e}{|u-v|}} \right\}_+ .
\end{aligned}$$

Standard arguments yield the weak convergence of the full statistic (Billingsley, 1999, Thm. 3.2).

The second claim is a direct consequence of Proposition 2.5.

Proof of Theorem 3.3

Since $T_n^*(\mathbb{F}_0) \leq T_n^*(f_n)$ under $\mathbb{H}_0$, the proof is identical to Theorem 3.1 if we can show that $\widehat{\sigma}_n^2 \to \sigma^2$ in probability. To this end, we observe that

$$\begin{aligned}
&\left| \widehat{\sigma}_n^2 - \frac{1}{2(n-1)} \sum_{t=2}^{n} (\eta_t - \eta_{t-1})^2 \right| \\
&\leq \frac{1}{2(n-1)} \sum_{t=2}^{n} 2|\eta_t - \eta_{t-1}| \cdot |f_n(t) - f_n(t-1)| + \frac{1}{2(n-1)} \sum_{t=2}^{n} |f_n(t) - f_n(t-1)|^2
\end{aligned}$$

$$
\begin{aligned}
&\leq \sqrt{\frac{1}{(n-1)}\sum_{t=2}^{n}|\eta_t-\eta_{t-1}|^2}\sqrt{\frac{1}{(n-1)}\sum_{t=2}^{n}|f_n(t)-f_n(t-1)|^2}+o(1)\\
&=\mathcal{O}_P(1)\cdot o(1)+o(1),
\end{aligned}
$$

where the first term is bounded by the law of large numbers. This establishes the consistency of $\widehat{\sigma}_n^2$.

Proof of Proposition 3.4

We observe that

$$
\begin{aligned}
P\left(\exists I\in\mathcal{I}_n \text{ such that } f_n\in\mathbb{F}_0^I\right) &\leq P\left(\bigcup_{I:f_n\in\mathbb{F}_0^I}\{T_n^*(\mathbb{F}_0^I)>\widehat{\sigma}_n\cdot q_\alpha\}\right)\\
&\leq P\left(\sup_{I:f_n\in\mathbb{F}_0^I} T_n^*(\mathbb{F}_0^I)>\widehat{\sigma}_n\cdot q_\alpha\right)\\
&\leq P\left(T_n^*(\{f_n\})>\widehat{\sigma}_n\cdot q_\alpha\right),
\end{aligned}
$$

which is asymptotically less than α.

Proof of Theorem 3.5

The first claim is a consequence of Proposition 3.4. For the power statement, let k such that $\tau_k\in\mathcal{D}(z)$ and set $I_n(\tau_k)=[\tau_k-\frac{w_k}{2},\tau_k+\frac{w_k}{2}]$ for $w_k=w(\tau_k,z)$ such that $I_n(\tau_k)\cap\{\tau_1,\ldots,\tau_\kappa\}=\{\tau_k\}$. Then

$$
T_n^*\left(\mathbb{F}_{\text{const}}^{I_n(\tau_k)}\right)=\inf_{a\in\mathbb{R}}\sup_{[un,vn]\subset I_n(\tau_k)}\frac{|S_n(v)-S_n(u)-[v-u]a|}{\rho_2(v-u)}.
$$

Denote $\mu_k=\sum_{j=0}^{k-1}\delta_j=f_n(\tau_k-1)$, and assume without loss of generality that $\delta_k>0$. For any $a\in\mathbb{R}$, we have $a/\sqrt{n}\leq\mu_k+\frac{\delta_k}{2}$ or $a/\sqrt{n}\geq\mu_k+\frac{\delta_k}{2}$. Assume for now the latter case. Then

$$
\begin{aligned}
&\sup_{[un,vn]\subset I_n(\tau_k)}\frac{|S_n(v)-S_n(u)-[v-u]a|}{\rho_2(v-u)}\\
&\geq\frac{\left|S_n((\tau_k+\frac{w_k}{2})/n)-S_n(\tau_k/n)-\frac{w_k a}{2n}\right|}{\rho_2(\frac{w_k}{2n})}\\
&=\frac{\left|\frac{w_k}{2\sqrt{n}}(\mu_k+\delta_k)+\widetilde{S}_n((\tau_k+\frac{w_k}{2})/n)-\widetilde{S}_n(\tau_k/n)-\frac{w_k a}{2n}\right|}{\rho_2(\frac{w_k}{2n})}\\
&\geq\frac{\frac{w_k\delta_k}{4\sqrt{n}}}{\rho_2(\frac{w_k}{2n})}-|\widetilde{S}_n|_{\rho_2}
\end{aligned}
$$

$$\geq c \frac{\sqrt{w_k}\delta_k}{\sqrt{\log \frac{en}{2w_k}}} - |\widetilde{S}_n|_{\rho_2} \quad \geq \quad c\sqrt{z} - |\widetilde{S}_n|_{\rho_2},$$

for some small $c > 0$. If instead $a/\sqrt{n} \leq \mu_k + \frac{\delta_k}{2}$, we observe that

$$\begin{aligned}
&\sup_{[un,vn]\subset I_n(\tau_k)} \frac{|S_n(v) - S_n(u) - [v-u]a|}{\rho_2(v-u)} \\
&\geq \frac{\left|S_n(\tau_k/n) - S_n((\tau_k - \frac{w_k}{2})/n) - \frac{w_k a}{2n}\right|}{\rho_2(\frac{w_k}{2n})} \\
&\geq \frac{\left|\widetilde{S}_n(\tau_k/n) - \widetilde{S}_n((\tau_k - \frac{w_k}{2})/n) + \frac{w_k\mu_k}{2\sqrt{n}} - \frac{w_k a}{2n}\right|}{\rho_2(\frac{w_k}{2n})} \\
&\geq \frac{\frac{w_k\delta_k}{4\sqrt{n}}}{\rho_2(\frac{w_k}{2n})} - |\widetilde{S}_n|_{\rho_2} \quad \geq \quad c\sqrt{z} - |\widetilde{S}_n|_{\rho_2}.
\end{aligned}$$

By virtue of Theorem 2.2 and Corollary 2.7, the random variable $|\widetilde{S}_n|$ is stochastically bounded, and thus $T^*(\mathbb{F}_{\text{const}}^{I_n(\tau_k)}) \to \infty$ as $z \to \infty$, showing that $I_n(\tau_k) \in \mathcal{I}_n$ eventually. Note furthermore that $I_n(\tau_k) \cap \{\tau_1, \ldots, \tau_\kappa\} = \{\tau_k\}$. Thus,

$$\begin{aligned}
&P\left(\forall \tilde{\tau} \in \mathcal{D}(z)\ \exists I \in \mathcal{I}_n^\tau \text{ such that } I \cap \{\tau_1, \ldots, \tau_\kappa\} = \{\tilde{\tau}\}\right) \\
&\geq P\left(\forall \tau_k \in D(z)\ :\ T^*(\mathbb{F}_{\text{const}}^{I_n(\tau_k)}) \geq \widehat{\sigma}_n \cdot q_\alpha\right) \\
&\geq P\left(c \cdot \sqrt{z}\ >\ \widehat{\sigma}_n \cdot q_\alpha + |\widetilde{S}_n|_{\rho_2}\right),
\end{aligned}$$

which tends to one as $n \to \infty$ and $z \to \infty$.

Proof of Proposition 3.6

Because any $I \in \mathcal{I}$ contains at least one point $\tau \in \widetilde{\mathcal{D}}$, we may decompose

$$\bigcup_{\tau\in\widetilde{\mathcal{D}}} \widetilde{\mathcal{I}}(\tau)\ \subseteq \mathcal{I} \subseteq\ \bigcup_{\tau\in\widetilde{\mathcal{D}}} \mathcal{I}(\tau),$$

for

$$\mathcal{I}(\tau) = \{I \in \mathcal{I} \,|\, \tau \in I\}, \qquad \widetilde{\mathcal{I}}(\tau) = \{I \in \mathcal{I}(\tau) \text{ such that } |I| \leq r(\tau)\} \neq \emptyset.$$

Choosing one interval $\widehat{I}_1$ of shortest length, we find that $\widehat{I}_1 \in \widetilde{\mathcal{I}}(\tau_{(1)})$ for some $\tau_{(1)} \in \widetilde{\mathcal{D}}$. Since $\tau_{(1)} \in \widehat{I}_1$, the class $\mathcal{I}(\tau_{(1)})$ is removed from the candidate set, while all other $\widetilde{\mathcal{I}}(\tau_{(k)})$ are still relevant as they do not contain $\tau_{(1)}$. Thus, at the next step, we have

$$\bigcup_{\tau\in\widetilde{\mathcal{D}}\setminus\{\tau_{(1)}\}} \widetilde{\mathcal{I}}(\tau)\ \subseteq \mathcal{I}^1 \subseteq\ \bigcup_{\tau\in\widetilde{\mathcal{D}}\setminus\{\tau_{(1)}\}} \mathcal{I}(\tau).$$

Proceeding inductively, we find that $\widehat{I}_k \in \widetilde{\mathcal{I}}(\tau_{(k)})$ for some $\tau_{(k)} \in \widetilde{\mathcal{D}}$. This implies that the iteration stops after exactly $|\widetilde{\mathcal{I}}|$ steps, with the claimed guarantees.

Proof of Theorem 3.7

Lemma A.4. *Under the conditions of Theorem 3.7,*

$$\sup_{u\in[0,1]}\left|\widehat{\Sigma}_n(u)-\Sigma(u)\right| \xrightarrow{\mathbb{P}} 0,$$

and for some constant K, and all $c\in(0,1)$,

$$\sup_{|u-v|>c}\left|\widehat{\Sigma}_n(u)-\widehat{\Sigma}_n(v)\right| \le Kc+\mathcal{O}_P\left(\sqrt{\frac{v_n b_n}{n}}+b_n^{-1}\right).$$

Proof of Lemma A.4. It is sufficient to consider $u=\frac{i}{n}$ and $v=\frac{j}{n}$, as both claims of the Lemma readily extend via interpolation.

Define the terms

$$\widetilde{\Sigma}_n(u)=\frac{1}{2n}\sum_{t=b_n}^{\lfloor un\rfloor-b_n}\left(\frac{1}{\sqrt{b_n}}\sum_{i=0}^{b_n-1}(\eta_{t-i}-\eta_{t+1+i})\right)^2,$$
$$\Delta_n=\frac{1}{2n}\sum_{t=b_n}^{n-b_n}\left(\frac{1}{\sqrt{b_n}}\sum_{i=0}^{b_n-1}(f_n(t-i)-f_n(t+1+i))\right)^2,$$

for $u=\frac{i}{n}$, and interpolated in between. By expanding the square and applying the Cauchy-Schwarz inequality, we find that

$$\left|\widehat{\Sigma}_n(u)-\widetilde{\Sigma}_n(u)\right|\le\Delta_n+\sqrt{\widetilde{\Sigma}_n(1)\Delta_n}. \tag{11}$$

We will show that $\Delta_n\to 0$ and $\widetilde{\Sigma}_n(u)\to\Sigma(u)$ uniformly, which implies that (11) tends to zero. The term Δ_n may be bounded as

$$\begin{aligned}\Delta_n&\le\frac{1}{2n}\sum_{t=b_n}^{\lfloor un\rfloor-b_n}\left(\frac{1}{\sqrt{b_n}}\sum_{j=-b_n}^{b_n-1}|f_n(t+j+1)-f_n(t+j)|\right)^2\\&=\frac{1}{2n}\sum_{t=b_n}^{\lfloor un\rfloor-b_n}4b_n\left(\frac{1}{2b_n}\sum_{j=-b_n}^{b_n-1}|f_n(t+j+1)-f_n(t+j)|\right)^2\\&\le\frac{1}{2n}\sum_{t=b_n}^{\lfloor un\rfloor-b_n}2\sum_{j=-b_n}^{b_n-1}|f_n(t+j+1)-f_n(t+j)|^2\\&\le\frac{2b_n}{n}\sum_{t=2}^{n}|f_n(t)-f_n(t-1)|^2\le\frac{2b_n}{n}v_n,\end{aligned}$$

which tends to zero by assumption.

To handle $\widetilde{\Sigma}_n(u)$, we decompose

$$
\begin{aligned}
\widetilde{\Sigma}_n(u) &= \frac{1}{2n}\sum_{t=b_n}^{\lfloor un\rfloor - b_n}\left(\frac{1}{\sqrt{b_n}}\sum_{i=0}^{b_n-1}\eta_{t-i}\right)^2 + \frac{1}{2n}\sum_{t=b_n}^{\lfloor un\rfloor - b_n}\left(\frac{1}{\sqrt{b_n}}\sum_{i=0}^{b_n-1}\eta_{t+i}\right)^2 \\
&\quad + \frac{1}{n}\sum_{t=b_n}^{\lfloor un\rfloor - b_n}\left(\frac{1}{\sqrt{b_n}}\sum_{i=0}^{b_n-1}\eta_{t-i}\right)\left(\frac{1}{\sqrt{b_n}}\sum_{i=0}^{b_n-1}\eta_{t+i}\right) \\
&= \tfrac{1}{2}\widetilde{\Sigma}_n^-(u) + \tfrac{1}{2}\widetilde{\Sigma}_n^+(u) + \widetilde{\Sigma}_n^\pm(u).
\end{aligned}
$$

Theorem 5.1 of Mies and Steland (2023) shows that in the regime $1 \ll b_n \ll n$, and $\beta \geq 3$,

$$
\sup_{u\in[0,1]}\left|\widetilde{\Sigma}_n^-(u) - \frac{1}{n}\sum_{t=1}^{\lfloor un\rfloor - b_n}\sigma^2_{\infty,n}(t)\right| = \mathcal{O}_P\left(\sqrt{\tfrac{b_n}{n}} + \tfrac{1}{b_n}\right),
$$

$$
\text{where}\quad \sigma^2_{\infty,n}(t) = \sum_{h=-\infty}^{\infty}\operatorname{Cov}(G_{t,n}(\boldsymbol{\epsilon}_0), G_{t,n}(\boldsymbol{\epsilon}_h)).
$$

Assumption (LS-3) implies that $\sigma^2_{\infty,n}(t)$ is bounded, see (Mies, 2024, Prop. 1), and thus $\frac{1}{n}\sum_{t=1}^{\lfloor un\rfloor - b_n}\sigma^2_{\infty,n}(t) = \int_0^u \sigma^2_{\infty,n}(\lfloor vn\rfloor)\,dv + \mathcal{O}(b_n/n)$. As shown in the proof of (Mies, 2024, Lemma 4), Assumptions (LS-1) and (LS-3) imply the convergence $\int_0^u \sigma^2_{\infty,n}(\lfloor vn\rfloor)\,dv \to \int_0^u \sigma^2_\infty(v)\,dv$, and thus $\widehat{\Sigma}_n^-(u) \xrightarrow{\mathbb{P}} \Sigma(u)$. Upon an index shift, the same arguments apply to $\widetilde{\Sigma}_n^+(u)$.

To show that $\widetilde{\Sigma}_n^\pm(u) \to 0$, we write $\widetilde{\Sigma}_n^\pm(u) = \frac{1}{n}\sum_{t=b_n}^{n-b_n}\chi_{t+b_n,n}$, for $\chi_{t+b_n,n} = \frac{1}{b_n}\sum_{i,j=0}^{b_n-1}\eta_{t-i}\eta_{t+j}$. Since $\eta_t = G_{t,n}(\boldsymbol{\epsilon}_t)$, we may also write $\chi_{t,n} = H_{t,n}(\boldsymbol{\epsilon}_t)$ for some kernel $H_{t,n}$. Proceeding as in the proof of (Mies and Steland, 2023, Thm. 5.1), in particular equations (30) and (31), we can derive that

$$
\|H_{t,n}(\boldsymbol{\epsilon}_t) - H_{t,n}(\boldsymbol{\epsilon}_{t,h})\|_{L_2} \leq K\frac{[(h-b_n)\vee 1]^{1-\beta}}{\sqrt{b_n}}
$$

for some constant K, and

$$
\left\|\max_{k=1,\ldots,n}\sum_{t=1}^{k}(\chi_{t,n} - \mathbb{E}\chi_{t,n})\right\|_{L_2} = \mathcal{O}\left(\sqrt{nb_n}\right).
$$

Hence, $\sup_u|\widetilde{\Sigma}_n^\pm(u) - \mathbb{E}\widetilde{\Sigma}_n^\pm(u)| = \mathcal{O}_P(\sqrt{b_n/n})$. Moreover,

$$
\begin{aligned}
\mathbb{E}\chi_{t+b_n,n} &\leq \frac{1}{b_n}\sum_{i,j=0}^{b_n-1}|\operatorname{Cov}(\eta_{t-i},\eta_{t+j})| \\
&\leq \frac{1}{b_n}\sum_{i,j=0}^{b_n-1}(i+j+1)^{-\beta}
\end{aligned}
$$

$$\leq \frac{K}{n}\sum_{t=b_n}^{n-b_n}\frac{1}{b_n}\sum_{i=0}^{b_n-1}(i+1)^{1-\beta} \quad \leq K b_n^{1-\beta},$$

see (Mies, 2024, Prop. 1) for the bound on the autocovariance of η_t. Thus, $\sup_u \mathbb{E}(\widetilde{\Sigma}_n^{\pm}(u)) = \mathcal{O}(b_n^{1-\beta})$. Jointly, we find

$$\sup_{u\in[0,1]}\left|\widetilde{\Sigma}_n(u) - \frac{1}{n}\sum_{t=1}^{\lfloor un\rfloor - b_n}\sigma^2_{\infty,n}(t)\right| = \mathcal{O}_P\left(\sqrt{\frac{b_n}{n}} + b_n^{-1} + b_n^{1-\beta}\right) = \mathcal{O}_P\left(\sqrt{\frac{b_n}{n}} + b_n^{-1}\right),$$

since $\beta > 3$, and in combination with (LS-3),

$$\widetilde{\Sigma}_n(u) - \Sigma(u) \xrightarrow{\mathbb{P}} 0.$$

The latter convergence holds uniformly by monotonicity and continuity of the limit $\Sigma(u)$, though without a rate, since (LS-3) does not state a rate. Moreover, together with (11), we note that

$$\sup_{u\in[0,1]}\left|\widehat{\Sigma}_n(u) - \frac{1}{n}\sum_{t=1}^{\lfloor un\rfloor - b_n}\sigma^2_{\infty,n}(t)\right| = \mathcal{O}_P\left(\sqrt{\frac{v_n b_n}{n}} + b_n^{-1}\right).$$

Since $\sigma^2_{\infty,n}(t)$ is bounded, we find that

$$\sup_{|u-v|>c_n}\left|\widehat{\Sigma}_n(u) - \widehat{\Sigma}_n(v)\right| = \mathcal{O}_P\left(\sqrt{\frac{v_n b_n}{n}} + b_n^{-1} + c_n\right).$$

□

Proof of Theorem 3.7. In the remainder of this proof, all probabilities only concern the randomness of W, i.e. we work conditionally on $\widehat{\Sigma}_n$. For any $0 \leq a < b \leq 1$, introduce

$$\begin{aligned}
Z_n(a,b) &= \sup_{|u-v|\in[a,b]}\frac{|B(\widehat{\Sigma}_n(u)) - B(\widehat{\Sigma}_n(v))|}{\rho_2(|u-v|)} \\
Z(a,b) &= \sup_{|u-v|\in[a,b]}\frac{|B(\Sigma(u)) - B(\Sigma(v))|}{\rho_2(|u-v|)}.
\end{aligned}$$

The uniform consistency of $\widehat{\Sigma}_n(u)$ implies that $Z_n(c,1) \Rightarrow Z(c,1)$ for any fixed $c \in (0,1)$. Moreover, $Z(c,1) \Rightarrow Z(0,1)$ as $c \to 0$. Using standard arguments (Billingsley, 1999, Thm. 3.2), we obtain the thresholded weak convergence $(Z_n(c_n,1)\vee\tau) \Rightarrow (Z(0,1)\vee\tau)$ if we can show that, for any $\epsilon > 0$,

$$\begin{aligned}
0 &\overset{!}{=} \lim_{c\to 0}\limsup_{n\to\infty} P\left(|(Z_n(c_n,1)\vee\tau) - (Z_n(c,1)\vee\tau)| > \epsilon\right) \qquad (12) \\
&\leq \lim_{c\to 0}\limsup_{n\to\infty} P\left(Z_n(c_n,c) > \tau\right).
\end{aligned}$$

To this end, observe that

$$
\begin{aligned}
Z_n(c_n, c) &= \sup_{|u-v|\in[c_n,c]} \frac{|B(\widehat{\Sigma}_n(u)) - B(\widehat{\Sigma}_n(v))|}{\rho_2(|u-v|)} \\
&\le \sup_{|u-v|\in[0,c]} \frac{|B(\widehat{\Sigma}_n(u)) - B(\widehat{\Sigma}_n(v))|}{\rho_2(|\widehat{\Sigma}_n(u) - \widehat{\Sigma}(v)|)} \cdot \sup_{|u-v|\ge[c_n,1]} \frac{\rho_2(|\widehat{\Sigma}_n(u) - \widehat{\Sigma}_n(v)|)}{\rho_2(|u-v|)} \\
&\le \sup_{\substack{|u-v|\in[0,\hat{c}] \\ u,v\le\widehat{\Sigma}_n(1)}} \frac{|B(u) - B(v)|}{\rho_2(|u-v|)} \cdot \sup_{u,v\in[0,1]} \frac{\rho_2(\widehat{R}|u-v|)}{\rho_2(|u-v|)},
\end{aligned}
$$

for the random variables

$$
\begin{aligned}
\hat{c} &= \sup_{|u-v|\le c} \left|\widehat{\Sigma}_n(u) - \widehat{\Sigma}_n(v)\right| \le Kc + \mathcal{O}_P\left(\frac{v_n b_n}{n} + \frac{1}{b_n}\right), \\
\widehat{R} &= \sup_{|u-v|\ge c_n} \frac{|\widehat{\Sigma}_n(u) - \widehat{\Sigma}_n(v)|}{|u-v|} \le K + \mathcal{O}_P\left(\frac{v_n b_n}{nc_n^2} + \frac{1}{b_n c_n}\right).
\end{aligned}
$$

The $\mathcal{O}_P(\dots)$ terms vanish by virtue of our rate assumptions. Lévy's Theorem on the modulus of continuity of B yields, for any $\epsilon > 0$

$$
\begin{aligned}
&\lim_{c\to\infty} \limsup_{n\to\infty} P\left(\sup_{\substack{|u-v|\in[0,\hat{c}] \\ u,v\le\widehat{\Sigma}_n(1)}} \frac{|B(u) - B(v)|}{\rho_2(|u-v|)} > \sqrt{2}(1+\epsilon) \right) \\
&\le \lim_{c\to\infty} P\left(\sup_{\substack{|u-v|\in[0,2Kc] \\ u,v\le 2\Sigma(1)}} \frac{|B(u) - B(v)|}{\rho_2(|u-v|)} > \sqrt{2}(1+\epsilon) \right) = 0.
\end{aligned}
$$

Moreover, by Lemma A.2, the function $\zeta(R) = \sup_{u,v\in[0,1]} \frac{\rho_2(R|u-v|)}{\rho_2(|u-v|)}$ is well-defined and strictly increasing. Since $P(\widehat{R} > 2K) \to 0$, we find that

$$
P\left(\sup_{u,v\in[0,1]} \frac{\rho_2(\widehat{R}|u-v|)}{\rho_2(|u-v|)} > \zeta(2K) \right) \to 0.
$$

This yields

$$
\lim_{c\to\infty} \limsup_{n\to\infty} P\left(\sup_{\substack{|u-v|\in[0,\hat{c}] \\ u,v\le\widehat{\Sigma}_n(1)}} \frac{|B(u) - B(v)|}{\rho_2(|u-v|)} \sup_{u,v\in[0,1]} \frac{\rho_2(\widehat{R}|u-v|)}{\rho_2(|u-v|)} > \sqrt{2}(1+\epsilon)\cdot\zeta(2K) \right) = 0,
$$

which establishes (12) for any $\tau > \sqrt{2}\zeta(2K)$. Thus, we find for any $t > \tau$,

$$
P\left(|B(\widehat{\Sigma}_n(u))|_{\rho_2} > t \,\middle|\, \widehat{\Sigma}_n \right) \xrightarrow{\mathbb{P}} P\left(|B(\Sigma(u))|_{\rho_2} > t \right).
$$

Note that this indeed holds for all $t > \tau$, since the limit has a continuous distribution Lifshits (1984). This limiting continuity also implies the convergence of quantiles. □

Proof of Proposition 3.8

Consider the signal detection problem first. Let $c_n = \frac{a_n}{n}$ and $d_n = \frac{a_n + l_n}{n}$. By assumption 7, there exists $(u_n, v_n) \in \mathcal{G}$ such that $u_n \leq c_n < d_n \leq v_n$ and $|u_n - v_n| \leq K\frac{l_n}{n}$. Then

$$|S_n|_{\rho_2,\mathcal{G}} \geq \frac{|S_n(v_n) - S_n(u_n)|}{\rho_2(K|d_n - c_n|)}.$$

Similar to the proof of Theorem 3.1, we have $|\mathbb{E}(S_n(v_n) - S_n(u_n))| \geq \sqrt{n}|d_n - c_n||\mu_n|$ and $\mathrm{Var}(S_n(v_n) - S_n(u_n)) = \sqrt{|v_n - u_n|} \leq \sqrt{K|d_n - c_n|}$, which yields

$$\begin{aligned}\frac{|S_n(v_n) - S_n(u_n)|}{\rho_2(K|d_n - c_n|)} &\asymp \frac{\sqrt{n}|\mu_n|\sqrt{|d_n - c_n|}}{\sqrt{\log|d_n - c_n|}} \\ &\asymp \frac{|\mu_n|\sqrt{l_n}}{\sqrt{\log(n/l_n)}},\end{aligned}$$

which tends to infinity by assumption.

Now consider the change localization problem. Denote $w_k = w(\tau_k, z)$ for $\tau_k \in \mathcal{D}(z)$, as well as $c_n = \frac{\tau_k - \frac{w_k}{4K}}{n}$ and $d_n = \frac{\tau_k + \frac{w_k}{4K}}{n}$. By assumption (7), there are $(u_n, v_n) \in \mathcal{G}$ such that $u_n \leq a_n < b_n \leq v_n$ and $|u_n - v_n| \leq K|a_n - b_n| \leq w_k \leq L_k$. As in the proof of Theorem 3.5, we find for $J_n(\tau_k) = [nu_n, nv_n]$ that

$$\begin{aligned}T_n^*\left(\mathbb{F}_{\text{const}}^{J_n(\tau_k)}\right) &= \inf_{a\in\mathbb{R}} \sup_{\substack{[un,vn]\subset J_n(\tau_k)\\(u,v)\in\mathcal{G}}} \frac{|S_n(v) - S_n(u) - [v-u]a|}{\rho_2(v-u)} \\ &\geq \frac{\frac{w_k\delta_k}{4K\sqrt{n}}}{\rho_2(w_k/n)} - |\widetilde{S}_n|_{\rho_2} \qquad \geq c\sqrt{z} - |\widetilde{S}_n|_{\rho_2},\end{aligned}$$

which tends to infinity as $z \to \infty$, while $|\widetilde{S}_n|_{\rho_2}$ remains stochastically bounded. Since $|J_n(\tau_k)| \leq w(\tau_k, z)$, this completes the proof.

References

Arias-Castro, E., Candès, E. J., and Durand, A. (2011). Detection of an anomalous cluster in a network. The Annals of Statistics, 39(1).

Billingsley, P. (1999). Convergence of Probability Measures. John Wiley & Sons.

Chan, H. P. and Walther, G. (2013). Detection with the scan and the average likelihood ratio. Statistica Sinica, pages 409–428. Publisher: JSTOR.

Chatterjee, S. and Lafferty, J. (2019). Adaptive risk bounds in unimodal regression. Bernoulli, 25(1).

Cho, H. and Kirch, C. (2021). Data segmentation algorithms: Univariate mean change and beyond. Econometrics and Statistics, page S2452306221001234.

Cui, Y., Levine, M., and Zhou, Z. (2021). Estimation and inference of time-varying auto-covariance under complex trend: A difference-based approach. Electronic Journal of Statistics, 15(2).

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. Annals of Statistics, 25(1):1–37.

Dahlhaus, R., Richter, S., and Wu, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. Bernoulli, 25(2):1013–1044.

Dette, H., Eckle, T., and Vetter, M. (2020). Multiscale change point detection for dependent data. Scandinavian Journal of Statistics, 47(4):1243–1274.

Dette, H., Wu, W., and Zhou, Z. (2019). Change Point Analysis of Correlation in Non-stationary Time Series. Statistica Sinica, 29:611–643. arXiv: 1801.10478.

Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale Testing of Qualitative Hypotheses. The Annals of Statistics, 29(1).

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. The Annals of Statistics, 36(4).

Feng, O. Y., Chen, Y., Han, Q., Carroll, R. J., and Samworth, R. J. (2022). Nonparametric, tuning-free estimation of S-shaped functions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 84(4):1324–1352.

Frick, K., Munk, A., and Sieling, H. (2014). Multiscale change point inference. Journal of the Royal Statistical Society: Series B: Statistical Methodology, pages 495–580. Publisher: JSTOR.

Fryzlewicz, P. (2024a). Narrowest Significance Pursuit: inference for multiple change-points in linear models. Journal of American Statistical Association, 119(546):1633–1646. arXiv: 2009.05431.

Fryzlewicz, P. (2024b). Robust Narrowest Significance Pursuit: inference for multiple change-points in the median. Journal of Business & Economic Statistics, 42(4):1389–1402. arXiv:2109.02487 [stat].

Gavioli-Akilagun, S. and Fryzlewicz, P. (2025). Fast and optimal inference for change points in piecewise polynomials via differencing. Electronic Journal of Statistics, 19(1).

Glaz, J., Pozdnyakov, V., and Wallenstein, S., editors (2009). Scan Statistics: Methods and Applications. Birkhäuser Boston, Boston, MA.

Hamadouche, D. (1998). Weak convergence of smoothed empirical process in Hölder spaces. Statistics & Probability Letters, 36(4):393–400.

Hamadouche, D. (2000). Invariance principles in Holder spaces. Portugaliae Mathematica, 57(2):127–152. Publisher: Lisboa, Gazeta de matematica [etc.].

Khismatullina, M. and Vogt, M. (2020). Multiscale Inference and Long-Run Variance Estimation in Non-Parametric Regression with Time Series Errors. Journal of the Royal Statistical Society Series B: Statistical Methodology, 82(1):5–37.

König, C., Munk, A., and Werner, F. (2020). Multidimensional multiscale scanning in exponential families: Limit theory and statistical consequences. The Annals of Statistics, 48(2).

Lamperti, J. (1962). On convergence of stochastic processes. Transactions of the American Mathematical Society, 104(3):430–435.

Lifshits, M. A. (1984). Absolute continuity of functionals of "supremum" type for Gaussian processes. Journal of Soviet Mathematics, 27(5):3103–3112.

Liu, W., Xiao, H., and Wu, W. B. (2013). Probability and moment inequalities under dependence. Statistica Sinica, 23:1257–1272.

Mies, F. (2023). Functional estimation and change detection for nonstationary time series. Journal of the American Statistical Association, 118(542):1011–1022.

Mies, F. (2024). Strong Gaussian approximations with random multipliers. arXiv:2412.14346 [math].

Mies, F. and Steland, A. (2023). Sequential Gaussian approximation for nonstationary time series in high dimensions. Bernoulli, 29(4):3114–3140.

Niu, Y. S. and Zhang, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. The Annals of Applied Statistics, 6(3).

Pein, F., Sieling, H., and Munk, A. (2017). Heterogeneous change point inference. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 79(4):1207–1227. Publisher: [Royal Statistical Society, Wiley].

Pilliat, E., Carpentier, A., and Verzelen, N. (2023). Optimal multiple change-point detection for high-dimensional data. Electronic Journal of Statistics, 17(1):1240–1315. arXiv:2011.07818 [math, stat].

Pinelis, I. (1994). Optimum Bounds for the Distributions of Martingales in Banach Spaces. The Annals of Probability, 22(4):1679–1706.

Računauskas, A. and Suquet, C. (2004a). Necessary and Sufficient Condition for the Functional Central Limit Theorem in Hölder Spaces. Journal of Theoretical Probability, 17(1):221–243.

Računauskas, A. and Suquet, C. (2004b). Necessary and sufficient condition for the Lamperti invariance principle. Theory of Probability and Mathematical Statistics, 68:127–137.

Računauskas, A. and Suquet, C. (2009). Hölderian invariance principle for Hilbertian linear processes. ESAIM: Probability and Statistics, 13:261–275.

Računauskas, A. and Wendler, M. (2020). Convergence of U-processes in Hölder spaces with application to robust detection of a changed segment. Statistical Papers, 61(4):1409–1435.

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q. (2007). A Review and Comparison of Changepoint Detection Techniques for Climate Data. Journal of Applied Meteorology and Climatology, 46(6):900–915.

Rivera, C. and Walther, G. (2013). Optimal detection of a jump in the intensity of a Poisson process or in a density with likelihood ratio statistics. Scandinavian Journal of Statistics, 40(4):752–769.

Rohde, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. The Annals of Statistics, 36(3).

Schilling, R. L. (2021). Brownian Motion: a guide to random processes and stochastic calculus. De Gruyter, Berlin ; Boston, 3rd edition edition.

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).

Sharpnack, J. and Arias-Castro, E. (2016). Exact asymptotics for the scan statistic and fast alternatives. Electronic Journal of Statistics, 10(2).

Verzelen, N., Fromont, M., Lerasle, M., and Reynaud-Bouret, P. (2023). Optimal Change-Point Detection and Localization. The Annals of Statistics, 51(4):1586–1610. arXiv: 2010.11470.

Vladimirova, M., Girard, S., Nguyen, H., and Arbel, J. (2020). Sub-Weibull distributions: Generalizing sub-Gaussian and sub-Exponential properties to heavier tailed distributions. Stat, 9(1):e318.

Vogt, M. and Dette, H. (2015). Detecting gradual changes in locally stationary processes. Annals of Statistics, 43(2):713–740.

Walther, G. and Perry, A. (2022). Calibrating the Scan Statistic: Finite Sample Performance Versus Asymptotics. Journal of the Royal Statistical Society Series B: Statistical Methodology, 84(5):1608–1639.

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.

Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. Statistica Sinica, 21(3):1397–1413.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. Journal of the American Statistical Association, 108(502):726–740.