Monitoring Relevant Shifts in Functional Time Series

Zhuo Lin* Jiajing Sun† Wolfgang Karl Härdle‡ Meiting Zhu§

Abstract: In functional time series, classical change-detection tests can reject for shifts that are statistically real yet practically negligible. We develop an adjusted-range self-normalization framework for testing *relevant* hypotheses in weakly dependent functional time series. Our proposed procedure is fully functional—it works directly with the functional objects and avoids Karhunen–Loève truncation. We establish large-sample validity, propose a plug-in Monte Carlo calibration for the classical (degenerate) case of no relevant change, and derive Pitman local power results explaining why adjusted-range self-normalization can improve power relative to quadratic self-normalization. Simulations across dependence structures confirm a favorable size–power trade-off. We illustrate the method with an empirical application to Bitcoin options implied-volatility smiles constructed from executed trade-level data, where the monitor identifies a small number of economically meaningful changes and a simple trading-style illustration delivers economically significant profits per straddle unit.

Keywords: functional time series; relevant hypotheses; self-normalization; change-point detection; covariance operator.

JEL Codes: C12, C14, C32, C15, G13.

*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, No. 55 Zhongguancun East Road, Haidian District, Beijing 100190, China; and School of Economics and Management, University of Chinese Academy of Sciences, Zhongguancun Nanyitiao, Haidian District, Beijing 100190, China.

†MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, and School of Economics and Management, University of Chinese Academy of Sciences, Zhongguancun Nanyitiao, Haidian District, Beijing 100190, China. Corresponding author. Email: jiajing.sun@gmail.com.

‡IDA Institute of Digital Assets, Bucharest University of Economic Studies, 6 Piata Romana, District 1, Bucharest 010374, Romania; Department of Information Management and Finance, National Yang Ming Chiao Tung University, Room 401, Management Building 1, 1001 University Road, Hsinchu 300093, Taiwan; School of Informatics, University of Edinburgh, Informatics Forum, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom; School of Business and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.

§School of Economics, Xiamen University, Economics Building, 422 South Siming Road, Siming District, Xiamen, Fujian 361005, China.

1 Introduction

In many fields we increasingly observe more of the process rather than a single summary number. Advances in sensing, storage, and data engineering mean that variables once recorded as one value are now available as within-period trajectories: a daily temperature curve rather than a daily average, an intraday volatility profile rather than a closing value, or a full yield curve observed day after day. Treating these objects as functional observations indexed by time leads naturally to dependent functional time series. A side effect of this richer sampling is that "detecting a difference" becomes easier: with enough data, classical "is there any change?" tests will often reject even when the shift is tiny.

This is why relevant questions and corresponding tests matter. Curves and surfaces move around all the time in real data: daily temperature profiles drift—sometimes unevenly across the year, effectively reshaping the seasons (Xu et al., 2025); river-flow trajectories shift with hydrological regimes (Dette et al., 2020); yield curves twist (Lengwiler and Lenz, 2010; Bardsley et al., 2017); intraday risk or volatility profiles evolve (Müller et al., 2011); and option-implied functional objects are standard inputs for asset pricing and risk measurement (Nadler and Sancetta, 2023). In many applications, the practical question is more selective: is the change large enough to matter for decisions, policy, or scientific interpretation? A small shift in seasonal temperature patterns may be statistically detectable but irrelevant for impacts; modest changes in river-flow curves might not alter flood-risk classification; and in finance, a risk manager does not want an alert every time an implied volatility (IV) surface wiggles, but only when the shift is large enough to affect hedging, margin, or risk limits in a meaningful way.

This paper develops an adjusted-range-based self-normalized framework for exactly that type of question in dependent functional time series. We study relevant hypotheses: rather than testing a point null such as $d = 0$, we test whether a suitable distance d is small enough to be treated as negligible or large enough to matter, relative to a user-chosen threshold $\Delta > 0$. We focus on two practically important objects: changes in the mean function and changes in the covariance operator.

Two methodological considerations drive the approach. First, we aim to stay fully functional. A convenient strategy in functional data analysis is to project each curve onto a finite-dimensional subspace (often via functional principal components) and then apply multivariate time series tools to the resulting scores. While effective in many settings, Karhunen–Loève (KL)-based dimension reduction introduces a truncation choice that can be consequential when departures from stability are subtle: retaining too few components may miss a diffuse change, whereas retaining too many can wash out the signal by admitting high-order noise (see, e.g., Aue et al. (2009), Berkes et al. (2009), and Hörmann and Kokoszka (2010)). To ensure that our relevance conclusions do not depend on an additional tuning decision, our procedures remain fully functional

and avoid KL truncation; for general background on dependent functional data and functional time series we refer to Bosq (2000) and Horváth and Kokoszka (2012).

Second, temporal dependence creates a nuisance standardization problem. For dependent curves, many limit laws involve a long-run covariance operator, so critical values typically require heteroskedasticity- and autocorrelation-consistent (HAC) estimation or resampling. Both are workable, but they introduce tuning (bandwidth/lag choices) and can lead to size distortions in small and moderate samples; see, for example, Müller (2007). Self-normalization (SN) provides an alternative by replacing explicit long-run variance (LRV) estimation with a data-driven normalizer that is stochastically proportional to the unknown dependence scale, leading to asymptotically pivotal limit laws; see Shao (2010) and Shao and Zhang (2010) and the review in Shao (2015). In a fully functional setting with relevant hypotheses, an SN construction is developed in Dette et al. (2020). A recurring practical issue, however, is that the classical quadratic self-normalizer can be conservative in power in CUSUM/KS-type problems. Recently, Hong et al. (2024) and Hong et al. (2025) propose an adjusted-range-based SN scheme designed to retain good size while improving power, and we adapt that adjusted-range idea to relevant-hypothesis testing with functional observations.

Two additional points are worth emphasizing. First, SN is not unique: even in the functional relevant-hypothesis framework, Dette et al. (2020, Remark 5) discuss alternative self-normalizing factors. Appendix B.1 formalizes this non-uniqueness via a generic SN result, and it also records the implied-relevance-bound (inversion) interpretation and a unified pivotal Pitman local-power formula for comparing alternative normalizers. Our focus is on the adjusted-range normalizer proposed recently by Hong et al. (2024), which replaces the quadratic aggregation of the partial-sum bridge by its adjusted range. While the substitution is simple at the level of the formula, it is consequential for inference. Quadratic SN—built on partial-sum quadratic variation as in Shao (2010) and used in Dette et al. (2020)—often delivers accurate size, but it can display the familiar "better size but less power" trade-off in change-detection settings, particularly for CUSUM/KS-type statistics (Shao and Zhang, 2010). The adjusted-range normalizer was introduced precisely to mitigate such power losses and has been shown to improve power in KS-type structural-break tests while retaining good size and the tuning-free appeal of SN (Hong et al., 2024). Beyond testing, Hong et al. (2025) report that adjusted-range-based SN can yield noticeably narrower confidence intervals, pointing to more precise inference.

Second, the present paper shows that these benefits carry over to relevant-hypothesis testing with functional observations. We provide a formal Pitman local-power analysis at the relevance boundary $d = \Delta$, obtaining a Brownian-functional limit experiment that makes the role of the normalizer transparent and enables a direct comparison to the quadratic SN benchmark. In addition, we establish asymptotic validity of a moving-block bootstrap for the adjusted-range self-normalized statistics, giving a theoretically justi-

fied finite-sample refinement to the Monte Carlo calibration based on the limiting Brownian functional. In this sense, the contribution is not only technical but also methodological: it clarifies how the choice of self-normalizer affects the power–precision trade-off in relevant functional inference.

We study settings in which economically meaningful regime shifts show up as changes in the shape of an object observed repeatedly over time. Our main empirical application is the Bitcoin (BTC) options market: from executed trade-level data on Deribit, we construct a daily sequence of IV smiles (across moneyness) and IV surfaces (across moneyness and maturity). IV functions are a natural object in this context because they are central to option pricing and risk management, and their empirical and theoretical properties have been studied extensively—ranging from early evidence on IV curves (Dumas et al., 1998), to analyses of surface dynamics (Cont and Fonseca, 2002), semiparametric modeling strategies (Fengler, 2005), and factor-structure approaches such as the dynamic semiparametric factor model (DSFM) (Borak et al., 2005; Park et al., 2009; Brüggemann et al., 2008). Applying our relevance-based procedures to these IV objects yields sharper and more interpretable statements about when the market transitions between regimes and whether the detected changes are large enough to be economically meaningful. More broadly, this is precisely the type of problem in which functional methods have become increasingly relevant in economics and finance, since many core features take the form of curves or surfaces, including yield curves (Lengwiler and Lenz, 2010; Bardsley et al., 2017) and intraday volatility profiles (Müller et al., 2011). To facilitate a like-for-like comparison, we also consider the benchmark functional datasets studied by Dette et al. (2020); additional details are deferred to Appendix D.

Overall, the paper contributes along four complementary dimensions. First, we propose an adjusted-range-based SN approach for testing relevant hypotheses in dependent functional time series, covering one-sample, two-sample, change-point and covariance-operator problems in a unified way and without KL truncation. Second, we establish large-sample validity under weak dependence in the non-degenerate relevant regime $\Delta > 0$, develop a plug-in Monte Carlo calibration that restores asymptotic level control in the classical/degenerate case $\Delta = 0$, and provide an optional block-bootstrap refinement for the adjusted-range pivotal statistic. Third, we derive Pitman local-alternative results that yield a Brownian-functional representation of local power and clarify why the adjusted-range normalizer can deliver power gains relative to quadratic SN. Fourth, simulations across multiple dependence structures and empirical illustrations (including the BTC option-implied IV application) show a favorable size–power trade-off and demonstrate how the method sharpens relevance thresholds in practice.

The paper is structured as follows. Section 2 introduces the adjusted-range-based SN method for testing relevant hypotheses in functional time series. Section 3 studies relevant changes in the mean function, and Section 4 extends the discussion to relevant changes in covariance operators. Sections 5 and 6 report

simulation evidence and empirical illustrations. Section 7 concludes and discusses directions for future work.

2 Relevant hypothesis and the adjusted-range-based self-normalization

We follow the general setup of Dette et al. (2020). The functional observations are curves defined on a one-dimensional compact domain. We take $T = [0, 1]$, and work in the Hilbert space $L^2(T)$ of square-integrable functions on T, equipped with the inner product and norm $\langle f, g\rangle \stackrel{\text{def}}{=} \int_0^1 f(t)g(t)\,dt$, $\|f\| \stackrel{\text{def}}{=} \langle f, f\rangle^{1/2}$. If the original curves are observed on an interval $[a, b]$, one may map $[a, b]$ to $[0, 1]$ by a linear rescaling of the argument; this only changes notation and not the theory. We write $t \in T$ for the curve argument and $\lambda \in [0, 1]$ for the subsample fraction used in the SN construction.

Let $\{X_n\}_{n\in\mathbb{Z}}$ be a strictly stationary $L^2(T)$-valued time series with mean function $\mu \stackrel{\text{def}}{=} \mathsf{E}[X_0] \in L^2(T)$. The relevant hypotheses studied in this paper compare quadratic distances involving μ (and, later, differences of mean functions or covariance operators) to a user-chosen relevance threshold $\Delta > 0$. The classical point-null case $\Delta = 0$ is degenerate and is treated separately (see Appendix A).

2.1 One-sample problems

We begin with the one-sample setting because it neatly captures the main idea of relevance. Classical tests typically ask whether a parameter is exactly zero (for example, whether a mean function vanishes), but with sufficiently rich functional measurements and large samples, such tests tend to reject even for changes that are too small to matter in practice. This is closely related to Berkson's observation that any consistent test will eventually detect arbitrarily small departures from the null when the sample size is large enough (Berkson, 1938). In many applications, what we actually want is a decision at a practically meaningful scale: is the effect small enough to be regarded as negligible, or large enough to warrant scientific or economic attention? This leads to the framework of relevant (or precise) hypotheses, which is also central in the equivalence and noninferiority testing literature (Berger and Delampady, 1987; Wellek, 2010).

Formally, we consider hypotheses of the form

$$H_0 : d \leqslant \Delta \quad \text{versus} \quad H_1 : d > \Delta, \tag{1}$$

where d is a scalar measure of effect size and $\Delta > 0$ is a user-chosen relevance threshold. Interpreting Δ as a tolerance level, acceptance of H_0 corresponds to no relevant difference: the effect may be nonzero,

but it is not large enough (in the chosen metric) to matter for decisions or interpretation. Choosing Δ is application-specific and should be stated in the same units as d, reflecting what domain experts consider a practically meaningful deviation.

We focus on testing hypotheses about the parameter $d = \int_T \mu^2(t)\,\mathrm{d}t$, specifically,

$$H_0 : \int_T \mu^2(t)\,\mathrm{d}t \le \Delta \quad \text{versus} \quad H_1 : \int_T \mu^2(t)\,\mathrm{d}t > \Delta. \tag{2}$$

Example 1 (Temperature or risk-profile curves)**.** *Suppose $Y_n(t)$ is a within-year temperature curve (or an intraday risk/volatility curve) and $g(t)$ is a benchmark curve (e.g., a climatological normal, a long-run mean profile, or a model-implied reference). Define the deviation curves $X_n(t) = Y_n(t) - g(t)$. Then the mean function $\mu(t) = \mathsf{E}[X_n(t)]$ describes the systematic average deviation from the benchmark, and $d = \int_T \mu^2(t)\,dt = \|\mu\|^2$ summarizes the overall magnitude of that deviation in an L^2 sense (equivalently, $\sqrt{d}$ is an RMS-type (root-mean-square–type) effect size over $t \in T$). If domain knowledge suggests that RMS deviations below, say, δ (e.g., $\delta = 0.1^\circ C$ in climate applications, or a small volatility tolerance in finance) are practically negligible, one can set $\Delta = \delta^2$ and test* (1) *to decide whether the systematic shift is relevant.*

Example 2 (Transaction-cost relevance)**.** *In trading and risk-management settings, it is natural to treat very small shifts as irrelevant, because round-trip transaction costs (bid–ask spreads, fees, and slippage) can dominate any potential profit and loss (P&L) from rebalancing. A convenient way to choose the relevance threshold is to translate transaction costs into "volatility units": for a representative position (or vega-weighted portfolio) with total vega V and round-trip cost C (both in option-price units), an IV move of size δ changes the position value by approximately $V\delta$, so the break-even move is $\delta_{\mathrm{tc}} \approx C/V$. Setting $\Delta = \delta_{\mathrm{tc}}^2$ in* (1) *filters out shifts too small to justify trading.*

We define the partial sums as follows:

$$S_n(t,\lambda) \overset{\mathrm{def}}{=} \frac{1}{n} \sum_{j=1}^{\lfloor n\lambda \rfloor} X_j(t), \quad \lambda \in [0,1],$$

where $\lfloor x \rfloor$ denotes the integer part of x.

Note that $S_n(\cdot, 1) = \bar{X}_n$ is the sample mean function. Under the standing weak-dependence assumptions stated below (Assumptions 1–4), a functional law of large numbers (LLN) holds, i.e.

$$\|S_n(\cdot,1) - \mu\| = \|\bar{X}_n - \mu\| = o_{\mathrm{p}}(1).$$

Consequently,

$$\widehat{\mathbb{I}}_n = \int_T S_n^2(t,1)\,dt = \|S_n(\cdot,1)\|^2 \to_{\mathrm{P}} \|\mu\|^2 = \int_T \mu^2(t)\,dt,$$

so $\widehat{\mathbb{I}}_n$ is a consistent estimator of $d = \int_T \mu^2(t)\,dt$. We reject H_0 for large values of the statistic:

$$\widehat{\mathbb{I}}_n = \int_T S_n^2(t,1)\,\mathrm{d}t.$$

Dette et al. (2020) show that under certain technical assumptions, the asymptotic distribution of $\widehat{\mathbb{I}}_n$, when suitably standardized, converges in distribution to a normal distribution:

$$\sqrt{n}\left\{\widehat{\mathbb{I}}_n - \int_T \mu^2(t)\,\mathrm{d}t\right\} \xrightarrow{\mathcal{L}} N(0,\tau^2),$$

where "$\xrightarrow{\mathcal{L}}$" denotes convergence in distribution/law, and τ^2 is the LRV given by

$$\tau^2 = 4\int_T\int_T \mu(s)\mu(t)C(s,t)\,\mathrm{d}s\,\mathrm{d}t, \tag{3}$$

where $C(s,t)$, the long-run covariance operator, is defined as

$$C(s,t) = \mathrm{Cov}\{X_0(s), X_0(t)\} + \sum_{l=1}^{\infty}\mathrm{Cov}\{X_0(s), X_l(t)\} + \sum_{l=1}^{\infty}\mathrm{Cov}\{X_0(s), X_{-l}(t)\}.$$

Direct estimation of the LRV τ^2 in (3) is delicate for dependent functional data because it requires estimating a long-run covariance operator (and hence introduces tuning choices such as bandwidths or truncation lags). We therefore use SN and replace τ by a data-driven normalizer constructed from the same partial-sum process. The basic idea is to build a scale estimate from the subsample path of the statistic, so that the unknown dependence scale cancels in the resulting ratio (see Dette et al. (2020) for the fully functional quadratic SN construction in the relevant-hypothesis setting).

To make the notation transparent, define the centered subsample functional

$$G_n(\lambda) \stackrel{\mathrm{def}}{=} \int_T S_n^2(t,\lambda)\,dt - \lambda^2\int_T S_n^2(t,1)\,dt, \qquad \lambda\in[0,1], \tag{4}$$

where $S_n(t,\lambda) = n^{-1}\sum_{j=1}^{\lfloor n\lambda\rfloor} X_j(t)$. A quadratic ($\mathrm{L}^2$-type) self-normalizer aggregates $\{G_n(\lambda)\}$ via a probability measure ν on $(0,1)$:

$$\widehat{\mathbb{V}}_n \stackrel{\mathrm{def}}{=} \left[\int_0^1 G_n^2(\lambda)\,\nu(d\lambda)\right]^{1/2}. \tag{5}$$

Under the same assumptions and whenever $\tau^2 > 0$, one obtains the pivotal limit

$$\frac{\widehat{\mathbb{I}}_n - d}{\widehat{\mathbb{V}}_n} \xrightarrow{\mathcal{L}} \mathbb{W} \stackrel{\text{def}}{=} \frac{\mathbb{B}(1)}{\left[\int_0^1 \lambda^2 \{\mathbb{B}(\lambda) - \lambda \mathbb{B}(1)\}^2 \, \nu(d\lambda)\right]^{1/2}},$$

where $\mathbb{B}$ is standard Brownian motion on $[0,1]$. The corresponding level-α rejection rule for (2) is

$$\widehat{\mathbb{I}}_n > \Delta + q_{1-\alpha}(\mathbb{W}) \, \widehat{\mathbb{V}}_n, \tag{6}$$

where $q_{1-\alpha}(\mathbb{W})$ denotes the $(1-\alpha)$-quantile of $\mathbb{W}$.

Remark 1 (Monte Carlo (and bootstrap) calibration of $q_{1-\alpha}(\mathbb{W})$)**.** *The limit distribution $\mathbb{W}$ is pivotal for a fixed choice of the weighting measure ν, but it is non-Gaussian. In applications, we therefore approximate $q_{1-\alpha}(\mathbb{W})$ by Monte Carlo simulation of Brownian motion (equivalently, Brownian bridge) and a numerical evaluation of the integral in the denominator of $\mathbb{W}$. In particular, if ν is discrete with support $\{\lambda_i\}_{i=1}^m$ and weights $\{w_i\}_{i=1}^m$, then*

$$\mathbb{W} = \frac{\mathbb{B}(1)}{\left[\sum_{i=1}^m w_i \lambda_i^2 \{\mathbb{B}(\lambda_i) - \lambda_i \mathbb{B}(1)\}^2\right]^{1/2}},$$

so one can simulate $\{\mathbb{B}(\lambda_i)\}$ on the grid via Gaussian increments and tabulate the resulting quantiles. Because the law of $\mathbb{W}$ does not depend on unknown features of the data-generating process, this calibration needs to be carried out only once for each choice of (α, ν). As an optional finite-sample refinement, one may instead approximate the distribution of $(\widehat{\mathbb{I}}_n - d)/\widehat{\mathbb{V}}_n$ by a block (or multiplier) bootstrap applied to the centered curves. See Appendix B.8 for a justification.

We use (6) as a benchmark in Section 5. The contrast between the quadratic normalizer (5) and the adjusted-range normalizer below lies in how the subsample path $\{G_n(\lambda)\}$ is converted into a scale estimate: (5) uses a quadratic (L^2) functional, whereas the adjusted-range normalizer uses the range. Quadratic self-normalizers are known to be sensitive to irregularities that inflate second moments (e.g., strong persistence, heteroskedasticity, near-unit-root behavior, or outliers), which may translate into conservative critical values and power loss. Range statistics provide a robust alternative scale measure in dependent and heavy-tailed environments (Mandelbrot and Wallis, 1969; Mandelbrot, 1972; Mandelbrot, 1975); in structural-break testing for scalar and multivariate time series, adjusted-range SN has been shown to improve the size–power trade-off and to avoid non-monotonic power behavior of quadratic SN (Hong et al., 2024).

In contrast, our adjusted-range normalizer uses the range of the same subsample path,

$$\widehat{\mathbb{H}}_n \stackrel{\text{def}}{=} \max_{\lambda\in[0,1]} G_n(\lambda) - \min_{\lambda\in[0,1]} G_n(\lambda), \tag{7}$$

which is the functional analogue of an adjusted-range SN.

More generally, the self-normalizer is not unique. Appendix B.1 (Theorem B.1) provides a generic SN principle: any continuous, positively homogeneous functional of the centered subsample path yields an asymptotically pivotal self-normalized ratio (hence covering both (5) and (7)). It also shows that the relevant decision rule can be interpreted as the inversion of a one-sided self-normalized confidence bound for d, which leads naturally to the implied relevance bound (the largest Δ rejected at level α). Finally, the same appendix gives a unified pivotal Pitman local-power expression—depending only on Brownian motion—which provides a direct way to compare candidate normalizers (quadratic, adjusted-range, and others).

Under Assumptions 1–4, it can be shown that

$$\Big(\sqrt{n}(\widehat{\mathbb{I}}_n - d),\, \sqrt{n}\widehat{\mathbb{H}}_n\Big) \;\xrightarrow{\mathcal{L}}\; \Big(\tau\mathbb{B}(1),\, \tau\{\max_{\lambda\in[0,1]} U(\lambda) - \min_{\lambda\in[0,1]} U(\lambda)\}\Big),$$

where $\mathbb{B}$ is standard Brownian motion on $[0,1]$ and $U(\lambda) \stackrel{\text{def}}{=} \lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)$. Consequently, whenever $\tau^2 > 0$, the ratio

$$\frac{\widehat{\mathbb{I}}_n - d}{\widehat{\mathbb{H}}_n} \;\xrightarrow{\mathcal{L}}\; \mathbb{S},$$

where

$$\mathbb{S} \stackrel{\text{def}}{=} \frac{\mathbb{B}(1)}{\max_{\lambda\in[0,1]} U(\lambda) - \min_{\lambda\in[0,1]} U(\lambda)}. \tag{8}$$

has a pivotal (nuisance-parameter-free) limit distribution. This yields the level-α rejection rule for (2):

$$\widehat{\mathbb{I}}_n > \Delta + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_n, \tag{9}$$

where $q^*_{1-\alpha}(\mathbb{S})$ denotes the $(1-\alpha)$-quantile of $\mathbb{S}$.

Although $\mathbb{S}$ is defined through $\max_{\lambda\in[0,1]}$ and $\min_{\lambda\in[0,1]}$, in practice we evaluate these extrema on an equally spaced grid $0 < \lambda_1 < \cdots < \lambda_M \le 1$ and approximate

$$\max_{\lambda\in[0,1]} U(\lambda) \approx \max_{1\le i\le M} U(\lambda_i), \qquad \min_{\lambda\in[0,1]} U(\lambda) \approx \min_{1\le i\le M} U(\lambda_i).$$

Given B Monte Carlo repetitions, we simulate i.i.d. standard normals $Z_1^{(b)}, \ldots, Z_M^{(b)}$ and construct Brownian

motion on the grid by

$$\mathbb{B}^{(b)}(\lambda_j) = \sum_{i=1}^{j} \sqrt{\lambda_i - \lambda_{i-1}}\, Z_i^{(b)}, \qquad \lambda_0 \stackrel{\text{def}}{=} 0,$$

set $U^{(b)}(\lambda_j) = \lambda_j \mathbb{B}^{(b)}(\lambda_j) - \lambda_j^2 \mathbb{B}^{(b)}(1)$, and compute

$$\mathbb{S}^{(b)} = \frac{\mathbb{B}^{(b)}(1)}{\max_{1\le j\le M} U^{(b)}(\lambda_j) - \min_{1\le j\le M} U^{(b)}(\lambda_j)}.$$

The empirical $(1-\alpha)$-quantile of $\{\mathbb{S}^{(b)}\}_{b=1}^{B}$ is then used as $q^*_{1-\alpha}(\mathbb{S})$ in (9).

Remark 2 (Bootstrap refinement for the adjusted-range SN critical value)**.** *As an optional finite-sample refinement to the Monte Carlo calibration of $q^*_{1-\alpha}(\mathbb{S})$, one may approximate the distribution of $(\widehat{\mathbb{I}}_n - d)/\widehat{\mathbb{H}}_n$ by a block (or multiplier) bootstrap applied to the centered curves. With $\{X_j^*\}_{j=1}^n$ generated as in Appendix B.8, form $\widehat{\mathbb{I}}_n^*$ and $G_n^*(\lambda)$ analogously, set*

$$\widehat{\mathbb{H}}_n^* = \max_{\lambda\in[0,1]} G_n^*(\lambda) - \min_{\lambda\in[0,1]} G_n^*(\lambda), \qquad T^*_{n,\mathrm{AR}} = \frac{\widehat{\mathbb{I}}_n^* - \widehat{\mathbb{I}}_n}{\widehat{\mathbb{H}}_n^*},$$

and let $q^{\mathrm{boot}}_{1-\alpha}(\mathbb{S})$ be the conditional $(1-\alpha)$-quantile of $T^*_{n,\mathrm{AR}}$. Then $q^{\mathrm{boot}}_{1-\alpha}(\mathbb{S})$ consistently estimates $q^*_{1-\alpha}(\mathbb{S})$; see Theorem B.3.

The critical values of $\mathbb{S}$ can be approximated via Monte Carlo simulation of Brownian motion. In all numerical work we approximate the supremum and infimum in (8) by evaluating $U(\lambda)$ on the grid

$$\lambda_i = \frac{i}{20}, \qquad i = 1, \dots, 19,$$

as in Dette et al. (2020). For this grid, Monte Carlo simulation (based on 10^6 replications) yields $q^*_{0.90}(\mathbb{S}) \approx 2.432$, $q^*_{0.95}(\mathbb{S}) \approx 3.269$, and $q^*_{0.99}(\mathbb{S}) \approx 5.159$.

We work under the following weak-dependence framework for the stationary functional time series $\{X_j\}_{j\in\mathbb{Z}}$. These conditions are standard in the functional time series literature and are used, for example, in Berkes et al. (2013) and Horváth et al. (2014) (see also Dette et al. (2020) for their use in a relevant-hypothesis SN setting).

Assumption 1. *For every $j \in \mathbb{Z}$, we have $X_j = \mu + \eta_j$, where $\mu \in L^2([0,1])$ is a (time-invariant) mean function and $(\eta_j)_{j\in\mathbb{Z}}$ is a centered error process satisfying Assumptions 2–4.*

Assumption 1 separates the deterministic signal μ from random fluctuations. The centering $\mathsf{E}[\eta_j] = 0$ ensures that μ is the population mean function, which is the object entering the relevant hypotheses through quantities such as $\|\mu\|^2 = \int_0^1 \mu^2(t)\, dt$.

Assumption 2. *The sequence $(\eta_j)_{j\in\mathbb{Z}}$ consists of Bernoulli shifts. That is, there exist a measurable space $(\mathcal{S},\mathcal{A})$ and a measurable map $f:\mathcal{S}^\infty\to L^2([0,1])$ such that*

$$\eta_j = f\left(\varepsilon_j,\varepsilon_{j-1},\ldots\right) \quad \text{for every } j\in\mathbb{Z},$$

where $(\varepsilon_j)_{j\in\mathbb{Z}}$ are i.i.d. $\mathcal{S}$-valued random elements. Here $\mathcal{S}^\infty$ denotes the space of one-sided sequences

$$\mathcal{S}^\infty \stackrel{\text{def}}{=} \{(s_0,s_1,s_2,\ldots): s_i\in\mathcal{S}\},$$

equipped with the product σ-algebra. Moreover, $\varepsilon_j(t)=\varepsilon_j(t,\omega)$ is assumed jointly measurable in (t,ω) for all $j\in\mathbb{Z}$.

Assumption 2 is a flexible "causal" representation: each curve η_j is generated from an i.i.d. innovation sequence through a (possibly nonlinear) measurable map f. This immediately implies strict stationarity and covers a wide range of short-memory functional models (including many functional linear and nonlinear processes).

Assumption 3. *The moment condition $\mathsf{E}\,\|\eta_j\|^{2+\psi}<\infty$ holds for some $\psi\in(0,1)$.*

Assumption 3 rules out excessively heavy tails and provides a little more than second moments. This extra moment is a standard technical ingredient for establishing tightness and functional central limit theorems for partial-sum processes in $L^2([0,1])$.

Assumption 4. *The sequence $(\eta_j)_{j\in\mathbb{Z}}$ may be approximated by l-dependent sequences $(\eta_{j,l})_{j\in\mathbb{Z}}$ such that for some $\kappa>2+\psi$,*

$$\sum_{l=1}^{\infty}\left(\mathsf{E}\,\|\eta_0-\eta_{0,l}\|^{2+\psi}\right)^{1/\kappa}<\infty,$$

where $\eta_{j,l}$ is defined by

$$\eta_{j,l} = f\left(\varepsilon_j,\varepsilon_{j-1},\ldots,\varepsilon_{j-l+1},\varepsilon^*_{j,l}\right), \tag{10}$$

$$\varepsilon^*_{j,l} = \left(\varepsilon^*_{j,l,j-l},\varepsilon^*_{j,l,j-l-1},\ldots\right), \tag{11}$$

and the random variables $\varepsilon^*_{j,l,k}$ are i.i.d. copies of ε_0, independent of $(\varepsilon_j)_{j\in\mathbb{Z}}$.

Assumption 4 formalizes weak dependence via an l-dependent approximation. The truncated process $\eta_{j,l}$ depends only on the most recent l innovations $(\varepsilon_j,\ldots,\varepsilon_{j-l+1})$; the remote tail of the innovation sequence is replaced by an independent copy $\varepsilon^*_{j,l}$. The summability condition requires that the approximation error

$\|\eta_0 - \eta_{0,l}\|$ vanishes sufficiently fast as $l \to \infty$, meaning that the influence of very old innovations becomes negligible. This is one of the key conditions in our framework, ensuring that the l-dependent approximation is accurate enough to apply weak invariance principle (WIP) arguments and thereby obtain the self-normalized limit laws used in the paper.

Theorem 1. *Assume that $\Delta > 0$. Under Assumptions 1-4, the decision rule* (9) *satisfies*

$$\lim_{n\to\infty} \mathrm{P}\left\{\widehat{\mathbb{I}}_n > \Delta + q^*_{1-\alpha}(\mathbb{S})\widehat{\mathbb{H}}_n\right\} = \begin{cases} 0 & \text{if } \int_T \mu^2(t)\,\mathrm{d}t < \Delta, \\ \alpha & \text{if } \int_T \mu^2(t)\,\mathrm{d}t = \Delta \text{ and } \tau^2 > 0, \\ 1 & \text{if } \int_T \mu^2(t)\,\mathrm{d}t > \Delta. \end{cases}$$

The proof of Theorem 1 is relegated to Appendix B.2.

2.1.1 Pitman local alternatives and asymptotic power

Theorem 1 describes the dichotomous large-sample behavior of the test: the rejection probability converges to 0, α, or 1 depending on whether the signal $d = \int_T \mu^2(t)\,dt$ lies below, on, or above the threshold Δ. To make the size–power trade-off more transparent (and to connect directly to the power comparisons in Section 5), it is useful to study Pitman local alternatives drifting to the boundary at rate $n^{-1/2}$.

Theorem 2 (Pitman local alternatives for the one-sample relevant test)**.** *Assume the setting of Section 2.1 and suppose $\Delta > 0$ with $\tau^2 > 0$. Consider a sequence of local alternatives such that*

$$d_n \overset{\text{def}}{=} \int_T \mu_n^2(t)\,dt = \Delta + \frac{c}{\sqrt{n}}, \qquad c \in \mathbb{R} \text{ fixed.} \tag{12}$$

Then, under Assumptions 1–4, the rejection probability of the adjusted-range-based test (9) *satisfies*

$$\lim_{n\to\infty} \mathrm{P}\Big\{\widehat{\mathbb{I}}_n > \Delta + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_n\Big\} = \mathrm{P}\left\{\frac{\mathbb{B}(1) + c/\tau}{\mathcal{R}} > q^*_{1-\alpha}(\mathbb{S})\right\}, \tag{13}$$

where $\mathbb{B}$ is standard Brownian motion on $[0,1]$ and

$$\mathcal{R} \overset{\text{def}}{=} \max_{\lambda\in[0,1]}\{\lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)\} - \min_{\lambda\in[0,1]}\{\lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)\}.$$

In particular, the limit in (13) *equals α at $c = 0$ and is strictly increasing in c.*

The proof is given in Appendix B.3.

A convenient way to represent the Pitman drift in (12) is to fix a boundary mean function $\mu \in L^2(T)$

with $\|\mu\|^2 = \int_T \mu^2(t)\,dt = \Delta$ and perturb it locally by a (non-random) direction $h \in L^2(T)$ with $\|h\| < \infty$, namely

$$\mu_n(t) = \mu(t) + \frac{h(t)}{\sqrt{n}}.$$

Then, by expanding $\|\mu_n\|^2 = \langle \mu_n, \mu_n \rangle$,

$$d_n = \|\mu_n\|^2 = \|\mu\|^2 + \frac{2\langle \mu, h \rangle}{\sqrt{n}} + \frac{\|h\|^2}{n}, \qquad \text{so} \qquad d_n - \Delta = \frac{2\langle \mu, h \rangle}{\sqrt{n}} + \mathcal{O}(n^{-1}).$$

Thus, $c = 2\langle \mu, h \rangle$ is the signed local deviation beyond the relevance boundary along direction h (positive c corresponds to relevant alternatives, negative c to local nulls); only this scalar c enters the limiting rejection probability in (13). Equation (13) also provides a simple numerical route to the asymptotic local power curve: one can approximate the probability on the right-hand side of (13) by simulating Brownian motion and evaluating the corresponding Brownian functional. This avoids simulating any functional time series and clarifies, at the level of the limit experiment, why the adjusted-range scaling can yield steeper power curves (Section 5).

Figure 1 illustrates the asymptotic Pitman local power implied by (13). Both procedures attain the correct size at $\kappa = 0$ and exhibit power increasing monotonically toward one as the standardized local effect size $\kappa = c/\tau$ grows. Across the range shown, the adjusted-range normalization delivers uniformly higher local power—most visibly for moderate κ—which provides a theory-based explanation for the power gains observed in the finite-sample experiments. The plot pertains to the non-degenerate relevant regime ($\Delta > 0$, $\tau > 0$); the classical/degenerate case ($\Delta = 0$) requires the separate plug-in calibration discussed in Appendix A.

Remark 3 (Extensions beyond the one-sample problem)**.** *The local-power statement in Theorem 2 carries over essentially unchanged to the two-sample, change-point, and covariance-operator settings studied below, with the obvious substitution of the problem-specific scale parameter. In each case the proposed statistic admits a linear expansion around the boundary $d = \Delta$ and a joint functional central limit theorem of the form*

$$\big(\sqrt{n}(\widehat{d}_n - d), \sqrt{n}\widehat{H}_n\big) \Rightarrow (\tau\mathbb{B}(1), \tau\mathcal{R}),$$

with the same Brownian-motion range functional $\mathcal{R}$ as in Section 2.1.1 and a problem-specific scale parameter τ (e.g., τ_D in the two-sample problem, τ_{δ,θ_0} in the mean change-point problem, and τ_{D_C} in the covariance-operator problem). Consequently, under Pitman drifts $d_n = \Delta + c/\sqrt{n}$ one obtains the same limiting rejection probability as in (13) *with the appropriate replacement of τ.*

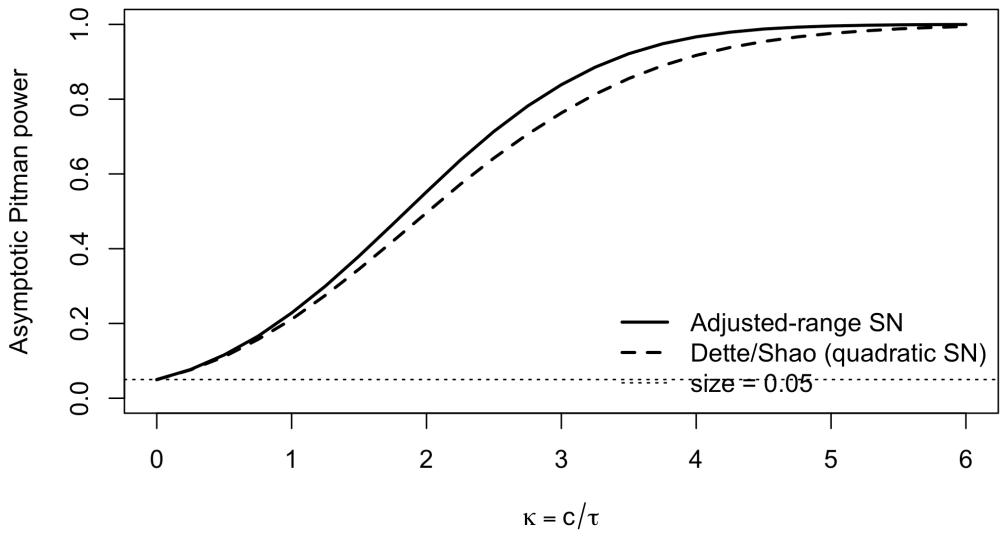


Figure 1: Asymptotic Pitman power curves for the relevant-hypothesis test under local alternatives $d_n = \Delta + c/\sqrt{n}$, plotted against $\kappa = c/\tau$ using the ν-grid $\{i/20 : i = 1, \ldots, 19\}$. Solid: adjusted-range normalization. Dashed: quadratic (Dette/Shao) SN. Dotted: $\alpha = 0.05$.

Finally, note that the Pitman drift in (12) is formulated in terms of the relevant distance $d_n = \|\mu_n\|^2$. When $\Delta > 0$, a drift of the form $d_n = \Delta + c/\sqrt{n}$ corresponds to an $L^2(T)$-perturbation of a boundary mean function of order $n^{-1/2}$ (see the expansion above). In the classical point-null case $\Delta = 0$, however, the same distance drift $d_n = c/\sqrt{n}$ translates into $\|\mu_n\| = c^{1/2} n^{-1/4}$, reflecting the quadratic nature of the L^2-energy functional. This is precisely the degenerate regime where the standard pivotal calibration breaks down and motivates the plug-in Monte Carlo calibration in Appendix A.

2.2 Two-sample problems

We now consider the two-sample setting, where the goal is to decide whether the average curves from two dependent functional time series differ by more than a practically meaningful amount. Typical examples include comparing (i) temperature or flow curves across two regions, (ii) yield curves across two market regimes, or (iii) option-implied objects (such as IV smiles) before and after a major event. As in the one-sample case, the threshold $\Delta > 0$ is chosen by the user to reflect what "large enough to matter" means in the application.

Let $\{X_j\}_{j\in\mathbb{Z}}$ and $\{Y_j\}_{j\in\mathbb{Z}}$ be two strictly stationary functional time series with values in $L^2(T)$. We observe finite stretches $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$. Write $\mu_1 = \mathsf{E}[X_1]$ and $\mu_2 = \mathsf{E}[Y_1]$ for the mean functions,

and let

$$D(t) \stackrel{\text{def}}{=} \mu_1(t) - \mu_2(t)$$

denote the mean difference. Our target parameter is the integrated squared distance $\int_T D^2(t)\,dt$, which measures the overall discrepancy between the two mean curves.

Following the one-sample construction, define the partial-sum difference process

$$D_{m,n}(t,\lambda) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j=1}^{\lfloor m\lambda\rfloor} X_j(t) - \frac{1}{n}\sum_{j=1}^{\lfloor n\lambda\rfloor} Y_j(t), \qquad \lambda \in [0,1].$$

Note that $D_{m,n}(t,1) = \bar{X}_m(t) - \bar{Y}_n(t)$, the usual difference of sample mean functions. Moreover,

$$\mathsf{E}\left[D_{m,n}(t,\lambda)\right] = \frac{\lfloor m\lambda\rfloor}{m}\mu_1(t) - \frac{\lfloor n\lambda\rfloor}{n}\mu_2(t) = \lambda D(t) + \mathscr{O}\big((m\wedge n)^{-1}\big),$$

where $a \wedge b$ denotes $\min(a,b)$. Thus, the λ–indexing makes $D_{m,n}(t,\lambda)$ behave like a "scaled" version of $D(t)$ plus fluctuations, which is exactly what the SN construction exploits.

We test the relevant hypotheses

$$H_0 : \int_T D^2(t)\,\mathrm{d}t \le \Delta \quad \text{versus} \quad H_1 : \int_T D^2(t)\,\mathrm{d}t > \Delta, \tag{14}$$

where Δ is a pre-specified relevance threshold.

Analogously to the one-sample case, define

$$\widehat{\mathbb{D}}_{m,n} \stackrel{\text{def}}{=} \int_T \{D_{m,n}(t,1)\}^2\,\mathrm{d}t$$

as the natural estimator of $\int_T D^2(t)\,dt$, and use the adjusted-range self-normalizer

$$\begin{aligned}\widehat{\mathbb{H}}_{m,n} = &\max_{\lambda\in[0,1]}\left\{\int_T \{D_{m,n}(t,\lambda)\}^2\,\mathrm{d}t - \lambda^2\int_T \{D_{m,n}(t,1)\}^2\,\mathrm{d}t\right\}\\ &- \min_{\lambda\in[0,1]}\left\{\int_T \{D_{m,n}(t,\lambda)\}^2\,\mathrm{d}t - \lambda^2\int_T \{D_{m,n}(t,1)\}^2\,\mathrm{d}t\right\}.\end{aligned} \tag{15}$$

We reject the null hypothesis of no relevant difference if

$$\widehat{\mathbb{D}}_{m,n} > \Delta + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_{m,n}, \tag{16}$$

where $q^*_{1-\alpha}(\mathbb{S})$ is the $(1-\alpha)$-quantile of the pivotal statistic $\mathbb{S}$ defined in (8).

In addition to the weak-dependence and moment conditions imposed in the one-sample setting for each series, we require the following two-sample conditions.

Assumption 5. *The sample sizes satisfy $m \to \infty$ and $n \to \infty$ with $\frac{m}{m+n} \to \rho$ for some $\rho \in (0,1)$.*

Assumption 5 rules out an extreme imbalance where one sample is negligible. The limit $\rho \in (0,1)$ ensures that both samples contribute to the asymptotic fluctuation scale, and it is the reason the LRV below carries weights $1/\rho$ and $1/(1-\rho)$.

Assumption 6. *The processes $\{X_j\}_{j\in\mathbb{Z}}$ and $\{Y_j\}_{j\in\mathbb{Z}}$ are independent, and each process satisfies the one-sample assumptions (Assumptions 1–4) with mean functions $\mathsf{E}[X_1] = \mu_1$ and $\mathsf{E}[Y_1] = \mu_2$.*

Assumption 6 is a clean way to separate "within-series" dependence from "between-series" dependence: each sample may be temporally dependent (weakly, in the same Bernoulli-shift / l-approximation sense as before), but the two samples do not interact. This independence makes the long-run covariance of the difference essentially additive across the two samples.

We measure the relevant fluctuation scale through

$$\tau_D^2 = 4 \int_T \int_T D(s)D(t) \left\{ \frac{1}{\rho} C_X(s,t) + \frac{1}{1-\rho} C_Y(s,t) \right\} \mathrm{d}s\, \mathrm{d}t, \tag{17}$$

where C_X and C_Y are the long-run covariance operators of $\{X_j\}$ and $\{Y_j\}$, defined as in the one-sample case (i.e. by summing lag covariances over $h \in \mathbb{Z}$).

Theorem 3. *Assume that $\Delta > 0$. Under Assumptions 5 and 6, the decision rule* (16) *satisfies*

$$\lim_{m,n\to\infty} \mathrm{P}\left\{ \widehat{\mathbb{D}}_{m,n} > \Delta + q^*_{1-\alpha}(\mathbb{S})\, \widehat{\mathbb{H}}_{m,n} \right\} = \begin{cases} 0 & \text{if } \int_T D^2(t)\,\mathrm{d}t < \Delta, \\ \alpha & \text{if } \int_T D^2(t)\,\mathrm{d}t = \Delta \text{ and } \tau_D^2 > 0 \text{ (see (17))}, \\ 1 & \text{if } \int_T D^2(t)\,\mathrm{d}t > \Delta. \end{cases}$$

The proof of Theorem 3 is given in Appendix B.4.

Similar to the one-sample case, the pivotal calibration underlying (16) breaks down when $\Delta = 0$, because the linearization behind the Brownian-motion limit requires $\tau_D^2 > 0$. A plug-in Monte Carlo calibration can be implemented exactly as in Appendix A, using an estimator of the long-run covariance operator $\frac{1}{\rho} C_X + \frac{1}{1-\rho} C_Y$ that appears in τ_D^2.

Finally, a Pitman local-alternative result analogous to Theorem 2 holds: if $\int_T D^2(t)\,dt = \Delta + c/\sqrt{m+n}$

for fixed c, then the rejection probability converges to

$$\mathrm{P}\Big\{(\mathbb{B}(1) + c/\tau_D)/\mathcal{R} > q^*_{1-\alpha}(\mathbb{S})\Big\},$$

with τ replaced by τ_D.

3 Relevant change in the mean function

A natural next step is the change-point problem, where the typical shape of a curve shifts at an unknown time. In many applications the mean function summarizes the baseline within-period pattern: for example, the average annual temperature curve (climate), the mean yearly hydrograph (hydrology), the average yield curve (macro-finance), or an average intraday volatility/risk profile (finance). Detecting such changes is a classical goal of change-point analysis; in the functional setting, mean-change tests are also a workhorse because a number of other structural questions can often be rewritten as a mean-change problem for a transformed sequence (e.g., changes in second-order structure correspond to mean changes of tensor-product objects); see, for instance, Carlstein et al. (1994) for general discussion and Berkes et al. (2009) and Aston and Kirch (2012) for functional change-point perspectives.

In this section we focus on relevant changes in the mean: rather than asking whether the mean changes at all, we ask whether the change is large enough to matter relative to a user-chosen threshold $\Delta > 0$. This is the relevant-hypothesis analogue of the usual mean-break problem and is especially important in modern high-resolution settings where even tiny shifts can become statistically detectable with enough data.

We consider data generated from the following triangular-array change-point model:

$$X_i = \begin{cases} \mu + f_1\left(\varepsilon_i, \varepsilon_{i-1}, \ldots\right) & \text{if } i \leqslant N\theta_0, \\ \mu + \delta + f_2\left(\varepsilon_i, \varepsilon_{i-1}, \ldots\right) & \text{if } i > N\theta_0, \end{cases} \tag{18}$$

where μ is the baseline mean function, $\delta \in L^2(T)$ is the mean shift after the break, and $\theta_0 \in (0,1)$ is the (unknown) break fraction. When θ_0 is known, the problem reduces to a two-sample comparison between the pre- and post-break segments, as discussed in Section 2.2. The filters f_1 and f_2 allow the short-run dynamics and/or the distribution of the innovations to change at the break as well, not only the mean. Throughout, we interpret the split point as $\lfloor N\theta_0 \rfloor$ when an integer index is needed.

As before, $(\varepsilon_j)_{j\in\mathbb{Z}}$ is an i.i.d. sequence of $\mathcal{S}$-valued random elements, and each $\varepsilon_j(t,\omega)$ is jointly measurable (as in Assumption 2). The processes $\big(f_1(\varepsilon_i, \varepsilon_{i-1}, \ldots)\big)_{i\in\mathbb{Z}}$ and $\big(f_2(\varepsilon_i, \varepsilon_{i-1}, \ldots)\big)_{i\in\mathbb{Z}}$ are assumed to satisfy

the same type of weak-dependence and moment conditions as in Assumptions 3–4 (applied separately to the pre- and post-break regimes).

We aim to test whether the mean shift is relevant in $L^2(T)$-magnitude:

$$H_0 : \int_T \delta^2(t)\,\mathrm{d}t \leq \Delta \quad \text{versus} \quad H_1 : \int_T \delta^2(t)\,\mathrm{d}t > \Delta, \tag{19}$$

where $\Delta > 0$ is a practitioner-chosen threshold encoding the smallest mean change that is considered practically meaningful. For example, in climate applications δ represents a shift in the seasonal temperature curve and Δ can be chosen to reflect an impact-relevant change in seasonal amplitude or timing; in hydrology, δ may correspond to a change in typical flow regimes that is large enough to affect flood classification; in macro/finance, δ can represent a regime shift in an average curve (e.g., yield-curve level/slope) or in an intraday profile that changes operational decisions.

The break fraction θ_0 is unknown. A convenient estimator is obtained by maximizing a contrast based on the integrated squared difference of the two segment means. Specifically, for a trimming parameter $\varepsilon \in (0, 1/2)$ define

$$\widehat{\theta} \stackrel{\text{def}}{=} \frac{1}{N}\ \operatorname{argmax}_{\lfloor N\varepsilon \rfloor + 1 \leqslant k \leqslant N - \lfloor N\varepsilon \rfloor} \widehat{f}(k) = \theta_0 + o_{\mathrm{p}}\left(N^{-1/2}\right), \tag{20}$$

where $\widehat{f}(0) = \widehat{f}(N) = 0$ and, for $k = 1, \ldots, N-1$,

$$\widehat{f}(k) \stackrel{\text{def}}{=} \frac{k}{N}\left(1 - \frac{k}{N}\right) \int_T \left\{ \frac{1}{k} \sum_{j=1}^{k} X_j(t) - \frac{1}{N-k} \sum_{j=k+1}^{N} X_j(t) \right\}^2 \mathrm{d}t.$$

The trimming excludes extremely short segments near the boundaries, where mean differences are too noisy to locate the break reliably. A proof of the rate in (20) can be found, for example, in Proposition 1 of Dette et al. (2020) under conditions matching our standing assumptions.

For any $\theta \in [1/N, 1)$ and $\lambda \in [0, 1]$, define the segment-wise partial-mean contrast

$$D_N^{\mathrm{cp}}(t, \lambda, \theta) \stackrel{\text{def}}{=} \frac{1}{\lfloor N\theta \rfloor} \sum_{j=1}^{\lfloor \lambda \lfloor N\theta \rfloor \rfloor} X_j(t) - \frac{1}{N - \lfloor N\theta \rfloor} \sum_{j=\lfloor N\theta \rfloor + 1}^{\lfloor N\theta \rfloor + \lfloor \lambda (N - \lfloor N\theta \rfloor) \rfloor} X_j(t).$$

The statistic $\widehat{\mathbb{D}}_N^{\mathrm{cp}}$ estimates the squared L^2-magnitude of the mean shift, while the adjusted-range normalizer $\widehat{\mathbb{H}}_N^{\mathrm{cp}}$ removes the unknown long-run dependence scale:

$$\widehat{\mathbb{D}}_N^{\mathrm{cp}} = \int_T D_N^{\mathrm{cp}}(t, 1, \widehat{\theta})^2 \mathrm{d}t,$$

$$\widehat{\mathbb{H}}_N^{\mathrm{cp}} = \max_{\lambda\in[0,1]} \left\{ \int_T D_N^{\mathrm{cp}}(t,\lambda,\widehat{\theta})^2 \mathrm{d}t - \lambda^2 \int_T D_N^{\mathrm{cp}}(t,1,\widehat{\theta})^2 \mathrm{d}t \right\} - \min_{\lambda\in[0,1]} \left\{ \int_T D_N^{\mathrm{cp}}(t,\lambda,\widehat{\theta})^2 \mathrm{d}t - \lambda^2 \int_T D_N^{\mathrm{cp}}(t,1,\widehat{\theta})^2 \mathrm{d}t \right\}. \tag{21}$$

The adjusted-range self-normalized test rejects H_0 in (19) whenever

$$\widehat{\mathbb{D}}_N^{\mathrm{cp}} > \Delta + q^*_{1-\alpha}(\mathbb{S})\widehat{\mathbb{H}}_N^{\mathrm{cp}}, \tag{22}$$

where $q^*_{1-\alpha}(\mathbb{S})$ is the $(1-\alpha)$-quantile of the same pivotal limit distribution used in the one- and two-sample settings.

To state the asymptotic behavior, define the long-run covariance kernels of the pre- and post-break error processes $\eta_i^{(j)} \stackrel{\mathrm{def}}{=} f_j(\varepsilon_i, \varepsilon_{i-1}, \ldots)$, $j = 1, 2$, by

$$K_j(s,t) = \sum_{h\in\mathbb{Z}} \mathrm{Cov}\left\{ \eta_0^{(j)}(s), \eta_h^{(j)}(t) \right\}, \quad j = 1, 2,$$

and the corresponding LRV parameter

$$\tau^2_{\delta,\theta_0} = 4 \int_T \int_T \delta(s)\delta(t) \left\{ \frac{1}{\theta_0} K_1(s,t) + \frac{1}{1-\theta_0} K_2(s,t) \right\} \mathrm{d}s\, \mathrm{d}t.$$

Theorem 4. *Assume that $\Delta > 0$. Under the data-generating process in* (18)*, with $\theta_0 \in (\varepsilon, 1-\varepsilon)$, and assuming the conditions stated below hold for model* (18)*, the decision rule* (22) *satisfies*

$$\lim_{N\to\infty} \mathrm{P}\left\{ \widehat{\mathbb{D}}_N^{\mathrm{cp}} > \Delta + q^*_{1-\alpha}(\mathbb{S})\widehat{\mathbb{H}}_N^{\mathrm{cp}} \right\} = \begin{cases} 0 & \text{if } \int_T \delta^2(t)\,\mathrm{d}t < \Delta, \\ \alpha & \text{if } \int_T \delta^2(t)\,\mathrm{d}t = \Delta \text{ and } \tau^2_{\delta,\theta_0} > 0, \\ 1 & \text{if } \int_T \delta^2(t)\,\mathrm{d}t > \Delta. \end{cases}$$

The proof of Theorem 4 is given in Appendix B.5.

Similarly to the one-sample and two-sample cases, the pivotal calibration underlying (22) breaks down when $\Delta = 0$. A plug-in Monte Carlo calibration analogous to Appendix A can be obtained by simulating the corresponding Gaussian limit process using an estimator of the relevant long-run covariance operator.

4 Relevant change in the covariance

Mean shifts are only part of the story. In many applications the more consequential changes show up in the second-order structure: how curves vary around their mean and how different parts of the curve co-move. In finance, this is essentially the "correlation risk": when the covariance structure changes, diversification

can suddenly stop working, hedges can misbehave, and risk forecasts based on yesterday's dependence can become misleading. In macroeconomics, changes in covariance can signal that sectors or countries have started to move together (or apart), which matters for forecasting and policy analysis. In environmental and climate settings, a covariance shift may indicate a regime change in variability rather than a change in average levels. And in all of these examples, it is rarely useful to react to every statistically detectable movement: with long samples (or high-frequency data) classical "is there any change?" tests tend to reject even for tiny shifts. Relevant hypotheses are designed to target the decision-relevant question: is the covariance change large enough to matter relative to a user-chosen threshold?

Technically, covariance-operator inference is more difficult than mean inference because it is quadratic in the data. In our functional setting the basic building block becomes the tensor-product process $(X_j - \mu) \otimes (X_j - \mu)$, which lives in $L^2(T^2)$ and has $\|(X_j - \mu) \otimes (X_j - \mu)\|_{L^2(T^2)} = \|X_j - \mu\|^2_{L^2(T)}$. As a result, one needs one extra layer of moment and weak-dependence control. We therefore keep Assumptions 1–2 unchanged and strengthen the moment and m-approximation conditions as follows (these are the covariance-operator analogues of Assumptions 3–4).

Assumption 7. $\mathsf{E}\,\|\eta_j\|^{4+\psi} < \infty$ *for some* $\psi \in (0, 1)$.

Assumption 8. *The sequence* $(\eta_j)_{j\in\mathbb{Z}}$ *can be approximated by* l*-dependent sequences* $(\eta_{j,l})_{j\in\mathbb{Z}}$ *in the sense that for some* $\kappa > 4 + \psi$,

$$\sum_{l=1}^{\infty} \Big(\mathsf{E}\,\|\eta_0 - \eta_{0,l}\|^{4+\psi} \Big)^{1/\kappa} < \infty, \tag{23}$$

where $\eta_{j,l}$ *is defined by*

$$\eta_{j,l} = f\big(\varepsilon_j, \varepsilon_{j-1}, \ldots, \varepsilon_{j-l+1}, \varepsilon^*_{j,l}\big),$$
$$\varepsilon^*_{j,l} = \big(\varepsilon^*_{j,l,j-l}, \varepsilon^*_{j,l,j-l-1}, \ldots\big),$$

and the random variables $\varepsilon^*_{j,l,k}$ *are i.i.d. copies of* ε_0*, independent of the innovation sequence* $(\varepsilon_j)_{j\in\mathbb{Z}}$.

Assumption 7 is a strengthened moment condition that is natural once we move from means to covariance objects: because the covariance operator is built from products of curves, finite fourth (plus a little) moments ensure that the relevant quadratic functionals are well behaved. Assumption 8 is a standard short-range dependence condition stated in an "m-approximability" form: $\eta_{j,l}$ depends only on the most recent l innovations, and (23) requires that this truncation approximates the full process sufficiently fast in $L^{4+\psi}$. Many common functional time series models (e.g. linear processes and functional ARMA-type constructions) satisfy conditions of this type.

4.1 Two-sample problem

Given two samples $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ from independent, strictly stationary functional time series $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$, we consider relevant hypotheses for differences in their (lag-0) covariance operators. Let $\mu_X(t) = \mathsf{E}[X_0(t)]$ and $\mu_Y(t) = \mathsf{E}[Y_0(t)]$, and define the covariance kernels

$$C^X(s,t) \stackrel{\text{def}}{=} \mathsf{E}\Big[\{X_0(s) - \mu_X(s)\}\{X_0(t) - \mu_X(t)\}\Big], \qquad C^Y(s,t) \text{ defined analogously.}$$

We quantify the discrepancy by the squared Hilbert–Schmidt distance

$$d_C \stackrel{\text{def}}{=} \int_T \int_T D_C^2(s,t)\,\mathrm{d}s\,\mathrm{d}t, \qquad D_C(s,t) \stackrel{\text{def}}{=} C^X(s,t) - C^Y(s,t),$$

and test the relevant hypotheses

$$H_0^C : d_C \leqslant \Delta \quad \text{versus} \quad H_1^C : d_C > \Delta, \tag{24}$$

where $\Delta > 0$ is a user-chosen relevance threshold.

To build a self-normalized statistic, we construct subsample covariance-difference kernels. For $\lambda \in [0,1]$, define the subsample means

$$\bar{X}_m(\lambda)(u) \stackrel{\text{def}}{=} \frac{1}{\lfloor m\lambda \rfloor \vee 1} \sum_{i=1}^{\lfloor m\lambda \rfloor} X_i(u), \qquad \bar{Y}_n(\lambda)(u) \stackrel{\text{def}}{=} \frac{1}{\lfloor n\lambda \rfloor \vee 1} \sum_{i=1}^{\lfloor n\lambda \rfloor} Y_i(u),$$

where $a \vee b$ denotes $\max(a,b)$; and the partial-sum covariance-difference process

$$\begin{aligned} D_{m,n}(s,t,\lambda) \stackrel{\text{def}}{=} & \frac{1}{m-1} \sum_{j=1}^{\lfloor m\lambda \rfloor} \Big(X_j(s) - \bar{X}_m(\lambda)(s)\Big)\Big(X_j(t) - \bar{X}_m(\lambda)(t)\Big) \\ & - \frac{1}{n-1} \sum_{j=1}^{\lfloor n\lambda \rfloor} \Big(Y_j(s) - \bar{Y}_n(\lambda)(s)\Big)\Big(Y_j(t) - \bar{Y}_n(\lambda)(t)\Big). \end{aligned}$$

The scaling by $m-1$ and $n-1$ is chosen so that $\mathsf{E}[D_{m,n}(\cdot,\cdot,\lambda)]$ is approximately linear in λ, matching the SN logic used in the mean problems.

We reject H_0^C in (24) if

$$\widehat{\mathbb{D}}_{m,n}^C > \Delta + q_{1-\alpha}^*(\mathbb{S})\,\widehat{\mathbb{H}}_{m,n}^C, \tag{25}$$

where $q^*_{1-\alpha}(\mathbb{S})$ is the $(1-\alpha)$-quantile of the pivotal statistic $\mathbb{S}$ defined in (8),

$$\widehat{\mathbb{D}}^C_{m,n} = \int_T \int_T D^2_{m,n}(s,t,1)\,\mathrm{d}s\,\mathrm{d}t,$$

and the adjusted-range self-normalizer is

$$\begin{aligned}\widehat{\mathbb{H}}^C_{m,n} = &\max_{\lambda\in[0,1]} \left\{ \int_T \int_T D^2_{m,n}(s,t,\lambda)\,\mathrm{d}s\,\mathrm{d}t - \lambda^2 \int_T \int_T D^2_{m,n}(s,t,1)\,\mathrm{d}s\,\mathrm{d}t \right\} \\ &- \min_{\lambda\in[0,1]} \left\{ \int_T \int_T D^2_{m,n}(s,t,\lambda)\,\mathrm{d}s\,\mathrm{d}t - \lambda^2 \int_T \int_T D^2_{m,n}(s,t,1)\,\mathrm{d}s\,\mathrm{d}t \right\}. \end{aligned} \tag{26}$$

For the asymptotic variance, let ρ be as in Assumption 5 and define the long-run covariance kernels of the tensor-product processes $\{(X_t-\mu_X)\otimes(X_t-\mu_X)\}$ and $\{(Y_t-\mu_Y)\otimes(Y_t-\mu_Y)\}$ by, for $z=(s,t)$ and $z'=(s',t')$,

$$C_X\{z,z'\} \stackrel{\text{def}}{=} \sum_{h\in\mathbb{Z}} \mathrm{Cov}\Big(\{X_0(s)-\mu_X(s)\}\{X_0(t)-\mu_X(t)\}, \{X_h(s')-\mu_X(s')\}\{X_h(t')-\mu_X(t')\}\Big),$$

and $C_Y\{z,z'\}$ analogously. Then set

$$\tau^2_{D_C} = 4\int_T\int_T\int_T\int_T D_C(s,t)\,D_C(s',t') \left[\frac{1}{\rho} C_X\{(s,t),(s',t')\} + \frac{1}{1-\rho} C_Y\{(s,t),(s',t')\}\right] \mathrm{d}s\,\mathrm{d}t\,\mathrm{d}s'\,\mathrm{d}t'.$$

Theorem 5. *Assume that $\Delta > 0$ and that Assumption 5 holds. Suppose that $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ are independent, satisfy Assumptions 1–2 (with means μ_X, μ_Y and error processes η^X_j, η^Y_j), and that η^X_j and η^Y_j satisfy the strengthened Assumptions 7–8. Then the decision rule* (25) *satisfies, as $m,n\to\infty$,*

$$\lim_{m,n\to\infty} \mathrm{P}\left\{\widehat{\mathbb{D}}^C_{m,n} > \Delta + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}^C_{m,n}\right\} = \begin{cases} 0 & \text{if } d_C < \Delta, \\ \alpha & \text{if } d_C = \Delta \text{ and } \tau^2_{D_C} > 0, \\ 1 & \text{if } d_C > \Delta. \end{cases}$$

The proof of Theorem 5 is given in Appendix B.6.

As in the mean-function settings, the pivotal calibration underlying (25) breaks down when $\Delta = 0$. A plug-in Monte Carlo calibration analogous to Appendix A can be implemented by estimating the long-run covariance kernel of the tensor-product process and simulating the corresponding Gaussian limit process.

4.2 Change-point problem

We now consider a relevant change-point problem for the covariance structure of a single dependent functional time series. The data are assumed to follow the triangular-array model

$$X_i = \begin{cases} \mu + f_1\left(\varepsilon_i, \varepsilon_{i-1}, \ldots\right) & \text{if } i \leqslant N\theta_0, \\ \mu + f_2\left(\varepsilon_i, \varepsilon_{i-1}, \ldots\right) & \text{if } i > N\theta_0, \end{cases} \tag{27}$$

where $\mu \in L^2(T)$ is an unknown mean function, $\theta_0 \in (0,1)$ is an unknown change-point location, and the dependence structure (and hence second-order dynamics) may differ before and after the break through the filters f_1 and f_2. Define $\eta_i^{(k)} \stackrel{\text{def}}{=} f_k\left(\varepsilon_i, \varepsilon_{i-1}, \ldots\right)$ for $k = 1, 2$, and denote by C_1 and C_2 the lag-zero covariance kernels (operators) of the process before and after the break, respectively. We are interested in the relevant hypotheses

$$H_0^C : d_C^{\text{cp}} = \int_T \int_T \left\{C_1(s,t) - C_2(s,t)\right\}^2 \, \mathrm{d}s \, \mathrm{d}t \leqslant \Delta \quad \text{versus} \quad H_1^C : d_C^{\text{cp}} > \Delta, \tag{28}$$

where $\Delta > 0$ is a user-chosen relevance threshold. In applications (e.g., risk management or macro–finance), changes in covariance matter because they signal changes in co-movement and diversification benefits, but acting on such shifts is typically only warranted when the change is large enough to affect decisions in a meaningful way.

Remark 4 (Choosing Δ for covariance changes)**.** *The distance $d_C^{\text{cp}} = \int_T \int_T \{C_1(s,t) - C_2(s,t)\}^2 \, \mathrm{d}s \, \mathrm{d}t$ (the squared Hilbert–Schmidt distance) is measured in squared covariance units. A practical way to calibrate Δ is to translate it into changes in the variance of economically relevant linear functionals. For any weight function $g \in L^2(T)$ with $\|g\| = 1$, consider the scalar exposure $L_g(X) = \langle X, g \rangle$. Its variance under covariance kernel C is $\mathrm{Var}\{L_g(X)\} = \int_T \int_T C(s,t) g(s) g(t) \, \mathrm{d}s \, \mathrm{d}t$. Hence the pre/post variance difference satisfies*

$$\left|\mathrm{Var}_{\text{post}}\{L_g(X)\} - \mathrm{Var}_{\text{pre}}\{L_g(X)\}\right| = \left| \int_T \int_T \{C_2(s,t) - C_1(s,t)\} g(s) g(t) \, \mathrm{d}s \, \mathrm{d}t \right| \leq \sqrt{d_C^{\text{cp}}},$$

by the Cauchy–Schwarz inequality. Therefore, choosing $\Delta = \delta^2$ means that under H_0^C all unit-norm exposures have variance changes bounded by δ, while H_1^C flags covariance shifts large enough to alter risk in at least some directions. In practice, g can be taken to represent key portfolios or risk factors (e.g. level/slope/curvature loadings), and δ can be set by a risk budget (margin/VaR) or other decision-relevant tolerances.

We first estimate θ_0 using a covariance CUSUM-type criterion. For integers $1 \leq l \leq k \leq N$ define the

sample mean

$$\bar{X}_{l:k}(t) \stackrel{\text{def}}{=} \frac{1}{k-l+1}\sum_{i=l}^{k} X_i(t),$$

and the sample covariance kernel

$$\widehat{C}_{l:k}(s,t) \stackrel{\text{def}}{=} \frac{1}{k-l}\sum_{j=l}^{k}\left\{X_j(s)-\bar{X}_{l:k}(s)\right\}\left\{X_j(t)-\bar{X}_{l:k}(t)\right\}, \qquad (k-l\geq 1),$$

so that $\widehat{C}_{1:k}$ and $\widehat{C}_{k+1:N}$ correspond to the usual sample covariance kernels on the first k and last $N-k$ observations (whenever the denominators are well-defined).

Fix a trimming constant $\varepsilon \in \left(0, \frac{1}{2}\right)$ and define

$$\widehat{\theta}^{\text{cov}} \stackrel{\text{def}}{=} \frac{1}{N}\,\text{argmax}_{\lfloor N\varepsilon\rfloor+1\leqslant k\leqslant N-\lfloor N\varepsilon\rfloor}\,\widehat{f}^{\text{cov}}(k),$$

where $\widehat{f}^{\text{cov}}(0)=\widehat{f}^{\text{cov}}(N)=0$ and, for $k=1,\ldots,N-1$,

$$\widehat{f}^{\text{cov}}(k) \stackrel{\text{def}}{=} \frac{k}{N}\left(1-\frac{k}{N}\right)\int_T\int_T\left\{\widehat{C}_{1:k}(s,t)-\widehat{C}_{k+1:N}(s,t)\right\}^2 \mathrm{d}s\,\mathrm{d}t.$$

The trimming parameter ε is a tuning constant that prevents the maximization from using break candidates too close to the sample boundaries, where covariance estimates are unstable because they are based on very short subsamples. In our numerical work we set $\varepsilon = 0.1$, but any fixed choice in $(0, 1/2)$ is allowed by the asymptotic theory.

For an arbitrary $\theta \in [2/N,\, 1-1/N)$, set $k_\theta \stackrel{\text{def}}{=} \lfloor N\theta\rfloor$ and, for $\lambda \in [0,1]$, let

$$k_1(\lambda,\theta) \stackrel{\text{def}}{=} \left\lfloor \lambda k_\theta \right\rfloor \vee 1, \qquad k_2(\lambda,\theta) \stackrel{\text{def}}{=} \left\lfloor \lambda(N-k_\theta)\right\rfloor \vee 1.$$

Define the (scaled) partial-sample covariance difference process

$$\begin{aligned} D_N^{\text{cp,cov}}(s,t,\lambda,\theta) \stackrel{\text{def}}{=} &\frac{1}{k_\theta-1}\sum_{j=1}^{k_1(\lambda,\theta)}\left\{X_j(s)-\bar{X}_{1:k_1(\lambda,\theta)}(s)\right\}\left\{X_j(t)-\bar{X}_{1:k_1(\lambda,\theta)}(t)\right\} \\ &-\frac{1}{N-k_\theta-1}\sum_{j=k_\theta+1}^{k_\theta+k_2(\lambda,\theta)}\left\{X_j(s)-\bar{X}_{k_\theta+1:k_\theta+k_2(\lambda,\theta)}(s)\right\}\left\{X_j(t)-\bar{X}_{k_\theta+1:k_\theta+k_2(\lambda,\theta)}(t)\right\}. \end{aligned}$$

Using the plug-in $\widehat{\theta}^{\mathrm{cov}}$, define

$$\begin{aligned}
\widehat{\mathbb{D}}_N^{\mathrm{cp,cov}} &= \int_T \int_T D_N^{\mathrm{cp,cov}}(s,t,1,\widehat{\theta}^{\mathrm{cov}})^2 \,\mathrm{d}s\,\mathrm{d}t, \\
\widehat{\mathbb{H}}_N^{\mathrm{cp,cov}} &= \max_{\lambda\in[0,1]} \left\{ \int_T \int_T D_N^{\mathrm{cp,cov}}(s,t,\lambda,\widehat{\theta}^{\mathrm{cov}})^2 \,\mathrm{d}s\,\mathrm{d}t - \lambda^2 \int_T \int_T D_N^{\mathrm{cp,cov}}(s,t,1,\widehat{\theta}^{\mathrm{cov}})^2 \,\mathrm{d}s\,\mathrm{d}t \right\} \\
&\quad - \min_{\lambda\in[0,1]} \left\{ \int_T \int_T D_N^{\mathrm{cp,cov}}(s,t,\lambda,\widehat{\theta}^{\mathrm{cov}})^2 \,\mathrm{d}s\,\mathrm{d}t - \lambda^2 \int_T \int_T D_N^{\mathrm{cp,cov}}(s,t,1,\widehat{\theta}^{\mathrm{cov}})^2 \,\mathrm{d}s\,\mathrm{d}t \right\}. \qquad (29)
\end{aligned}$$

We reject H_0^C in (28) whenever

$$\widehat{\mathbb{D}}_N^{\mathrm{cp,cov}} > \Delta + q_{1-\alpha}^*(\mathbb{S})\, \widehat{\mathbb{H}}_N^{\mathrm{cp,cov}}, \qquad (30)$$

where $q_{1-\alpha}^*(\mathbb{S})$ is the $(1-\alpha)$-quantile of the pivotal limit variable $\mathbb{S}$ defined in (8). The use of this critical value is asymptotically correct because, under the conditions of the theorem below, the self-normalized statistic $(\widehat{\mathbb{D}}_N^{\mathrm{cp,cov}} - d_C^{\mathrm{cp}})/\widehat{\mathbb{H}}_N^{\mathrm{cp,cov}}$ converges in distribution to $\mathbb{S}$.

Let $\delta_C(s,t) \overset{\mathrm{def}}{=} C_2(s,t) - C_1(s,t)$ and, for $j = 1, 2$, let K_j denote the long-run covariance kernel of the tensor-product process $\{\eta_i^{(j)} \otimes \eta_i^{(j)}\}_{i\in\mathbb{Z}}$. Define

$$\tau^2_{\delta_C,\theta_0} = 4 \int_T \int_T \int_T \int_T \delta_C(s,t)\, \delta_C(s',t') \left[\frac{1}{\theta_0} K_1\big(\{(s,t),(s',t')\}\big) + \frac{1}{1-\theta_0} K_2\big(\{(s,t),(s',t')\}\big) \right] \mathrm{d}s\,\mathrm{d}t\,\mathrm{d}s'\,\mathrm{d}t'.$$

Theorem 6. *Assume that $\Delta > 0$ and that the data are generated according to model (27) with $\theta_0 \in (\varepsilon, 1-\varepsilon)$. Suppose that the innovation sequence $\{\varepsilon_j\}_{j\in\mathbb{Z}}$ and the filters f_1, f_2 satisfy Assumption 2 and that the corresponding error processes satisfy the strengthened moment and approximation conditions Assumptions 7 and 8. Then the decision rule (30) satisfies*

$$\lim_{N\to\infty} \mathrm{P}\left\{ \widehat{\mathbb{D}}_N^{\mathrm{cp,cov}} > \Delta + q_{1-\alpha}^*(\mathbb{S})\, \widehat{\mathbb{H}}_N^{\mathrm{cp,cov}} \right\} = \begin{cases} 0 & \text{if } d_C^{\mathrm{cp}} < \Delta, \\ \alpha & \text{if } d_C^{\mathrm{cp}} = \Delta \text{ and } \tau^2_{\delta_C,\theta_0} > 0, \\ 1 & \text{if } d_C^{\mathrm{cp}} > \Delta. \end{cases}$$

The proof of Theorem 6 is given in Appendix B.7.

Similarly to the mean-based problems, the pivotal calibration in (30) breaks down in the classical/degenerate case $\Delta = 0$, and a plug-in Monte Carlo calibration can be implemented by simulating the corresponding Gaussian limit process using an estimator of the relevant long-run covariance kernel of the tensor-product process.

5 Simulation studies

This section reports Monte Carlo evidence on the finite-sample behavior of the proposed adjusted-range-based tests for relevant hypotheses. We cover the one-sample, two-sample, change-point, and covariance-operator settings, and we compare our method to the quadratic SN benchmark of Dette et al. (2020). Unless stated otherwise, all rejection probabilities are based on 1,000 replications with nominal level $\alpha = 0.05$.

To facilitate a strict apples-to-apples comparison with Dette et al. (2020), all data-generating processes, tuning parameters, and discretization choices in this section coincide with theirs; the only change is to replace their quadratic SN normalizer by the adjusted-range SN normalizer proposed in this paper. The resulting power gains are in line with the Pitman local-power dominance of the adjusted-range pivot over the quadratic pivot established in Section 2.1.1 (see also Figure 1).

All quadratic SN normalizers in this paper involve an integral with respect to a probability measure ν on $(0,1)$, while the adjusted-range normalizers involve a supremum and infimum over $\lambda \in (0,1)$. Following Dette et al. (2020), we fix the discretization used to implement both self-normalizers. Specifically, we take

$$\nu \;=\; \frac{1}{19}\sum_{i=1}^{19} \delta_{i/20},$$

where δ_λ denotes the Dirac measure at $\lambda \in (0,1)$, and we evaluate all suprema/infima in the adjusted-range normalizers on the same grid

$$\lambda_i = \frac{i}{20}, \qquad i = 1, \ldots, 19.$$

As an illustration, for the one-sample mean problem the subsample functional $G_n(\lambda)$ defined in (4) yields the implemented adjusted-range and quadratic normalizers

$$\widehat{\mathbb{H}}_n = \max_{1\le i\le 19} G_n(\lambda_i) - \min_{1\le i\le 19} G_n(\lambda_i),$$
$$\widehat{\mathbb{V}}_n = \left\{ \frac{1}{19}\sum_{i=1}^{19} G_n^2(\lambda_i) \right\}^{1/2}.$$

All other normalizers (two-sample, change-point, and covariance-operator settings) are discretized in the same way, replacing $G_n(\lambda)$ by the corresponding subsample-path functional in each problem.

5.1 One-sample problems

We generate functional observations according to

$$X_j(t) = \mu(t) + \varepsilon_j(t), \qquad j = 1, \dots, n,$$

where the mean function is

$$\mu(t) = \sqrt{2\delta}\sin(2\pi t). \tag{31}$$

By construction, $\int_0^1 \mu^2(t)\,dt = \delta$, so in the one-sample relevant test (2) the boundary of the null hypothesis corresponds to $\delta = \Delta$. We set $\Delta = 0.02$ and study the rejection probability as δ varies around this boundary.

We consider the three error structures used in Dette et al. (2020) (see also Aue et al. (2015, Sections 6.3–6.4)): (i) independent errors, (ii) functional MA(1) errors, and (iii) independent Brownian bridges. For cases (i)–(ii), we work on the finite-dimensional spline space $\mathbb{H} = \operatorname{span}\{\phi_1, \dots, \phi_D\} \subset L^2([0,1])$ generated by a D-dimensional B-spline basis (with $D = 21$), and define i.i.d. innovations

$$\eta_j(t) = \sum_{k=1}^{D} N_{k,j}\phi_k(t), \qquad j = 1, \dots, n,$$

where $\{N_{k,j}\}$ are independent Gaussian random variables with

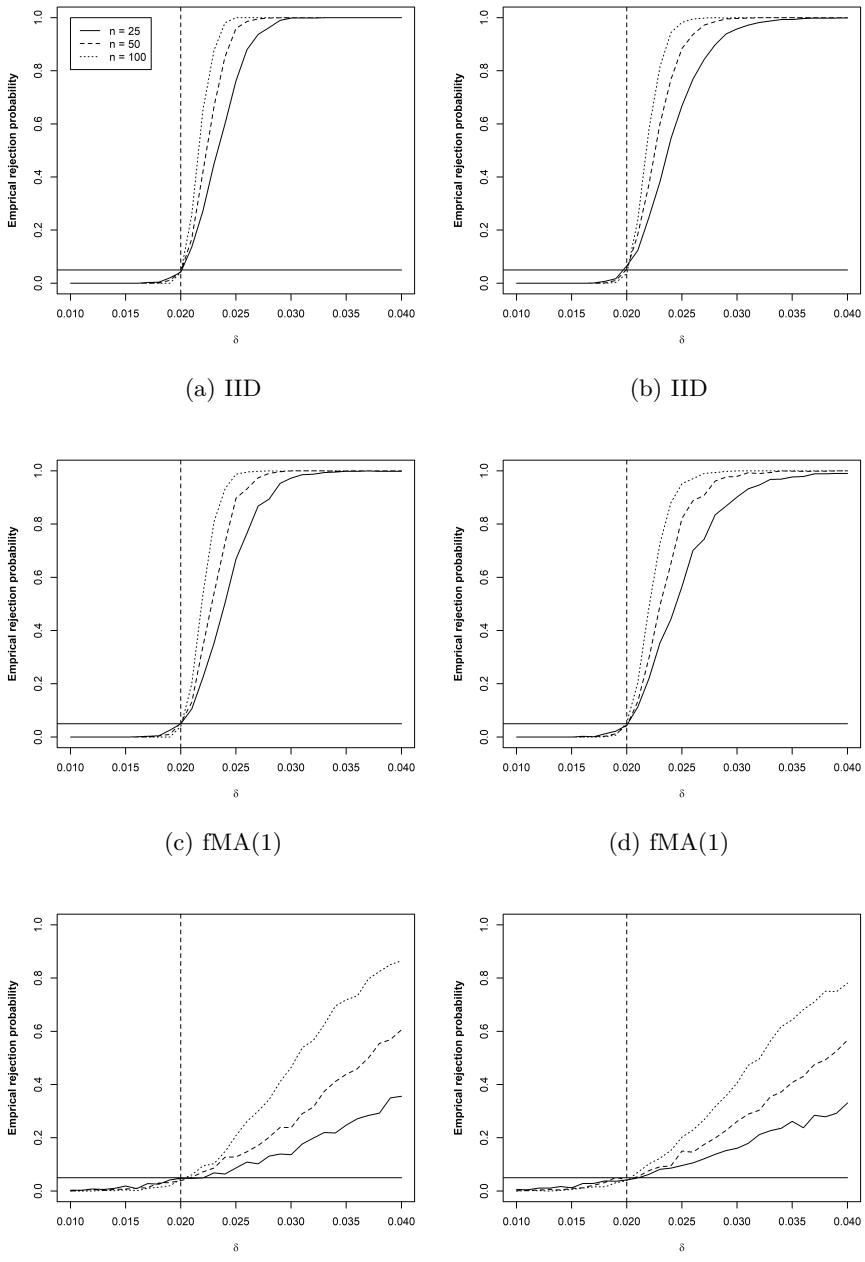
$$N_{k,j} \sim N(0, \sigma_k^2), \qquad \sigma_k^2 = \frac{1}{k^2}, \qquad k = 1, \dots, D.$$

In the IID case we set $\varepsilon_j = \eta_j$. In the functional MA(1) case we use

$$\varepsilon_j = \eta_j + \Theta\,\eta_{j-1}, \qquad j \in \mathbb{Z}, \tag{32}$$

where $\Theta : \mathbb{H} \to \mathbb{H}$ is a random linear operator represented by a random matrix $\Theta = (\Theta_{k\ell})_{k,\ell=1}^{D} \in \mathbb{R}^{D\times D}$ in the B-spline basis. Following Dette et al. (2020), the entries are Gaussian with mean 0 and standard deviation $\kappa\,\sigma_k\sigma_\ell$, and the scaling factor κ is chosen so that the induced spectral norm $\|\Theta\|_{\mathrm{op}}$ equals 0.7. The operator Θ is regenerated in every Monte Carlo replication. Finally, in the Brownian-bridge case we take ε_j to be independent Brownian bridges on $[0,1]$.

Figure 2 reports the simulated rejection probabilities of the adjusted-range test (9) (left column) and the quadratic SN benchmark (6) (right column). In all three error settings, both procedures are close to the nominal level at $\delta = \Delta$, while the adjusted-range normalizer delivers noticeably steeper power curves in small and moderate samples, especially under dependence.



(a) IID (b) IID

(c) fMA(1) (d) fMA(1)

(e) Brownian bridge (f) Brownian bridge

Figure 2: One-sample relevant test (2) with $\Delta = 0.02$ and mean function (31). Rows correspond to the three innovation models (IID, fMA(1), and Brownian-bridge errors). Left column: adjusted-range SN test (9). Right column: quadratic SN benchmark (6). In each panel, the solid/dashed/dotted curves correspond to $n = 25, 50, 100$, respectively, and the vertical dotted line marks the relevance boundary $\delta = \Delta$.

5.2 Two-sample problems

We consider two independent samples $\{X_1, \dots, X_m\}$ and $\{Y_1, \dots, Y_n\}$ with $\mathsf{E}[X_1] = \mu_1$ and $\mathsf{E}[Y_1] = \mu_2$, and we test the relevant two-sample hypothesis (14). We set $\mu_1(t) \equiv 0$ and

$$\mu_2(t) = a\, t(1-t),$$

so that the mean difference $D(t) = \mu_1(t) - \mu_2(t) = -a\, t(1-t)$ vanishes at the endpoints and satisfies

$$\int_0^1 D^2(t)\, dt = \frac{a^2}{30}.$$

We set $\Delta = (0.2^2)/30$, so that $a = 0.2$ lies on the boundary of the null, and we evaluate the rejection probability as a varies.

We use the same three error structures as in the one-sample experiments (IID, fMA(1), and Brownian-bridge errors), generated exactly as in Dette et al. (2020). In the fMA(1) case, the two samples are generated independently; in particular, they use independent innovation sequences and (independently generated) random operators Θ (with $\|\Theta\|_{\mathrm{op}} = 0.7$) in each Monte Carlo replication.

Figure 3 shows the rejection probabilities of the adjusted-range two-sample test (16) (left column) together with its quadratic SN analogue used as a benchmark (right column). Across the three dependence settings, the adjusted-range version achieves higher power while maintaining good size control.

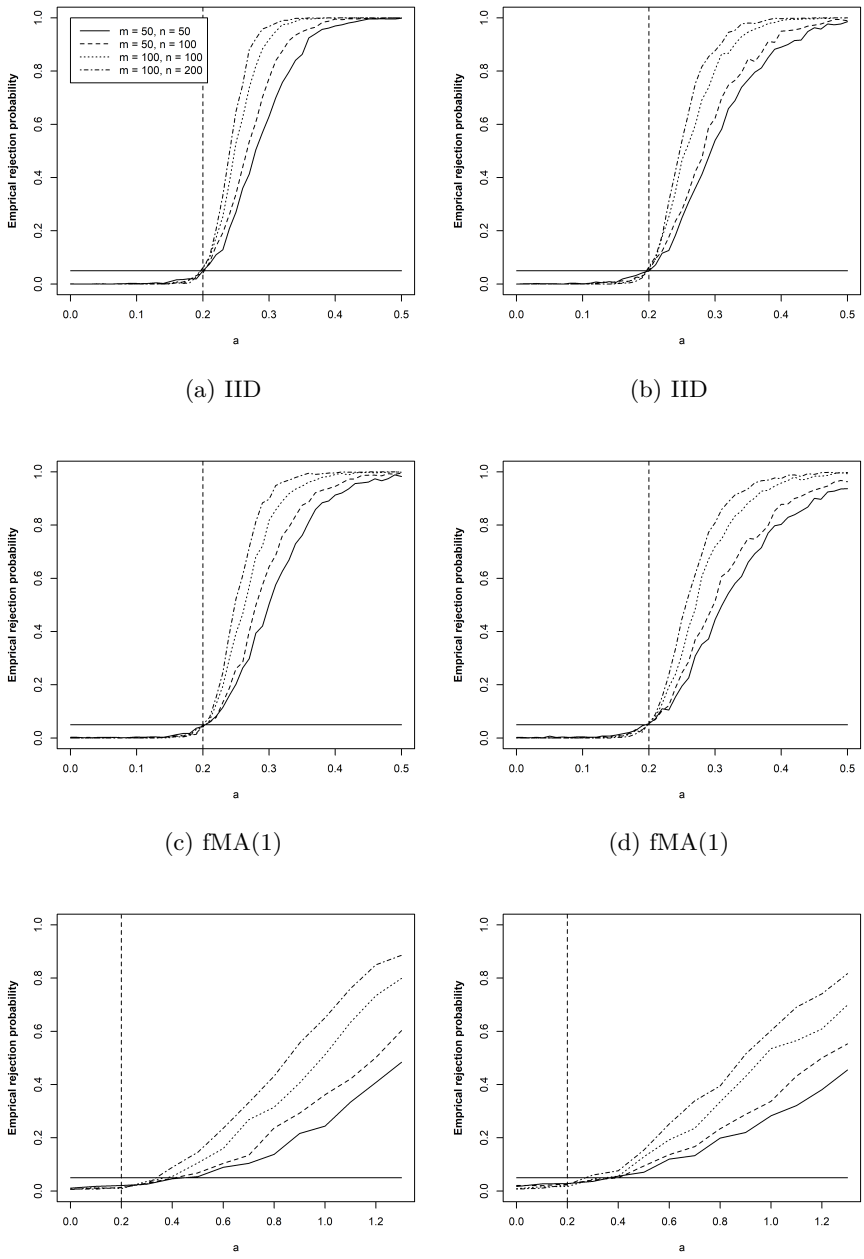
5.3 Change-point problems

We next study the mean change-point model (18) with $\theta_0 = 0.5$, $\mu \equiv 0$, and

$$\delta(t) = a\, t(1-t).$$

The error process is i.i.d. and generated as in the IID design of Section 5.1 (i.e. $\varepsilon_i = \eta_i$). As in the two-sample design, $\int_0^1 \delta^2(t)\, dt = a^2/30$, so a threshold of the form $\Delta = (a_0^2)/30$ corresponds to declaring changes with amplitude $a > a_0$ as practically relevant. We report results for $\Delta = (0.1^2)/30$, $\Delta = (0.2^2)/30$, and $\Delta = (0.3^2)/30$. Following Dette et al. (2020), data are generated on the grid $a = 0, 0.02, \dots, 0.50$ and rejection probabilities are computed as functions of a.

The trimming parameter in the change-point estimator (20) is set to $\varepsilon = 0.05$, so that at least 5% of the sample is used on each side of a candidate split; this avoids extremely short segments where mean and covariance estimates can be unstable.



(a) IID (b) IID

(c) fMA(1) (d) fMA(1)

(e) Brownian bridge (f) Brownian bridge

Figure 3: Two-sample relevant test (14) with $\Delta = (0.2^2)/30$ and $\mu_2(t) = a\,t(1-t)$ (so that $\int_0^1 D^2(t)\,dt = a^2/30$). Rows correspond to IID, fMA(1), and Brownian-bridge innovations. Left column: adjusted-range SN test (16). Right column: quadratic SN benchmark. Line types correspond to $(m, n) = (50, 50), (50, 100), (100, 100), (100, 200)$ (as in the inset legend), and the vertical dotted line marks the boundary $a = 0.2$ (i.e. $\int_0^1 D^2 = \Delta$).

Figure 4 reports rejection probabilities of the adjusted-range change-point test (22) (left column) and the quadratic SN benchmark (right column) as a varies. The adjusted-range version again yields higher power. The only noticeable exception occurs for the most stringent threshold $\Delta = (0.1^2)/30$, where both methods can be slightly oversized in small samples because estimating θ_0 becomes noisy when the signal is close to the boundary. Figure 5 illustrates this effect by showing the distribution of $\widehat{\theta}$ for different values of a.

We also consider dependent errors with a possible variance change at the break, using the same fMA(1) construction as in (32) (with $\|\Theta\|_{\mathrm{op}} = 0.7$ and Θ regenerated in each replication). Specifically, we study

$$X_i(t) = \delta(t)\mathbf{1}(i > N\theta_0) + \varepsilon_i(t), \tag{33}$$

$$X_i(t) = \delta(t)\mathbf{1}(i > N\theta_0) + \varepsilon_i(t)\mathbf{1}(i \leq N\theta_0) + \sqrt{3}\,\varepsilon_i(t)\mathbf{1}(i > N\theta_0), \tag{34}$$

where $\varepsilon_i(t)$ is generated from the fMA(1) process in (32). Model (33) isolates the impact of dependence (pure mean change), while (34) combines a mean shift with a higher-variance post-break regime. Following Dette et al. (2020), we report results for the threshold $\Delta = (0.3^2)/30$ (boundary $a = 0.3$) and $N \in \{200, 500\}$. Figure 6 shows that the adjusted-range procedure retains a favorable size–power trade-off in both scenarios.

5.4 Covariance operator problems

We now report simulations for relevant differences in covariance operators. Let $D = 21$ and let $\{b_\ell\}_{\ell=1}^{D}$ denote the orthonormal basis functions defined by $b_1(t) \equiv 1$ and

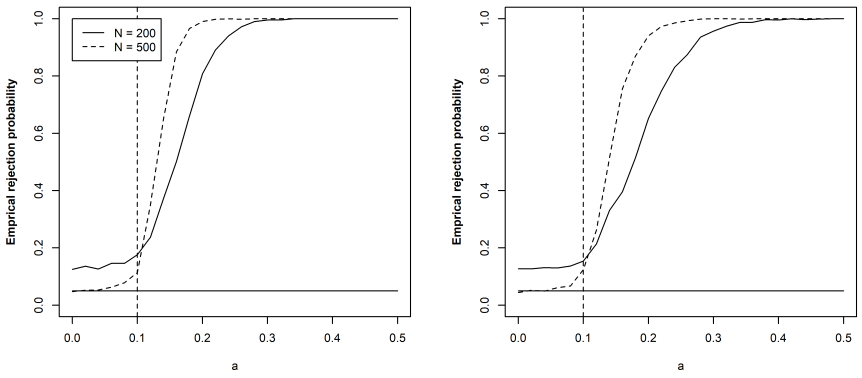
$$b_\ell(t) = \begin{cases} \sqrt{2}\,\sin(\ell\pi t), & \ell \text{ even}, \\ \sqrt{2}\,\cos((\ell-1)\pi t), & \ell > 1 \text{ odd}. \end{cases}$$

For each $\ell = 1, \ldots, D$, let $\{N_{\ell,j}\}$ and $\{N'_{\ell,i}\}$ be independent Gaussian sequences with

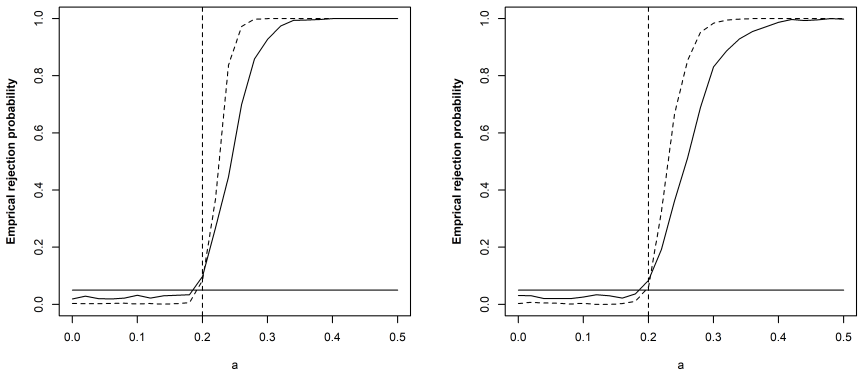
$$N_{\ell,j} \sim N(0, \sigma_\ell^2), \qquad N'_{\ell,i} \sim N(0, \sigma_\ell^2),$$

independent across ℓ and time indices. We consider the two variance scenarios (A and B) from Dette et al. (2020):

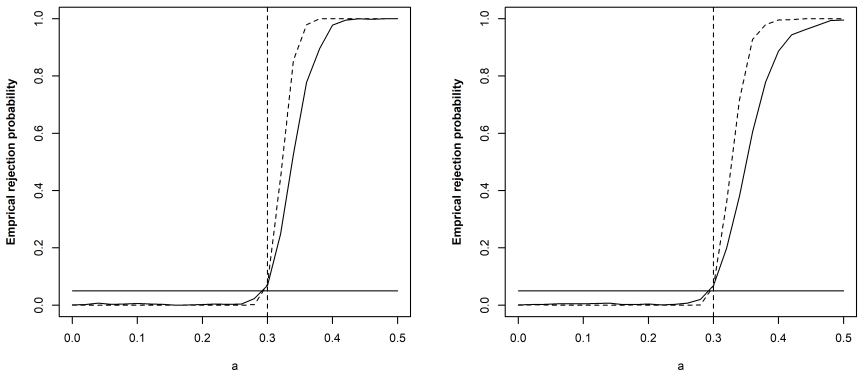
$$\text{Scenario A: } \sigma_\ell^2 = \frac{1}{\ell^2}, \qquad \text{Scenario B: } \sigma_\ell^2 = 1.2^{-2\ell}, \qquad \ell = 1, \ldots, D.$$



(a) $\Delta = (0.1^2)/30$ (b) $\Delta = (0.1^2)/30$

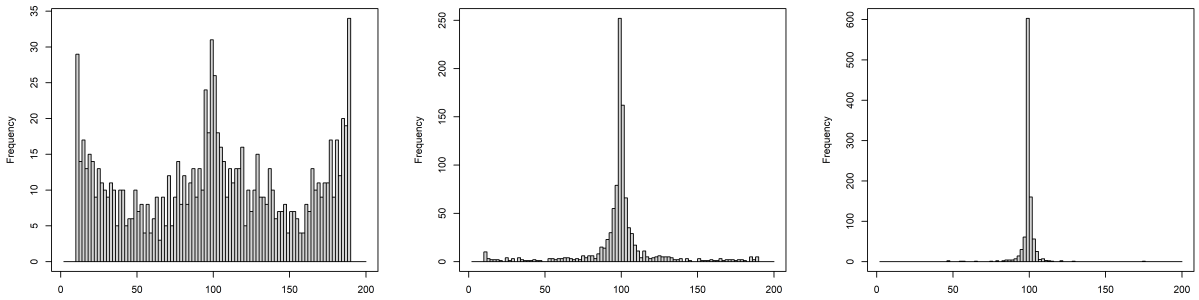


(c) $\Delta = (0.2^2)/30$ (d) $\Delta = (0.2^2)/30$



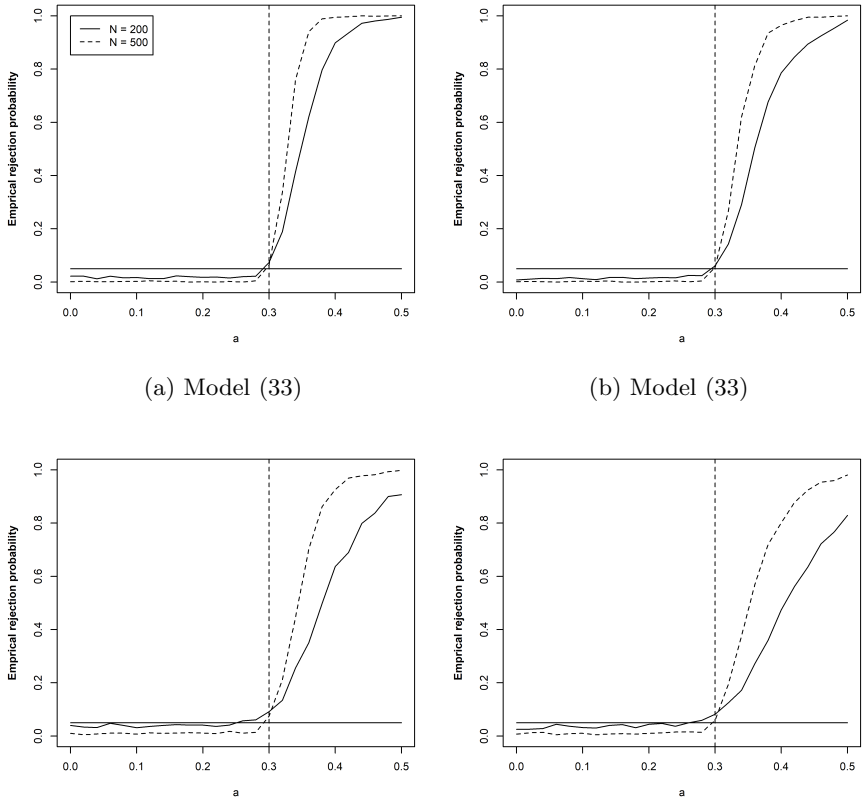
(e) $\Delta = (0.3^2)/30$ (f) $\Delta = (0.3^2)/30$

Figure 4: Mean change-point problem (19) with $\delta(t) = a\,t(1-t)$. Rows correspond to $\Delta = (0.1^2)/30$, $(0.2^2)/30$, and $(0.3^2)/30$ (so that the boundary of H_0 is at $a = 0.1, 0.2, 0.3$, respectively). Left column: adjusted-range SN test (22). Right column: quadratic SN benchmark. In each panel, solid/dashed curves correspond to $N = 200$ and $N = 500$, respectively; vertical dotted lines mark the relevance boundary ($a = 0.1, 0.2, 0.3$ by row).



(a) $a = 0.1$ (b) $a = 0.2$ (c) $a = 0.3$

Figure 5: Distribution of the change-point estimator (20) in the mean change-point model (18) with $\theta_0 = 0.5$ and $\delta(t) = a\,t(1-t)$. When a is close to the relevance boundary, $\widehat{\theta}$ is more dispersed, which can translate into mild size distortions for very small Δ in small samples.



(a) Model (33) (b) Model (33)

(c) Model (34) (d) Model (34)

Figure 6: Mean change-point problem with fMA(1) dependence and $\Delta = (0.3^2)/30$ (boundary $a = 0.3$). Rows correspond to the data-generating processes (33) (pure mean change) and (34) (mean change with a variance increase by a factor of 3 after the break). Left column: adjusted-range SN test (22). Right column: quadratic SN benchmark. Solid/dashed curves correspond to $N = 200$ and $N = 500$, respectively.

5.4.1 Two-sample problem

For brevity, we display only results for fMA(1) processes defined by

$$X_j = \eta_j + \kappa\,\eta_{j-1}, \qquad Y_i = \tilde{\eta}_i + \kappa\,\tilde{\eta}_{i-1}, \qquad j = 1,\dots,m, \quad i = 1,\dots,n,$$

with $\kappa = 0.7$, where the error processes are

$$\eta_j(t) = \sum_{\ell=1}^{D} N_{\ell,j}\, b_\ell(t), \qquad \tilde{\eta}_i(t) = a \sum_{\ell=1}^{D} N'_{\ell,i}\, b_\ell(t).$$

In this construction, Y_i is a multiple of X_j in distribution, and the (lag-0) covariance operators satisfy

$$C^X(s,t) = (1+\kappa^2) \sum_{\ell=1}^{D} \sigma_\ell^2\, b_\ell(s) b_\ell(t), \qquad C^Y(s,t) = a^2\, C^X(s,t).$$

Hence $D_C(s,t) = C^X(s,t) - C^Y(s,t) = (1-a^2)C^X(s,t)$ and the squared Hilbert–Schmidt distance equals

$$d_C = \int_0^1 \int_0^1 D_C^2(s,t)\, ds\, dt = (1-a^2)^2\,(1+\kappa^2)^2 \sum_{\ell=1}^{D} \sigma_\ell^4.$$

We set

$$\Delta = \left(1 - 1.5^2\right)^2 (1+0.7^2)^2 \sum_{\ell=1}^{D} \sigma_\ell^4,$$

so that $a = 1.5$ lies on the boundary of H_0^C in (24). Figure 7 shows rejection probabilities for the two variance scenarios (A and B) and illustrates that the adjusted-range procedure achieves a favorable power improvement while keeping size close to nominal.

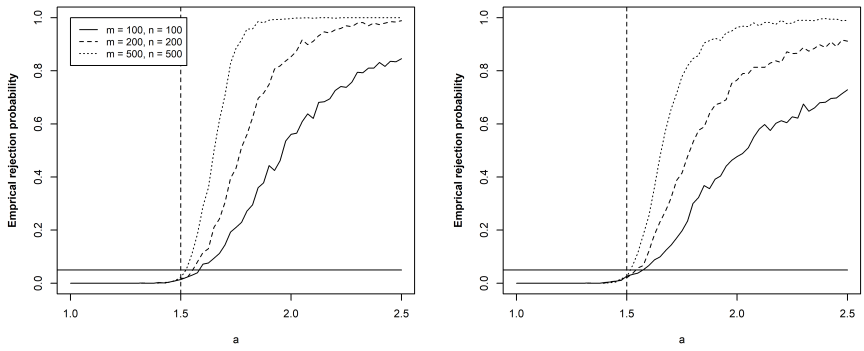
5.4.2 Change-point problem

Let $\{X'_j\}_{j=1}^N$ be an fMA(1) process generated as above:

$$X'_j = \eta_j + \kappa\,\eta_{j-1}, \qquad \kappa = 0.7, \qquad \eta_j(t) = \sum_{\ell=1}^{D} N_{\ell,j}\, b_\ell(t),$$

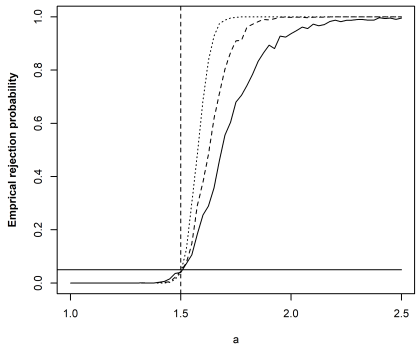
with the same basis $\{b_\ell\}$ and the same variance scenarios for $\{\sigma_\ell^2\}$. The observed sample $\{X_j\}_{j=1}^N$ is defined by

$$X_j = \begin{cases} X'_j, & j \le \lfloor N\theta_0 \rfloor, \\ a\,X'_j, & j > \lfloor N\theta_0 \rfloor, \end{cases} \qquad \theta_0 = 0.5,$$

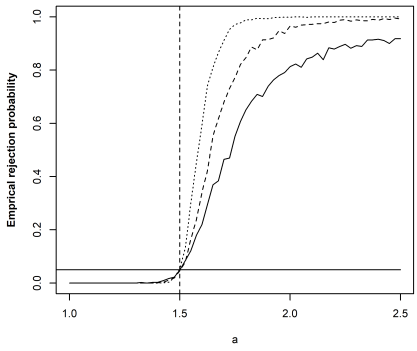


(a) Scenario A

(b) Scenario A



(c) Scenario B



(d) Scenario B

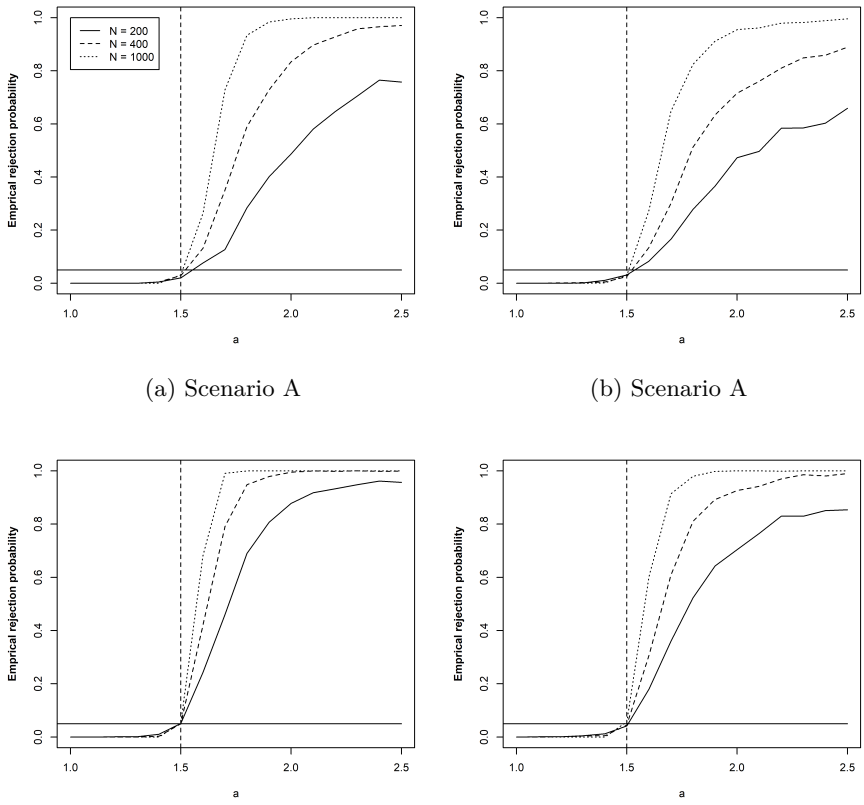
Figure 7: Two-sample covariance-operator test (24) with threshold $\Delta = (1 - 1.5^2)^2(1 + 0.7^2)^2 \sum_{\ell=1}^{D} \sigma_\ell^4$ (so that $a = 1.5$ is on the boundary of H_0^C). Rows correspond to the two variance scenarios. Left column: adjusted-range SN test (25). Right column: quadratic SN benchmark.

so that the covariance operator is multiplied by a^2 after the break. We again set

$$\Delta = \left(1 - 1.5^2\right)^2 (1 + 0.7^2)^2 \sum_{\ell=1}^{D} \sigma_\ell^4,$$

so that $a = 1.5$ corresponds to the relevance boundary in (28). Figure 8 shows the rejection probabilities under the same two variance scenarios and confirms the power advantage of the adjusted-range normalizer in the change-point setting as well.

Remark 5 (Additional stress tests and subspace covariance changes)**.** *We also experimented with additional stress-test designs featuring highly persistent and volatile functional time series, for example near-unit-root functional autoregressive dynamics with time-varying volatility (stochastic-volatility innovations). Such settings are particularly challenging for finite-sample inference. Dette et al. (2020) similarly observes that size control deteriorates as dependence strengthens.*



(a) Scenario A (b) Scenario A

(c) Scenario B (d) Scenario B

Figure 8: Covariance change-point test (28) with threshold $\Delta = (1 - 1.5^2)^2(1 + 0.7^2)^2 \sum_{\ell=1}^{D} \sigma_\ell^4$ (so that $a = 1.5$ is on the boundary of H_0^C). Rows correspond to the two variance scenarios. Left column: adjusted-range SN test (30). Right column: quadratic SN benchmark.

For covariance operators, we additionally considered "subspace" changes in which only the leading K_0 directions are perturbed, in contrast to the global multiplicative covariance changes used in Dette et al. (2020). As expected, size remains close to nominal, but power is lower than under global rescaling (i.e., the benchmark design of Dette et al. (2020) reported here). Nevertheless, the proposed adjusted-range SN test typically remains more powerful than the quadratic SN benchmark. Additional results are available upon request.

6 Empirical analysis

This section applies the proposed relevant mean change-point methodology to BTC option-IV smiles and gives a simple trading illustration. The empirical workflow has two components. First, we construct a daily functional time series of constant-maturity BTC IV smiles from executed Deribit option trades. Second, we

run a rolling-window relevant change-point monitor to (i) estimate the most likely structural break within each window and (ii) test whether the associated mean-smile shift is economically material relative to a user-chosen RMS threshold.

6.1 Data and construction of daily constant-maturity smiles

We work with executed BTC option trades from Deribit and merge them with a daily BTCUSDT close from Binance. For each trading day t with sufficient liquidity, we build a 30-day constant-maturity IV smile as a function of log-moneyness

$$u = \log(K/S_t),$$

where K is the option strike and S_t is the spot close. IVs are treated in decimal units (e.g., 0.60 corresponds to 60 vol points), so that "one vol point" corresponds to 0.01 in decimals. Table 1 summarizes the main data sources and construction choices.

Table 1: Data sources and construction of daily 30-day BTC IV smiles (six-month sample).

Component	Source	Processing / definition
Option trades	Deribit public trade feed	Executed BTC option trades; parse instrument name into expiry date, strike K, and type (call/put); retain trade-reported IV and trade amount.
Spot price S_t	Binance BTCUSDT daily close	Merged by calendar date; used for log-moneyness $u = \log(K/S_t)$ and for delta hedging in the trading illustration.
Sample window	—	20250801 to 20260130 (inclusive); after liquidity filters the sample contains $N = 72$ daily smiles.
Maturity filter	—	Constant-maturity target $\tau = 30$ days; keep trades with time-to-expiry $\tau \in [23, 37]$ days (i.e. $\lvert\tau - 30\rvert \leq 7$ calendar days).
OTM filter	—	Keep out-of-the-money quotes: puts for $u < 0$ and calls for $u > 0$.
Moneyness support	—	Restrict to $u \in [-0.35, 0.35]$.
Grid, binning, smoothing	—	Use an evenly spaced grid of $J = 61$ points in u. Within each day, bin trades using midpoints between grid values, compute an amount-weighted mean IV per non-empty bin, then fit a weighted smoothing spline across u and evaluate it on the fixed grid. Drop days with fewer than 50 trades or fewer than 10 non-empty bins.

The outcome is a daily functional time series $\{X_t(u)\}$ observed on the common grid $\{u_j\}_{j=1}^{J}$. We treat each discretized smile as a curve in $L^2([0,1])$ after linear rescaling of the grid (as discussed in Section 2), and approximate L^2 integrals by trapezoidal weights on the grid.

6.2 Rolling-window relevant change-point monitoring

Let $X_t(u)$ denote the daily IV smile. For each rolling window of length $W = 30$ smile observations, ending at date t, we estimate a single break location by maximizing the integrated squared mean-difference criterion in (20) (Section 3), using trimming $\varepsilon = 0.10$ to avoid candidate splits too close to the window boundaries. Denote the maximizing index by $\widehat{k}_t$ and the corresponding calendar break date by $\widehat{t}_{cp,t}$.

Within the same window, we estimate the squared L^2 magnitude of the pre/post mean-smile difference by

$$\widehat{\mathbb{D}}_t = \int_0^1 \Big\{ \bar{X}_{t,\,\mathrm{pre}}(u) - \bar{X}_{t,\,\mathrm{post}}(u) \Big\}^2 \,\mathrm{d}u,$$

where $\bar{X}_{t,\,\mathrm{pre}}$ and $\bar{X}_{t,\,\mathrm{post}}$ are the sample mean smiles before and after $\widehat{k}_t$ within the window, and the integral is approximated on the grid. We report the associated RMS effect size on the IV scale as $100\sqrt{\widehat{\mathbb{D}}_t}$ (vol points).

Inference is based on the self-normalized relevant change-point decision rule (22). Specifically, for each window we compute two self-normalized statistics: (i) the adjusted-range normalizer $\widehat{\mathbb{H}}_t$ in (21) and (ii) the quadratic (Dette-type) self-normalizer (implemented as the window analogue of (5)), denoted $\widehat{\mathbb{V}}_t$. For critical values we use the pivotal $(1-\alpha)$-quantiles of the corresponding limit variables: $q^*_{1-\alpha}(\mathbb{S})$ for the adjusted-range pivot $\mathbb{S}$ in (8), and $q_{1-\alpha}(\mathbb{W})$ for the quadratic pivot $\mathbb{W}$ in (6) (see, e.g., Dette et al. (2020, Table 1) for tabulated values). We set $\alpha = 0.10$ and use an RMS relevance threshold of $\Delta_{\mathrm{rms}} = 1$ vol point, which corresponds to $\Delta = (\Delta_{\mathrm{rms}}/100)^2 = 10^{-4}$ in the underlying (decimal) units.

To summarize the rolling results in an economically interpretable way, we compute the *implied relevance boundary* for each monitor variant:

$$\widehat{\Delta}^{\star}_{t,\mathrm{AR}} = \max\big\{0,\ \widehat{\mathbb{D}}_t - q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_t\big\}, \qquad \widehat{\Delta}^{\star}_{t,Q} = \max\big\{0,\ \widehat{\mathbb{D}}_t - q_{1-\alpha}(\mathbb{W})\,\widehat{\mathbb{V}}_t\big\}.$$

For any candidate threshold Δ, monitor $j \in \{\mathrm{AR}, Q\}$ rejects the relevant null whenever $\Delta < \widehat{\Delta}^{\star}_{t,j}$; hence $\widehat{\Delta}^{\star}_{t,j}$ can be interpreted as the largest relevance threshold that would still be rejected in that window.

Table 2 reports the windows where at least one of the two self-normalizers rejects at $\Delta_{\mathrm{rms}} = 1$ vol point. The estimated break dates cluster around 20251008 (an upward shift in smile level) and 20251203 (a downward shift). We label the direction as "Up" if the post-break mean smile is higher on average across the grid (and "Down" otherwise). The two self-normalizers can be complementary in short samples: the quadratic monitor rejects for the 20251201 window end, the adjusted-range monitor rejects for the 20251227 window end, and both monitors reject strongly for the 20260107 window end, with implied RMS relevance boundaries exceeding 4 vol points.

Because the constant-maturity smile series is observed on an irregular calendar grid with several multi-week gaps (after liquidity/quality filtering), many $W = 30$ rolling windows contain an imbalanced split around the estimated break, which inflates the self-normalizer and reduces power. Consequently, only the few windows where the break is well inside the window and supported by enough observations on both sides exceed the self-normalized critical value, yielding a small number of rejections clustered around two break episodes. For transparency, the rolling-window output for all windows is summarized in Appendix C (Table C.1).

Table 2: Rolling-window relevant breaks ($W = 30$, $\varepsilon = 0.10$, $\alpha = 0.10$, $\Delta_{\text{rms}} = 1$ vol point). "RMS shift" is $100\sqrt{\widehat{\mathbb{D}}_t}$ (vol points). "RMS boundary" is $100\sqrt{\widehat{\Delta}_t^{\star}}$ (vol points).

Window end	$\widehat{t}_{cp}$	Direction	RMS shift	RMS boundary$_{AR}$	RMS boundary$_{Q}$	Reject$_{AR}$	Reject$_{Q}$
20251201	20251008	Up	12.25	0.00	2.82	0	1
20251227	20251008	Up	9.95	2.14	0.00	1	0
20260107	20251203	Down	5.27	4.28	4.35	1	1

Notes. We evaluate the relevant monitor over a grid of relevance thresholds $\Delta \geq \Delta_{\min} = 10^{-4}$ (one RMS vol point). If the monitor does not reject even at $\Delta_{\min}$, then no positive relevance threshold is rejected in that window; in such cases we report the RMS boundary as "0.00". Reject indicators equal 1 if the monitor rejects the relevant null at the stated $(\alpha, \Delta_{\text{rms}})$ and 0 otherwise.

6.3 Trading illustration: break-direction straddle

We next use the rolling relevant test as an event detector and translate it into a stylized, low-turnover trading rule. On any date t where at least one of the two monitors rejects the relevant null at $\Delta_{\text{rms}} = 1$ vol point (i.e., Reject$_{AR} = 1$ or Reject$_{Q} = 1$ in Table 2), we enter a delta-hedged at-the-money (ATM) straddle with a target maturity of $\tau = 30$ days. The direction of the position is tied to the estimated break direction within the window: for an upward mean-smile shift we go long the straddle (seeking to benefit from an IV mark-up), while for a downward shift we short the straddle (seeking to benefit from an IV mark-down). After a signal, we keep the position for $H = 3$ subsequent smile observations unless a new signal overwrites the position.

P&L is computed per one straddle unit, where one unit is defined as one Black–Scholes call plus one Black–Scholes put on BTC with the same strike K and maturity τ. On entry date t we set K to the nearest multiple of 1000 USD to the spot close S_t, which approximates an ATM straddle. We obtain an IV input by interpolating the daily smile at the corresponding log-moneyness $u_t = \log(K/S_t)$, denoted σ_t. Using Black–Scholes with zero rates, the model straddle value is

$$V_t = C_{\text{BS}}(S_t, K, \tau, \sigma_t) + P_{\text{BS}}(S_t, K, \tau, \sigma_t).$$

We delta-hedge at entry using the model straddle delta $\Delta_t^{\text{str}} = \Delta_{\text{BS}}^{\text{call}} + \Delta_{\text{BS}}^{\text{put}}$ and take a spot hedge of

$h_t = -\text{pos}_t \, \Delta_t^{\text{str}}$ units of BTC, where $\text{pos}_t \in \{-1, +1\}$ denotes the straddle direction (short/long).

Let $\{d_t\}_{t=1}^{T}$ denote the sequence of calendar dates on which an IV-smile observation is available (after applying the liquidity filters used to construct the smile panel). Hence the spacing $d_{t+1} - d_t$ can be irregular. We index these available observations by t, so $t+1$ denotes the next available smile in the panel.

At the end of each observation date d_t we compute the rolling monitor using only information up to d_t. Whenever the strategy takes a non-zero position, it is entered immediately on the same observation date d_t (i.e., the window-end signal date, not the in-window estimated change-point date $\widehat{d}_{cp,t}$), so the backtest is causal and does not use future smiles to form signals.

Because smiles are only observed on this irregular grid, we implement the backtest in observation time: positions are marked to the next available observation d_{t+1} using the next available smile-IV, and each step $t \to t+1$ is treated as a one-trading-day roll in maturity, i.e. the exit valuation uses remaining maturity $\tau - 1$ (one day shorter than at entry), regardless of the calendar gap $d_{t+1} - d_t$. This choice avoids introducing additional assumptions (e.g., interpolation) about the smile on non-observation days.

The one-step hedged P&L is therefore

$$\text{P\&L}_{t \to t+1} = \text{pos}_t \, (V_{t+1} - V_t) + h_t \, (S_{t+1} - S_t) - \text{TC}_{t \to t+1},$$

where proportional transaction costs are charged on both the option premium and the spot hedge. We use 10 bps on the straddle premium (entry and exit) and 2 bps on the spot notional (entry and exit) here.

Figure 9 plots the resulting cumulative P&L in USD per one straddle unit. Over the six-month sample, the strategy trades only on a small set of signal days (because signals are gated by the relevance threshold) and accumulates approximately \$1,300 per straddle unit, with the largest gains concentrated around the early-January downward-break episode. Note that this trading exercise is included purely as an illustration of how relevance-based monitoring can be operationalized; it abstracts from microstructure frictions and does not constitute an out-of-sample performance claim.

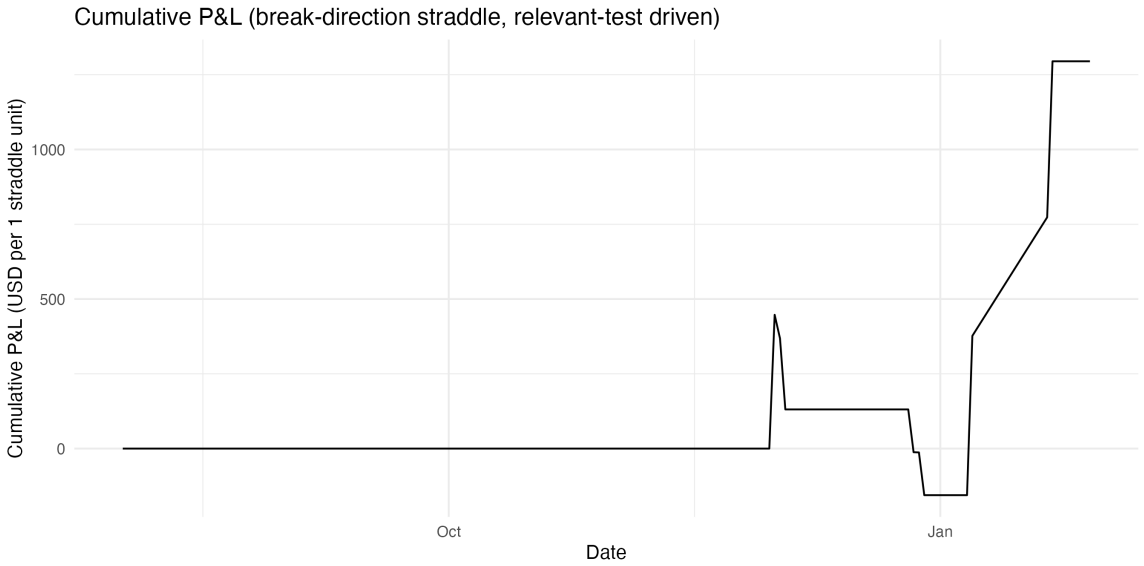


Figure 9: Cumulative P&L for the break-direction straddle strategy driven by the rolling relevant test ($W = 30$, $\Delta_{\mathrm{rms}} = 1$ vol point, $\alpha = 0.10$). P&L is reported in USD per one straddle unit (one call plus one put) after proportional option and spot transaction costs.

7 Conclusion

Classical change tests in time series are very good at answering the question "is there any change at all?" The problem is that, with modern high-resolution data, the answer is often "yes" even when the difference is too small to matter. In many empirical settings—climate monitoring, hydrological risk assessment, and financial risk management are prime examples—what we actually need is an inference tool that separates detectable changes from decision-relevant changes. This paper takes that "so what" perspective seriously and develops a self-normalized framework for testing relevant hypotheses in dependent functional time series: we test whether a distance between mean functions or covariance operators is negligible or large enough to matter relative to a user-chosen threshold $\Delta > 0$.

Methodologically, our procedures remain fully functional. They avoid Karhunen–Loève truncation and therefore do not require choosing the number of principal components, a tuning decision that can materially affect size and power when the signal is not aligned with the leading modes of variation. Within this fully functional setup, we provide a unified treatment of one-sample and two-sample relevant tests, change-point problems, and relevant changes in covariance operators.

A key message is that the adjusted-range-based normalizer is not just a cosmetic replacement of the quadratic self-normalizer. It leads to a different self-normalized calibration and, in practice, a clearer separation between "small but statistically detectable" and "large enough to be relevant." Across a broad collection of dependence scenarios in our simulations, the adjusted-range procedures exhibit a more favor-

able size–power trade-off: they keep rejection probabilities close to the nominal level under the relevant null, while delivering steeper power curves once the effect crosses the relevance boundary. Importantly, this improvement is not merely a simulation artifact. Our Pitman local-alternative theory under boundary drifts $d_n = \Delta + c/\sqrt{n}$ yields an explicit Brownian-functional representation of the limiting rejection probability, making transparent why the adjusted-range scaling can dominate quadratic SN in local power.

We also clarify an important boundary case for relevance-based SN. When $\Delta = 0$, the relevant framework collapses to a classical point null, and the usual pivotal ratio argument can break down because the long-run dependence structure may re-enter the limit law. To cover this benchmark situation with minimal changes to the overall testing logic, we propose a plug-in Monte Carlo calibration based on an estimator of the long-run covariance operator and establish asymptotic level control for the resulting procedure. This provides a practically usable fallback whenever the problem is (near-)degenerate.

In the BTC options illustration, expressing the relevance threshold Δ in RMS IV units (vol points) turns the statistical question into an economically interpretable one. Using executed Deribit trades, we construct constant-maturity IV smiles and apply the rolling-window relevant mean change-point test of Section 3. The resulting monitor produces only a small number of signals, suggesting that relevance-based procedures can filter out "small but statistically detectable" variation and instead highlight shifts that are large enough to be economically meaningful and plausibly offset transaction costs and other market frictions. A break-direction, delta-hedged straddle experiment then illustrates how the same rolling relevance monitor can be used as an event detector and translated into a trading rule with positive model-implied P&L net of costs in this sample.

Additional empirical evidence is provided in Appendix D, where we revisit two benchmark functional datasets studied in Dette et al. (2020) for an apples-to-apples comparison. In the annual temperature two-sample example, the adjusted-range-based and quadratic SN procedures yield similar cut-off locations, but the adjusted-range SN delivers a slightly sharper decision boundary in Δ. In the river-flow change-point example, the relevance boundary shifts more noticeably: at the 10% level the adjusted-range-based procedure rejects up to $\Delta = 0.45$ (not for $\Delta \geq 0.46$), whereas the quadratic benchmark rejects only around $\Delta \approx 0.72$ (not for $\Delta \geq 0.73$). Thus, the proposed adjusted-range-based method can yield more cautious relevance statements—helpful when the goal is to avoid overstating effect sizes under dependence while still maintaining power once changes become genuinely relevant.

Several extensions look promising. A first direction is to develop sequential/online versions of the adjusted-range SN procedures for functional time series, enabling real-time monitoring with controlled false-alarm probabilities under dependence. A second direction is to extend the framework to multiple relevant change points using segmentation or multiscale strategies while preserving relevance-based interpretation (i.e., splitting only when the estimated effect exceeds Δ). Closely related, and practically important, is to

develop methods that not only detect how many change points occur, but also locate them accurately for functional observations—that is, to provide consistent dating/localization of multiple structural breaks in curves (and, ideally, uncertainty quantification for the estimated break locations). Third, it would be valuable to broaden the class of relevant functionals beyond mean and covariance operators—for instance, lagged covariance operators or spectral characteristics, where standardization is even more challenging. Finally, further work on robustness to heavy tails and time-varying volatility, as well as extensions to spatio-temporal functional processes and high-dimensional functional panels, would expand the scope of adjusted-range SN in modern functional econometrics and environmental applications.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (NSFC Grant No. 72173120); the Ministry of Education of Taiwan (Yushan Fellowship, 玉山學者); the project "IDA Institute for Digital Assets" (CF166/15.11.2022; Contract No. CN760046/23.05.2023), financed under Romania's National Recovery and Resilience Plan (Apel nr. PNRR-III-C9-2022-25I8); the Marie Skłodowska-Curie Actions under the European Union's Horizon Europe research and innovation programme (Industrial Doctoral Network on Digital Finance, DIGITAL, Project No. 101119635); and COST Action CA19130 and COST Action CA21163, funded by COST (European Cooperation in Science and Technology). We thank the participants of our contributed session at the 2025 World Congress of the Econometric Society (ESWC 2025) for valuable comments and suggestions.

References

Aston, J. A. D. and C. Kirch (2012). "Detecting and estimating changes in dependent functional data". In: *Journal of Multivariate Analysis* 109, pp. 204–220. DOI: 10.1016/j.jmva.2012.03.006.

Aue, A., S. Hörmann, L. Horváth, M. Reimherr, et al. (2009). "Break detection in the covariance structure of multivariate time series models". In: *The Annals of Statistics* 37.6B, pp. 4046–4087. DOI: 10.1214/09-AOS707.

Bardsley, P., L. Horváth, P. Kokoszka, and G. Young (2017). "Change point tests in functional factor models with application to yield curves". In: *The Econometrics Journal* 20.1, pp. 86–117. DOI: 10.1111/ectj.12075.

Berger, J. O. and M. Delampady (1987). "Testing Precise Hypotheses". In: *Statistical Science* 2.3, pp. 317–335. DOI: 10.1214/ss/1177013238.

Berkes, I., R. Gabrys, L. Horváth, and P. Kokoszka (2009). "Detecting changes in the mean of functional observations". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 71.5, pp. 927–946. DOI: 10.1111/j.1467-9868.2009.00713.x.

Berkes, I., L. Horváth, and G. Rice (2013). "Weak invariance principles for sums of dependent random functions". In: *Stochastic Processes and their Applications* 123.2, pp. 385–403. DOI: 10.1016/j.spa.2012.10.003.

Berkson, J. (1938). "Some Difficulties of Interpretation Encountered in the Application of the Chi-Square Test". In: *Journal of the American Statistical Association* 33.203, pp. 526–536. DOI: 10.1080/01621459.1938.10502329.

Borak, S., M. R. Fengler, and W. K. Härdle (2005). "DSFM Fitting of Implied Volatility Surfaces". In: *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications (ISDA 2005)*. IEEE Computer Society Order Number P2286; Library of Congress Number 2005930524. Wroclaw, Poland: IEEE Computer Society, pp. 526–531. DOI: 10.1109/ISDA.2005.40.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Vol. 149. Lecture Notes in Statistics. New York, NY: Springer. DOI: 10.1007/978-1-4612-1154-9.

Brüggemann, R., W. Härdle, J. Mungo, and C. Trenkler (2008). "VAR modeling for dynamic loadings driving volatility strings". In: *Journal of Financial Econometrics* 6.3, pp. 361–381. DOI: 10.1093/jjfinec/nbn004.

Carlstein, E. G., H.-G. Müller, and D. Siegmund (1994). *Change-point Problems*. California, USA: Institute of Mathematical Sciences.

Cont, R. and J. da Fonseca (2002). "Dynamics of Implied Volatility Surfaces". In: *Quantitative Finance* 2.1, pp. 45–60. DOI: 10.1088/1469-7688/2/1/304.

Dette, H., K. Kokot, and S. Volgushev (2020). "Testing relevant hypotheses in functional time series via self-normalization". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.3, pp. 629–660. DOI: 10.1111/rssb.12370.

Dumas, B., J. Fleming, and R. E. Whaley (1998). "Implied Volatility Functions: Empirical Tests". In: *The Journal of Finance* 53.6, pp. 2059–2106. DOI: 10.1111/0022-1082.00083.

Fengler, M. R. (2005). *Semiparametric Modeling of Implied Volatility*. Springer Finance. Berlin, Heidelberg: Springer. DOI: 10.1007/3-540-30591-2.

Fremdt, S., L. Horváth, P. Kokoszka, and J. G. Steinebach (2014). "Functional data analysis with increasing number of projections". In: *Journal of Multivariate Analysis* 124, pp. 313–332. DOI: 10.1016/j.jmva.2013.11.009.

Hong, Y., Z. Lin, O. Linton, W. Newey, and J. Sun (2025). "Confidence interval estimation and hypothesis testing using the adjusted-range-based self-normalization approach". Working paper.

Hong, Y., O. Linton, B. McCabe, J. Sun, and S. Wang (2024). "Kolmogorov–Smirnov type testing for structural breaks: A new adjusted-range based self-normalization approach". In: *Journal of Econometrics* 238.2, p. 105603. DOI: 10.1016/j.jeconom.2023.105603.

Hörmann, S. and P. Kokoszka (2010). "Weakly dependent functional data". In: *The Annals of Statistics* 38.3, pp. 1845–1884. DOI: 10.1214/09-AOS768.

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics. New York, NY: Springer. DOI: 10.1007/978-1-4614-3655-3.

Horváth, L., P. Kokoszka, and R. Reeder (June 2012). "Estimation of the Mean of Functional Time Series and a Two-Sample Problem". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 75.1, pp. 103–122. DOI: 10.1111/j.1467-9868.2012.01032.x.

Horváth, L., P. Kokoszka, and G. Rice (2014). "Testing stationarity of functional time series". In: *Journal of Econometrics* 179.1, pp. 66–82. DOI: 10.1016/j.jeconom.2013.11.002.

Lengwiler, Y. and C. Lenz (2010). "Intelligible factors for the yield curve". In: *Journal of Econometrics* 157.2, pp. 481–491. DOI: 10.1016/j.jeconom.2010.04.001.

Mandelbrot, B. B. (1972). "Statistical Methodology for Nonperiodic Cycles: From the Covariance to R/S Analysis". In: *Annals of Economic and Social Measurement* 1.3, pp. 259–290.

— (1975). "Limit Theorems on the Self-Normalized Range for Weakly and Strongly Dependent Processes". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 31.4, pp. 271–285. DOI: 10.1007/BF00532867.

Mandelbrot, B. B. and J. R. Wallis (1969). "Computer Experiments with Fractional Gaussian Noises: Part 1, Averages and Variances". In: *Water Resources Research* 5.1, pp. 228–241. DOI: 10.1029/WR005i001p00228.

Müller, H.-G., R. Sen, and U. Stadtmüller (2011). "Functional data analysis for volatility". In: *Journal of Econometrics* 165.2, pp. 233–245. DOI: 10.1016/j.jeconom.2011.08.002.

Müller, U. K. (2007). "A theory of robust long-run variance estimation". In: *Journal of Econometrics* 141.2, pp. 1331–1352. DOI: 10.1016/j.jeconom.2007.01.019.

Nadler, P. and A. Sancetta (2023). "Empirical asset pricing with functional factors". In: *Journal of Financial Econometrics* 21.4, pp. 1258–1281. DOI: 10.1093/jjfinec/nbac003.

Paparoditis, E. and T. Sapatinas (2016). "Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data". In: *Biometrika* 103.3, pp. 727–733. DOI: 10.1093/biomet/asw033.

Park, B. U., E. Mammen, W. Härdle, and S. Borak (2009). "Time series modelling with semiparametric factor dynamics". In: *Journal of the American Statistical Association* 104.485, pp. 284–298. DOI: 10.1198/jasa.2009.0105.

Shao, X. (2010). "A self-normalized approach to confidence interval construction in time series". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 72.3, pp. 343–366. DOI: 10.1111/j.1467-9868.2009.00737.x.

— (2015). "Self-Normalization for Time Series: A Review of Recent Developments". In: *Journal of the American Statistical Association* 110.512, pp. 1797–1817. DOI: 10.1080/01621459.2015.1050493.

Shao, X. and X. Zhang (2010). "Testing for change points in time series". In: *Journal of the American Statistical Association* 105.491, pp. 1228–1240. DOI: 10.1198/jasa.2010.tm10103.

Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC. DOI: 10.1201/EBK1439808184.

Xu, X., L. Chen, and W. K. Härdle (2025). "The Evolving Shape of the Seasons: Quantifying Asymmetric Global Warming via a Nonparametric Functional Data Model". Working paper.

A The classical case $\Delta = 0$ and plug-in Monte Carlo calibration

The test (9) is designed for relevant hypotheses with $\Delta > 0$ and relies on the cancellation of the LRV through the ratio structure. When $\Delta = 0$, the null reduces to the point hypothesis $H_0 : \mu \equiv 0$, for which the quantity τ^2 in (3) degenerates to zero. In this case, the pivotal limit in (8) is no longer valid (cf. Remark 2.1 in Dette et al. (2020)). More precisely, if $\mu \equiv 0$, then both $\widehat{\mathbb{I}}_n$ and $\widehat{\mathbb{H}}_n$ are of smaller order and the ratio converges to a non-pivotal limit that depends on the long-run covariance operator.

To address this case with minimal changes to the overall testing logic, we propose a plug-in Monte Carlo calibration based on an estimator $\widehat{C}$ of the long-run covariance operator C.

Theorem A.1 (Limit law under $\Delta = 0$)**.** *Assume $\mu \equiv 0$ and Assumptions 1–4. Let $\Gamma(t, \lambda)$ denote the centered Gaussian process appearing in the functional central limit theorem for $\sqrt{n}S_n(\cdot, \lambda)$, with covariance*

$$\mathrm{Cov}\{\Gamma(t, \lambda), \Gamma(s, \lambda')\} = (\lambda \wedge \lambda')\, C(s, t).$$

Then,

$$\frac{\widehat{\mathbb{I}}_n}{\widehat{\mathbb{H}}_n} \xrightarrow{\mathcal{L}} \widetilde{\mathbb{S}}(C) \stackrel{\mathrm{def}}{=} \frac{\int_T \Gamma^2(t,1)\, dt}{\max_{\lambda\in[0,1]} \Big\{ \int_T \Gamma^2(t,\lambda)\, dt - \lambda^2 \int_T \Gamma^2(t,1)\, dt \Big\} - \min_{\lambda\in[0,1]} \Big\{ \int_T \Gamma^2(t,\lambda)\, dt - \lambda^2 \int_T \Gamma^2(t,1)\, dt \Big\}}. \tag{A.1}$$

The distribution of $\widetilde{\mathbb{S}}(C)$ depends on C and is therefore not pivotal.

Proof. Recall

$$S_n(t,\lambda) = \frac{1}{n} \sum_{j=1}^{\lfloor n\lambda \rfloor} X_j(t), \qquad \widehat{\mathbb{I}}_n(\lambda) = \int_T S_n^2(t,\lambda)\, dt, \qquad \widehat{\mathbb{I}}_n = \widehat{\mathbb{I}}_n(1),$$

and

$$G_n(\lambda) = \widehat{\mathbb{I}}_n(\lambda) - \lambda^2 \widehat{\mathbb{I}}_n(1), \qquad \widehat{\mathbb{H}}_n = \max_{\lambda\in[0,1]} G_n(\lambda) - \min_{\lambda\in[0,1]} G_n(\lambda).$$

Under Assumptions 1–4, the functional invariance principle holds for the centered partial sums (see, e.g., the discussion around (B.9) in Appendix B):

$$\Big(\sqrt{n}\, S_n(\cdot,\lambda)\Big)_{\lambda\in[0,1]} = \Big(\frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor n\lambda \rfloor} X_j\Big)_{\lambda\in[0,1]} \;\Rightarrow\; \big(\Gamma(\cdot,\lambda)\big)_{\lambda\in[0,1]}$$

in $C([0,1]; L^2(T))$, where Γ is a centered Gaussian process with covariance $\mathrm{Cov}\{\Gamma(t, \lambda), \Gamma(s, \lambda')\} = (\lambda \wedge \lambda')\, C(s, t)$.

Define the mapping $\Phi : C([0,1]; L^2(T)) \to C([0,1])$ by

$$(\Phi x)(\lambda) = \|x(\lambda)\|^2 = \int_T x^2(t,\lambda)\, dt.$$

This map is continuous under the sup-norm because for all x, y and all λ,

$$\big|\|x(\lambda)\|^2 - \|y(\lambda)\|^2\big| \le (\|x(\lambda)\| + \|y(\lambda)\|)\, \|x(\lambda) - y(\lambda)\| \le (\|x\|_\infty + \|y\|_\infty)\, \|x - y\|_\infty,$$

hence $\|\Phi x - \Phi y\|_\infty \le (\|x\|_\infty + \|y\|_\infty)\|x - y\|_\infty$. Therefore, by the continuous mapping theorem,

$$\Big(n\,\widehat{\mathbb{I}}_n(\lambda)\Big)_{\lambda\in[0,1]} = \Big(\int_T (\sqrt{n}\, S_n(t,\lambda))^2\, dt\Big)_{\lambda\in[0,1]} = \Phi\big(\sqrt{n} S_n\big) \;\Rightarrow\; \Big(\int_T \Gamma^2(t,\lambda)\, dt\Big)_{\lambda\in[0,1]}$$

in $C([0,1])$.

Write

$$\widetilde{G}_n(\lambda) \stackrel{\text{def}}{=} n\, G_n(\lambda) = n\,\widehat{\mathbb{I}}_n(\lambda) - \lambda^2\, n\,\widehat{\mathbb{I}}_n(1).$$

By Step 2 and another application of the continuous mapping theorem,

$$\big(\widetilde{G}_n(\lambda)\big)_{\lambda\in[0,1]} \;\Rightarrow\; \big(\widetilde{G}(\lambda)\big)_{\lambda\in[0,1]},$$

where

$$\widetilde{G}(\lambda) \stackrel{\text{def}}{=} \int_T \Gamma^2(t,\lambda)\, dt - \lambda^2 \int_T \Gamma^2(t,1)\, dt.$$

Since $\Gamma(\cdot,\lambda)$ is continuous in λ as an $L^2(T)$-valued process, the map $\lambda \mapsto \int_T \Gamma^2(t,\lambda)\, dt$ is continuous a.s., hence $\widetilde{G} \in C([0,1])$ a.s. The range functional $\mathcal{R} : C([0,1]) \to \mathbb{R}$, $\mathcal{R}(f) = \max_{\lambda\in[0,1]} f(\lambda) - \min_{\lambda\in[0,1]} f(\lambda)$, is continuous, so

$$n\,\widehat{\mathbb{H}}_n = \mathcal{R}(\widetilde{G}_n) \;\Rightarrow\; \mathcal{R}(\widetilde{G}).$$

Moreover, $\mathcal{R}(\widetilde{G}) > 0$ almost surely whenever C is not the zero operator; indeed, $\widetilde{G}(0) = \widetilde{G}(1) = 0$ while $\widetilde{G}(\lambda)$ is non-degenerate for any fixed $\lambda \in (0,1)$ under a non-degenerate long-run covariance.

Combine Steps 2–3:

$$\frac{\widehat{\mathbb{I}}_n}{\widehat{\mathbb{H}}_n} = \frac{n\,\widehat{\mathbb{I}}_n(1)}{n\,\widehat{\mathbb{H}}_n} \;\Rightarrow\; \frac{\int_T \Gamma^2(t,1)\, dt}{\max_{\lambda\in[0,1]} \widetilde{G}(\lambda) - \min_{\lambda\in[0,1]} \widetilde{G}(\lambda)} = \widetilde{\mathbb{S}}(C),$$

which is exactly (A.1). The limit depends on C through Γ, hence is not pivotal. □

A.1 Plug-in Monte Carlo critical values

Because the limit distribution in (A.1) depends on the unknown long-run covariance operator C, we approximate the critical value $q^0_{1-\alpha}(C)$ via a plug-in Monte Carlo scheme based on an estimator $\widehat{C}$. Let $\widehat{C}$ be any consistent estimator of C (for instance, a kernel/HAC estimator applied to the centered curves). Choose a grid $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_M = 1$ (e.g., $\lambda_i = i/20$). For each Monte Carlo repetition $b = 1, \ldots, B$, generate i.i.d. centered Gaussian elements $Z_1^{(b)}, \ldots, Z_M^{(b)} \in L^2(T)$ with covariance operator $\widehat{C}$ and construct a discretized $L^2(T)$-valued Brownian motion in λ,

$$\widehat{\Gamma}^{(b)}(t, \lambda_j) \stackrel{\text{def}}{=} \sum_{i=1}^{j} \sqrt{\lambda_i - \lambda_{i-1}}\, Z_i^{(b)}(t), \qquad j = 0, 1, \ldots, M,$$

with the convention $\widehat{\Gamma}^{(b)}(t, \lambda_0) \equiv 0$. From this path define the scalar process (which vanishes at $lamda = 0$ and $lamda = 1$)

$$\widehat{\widetilde{G}}^{(b)}(\lambda_j) \stackrel{\text{def}}{=} \int_T \{\widehat{\Gamma}^{(b)}(t, \lambda_j)\}^2\, dt - \lambda_j^2 \int_T \{\widehat{\Gamma}^{(b)}(t, 1)\}^2\, dt, \qquad j = 0, 1, \ldots, M,$$

and compute the corresponding draw of the limit functional,

$$\widetilde{\mathbb{S}}^{(b)}(\widehat{C}) \stackrel{\text{def}}{=} \frac{\int_T \{\widehat{\Gamma}^{(b)}(t, 1)\}^2\, dt}{\max_{0 \le j \le M} \widehat{\widetilde{G}}^{(b)}(\lambda_j) - \min_{0 \le j \le M} \widehat{\widetilde{G}}^{(b)}(\lambda_j)}.$$

Finally, let $q^0_{1-\alpha}(\widehat{C})$ be the empirical $(1-\alpha)$-quantile of $\{\widetilde{\mathbb{S}}^{(b)}(\widehat{C})\}_{b=1}^B$.

A.2 Decision rule for $\Delta = 0$

We reject $H_0 : \mu \equiv 0$ in favor of $H_1 : \int_T \mu^2(t)\, dt > 0$ whenever

$$\widehat{\mathbb{I}}_n > q^0_{1-\alpha}(\widehat{C})\, \widehat{\mathbb{H}}_n. \tag{A.2}$$

Theorem A.2 (Asymptotic level for the plug-in test)**.** *Assume $\mu \equiv 0$ and Assumptions 1–4. Because $\int_T \int_T \left\{\widehat{C}(t,s) - C(t,s)\right\}^2 \mathrm{d}t\mathrm{d}s \to 0$ in probability (cf. Eq. (2.15) in Horváth et al. (2012)), the test* (A.2) *has asymptotic level α, that is*

$$\lim_{n \to \infty} \mathrm{P}\big\{\widehat{\mathbb{I}}_n > q^0_{1-\alpha}(\widehat{C})\, \widehat{\mathbb{H}}_n\big\} = \alpha.$$

Moreover, if $\int_T \mu^2(t)\, dt > 0$ and $\tau^2 > 0$, then the test is consistent.

The proof of Theorem A.2 follows from Theorem A.1, continuity of the Gaussian law in the covariance

operator (cf. the Gaussian approximation step in Berkes et al. (2013)), and Slutsky's theorem. It is, therefore, omitted for brevity.

Remark 6 (Near-degenerate regimes)**.** *Because $d = \int_T \mu^2$ is quadratic in μ, the natural local alternatives around the classical null $\mu \equiv 0$ scale as $\mu_n = n^{-1/4}h$, for which $d_n = \|h\|^2/\sqrt{n}$. In this regime the leading term in $\widehat{\mathbb{I}}_n$ is no longer linear and the long-run covariance operator enters the limit law through* (A.1)*. The plug-in Monte Carlo calibration in* (A.2) *is therefore a natural way to obtain valid critical values uniformly in such near-degenerate settings.*

B Proof of main theorems

In this appendix, we use $G_n \Rightarrow G$ to denote weak convergence in the space $\ell^\infty([0,1])$ (or $C([0,1])$, as appropriate) for stochastic processes. Most of the argument follows the same invariance-principle route as in the quadratic SN case; the only change is in the continuous-mapping step for the adjusted-range functional.

B.1 A general self-normalization principle, implied relevance bounds, and local-power comparison

SN in relevant-hypothesis testing is not unique; see Dette et al. (2020, Remark 5) for discussion in the quadratic SN setting. This subsection records a generic principle showing that any continuous, positively homogeneous functional of the centered subsample path yields a pivotal self-normalized limit. It also clarifies that the relevant test can be interpreted as an inversion of a one-sided confidence bound for the effect size d, and it provides a unified local-power formula for comparing alternative self-normalizers.

Let $d \in \mathbb{R}$ denote a scalar parameter and let $\widehat{d}_n$ be an estimator admitting a subsample path $\{\widehat{d}_n(\lambda)\}_{\lambda\in[0,1]}$ with $\widehat{d}_n = \widehat{d}_n(1)$. Define the centered subsample path

$$A_n(\lambda) \stackrel{\text{def}}{=} \widehat{d}_n(\lambda) - \lambda^2 \widehat{d}_n(1), \qquad \lambda \in [0,1], \tag{B.1}$$

and let $\rho : \ell^\infty([0,1]) \to [0,\infty)$ be a functional. We define the associated self-normalizer by

$$\widehat{H}_{n,\rho} \stackrel{\text{def}}{=} \rho(A_n). \tag{B.2}$$

For the one-sample mean problem, $A_n(\lambda) = G_n(\lambda)$ in (4). The two-sample, change-point, and covariance-operator settings use the same construction with the corresponding subsample statistic $\widehat{d}_n(\lambda)$.

Theorem B.1 (Generic SN principle)**.** *Assume $\tau > 0$ and suppose that, in $\mathbb{R} \times \ell^\infty([0,1])$,*

$$\Big(\sqrt{n}(\widehat{d}_n - d),\, \sqrt{n}A_n\Big) \Rightarrow \Big(\tau\mathbb{B}(1),\, \tau U\Big), \tag{B.3}$$

where $\mathbb{B}$ is standard Brownian motion on $[0,1]$ and $U(\lambda) \stackrel{\text{def}}{=} \lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)$. Let ρ be continuous on $C([0,1])$ under the supremum norm and positively homogeneous, i.e., $\rho(cf) = c\rho(f)$ for all $c \geq 0$ and $f \in C([0,1])$. Assume further that $\rho(U) > 0$ almost surely. Define the pivotal limit variable

$$\mathbb{T}_\rho \stackrel{\text{def}}{=} \frac{\mathbb{B}(1)}{\rho(U)}, \qquad q_{1-\alpha}(\mathbb{T}_\rho) \stackrel{\text{def}}{=} \inf\{x : \mathrm{P}(\mathbb{T}_\rho \leq x) \geq 1-\alpha\}.$$

(i) Pivotal ratio limit.

$$\frac{\widehat{d}_n - d}{\widehat{H}_{n,\rho}} \;\xrightarrow{\mathcal{L}}\; \mathbb{T}_\rho.$$

(ii) Inversion and implied relevance bound. Let

$$L_{n,\alpha}^{(\rho)} \stackrel{\text{def}}{=} \widehat{d}_n - q_{1-\alpha}(\mathbb{T}_\rho)\, \widehat{H}_{n,\rho}.$$

Then $[L_{n,\alpha}^{(\rho)}, \infty)$ is an asymptotic $(1-\alpha)$ one-sided confidence set for d:

$$\lim_{n\to\infty} \mathrm{P}\{d \geq L_{n,\alpha}^{(\rho)}\} = 1 - \alpha.$$

Moreover, the level-α relevant test based on ρ,

$$\text{reject } H_0 : d \leq \Delta \quad \Longleftrightarrow \quad \widehat{d}_n > \Delta + q_{1-\alpha}(\mathbb{T}_\rho)\, \widehat{H}_{n,\rho},$$

can be written equivalently as $L_{n,\alpha}^{(\rho)} > \Delta$. Thus $L_{n,\alpha}^{(\rho)}$ is the implied relevance bound (largest threshold Δ rejected at level α). If the application reports the effect in RMS units $\sqrt{d}$ (e.g. IV "vol points"), then

$$\liminf_{n\to\infty} \mathrm{P}\Big\{\sqrt{d} \geq \sqrt{\max\{L_{n,\alpha}^{(\rho)}, 0\}}\Big\} \;\geq\; 1 - \alpha.$$

(iii) Pitman local power for generic SN. Consider Pitman drifts $d_n = \Delta + c/\sqrt{n}$ with $c \in \mathbb{R}$ fixed. Then the corresponding level-α test based on ρ satisfies

$$\lim_{n\to\infty} \mathrm{P}\Big\{\widehat{d}_n > \Delta + q_{1-\alpha}(\mathbb{T}_\rho)\, \widehat{H}_{n,\rho}\Big\} = \mathrm{P}\left\{\frac{\mathbb{B}(1) + c/\tau}{\rho(U)} > q_{1-\alpha}(\mathbb{T}_\rho)\right\},$$

which is increasing in c and equals α at $c = 0$.

Proof. By positive homogeneity, $\sqrt{n}\,\widehat{H}_{n,\rho} = \rho(\sqrt{n}A_n)$. The joint convergence (B.3) and the continuous mapping theorem yield $\sqrt{n}\,\widehat{H}_{n,\rho} \Rightarrow \rho(\tau U) = \tau\rho(U)$, while $\sqrt{n}(\widehat{d}_n - d) \Rightarrow \tau\mathbb{B}(1)$. Part (i) follows from Slutsky's theorem (using $\rho(U) > 0$ a.s.). Part (ii) is an algebraic rearrangement of the event $\{(\widehat{d}_n - d)/\widehat{H}_{n,\rho} \leq q_{1-\alpha}(\mathbb{T}_\rho)\}$ combined with part (i). For part (iii), write

$$\sqrt{n}(\widehat{d}_n - \Delta) = \sqrt{n}(\widehat{d}_n - d_n) + c,$$

and apply (B.3) with d replaced by d_n to obtain the limit $\tau\mathbb{B}(1)+c$ for the numerator, while $\sqrt{n}\,\widehat{H}_{n,\rho} \Rightarrow \tau\rho(U)$ as above. □

Remark 7 (Examples of ρ and a local-power comparison criterion)**.** *Two choices of ρ studied in the paper are*

$$\rho_{2,\nu}(f) \stackrel{\text{def}}{=} \Big\{ \int_0^1 f^2(\lambda)\,\nu(d\lambda) \Big\}^{1/2}, \qquad \rho_{\text{range}}(f) \stackrel{\text{def}}{=} \sup_{\lambda\in[0,1]} f(\lambda) - \inf_{\lambda\in[0,1]} f(\lambda).$$

With $\rho = \rho_{2,\nu}$, $\mathbb{T}_\rho$ coincides with $\mathbb{W}$ in (6); with $\rho = \rho_{\text{range}}$, it coincides with $\mathbb{S}$ in (8) (and $q_{1-\alpha}(\mathbb{T}_\rho)$ corresponds to $q^*_{1-\alpha}(\mathbb{S})$ in the main text). Other continuous homogeneous functionals (e.g. $\rho_\infty(f) = \sup_\lambda |f(\lambda)|$ or $\rho_{1,\nu}(f) = \int_0^1 |f(\lambda)|\,\nu(d\lambda)$) are also covered by Theorem B.1 whenever $\rho(U) > 0$ a.s.

Part (iii) in Theorem B.1 provides a fully pivotal way to compare candidate self-normalizers: for fixed (α, ρ) the entire Pitman local-power curve depends only on Brownian motion. Hence it can be evaluated once by Monte Carlo simulation (as in Section 2.1.1), which provides a practical design criterion for selecting ρ based on the range of effect sizes deemed practically relevant.

B.2 Proof of Theorem 1

Proof. Write

$$d \stackrel{\text{def}}{=} \int_T \mu^2(t)\,dt, \qquad \widehat{\mathbb{I}}_n(\lambda) \stackrel{\text{def}}{=} \int_T S_n^2(t,\lambda)\,dt, \qquad \widehat{\mathbb{I}}_n \stackrel{\text{def}}{=} \widehat{\mathbb{I}}_n(1),$$

and define the centered process

$$Z_n(\lambda) \stackrel{\text{def}}{=} \sqrt{n}\Big(\widehat{\mathbb{I}}_n(\lambda) - \lambda^2 d\Big), \qquad \lambda \in [0,1].$$

Under Assumptions 1–4, it can be shown (see Dette et al. (2020, Eq. (2.11) and the proof of Theorem 1)) that, in $\ell^\infty([0,1])$,

$$\big(Z_n(\lambda)\big)_{\lambda\in[0,1]} \;\Rightarrow\; \big(\lambda\,\tau\,\mathbb{B}(\lambda)\big)_{\lambda\in[0,1]}, \tag{B.4}$$

where $\mathbb{B}$ is standard Brownian motion on $[0,1]$ and

$$\tau^2 = 4\int_T\int_T \mu(s)\mu(t)\, C(s,t)\, ds\, dt$$

with C the long-run covariance kernel (cf. Dette et al. (2020, Eqs. (2.4)–(2.5))).

Taking $\lambda = 1$ in (B.4) yields

$$\sqrt{n}\big(\widehat{\mathbb{I}}_n - d\big) = Z_n(1) \;\Rightarrow\; \tau\,\mathbb{B}(1). \tag{B.5}$$

For $\lambda \in [0,1]$ define

$$A_n(\lambda) \stackrel{\text{def}}{=} \widehat{\mathbb{I}}_n(\lambda) - \lambda^2 \widehat{\mathbb{I}}_n(1),$$

so that the adjusted-range self-normalizer is

$$\widehat{\mathbb{H}}_n = \max_{\lambda\in[0,1]} A_n(\lambda) - \min_{\lambda\in[0,1]} A_n(\lambda).$$

By simple algebra,

$$A_n(\lambda) = \Big(\widehat{\mathbb{I}}_n(\lambda) - \lambda^2 d\Big) - \lambda^2\Big(\widehat{\mathbb{I}}_n(1) - d\Big),$$

and hence

$$\sqrt{n}A_n(\lambda) = Z_n(\lambda) - \lambda^2 Z_n(1).$$

Combining (B.4) and (B.5) and applying the continuous mapping theorem (the mapping $x(\cdot) \mapsto x(\lambda) - \lambda^2 x(1)$ is continuous on $\ell^\infty([0,1])$) gives, in $\ell^\infty([0,1])$,

$$\big(\sqrt{n}A_n(\lambda)\big)_{\lambda\in[0,1]} \;\Rightarrow\; \big(\tau\, U(\lambda)\big)_{\lambda\in[0,1]}, \qquad U(\lambda) \stackrel{\text{def}}{=} \lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1).$$

Since $\mathbb{B}$ has continuous paths a.s., U is continuous a.s., and the functionals $f \mapsto \max_{\lambda\in[0,1]} f(\lambda)$ and $f \mapsto \min_{\lambda\in[0,1]} f(\lambda)$ are continuous on $C([0,1])$ under the sup-norm. Therefore another application of the continuous mapping theorem yields

$$\sqrt{n}\,\widehat{\mathbb{H}}_n \;\Rightarrow\; \tau\Big(\max_{\lambda\in[0,1]} U(\lambda) - \min_{\lambda\in[0,1]} U(\lambda)\Big). \tag{B.6}$$

Moreover, if $\tau^2 > 0$, then $\max_\lambda U(\lambda) - \min_\lambda U(\lambda) > 0$ almost surely (because U is not almost surely constant), so the limit in (B.6) is positive a.s.

If $\tau^2 > 0$, Slutsky's theorem and (B.5)–(B.6) give

$$\frac{\widehat{\mathbb{I}}_n - d}{\widehat{\mathbb{H}}_n} = \frac{\sqrt{n}(\widehat{\mathbb{I}}_n - d)}{\sqrt{n}\widehat{\mathbb{H}}_n} \;\Rightarrow\; \frac{\mathbb{B}(1)}{\max_{\lambda\in[0,1]} U(\lambda) - \min_{\lambda\in[0,1]} U(\lambda)} \;\equiv\; \mathbb{S},$$

which is exactly the pivotal limit used to define $q^*_{1-\alpha}(\mathbb{S})$.

Recall the decision rule (9): reject $H_0 : d \le \Delta$ if

$$\widehat{\mathbb{I}}_n > \Delta + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_n.$$

Case 1: $d < \Delta$. Since $\widehat{\mathbb{I}}_n \to d$ in probability and $\widehat{\mathbb{H}}_n = \mathscr{O}_{\mathrm{p}}(n^{-1/2})$ by (B.6), we have

$$\Delta + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_n = \Delta + o_{\mathrm{p}}(1) \;\xrightarrow{\mathrm{p}}\; \Delta.$$

Hence

$$\mathrm{P}\Big(\widehat{\mathbb{I}}_n > \Delta + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_n\Big) = \mathrm{P}\Big(\widehat{\mathbb{I}}_n - \Delta - q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_n > 0\Big) \;\to\; 0,$$

because the difference converges in probability to $d - \Delta < 0$.

Case 2: $d = \Delta$ and $\tau^2 > 0$. On the event $\{\widehat{\mathbb{H}}_n > 0\}$ (which has probability tending to one when $\tau^2 > 0$), the rejection event is equivalent to

$$\frac{\widehat{\mathbb{I}}_n - d}{\widehat{\mathbb{H}}_n} > q^*_{1-\alpha}(\mathbb{S}).$$

By Step 3, the ratio converges in distribution to $\mathbb{S}$, and therefore

$$\lim_{n\to\infty} \mathrm{P}\Big(\widehat{\mathbb{I}}_n > d + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_n\Big) = \mathrm{P}\big(\mathbb{S} > q^*_{1-\alpha}(\mathbb{S})\big) = \alpha.$$

Case 3: $d > \Delta$. Exactly as in Case 1, $\widehat{\mathbb{I}}_n \to d$ and $\Delta + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_n = \Delta + o_{\mathrm{p}}(1)$, so

$$\mathrm{P}\Big(\widehat{\mathbb{I}}_n > \Delta + q^*_{1-\alpha}(\mathbb{S})\,\widehat{\mathbb{H}}_n\Big) \;\to\; 1,$$

because the left-hand side converges to $d > \Delta$.

Putting the three cases together proves the theorem. □

B.3 Proof of Theorem 2

Proof. Fix $c \in \mathbb{R}$ and consider the local alternatives in (12), $d_n = \Delta + c/\sqrt{n}$. For clarity we write the triangular-array representation

$$X_{j,n} = \mu_n + \eta_j, \qquad j = 1, \ldots, n,$$

where the centered stationary error process $(\eta_j)_{j\in\mathbb{Z}}$ satisfies Assumptions 2–4 (and thus does not depend on n). Define for $\lambda \in [0,1]$

$$S_n(t,\lambda) = \frac{1}{n}\sum_{j=1}^{\lfloor n\lambda \rfloor} X_{j,n}(t) = \lambda\mu_n(t) + \widetilde{S}_n(t,\lambda), \qquad \widetilde{S}_n(t,\lambda) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{j=1}^{\lfloor n\lambda \rfloor} \eta_j(t).$$

Let

$$A_n(\lambda) \stackrel{\text{def}}{=} \int_T S_n^2(t,\lambda)\,dt \quad \text{so that} \quad \widehat{\mathbb{I}}_n = A_n(1).$$

By expanding the square we obtain, for each $\lambda \in [0,1]$,

$$\begin{aligned} A_n(\lambda) &= \int_T \{\lambda\mu_n(t) + \widetilde{S}_n(t,\lambda)\}^2\,dt \\ &= \lambda^2 \int_T \mu_n^2(t)\,dt + 2\lambda \int_T \mu_n(t)\widetilde{S}_n(t,\lambda)\,dt + \int_T \widetilde{S}_n^2(t,\lambda)\,dt \\ &= \lambda^2 d_n + 2\lambda\langle \mu_n, \widetilde{S}_n(\cdot,\lambda)\rangle + \|\widetilde{S}_n(\cdot,\lambda)\|^2. \end{aligned} \tag{B.7}$$

By the WIP (e.g., Theorem 1.1 in Berkes et al., 2013), $\{S_n\}$ is tight in $D_H[0,1]$. Consequently, $\sup_\lambda \left\|\sum_{j\le n\lambda} \eta_j\right\| = \mathscr{O}_{\mathrm{p}}(\sqrt{n})$, and hence

$$\sup_{\lambda\in[0,1]} \sqrt{n}\,\|\widetilde{S}_n(\cdot,\lambda)\|^2 = \sup_{\lambda\in[0,1]} \frac{1}{n^{3/2}} \Big\| \sum_{j\le\lfloor n\lambda\rfloor} \eta_j \Big\|^2 = \mathscr{O}_{\mathrm{p}}(n^{-1/2}) = o_{\mathrm{p}}(1).$$

Therefore, uniformly in $\lambda \in [0,1]$,

$$\sqrt{n}\{A_n(\lambda) - \lambda^2 d_n\} = 2\lambda \left\langle \mu_n, \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor n\lambda\rfloor} \eta_j \right\rangle + o_{\mathrm{p}}(1). \tag{B.8}$$

Under Assumptions 1–4 the functional invariance principle for the centered partial sums holds, so $\{n^{-1/2}\sum_{j\le\lfloor n\lambda\rfloor}\eta_j\}_{\lambda\in[0,1]}$ converges weakly in $\ell^\infty([0,1];L^2(T))$ to a centered Gaussian process $\{\Gamma(\cdot,\lambda)\}_{\lambda\in[0,1]}$ with covariance $\mathrm{Cov}\{\Gamma(t,\lambda),\Gamma(s,\lambda')\} = (\lambda\wedge\lambda')\,C(s,t)$. Consequently, the real-valued process $\{\langle\mu_n,\Gamma(\cdot,\lambda)\rangle\}_{\lambda\in[0,1]}$

is Gaussian with covariance

$$\mathrm{Cov}\{\langle \mu_n, \Gamma(\cdot,\lambda)\rangle, \langle \mu_n, \Gamma(\cdot,\lambda')\rangle\} = (\lambda \wedge \lambda')\,\langle \mu_n, C\mu_n\rangle.$$

Define

$$\tau_n^2 \stackrel{\text{def}}{=} 4\langle \mu_n, C\mu_n\rangle = 4\int_T\int_T \mu_n(s)\mu_n(t)C(s,t)\,ds\,dt,$$

and assume (as in the theorem statement) that $\tau_n \to \tau > 0$. Then there exists a standard Brownian motion $\mathbb{B}$ on $[0,1]$ such that

$$2\langle \mu_n, \Gamma(\cdot,\lambda)\rangle \;\stackrel{d}{=}\; \tau_n\,\mathbb{B}(\lambda), \qquad \lambda \in [0,1],$$

where $\stackrel{d}{=}$ denotes equality in distribution/law. Combining this with (B.8) and the continuous mapping theorem yields in $\ell^\infty([0,1])$,

$$\Big\{\sqrt{n}\big(A_n(\lambda) - \lambda^2 d_n\big)\Big\}_{\lambda\in[0,1]} \;\Rightarrow\; \Big\{\lambda\,\tau\,\mathbb{B}(\lambda)\Big\}_{\lambda\in[0,1]}. \tag{B.9}$$

In particular, at $\lambda = 1$ we have

$$\sqrt{n}\big(\widehat{\mathbb{I}}_n - d_n\big) = \sqrt{n}\big(A_n(1) - d_n\big) \;\xrightarrow{\mathcal{L}}\; \tau\,\mathbb{B}(1). \tag{B.10}$$

Recall that

$$\widehat{\mathbb{H}}_n = \max_{\lambda\in[0,1]}\{A_n(\lambda) - \lambda^2 A_n(1)\} - \min_{\lambda\in[0,1]}\{A_n(\lambda) - \lambda^2 A_n(1)\}.$$

Define the $\sqrt{n}$-scaled bridge-type process

$$Z_n(\lambda) \stackrel{\text{def}}{=} \sqrt{n}\{A_n(\lambda) - \lambda^2 A_n(1)\}.$$

Using $A_n(1) = d_n + \mathscr{O}_{\mathrm{p}}(n^{-1/2})$ and (B.9)–(B.10),

$$Z_n(\lambda) = \sqrt{n}\{A_n(\lambda) - \lambda^2 d_n\} - \lambda^2\sqrt{n}\{A_n(1) - d_n\} \;\Rightarrow\; \tau\,U(\lambda),$$

in $\ell^\infty([0,1])$, where

$$U(\lambda) \stackrel{\text{def}}{=} \lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1).$$

The mapping $x \mapsto \max_\lambda x(\lambda) - \min_\lambda x(\lambda)$ is continuous (and Lipschitz) under the supremum norm on

$\ell^\infty([0,1])$. Therefore, by the continuous mapping theorem,

$$\sqrt{n}\,\widehat{\mathbb{H}}_n = \max_\lambda Z_n(\lambda) - \min_\lambda Z_n(\lambda) \xrightarrow{\mathcal{L}} \tau\,\mathcal{R}, \tag{B.11}$$

with

$$\mathcal{R} \stackrel{\text{def}}{=} \max_{\lambda\in[0,1]} U(\lambda) - \min_{\lambda\in[0,1]} U(\lambda).$$

Joint convergence of (B.10) and (B.11) follows from (B.9) by another application of the continuous mapping theorem.

Let $q = q^*_{1-\alpha}(\mathbb{S})$. Under $d_n = \Delta + c/\sqrt{n}$, the rejection event is

$$\begin{aligned}\Big\{\widehat{\mathbb{I}}_n > \Delta + q\,\widehat{\mathbb{H}}_n\Big\} &= \Big\{\widehat{\mathbb{I}}_n - d_n > \Delta - d_n + q\,\widehat{\mathbb{H}}_n\Big\} \\ &= \Big\{\sqrt{n}(\widehat{\mathbb{I}}_n - d_n) > -c + q\,\sqrt{n}\,\widehat{\mathbb{H}}_n\Big\}.\end{aligned}$$

Taking limits and using the joint convergence gives

$$\lim_{n\to\infty} \mathrm{P}\Big\{\widehat{\mathbb{I}}_n > \Delta + q\,\widehat{\mathbb{H}}_n\Big\} = \mathrm{P}\Big\{\tau\mathbb{B}(1) > -c + q\,\tau\mathcal{R}\Big\} = \mathrm{P}\Big\{\frac{\mathbb{B}(1) + c/\tau}{\mathcal{R}} > q\Big\},$$

which is exactly (13).

Finally, at $c = 0$ the right-hand side equals α by definition of $q^*_{1-\alpha}(\mathbb{S})$ for $\mathbb{S} = \mathbb{B}(1)/\mathcal{R}$. Moreover, $\mathcal{R} > 0$ almost surely, and for each sample path the indicator $\mathbf{1}\{(\mathbb{B}(1)+c/\tau)/\mathcal{R} > q\}$ is strictly increasing in c, which implies that the limit power is strictly increasing in c. □

B.4 Proof of Theorem 3

Proof. Set $N \stackrel{\text{def}}{=} m + n$ and $\rho_N \stackrel{\text{def}}{=} m/N$, so $\rho_N \to \rho \in (0,1)$ by Assumption 5. Let $d \stackrel{\text{def}}{=} \int_T D^2(t)\,dt$. Write $X_j = \mu_1 + \eta_j^{(X)}$ and $Y_j = \mu_2 + \eta_j^{(Y)}$, where the two centered error sequences satisfy the weak-dependence/moment conditions inherited from Assumptions 1–4, and are independent by Assumption 6.

Define the centered process

$$\widetilde{D}_{m,n}(t,\lambda) \stackrel{\text{def}}{=} D_{m,n}(t,\lambda) - \lambda D(t), \qquad \lambda \in [0,1].$$

By applying Berkes et al. (2013, Theorem 1.1) to each centered error sequence $\{\eta_j^{(X)}\}$ and $\{\eta_j^{(Y)}\}$, the corresponding $L^2(T)$-valued partial-sum processes converge weakly in $D_{L^2(T)}[0,1]$ (indeed in $C_{L^2(T)}[0,1]$).

Using independence of the samples and $\rho_N \to \rho$, Slutsky's theorem yields

$$\Big\{\sqrt{N}\,\widetilde{D}_{m,n}(\cdot,\lambda)\Big\}_{\lambda\in[0,1]} \Rightarrow \Big\{\Gamma_D(\cdot,\lambda)\Big\}_{\lambda\in[0,1]},$$

where $\Gamma_D(\cdot,\lambda)$ is a centered Gaussian element in $L^2(T)$ with covariance structure

$$\mathrm{Cov}\{\Gamma_D(s,\lambda),\Gamma_D(t,\lambda')\} = (\lambda\wedge\lambda')\Big\{\tfrac{1}{\rho}C_X(s,t) + \tfrac{1}{1-\rho}C_Y(s,t)\Big\}.$$

For $\lambda \in [0,1]$ define

$$\widehat{\mathbb{D}}_{m,n}(\lambda) \overset{\text{def}}{=} \int_T D_{m,n}^2(t,\lambda)\,dt \qquad \text{so that} \qquad \widehat{\mathbb{D}}_{m,n} = \widehat{\mathbb{D}}_{m,n}(1).$$

Using $D_{m,n}(\cdot,\lambda) = \lambda D(\cdot) + \widetilde{D}_{m,n}(\cdot,\lambda)$ we expand

$$\begin{aligned}\widehat{\mathbb{D}}_{m,n}(\lambda) - \lambda^2 d &= \int_T \Big\{\lambda D(t) + \widetilde{D}_{m,n}(t,\lambda)\Big\}^2 dt - \lambda^2\int_T D^2(t)dt \\ &= 2\lambda\int_T D(t)\,\widetilde{D}_{m,n}(t,\lambda)\,dt \;+\; \int_T \widetilde{D}_{m,n}^2(t,\lambda)\,dt.\end{aligned}$$

Since $\{\sqrt{N}\,\widetilde{D}_{m,n}(\cdot,\lambda)\}_{\lambda\in[0,1]}$ converges weakly in $D_{L^2(T)}[0,1]$ by Step 1, it is tight in this path space. In particular,

$$\sup_{\lambda\in[0,1]} \big\|\sqrt{N}\,\widetilde{D}_{m,n}(\cdot,\lambda)\big\| = \mathscr{O}_p(1),$$

and hence

$$\sup_{\lambda\in[0,1]} \big\|\widetilde{D}_{m,n}(\cdot,\lambda)\big\| = \mathscr{O}_p(N^{-1/2}).$$

Thus, we have

$$\sup_{\lambda\in[0,1]} \sqrt{N}\int_T \widetilde{D}_{m,n}^2(t,\lambda)\,dt = \sqrt{N}\sup_{\lambda\in[0,1]} \|\widetilde{D}_{m,n}(\cdot,\lambda)\|^2 = \mathscr{O}_p(N^{-1/2}) = o_p(1).$$

Therefore, in $\ell^\infty([0,1])$,

$$Z_{m,n}(\lambda) \;\overset{\text{def}}{=}\; \sqrt{N}\Big\{\widehat{\mathbb{D}}_{m,n}(\lambda) - \lambda^2 d\Big\} = 2\lambda\int_T D(t)\,\sqrt{N}\,\widetilde{D}_{m,n}(t,\lambda)\,dt \;+\; o_{\mathrm{p}}(1),$$

and by the continuous mapping theorem,

$$\{Z_{m,n}(\lambda)\}_{\lambda\in[0,1]} \Rightarrow \Big\{2\lambda\int_T D(t)\,\Gamma_D(t,\lambda)\,dt\Big\}_{\lambda\in[0,1]}.$$

The limit on the right is a centered Gaussian process with covariance

$$\mathrm{Cov}\Big(2\lambda\int_T D(t)\Gamma_D(t,\lambda)dt,\ 2\lambda'\int_T D(s)\Gamma_D(s,\lambda')ds\Big) = (\lambda\wedge\lambda')\,\lambda\lambda'\,\tau_D^2,$$

where τ_D^2 is exactly the constant defined in the text. Hence we may represent the limit as

$$\{Z_{m,n}(\lambda)\}_{\lambda\in[0,1]} \Rightarrow \{\lambda\,\tau_D\,\mathbb{B}(\lambda)\}_{\lambda\in[0,1]},$$

with $\mathbb{B}$ standard Brownian motion on $[0,1]$. In particular,

$$\sqrt{N}\big(\widehat{\mathbb{D}}_{m,n} - d\big) = Z_{m,n}(1) \Rightarrow \tau_D\,\mathbb{B}(1).$$

Recall

$$\widehat{\mathbb{H}}_{m,n} = \max_{\lambda\in[0,1]}\Big\{\widehat{\mathbb{D}}_{m,n}(\lambda) - \lambda^2\widehat{\mathbb{D}}_{m,n}(1)\Big\} - \min_{\lambda\in[0,1]}\Big\{\widehat{\mathbb{D}}_{m,n}(\lambda) - \lambda^2\widehat{\mathbb{D}}_{m,n}(1)\Big\}.$$

Using $\widehat{\mathbb{D}}_{m,n}(\lambda) = \lambda^2 d + N^{-1/2}Z_{m,n}(\lambda) + o_{\mathrm{p}}(N^{-1/2})$ and $\widehat{\mathbb{D}}_{m,n}(1) = d + N^{-1/2}Z_{m,n}(1) + o_{\mathrm{p}}(N^{-1/2})$, we obtain uniformly in λ,

$$\widehat{\mathbb{D}}_{m,n}(\lambda) - \lambda^2\widehat{\mathbb{D}}_{m,n}(1) = \frac{1}{\sqrt{N}}\Big\{Z_{m,n}(\lambda) - \lambda^2 Z_{m,n}(1)\Big\} + o_{\mathrm{p}}(N^{-1/2}).$$

Therefore,

$$\sqrt{N}\,\widehat{\mathbb{H}}_{m,n} = \max_{\lambda\in[0,1]}\{Z_{m,n}(\lambda) - \lambda^2 Z_{m,n}(1)\} - \min_{\lambda\in[0,1]}\{Z_{m,n}(\lambda) - \lambda^2 Z_{m,n}(1)\} + o_{\mathrm{p}}(1).$$

Since Brownian paths are almost surely continuous, the map $x \mapsto \max_\lambda(x(\lambda)-\lambda^2 x(1)) - \min_\lambda(x(\lambda)-\lambda^2 x(1))$ is a.s. continuous at the limit process, so another application of the continuous mapping theorem yields

$$\sqrt{N}\,\widehat{\mathbb{H}}_{m,n} \Rightarrow \tau_D\left(\max_{\lambda\in[0,1]}\{\lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)\} - \min_{\lambda\in[0,1]}\{\lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)\}\right) =: \tau_D\,\mathcal{R}.$$

Consequently, by Slutsky,

$$\frac{\widehat{\mathbb{D}}_{m,n} - d}{\widehat{\mathbb{H}}_{m,n}} = \frac{\sqrt{N}(\widehat{\mathbb{D}}_{m,n} - d)}{\sqrt{N}\widehat{\mathbb{H}}_{m,n}} \Rightarrow \frac{\mathbb{B}(1)}{\mathcal{R}} =: \mathbb{S},$$

which is pivotal.

(i) If $d < \Delta$, then $\widehat{\mathbb{D}}_{m,n} \to d$ in probability and $\widehat{\mathbb{H}}_{m,n} = \mathscr{O}_{\mathrm{p}}(N^{-1/2})$, so $\Delta + q^*_{1-\alpha}(\mathbb{S})\widehat{\mathbb{H}}_{m,n} \to \Delta$. Hence $\mathrm{P}\{\widehat{\mathbb{D}}_{m,n} > \Delta + q^*_{1-\alpha}(\mathbb{S})\widehat{\mathbb{H}}_{m,n}\} \to 0$.

(ii) If $d > \Delta$, the same reasoning gives $\mathrm{P}\{\widehat{\mathbb{D}}_{m,n} > \Delta + q^*_{1-\alpha}(\mathbb{S})\widehat{\mathbb{H}}_{m,n}\} \to 1$.

(iii) If $d = \Delta$ and $\tau_D^2 > 0$, then

$$\mathrm{P}\Big\{\widehat{\mathbb{D}}_{m,n} > \Delta + q^*_{1-\alpha}(\mathbb{S})\widehat{\mathbb{H}}_{m,n}\Big\} = \mathrm{P}\Big\{\sqrt{N}(\widehat{\mathbb{D}}_{m,n} - d) > q^*_{1-\alpha}(\mathbb{S})\,\sqrt{N}\widehat{\mathbb{H}}_{m,n}\Big\} \to \mathrm{P}\Big\{\mathbb{S} > q^*_{1-\alpha}(\mathbb{S})\Big\} = \alpha,$$

which completes the proof. □

B.5 Proof of Theorem 4

Proof. Write

$$d \overset{\text{def}}{=} \int_T \delta^2(t)\,dt.$$

For $\theta \in (\varepsilon, 1-\varepsilon)$ and $\lambda \in [0,1]$, define the integrated squared partial-sum contrast

$$\mathbb{D}_N^{\mathrm{cp}}(\lambda, \theta) \overset{\text{def}}{=} \int_T \Big(D_N^{\mathrm{cp}}(t, \lambda, \theta)\Big)^2\,dt,$$

so that, by construction,

$$\widehat{\mathbb{D}}_N^{\mathrm{cp}} = \mathbb{D}_N^{\mathrm{cp}}(1, \widehat{\theta}), \qquad \widehat{\mathbb{H}}_N^{\mathrm{cp}} = \max_{\lambda\in[0,1]} \Big\{\mathbb{D}_N^{\mathrm{cp}}(\lambda, \widehat{\theta}) - \lambda^2 \mathbb{D}_N^{\mathrm{cp}}(1, \widehat{\theta})\Big\} - \min_{\lambda\in[0,1]} \Big\{\mathbb{D}_N^{\mathrm{cp}}(\lambda, \widehat{\theta}) - \lambda^2 \mathbb{D}_N^{\mathrm{cp}}(1, \widehat{\theta})\Big\}.$$

Define the process

$$Z_N(\lambda, \theta_0) \overset{\text{def}}{=} \sqrt{N}\Big(\mathbb{D}_N^{\mathrm{cp}}(\lambda, \theta_0) - \lambda^2 d\Big), \qquad \lambda \in [0,1].$$

Under the model (18) and the conditions imposed below it (Bernoulli-shift type dependence with the moment and l-approximation conditions), a functional invariance principle applies on each segment. Proceeding as in the standard SN proof for relevant hypotheses in the change-point mean problem, one obtains in $\ell^\infty([0,1])$ that

$$\big\{Z_N(\lambda, \theta_0)\big\}_{\lambda\in[0,1]} \xrightarrow{\mathcal{L}} \big\{\lambda\,\tau_{\delta,\theta_0}\,\mathbb{B}(\lambda)\big\}_{\lambda\in[0,1]}, \tag{B.12}$$

where $\mathbb{B}$ is standard Brownian motion on $[0,1]$ and τ^2_{δ,θ_0} is defined in the display preceding Theorem 4. (Technically, this step is obtained by linearizing $\mathbb{D}_N^{\mathrm{cp}}(\lambda, \theta_0)$ around $\lambda^2 d$ and applying the WIP to the resulting inner products with δ.)

By (20) we have $\widehat{\theta} = \theta_0 + o_{\mathrm{p}}(N^{-1/2})$. Using this rate and the fact that the number of observations that are "misallocated" across the two segments when replacing θ_0 by $\widehat{\theta}$ is of order $N|\widehat{\theta} - \theta_0| = o_{\mathrm{p}}(\sqrt{N})$, one can show the uniform negligibility

$$\sup_{\lambda\in[0,1]} \Big|Z_N(\lambda, \widehat{\theta}) - Z_N(\lambda, \theta_0)\Big| = o_{\mathrm{p}}(1), \tag{B.13}$$

where $Z_N(\lambda,\widehat{\theta}) \stackrel{\text{def}}{=} \sqrt{N}\big(\mathbb{D}_N^{\text{cp}}(\lambda,\widehat{\theta}) - \lambda^2 d\big)$. Combining (B.12) and (B.13) yields

$$\big\{Z_N(\lambda,\widehat{\theta})\big\}_{\lambda\in[0,1]} \;\xrightarrow{\mathcal{L}}\; \big\{\lambda\,\tau_{\delta,\theta_0}\,\mathbb{B}(\lambda)\big\}_{\lambda\in[0,1]} \qquad \text{in } \ell^\infty([0,1]). \tag{B.14}$$

Note that

$$\sqrt{N}\big(\widehat{\mathbb{D}}_N^{\text{cp}} - d\big) = Z_N(1,\widehat{\theta}).$$

Moreover, by the algebra

$$\sqrt{N}\Big(\mathbb{D}_N^{\text{cp}}(\lambda,\widehat{\theta}) - \lambda^2\mathbb{D}_N^{\text{cp}}(1,\widehat{\theta})\Big) = Z_N(\lambda,\widehat{\theta}) - \lambda^2 Z_N(1,\widehat{\theta}),$$

we can rewrite the scaled adjusted-range normalizer as the continuous functional

$$\sqrt{N}\,\widehat{\mathbb{H}}_N^{\text{cp}} = \max_{\lambda\in[0,1]}\Big\{Z_N(\lambda,\widehat{\theta}) - \lambda^2 Z_N(1,\widehat{\theta})\Big\} - \min_{\lambda\in[0,1]}\Big\{Z_N(\lambda,\widehat{\theta}) - \lambda^2 Z_N(1,\widehat{\theta})\Big\}.$$

Define for a function $z \in \ell^\infty([0,1])$ the mapping

$$\mathcal{T}(z) \;\stackrel{\text{def}}{=}\; \left(z(1),\; \max_{\lambda\in[0,1]}\{z(\lambda) - \lambda^2 z(1)\} - \min_{\lambda\in[0,1]}\{z(\lambda) - \lambda^2 z(1)\}\right).$$

Whenever the second component is strictly positive, $\mathcal{T}$ is continuous with respect to the supremum norm. If $\tau^2_{\delta,\theta_0} > 0$, then the limit process $z_\infty(\lambda) = \lambda\tau_{\delta,\theta_0}\mathbb{B}(\lambda)$ has continuous sample paths and is not almost surely constant, hence

$$\max_{\lambda\in[0,1]}\{z_\infty(\lambda) - \lambda^2 z_\infty(1)\} - \min_{\lambda\in[0,1]}\{z_\infty(\lambda) - \lambda^2 z_\infty(1)\} > 0 \quad \text{a.s.}$$

Therefore, by the continuous mapping theorem applied to (B.14),

$$\Big(\sqrt{N}(\widehat{\mathbb{D}}_N^{\text{cp}} - d),\; \sqrt{N}\widehat{\mathbb{H}}_N^{\text{cp}}\Big) = \mathcal{T}\Big(\{Z_N(\lambda,\widehat{\theta})\}_{\lambda\in[0,1]}\Big) \;\xrightarrow{\mathcal{L}}\; \Big(\tau_{\delta,\theta_0}\mathbb{B}(1),\; \tau_{\delta,\theta_0}\mathcal{R}\Big),$$

where

$$\mathcal{R} = \max_{\lambda\in[0,1]}\{\lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)\} - \min_{\lambda\in[0,1]}\{\lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)\}.$$

In particular,

$$\frac{\widehat{\mathbb{D}}_N^{\text{cp}} - d}{\widehat{\mathbb{H}}_N^{\text{cp}}} \;\xrightarrow{\mathcal{L}}\; \frac{\mathbb{B}(1)}{\mathcal{R}} = \mathbb{S},$$

where $\mathbb{S}$ is the pivotal statistic used in the calibration of (22).

Let $q^*_{1-\alpha}(\mathbb{S})$ denote the $(1-\alpha)$-quantile of $\mathbb{S}$.

(i) If $d < \Delta$: From the joint convergence above we have $\widehat{\mathbb{D}}_N^{\mathrm{cp}} = d + o_{\mathrm{p}}(1)$ and $\widehat{\mathbb{H}}_N^{\mathrm{cp}} = o_{\mathrm{p}}(1)$, hence $\Delta + q_{1-\alpha}^*(\mathbb{S})\widehat{\mathbb{H}}_N^{\mathrm{cp}} = \Delta + o_{\mathrm{p}}(1)$. Therefore

$$\mathrm{P}\Big\{\widehat{\mathbb{D}}_N^{\mathrm{cp}} > \Delta + q_{1-\alpha}^*(\mathbb{S})\widehat{\mathbb{H}}_N^{\mathrm{cp}}\Big\} \;\longrightarrow\; 0.$$

(ii) If $d = \Delta$ and $\tau_{\delta,\theta_0}^2 > 0$: Then

$$\mathrm{P}\Big\{\widehat{\mathbb{D}}_N^{\mathrm{cp}} > \Delta + q_{1-\alpha}^*(\mathbb{S})\widehat{\mathbb{H}}_N^{\mathrm{cp}}\Big\} = \mathrm{P}\left\{\frac{\widehat{\mathbb{D}}_N^{\mathrm{cp}} - d}{\widehat{\mathbb{H}}_N^{\mathrm{cp}}} > q_{1-\alpha}^*(\mathbb{S})\right\} \;\longrightarrow\; \mathrm{P}\{\mathbb{S} > q_{1-\alpha}^*(\mathbb{S})\} = \alpha.$$

(iii) If $d > \Delta$: Again $\widehat{\mathbb{D}}_N^{\mathrm{cp}} = d + o_{\mathrm{p}}(1)$ and $\Delta + q_{1-\alpha}^*(\mathbb{S})\widehat{\mathbb{H}}_N^{\mathrm{cp}} = \Delta + o_{\mathrm{p}}(1)$, so the rejection probability converges to 1.

This proves Theorem 4. □

B.6 Proof of Theorem 5

Proof. For $x \in L^2(T)$ write $(x \otimes x)(s,t) \stackrel{\mathrm{def}}{=} x(s)x(t)$, which we regard as an element of $L^2(T^2)$ with inner product

$$\langle A, B\rangle_{T^2} \stackrel{\mathrm{def}}{=} \int_T \int_T A(s,t)B(s,t)\,ds\,dt, \qquad \|A\|_{T^2}^2 \stackrel{\mathrm{def}}{=} \langle A, A\rangle_{T^2}.$$

Define the tensor-product observations

$$Z_j^X \stackrel{\mathrm{def}}{=} (X_j - \mu_X) \otimes (X_j - \mu_X), \qquad Z_j^Y \stackrel{\mathrm{def}}{=} (Y_j - \mu_Y) \otimes (Y_j - \mu_Y),$$

with means $C^X \stackrel{\mathrm{def}}{=} \mathsf{E}[Z_1^X]$ and $C^Y \stackrel{\mathrm{def}}{=} \mathsf{E}[Z_1^Y]$ (the covariance kernels), and set $D_C \stackrel{\mathrm{def}}{=} C^X - C^Y \in L^2(T^2)$ and $d_C \stackrel{\mathrm{def}}{=} \|D_C\|_{T^2}^2$.

Let $\tilde{Z}_j^X \stackrel{\mathrm{def}}{=} Z_j^X - C^X$ and $\tilde{Z}_j^Y \stackrel{\mathrm{def}}{=} Z_j^Y - C^Y$ denote centered sequences. For $\lambda \in [0,1]$ define the (idealized) partial-sum difference process

$$\tilde{D}_{m,n}(s,t,\lambda) \stackrel{\mathrm{def}}{=} \frac{1}{m-1}\sum_{j=1}^{\lfloor m\lambda\rfloor} Z_j^X(s,t) - \frac{1}{n-1}\sum_{j=1}^{\lfloor n\lambda\rfloor} Z_j^Y(s,t).$$

Then $\mathsf{E}[\tilde{D}_{m,n}(\cdot,\cdot,\lambda)] = \lambda D_C + \mathcal{O}\{(m \wedge n)^{-1}\}$ in $L^2(T^2)$.

Let $D_{m,n}(s,t,\lambda)$ be the empirical partial sum used in the definition of $\widehat{\mathbb{D}}_{m,n}^C$ and $\widehat{\mathbb{H}}_{m,n}^C$, i.e., the version using sample means within the first $\lfloor m\lambda\rfloor$ and $\lfloor n\lambda\rfloor$ curves. Under the moment and approximation

assumptions (Assumptions 1–2 and 7–8), the usual Cauchy–Schwarz calculations show that

$$\sup_{\lambda\in[0,1]} \big\|D_{m,n}(\cdot,\cdot,\lambda) - \tilde{D}_{m,n}(\cdot,\cdot,\lambda)\big\|_{T^2} = o_{\mathrm{p}}\big((m+n)^{-1/2}\big).$$

Consequently, if we define

$$\tilde{\mathbb{D}}^C_{m,n} \stackrel{\text{def}}{=} \int_T\int_T \tilde{D}^2_{m,n}(s,t,1)\,ds\,dt, \qquad \tilde{\mathbb{H}}^C_{m,n} \stackrel{\text{def}}{=} \max_{\lambda\in[0,1]}\Big\{\tilde{\mathbb{D}}^C_{m,n}(\lambda)-\lambda^2\tilde{\mathbb{D}}^C_{m,n}(1)\Big\} - \min_{\lambda\in[0,1]}\Big\{\tilde{\mathbb{D}}^C_{m,n}(\lambda)-\lambda^2\tilde{\mathbb{D}}^C_{m,n}(1)\Big\},$$

with $\tilde{\mathbb{D}}^C_{m,n}(\lambda) \stackrel{\text{def}}{=} \int_T\int_T \tilde{D}^2_{m,n}(s,t,\lambda)\,ds\,dt$, then

$$\sqrt{m+n}\big(\widehat{\mathbb{D}}^C_{m,n} - \tilde{\mathbb{D}}^C_{m,n}\big) = o_{\mathrm{p}}(1), \qquad \sqrt{m+n}\big(\widehat{\mathbb{H}}^C_{m,n} - \tilde{\mathbb{H}}^C_{m,n}\big) = o_{\mathrm{p}}(1).$$

Hence it suffices to establish the theorem with $(\tilde{\mathbb{D}}^C_{m,n}, \tilde{\mathbb{H}}^C_{m,n})$ in place of $(\widehat{\mathbb{D}}^C_{m,n}, \widehat{\mathbb{H}}^C_{m,n})$.

By Assumption 5, $m,n\to\infty$ with $m/(m+n)\to\rho\in(0,1)$. The strengthened Assumptions 7–8 ensure that the tensor-product sequences $\{\tilde{Z}^X_j\}$ and $\{\tilde{Z}^Y_j\}$ satisfy a functional invariance principle in $L^2(T^2)$ (see the tensor-product extension in the supplement of Dette et al. (2020), Eqs. (B.43)–(B.46)). In particular, there exist centered Gaussian elements $\Gamma_X(\cdot,\cdot,\lambda)$ and $\Gamma_Y(\cdot,\cdot,\lambda)$ in $L^2(T^2)$ such that

$$\Big\{\sqrt{m}\,\frac{1}{m-1}\sum_{j=1}^{\lfloor m\lambda\rfloor}\tilde{Z}^X_j\Big\}_{\lambda\in[0,1]} \Rightarrow \{\Gamma_X(\cdot,\cdot,\lambda)\}_{\lambda\in[0,1]}, \qquad \Big\{\sqrt{n}\,\frac{1}{n-1}\sum_{j=1}^{\lfloor n\lambda\rfloor}\tilde{Z}^Y_j\Big\}_{\lambda\in[0,1]} \Rightarrow \{\Gamma_Y(\cdot,\cdot,\lambda)\}_{\lambda\in[0,1]},$$

and the two limits are independent. Therefore, by Slutsky's lemma,

$$\mathbb{G}_{m,n}(\lambda) \stackrel{\text{def}}{=} \sqrt{m+n}\big(\tilde{D}_{m,n}(\cdot,\cdot,\lambda) - \lambda D_C\big) \Rightarrow \Gamma_C(\cdot,\cdot,\lambda) \stackrel{\text{def}}{=} \frac{1}{\sqrt{\rho}}\,\Gamma_X(\cdot,\cdot,\lambda) - \frac{1}{\sqrt{1-\rho}}\,\Gamma_Y(\cdot,\cdot,\lambda)$$

in $\ell^\infty([0,1];L^2(T^2))$.

Now use the identity $\|a+b\|^2_{T^2} - \|a\|^2_{T^2} = 2\langle a,b\rangle_{T^2} + \|b\|^2_{T^2}$ with $a=\lambda D_C$ and $b=\tilde{D}_{m,n}(\cdot,\cdot,\lambda)-\lambda D_C$ to get

$$\tilde{\mathbb{D}}^C_{m,n}(\lambda) - \lambda^2 d_C = 2\lambda\Big\langle D_C, \tilde{D}_{m,n}(\cdot,\cdot,\lambda) - \lambda D_C\Big\rangle_{T^2} + \Big\|\tilde{D}_{m,n}(\cdot,\cdot,\lambda) - \lambda D_C\Big\|^2_{T^2}.$$

Since $\sup_{\lambda\in[0,1]}\|\tilde{D}_{m,n}(\cdot,\cdot,\lambda)-\lambda D_C\|_{T^2} = \mathscr{O}_{\mathrm{p}}((m+n)^{-1/2})$, the last term is $o_{\mathrm{p}}((m+n)^{-1/2})$ uniformly in λ, and thus

$$\sqrt{m+n}\big(\tilde{\mathbb{D}}^C_{m,n}(\lambda) - \lambda^2 d_C\big) = 2\lambda\Big\langle D_C, \mathbb{G}_{m,n}(\lambda)\Big\rangle_{T^2} + o_{\mathrm{p}}(1), \qquad \text{uniformly in } \lambda\in[0,1]. \tag{B.15}$$

By the continuous mapping theorem, $2\langle D_C, \Gamma_C(\lambda)\rangle_{T^2}$ is a centered Gaussian process with covariance

$(\lambda \wedge \lambda')\,\tau_{D_C}^2$, where $\tau_{D_C}^2$ is exactly the constant defined in the text,

$$\tau_{D_C}^2 = 4\int_T\int_T\int_T\int_T D_C(s,t)D_C(s',t')\Big[\frac{1}{\rho}C_X\{(s,t),(s',t')\} + \frac{1}{1-\rho}C_Y\{(s,t),(s',t')\}\Big]\,ds\,dt\,ds'\,dt'.$$

Hence we can represent $2\langle D_C, \Gamma_C(\lambda)\rangle_{T^2} = \tau_{D_C}\,\mathbb{B}(\lambda)$ for a standard Brownian motion $\mathbb{B}$ on $[0,1]$, and combining this with (B.15) yields the weak convergence in $\ell^\infty([0,1])$,

$$\Big\{\sqrt{m+n}\big(\tilde{\mathbb{D}}_{m,n}^C(\lambda) - \lambda^2 d_C\big)\Big\}_{\lambda\in[0,1]} \Rightarrow \{\tau_{D_C}\lambda\mathbb{B}(\lambda)\}_{\lambda\in[0,1]}.$$

In particular,

$$\sqrt{m+n}\big(\tilde{\mathbb{D}}_{m,n}^C - d_C\big) = \sqrt{m+n}\big(\tilde{\mathbb{D}}_{m,n}^C(1) - d_C\big) \Rightarrow \tau_{D_C}\,\mathbb{B}(1). \tag{B.16}$$

Define

$$U(\lambda) \stackrel{\text{def}}{=} \lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1), \qquad \mathcal{R} \stackrel{\text{def}}{=} \max_{\lambda\in[0,1]} U(\lambda) - \min_{\lambda\in[0,1]} U(\lambda).$$

Using the decomposition

$$\tilde{\mathbb{D}}_{m,n}^C(\lambda) - \lambda^2\tilde{\mathbb{D}}_{m,n}^C(1) = \big(\tilde{\mathbb{D}}_{m,n}^C(\lambda) - \lambda^2 d_C\big) - \lambda^2\big(\tilde{\mathbb{D}}_{m,n}^C(1) - d_C\big),$$

the joint weak convergence established above and the continuous mapping theorem imply

$$\sqrt{m+n}\,\tilde{\mathbb{H}}_{m,n}^C \Rightarrow \tau_{D_C}\Big(\max_{\lambda\in[0,1]} U(\lambda) - \min_{\lambda\in[0,1]} U(\lambda)\Big) = \tau_{D_C}\,\mathcal{R}.$$

Together with (B.16), this yields

$$\Big(\sqrt{m+n}\big(\tilde{\mathbb{D}}_{m,n}^C - d_C\big),\ \sqrt{m+n}\tilde{\mathbb{H}}_{m,n}^C\Big) \Rightarrow \big(\tau_{D_C}\mathbb{B}(1),\ \tau_{D_C}\mathcal{R}\big),$$

and therefore (since $\mathcal{R} > 0$ almost surely) the ratio converges to the pivotal limit

$$\frac{\tilde{\mathbb{D}}_{m,n}^C - d_C}{\tilde{\mathbb{H}}_{m,n}^C} \Rightarrow \frac{\mathbb{B}(1)}{\mathcal{R}} = \mathbb{S}.$$

By Step 2, the same convergence holds with $(\widehat{\mathbb{D}}_{m,n}^C, \widehat{\mathbb{H}}_{m,n}^C)$.

If $d_C < \Delta$, then $\widehat{\mathbb{D}}_{m,n}^C \to d_C$ in probability, while $\widehat{\mathbb{H}}_{m,n}^C = o_{\mathrm{p}}(1)$, so $\mathrm{P}\{\widehat{\mathbb{D}}_{m,n}^C > \Delta + q_{1-\alpha}^*(\mathbb{S})\widehat{\mathbb{H}}_{m,n}^C\} \to 0$. If $d_C > \Delta$, the same reasoning gives convergence to 1.

If $d_C = \Delta$ and $\tau_{D_C}^2 > 0$, then

$$\mathrm{P}\{\widehat{\mathbb{D}}_{m,n}^C > \Delta + q_{1-\alpha}^*(\mathbb{S})\widehat{\mathbb{H}}_{m,n}^C\} = \mathrm{P}\left\{\frac{\widehat{\mathbb{D}}_{m,n}^C - d_C}{\widehat{\mathbb{H}}_{m,n}^C} > q_{1-\alpha}^*(\mathbb{S})\right\} + o(1) \to \mathrm{P}\{\mathbb{S} > q_{1-\alpha}^*(\mathbb{S})\} = \alpha,$$

which completes the proof. □

B.7 Proof of Theorem 6

Proof. Write $T^2 \stackrel{\text{def}}{=} T \times T$ and endow $L^2(T^2)$ with the usual inner product $\langle f, g\rangle_2 \stackrel{\text{def}}{=} \int_T \int_T f(s,t)g(s,t)\,ds\,dt$ and norm $\|f\|_2^2 = \langle f, f\rangle_2$. Let $D_C \stackrel{\text{def}}{=} C^X - C^Y$ and $d_C = \|D_C\|_2^2$.

Let $\eta_j^X \stackrel{\text{def}}{=} X_j - \mu_X$ and $\eta_j^Y \stackrel{\text{def}}{=} Y_j - \mu_Y$ and define the centered tensor-product processes

$$\xi_j^X(s,t) \stackrel{\text{def}}{=} \eta_j^X(s)\eta_j^X(t) - C^X(s,t), \qquad \xi_j^Y(s,t) \stackrel{\text{def}}{=} \eta_j^Y(s)\eta_j^Y(t) - C^Y(s,t),$$

viewed as random elements in $L^2(T^2)$. Following the same algebra as in the quadratic-SN covariance-operator case (one expands the sample-centering terms and uses moment bounds), one can replace $D_{m,n}(s,t,\lambda)$ by the simplified process

$$\widetilde{D}_{m,n}(s,t,\lambda) \stackrel{\text{def}}{=} \lambda D_C(s,t) + \frac{1}{m}\sum_{j=1}^{\lfloor m\lambda\rfloor} \xi_j^X(s,t) - \frac{1}{n}\sum_{j=1}^{\lfloor n\lambda\rfloor} \xi_j^Y(s,t),$$

in the sense that for any fixed $\varepsilon \in (0,1)$,

$$\sup_{\lambda\in[\varepsilon,1]} \big\|D_{m,n}(\cdot,\cdot,\lambda) - \widetilde{D}_{m,n}(\cdot,\cdot,\lambda)\big\|_2 = o_{\mathrm{p}}\big((m+n)^{-1/2}\big). \tag{B.17}$$

(For the corresponding argument in the quadratic-SN setting see the reduction step in the covariance-operator proofs, which relies on moment bounds for $\|\eta_j\|$ and shows that centering by partial-sample means is asymptotically negligible.)

Consequently, defining

$$\widetilde{\mathbb{D}}_{m,n}^C(\lambda) \stackrel{\text{def}}{=} \|\widetilde{D}_{m,n}(\cdot,\cdot,\lambda)\|_2^2, \qquad \widetilde{\mathbb{D}}_{m,n}^C \stackrel{\text{def}}{=} \widetilde{\mathbb{D}}_{m,n}^C(1),$$

the statistics $(\widehat{\mathbb{D}}_{m,n}^C, \widehat{\mathbb{H}}_{m,n}^C)$ may be replaced by their analogues built from $\widetilde{\mathbb{D}}_{m,n}^C(\lambda)$, up to $o_{\mathrm{p}}((m+n)^{-1/2})$ errors (uniformly over λ on $[\varepsilon,1]$).

Let $N \stackrel{\text{def}}{=} m+n$ and $\rho_N \stackrel{\text{def}}{=} m/N \to \rho \in (0,1)$. Define the process

$$Z_N(\lambda) \stackrel{\text{def}}{=} \sqrt{N}\Big(\widetilde{\mathbb{D}}^C_{m,n}(\lambda) - \lambda^2 d_C\Big), \qquad \lambda \in [0,1].$$

Expanding $\|\lambda D_C + R_{m,n}(\lambda)\|_2^2$ with $R_{m,n}(\lambda) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{j\le \lfloor m\lambda\rfloor} \xi_j^X - \frac{1}{n}\sum_{j\le\lfloor n\lambda\rfloor}\xi_j^Y$ gives

$$\widetilde{\mathbb{D}}^C_{m,n}(\lambda) - \lambda^2 d_C = 2\lambda\langle D_C, R_{m,n}(\lambda)\rangle_2 + \|R_{m,n}(\lambda)\|_2^2.$$

Under the strengthened moment and l-approximation assumptions (Assumptions 7–8) the remainder term $\sup_{\lambda\in[0,1]}\|R_{m,n}(\lambda)\|_2^2 = o_{\mathrm{p}}(N^{-1/2})$, so that

$$Z_N(\lambda) = 2\lambda\sqrt{N}\,\langle D_C, R_{m,n}(\lambda)\rangle_2 + o_{\mathrm{p}}(1), \quad \text{in } \ell^\infty([0,1]). \tag{B.18}$$

By the tensor-product invariance principle for Bernoulli shifts in $L^2(T^2)$ (applied to $\{\xi_j^X\}$ and $\{\xi_j^Y\}$ separately) and the assumed independence of $\{X_t\}$ and $\{Y_t\}$, the scalar processes

$$\Big\{\sqrt{m}\,\langle D_C, m^{-1}\sum_{j\le\lfloor m\lambda\rfloor}\xi_j^X\rangle_2\Big\}_{\lambda\in[0,1]}, \qquad \Big\{\sqrt{n}\,\langle D_C, n^{-1}\sum_{j\le\lfloor n\lambda\rfloor}\xi_j^Y\rangle_2\Big\}_{\lambda\in[0,1]}$$

converge weakly in $\ell^\infty([0,1])$ to independent scaled Brownian motions. Combining the scaling $\sqrt{N}/m \sim 1/(\sqrt{\rho}\sqrt{N})$ and $\sqrt{N}/n \sim 1/(\sqrt{1-\rho}\sqrt{N})$ yields

$$\{Z_N(\lambda)\}_{\lambda\in[0,1]} \;\Rightarrow\; \{\lambda\tau_{D_C}\,\mathbb{B}(\lambda)\}_{\lambda\in[0,1]},$$

where $\mathbb{B}$ is standard Brownian motion and $\tau^2_{D_C}$ is the constant defined in the text.

Note that $\sqrt{N}(\widetilde{\mathbb{D}}^C_{m,n} - d_C) = Z_N(1) + o_{\mathrm{p}}(1)$. Moreover, by definition,

$$\widetilde{\mathbb{H}}^C_{m,n} = \max_{\lambda\in[0,1]}\{\widetilde{\mathbb{D}}^C_{m,n}(\lambda) - \lambda^2\widetilde{\mathbb{D}}^C_{m,n}(1)\} - \min_{\lambda\in[0,1]}\{\widetilde{\mathbb{D}}^C_{m,n}(\lambda) - \lambda^2\widetilde{\mathbb{D}}^C_{m,n}(1)\}.$$

Using $\widetilde{\mathbb{D}}^C_{m,n}(\lambda) = \lambda^2 d_C + Z_N(\lambda)/\sqrt{N} + o_{\mathrm{p}}(N^{-1/2})$ uniformly in λ gives

$$\sqrt{N}\,\widetilde{\mathbb{H}}^C_{m,n} = \max_{\lambda\in[0,1]}\{Z_N(\lambda) - \lambda^2 Z_N(1)\} - \min_{\lambda\in[0,1]}\{Z_N(\lambda) - \lambda^2 Z_N(1)\} + o_{\mathrm{p}}(1).$$

The map $z \mapsto \big(z(1), \max_\lambda\{z(\lambda)-\lambda^2 z(1)\} - \min_\lambda\{z(\lambda)-\lambda^2 z(1)\}\big)$ is continuous on $C([0,1])$, hence the

continuous mapping theorem yields

$$\Big(\sqrt{N}(\widetilde{\mathbb{D}}^C_{m,n} - d_C),\ \sqrt{N}\widetilde{\mathbb{H}}^C_{m,n}\Big) \ \Rightarrow\ \Big(\tau_{D_C}\mathbb{B}(1),\ \tau_{D_C}\mathcal{R}\Big),$$

where $\mathcal{R} \stackrel{\text{def}}{=} \max_{\lambda\in[0,1]} U(\lambda) - \min_{\lambda\in[0,1]} U(\lambda)$ and $U(\lambda) = \lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)$. Therefore,

$$\frac{\widetilde{\mathbb{D}}^C_{m,n} - d_C}{\widetilde{\mathbb{H}}^C_{m,n}} \ \xrightarrow{\mathcal{L}}\ \mathbb{S} \stackrel{\text{def}}{=} \frac{\mathbb{B}(1)}{\mathcal{R}}.$$

By (B.17) and the preceding reductions, the same limit holds with $(\widehat{\mathbb{D}}^C_{m,n}, \widehat{\mathbb{H}}^C_{m,n})$.

If $d_C < \Delta$, then $\widehat{\mathbb{D}}^C_{m,n} \to d_C$ in probability while $\widehat{\mathbb{H}}^C_{m,n} = \mathscr{O}_{\mathrm{p}}(N^{-1/2})$, hence $\mathrm{P}\{\widehat{\mathbb{D}}^C_{m,n} > \Delta + q^*_{1-\alpha}(\mathbb{S})\widehat{\mathbb{H}}^C_{m,n}\} \to 0$. If $d_C > \Delta$, the same reasoning gives convergence to 1. If $d_C = \Delta$ and $\tau^2_{D_C} > 0$, then

$$\mathrm{P}\Big\{\widehat{\mathbb{D}}^C_{m,n} > \Delta + q^*_{1-\alpha}(\mathbb{S})\widehat{\mathbb{H}}^C_{m,n}\Big\} = \mathrm{P}\Big\{\frac{\sqrt{N}(\widehat{\mathbb{D}}^C_{m,n} - \Delta)}{\sqrt{N}\widehat{\mathbb{H}}^C_{m,n}} > q^*_{1-\alpha}(\mathbb{S})\Big\} \ \to\ \mathrm{P}\{\mathbb{S} > q^*_{1-\alpha}(\mathbb{S})\} = \alpha.$$

□

B.8 Bootstrap validity for quadratic and adjusted-range self-normalized statistics

This appendix provides a formal justification for the optional block-bootstrap refinement for the self-normalized pivots considered in the paper. We state bootstrap validity for the quadratic self-normalized pivot in Theorem B.2 and for the adjusted-range self-normalized pivot in Theorem B.3. Both results rely on the same bootstrap weak invariance principle for the centered partial-sum process; the only difference is whether the self-normalizer is the quadratic functional $f \mapsto \{\int_0^1 f^2(\lambda)\,\nu(d\lambda)\}^{1/2}$ or the range functional $f \mapsto \sup_{\lambda\in[0,1]} f(\lambda) - \inf_{\lambda\in[0,1]} f(\lambda)$.

Throughout, let $\langle f, g\rangle \stackrel{\text{def}}{=} \int_T f(t)g(t)\,dt$ and $\|f\|^2 = \langle f, f\rangle$ for $f, g \in L^2(T)$. Let ν be a probability measure on $(0, 1)$ and recall the quadratic self-normalizer $\widehat{\mathbb{V}}_n$ in (5).

Let $b = b_n$ be a block length such that $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$. Define the centered curves $\widetilde{X}_j(t) \stackrel{\text{def}}{=} X_j(t) - \bar{X}_n(t)$, $j = 1, \ldots, n$. Using circular indexing $X_{n+k} \equiv X_k$, draw i.i.d. starting indices $I_1, \ldots, I_{k_n}$ uniformly on $\{1, \ldots, n\}$, where $k_n \stackrel{\text{def}}{=} \lceil n/b_n \rceil$. Form a bootstrap residual sample by concatenating blocks of length b_n:

$$\widetilde{X}^*_{(r-1)b_n+s} \stackrel{\text{def}}{=} \widetilde{X}_{I_r+s-1}, \qquad s = 1, \ldots, b_n, \quad r = 1, \ldots, k_n,$$

and retain the first n curves $\{\widetilde{X}_j^*\}_{j=1}^n$. Finally, add back the sample mean and set

$$X_j^*(t) \stackrel{\text{def}}{=} \bar{X}_n(t) + \widetilde{X}_j^*(t), \qquad j = 1, \dots, n.$$

Let $\mathrm{P}^*(\cdot)$ and $\mathsf{E}^*(\cdot)$ denote probability and expectation under the bootstrap law conditional on the data. Define the bootstrap partial sums

$$S_n^*(t, \lambda) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=1}^{\lfloor n\lambda \rfloor} X_j^*(t), \qquad \lambda \in [0, 1],$$

and the bootstrap analogues

$$\widehat{\mathbb{I}}_n^*(\lambda) \stackrel{\text{def}}{=} \int_T S_n^{*2}(t, \lambda)\, dt, \qquad \widehat{\mathbb{I}}_n^* \stackrel{\text{def}}{=} \widehat{\mathbb{I}}_n^*(1),$$

$$G_n^*(\lambda) \stackrel{\text{def}}{=} \widehat{\mathbb{I}}_n^*(\lambda) - \lambda^2 \widehat{\mathbb{I}}_n^*(1), \qquad \widehat{\mathbb{V}}_n^* \stackrel{\text{def}}{=} \left[\int_0^1 G_n^{*2}(\lambda)\, \nu(d\lambda) \right]^{1/2}.$$

The bootstrap pivot is

$$T_n^* \stackrel{\text{def}}{=} \frac{\widehat{\mathbb{I}}_n^* - \widehat{\mathbb{I}}_n}{\widehat{\mathbb{V}}_n^*}.$$

Theorem B.2 (Asymptotic validity of the block bootstrap for the quadratic self-normalized pivot)**.** *Assume Assumptions 1–4 and $\tau^2 > 0$. Let $b_n \to \infty$ and $b_n/n \to 0$. Then, as $n \to \infty$,*

$$T_n^* \xrightarrow{\mathcal{L}^*} \mathbb{W} \qquad \text{in probability,} \tag{B.19}$$

where $\mathbb{W}$ is defined in Remark 1 and $\xrightarrow{\mathcal{L}^*}$ denotes weak convergence under the bootstrap law conditional on the data. Consequently, the bootstrap $(1-\alpha)$-quantile

$$q_{1-\alpha}^* \stackrel{\text{def}}{=} \inf\left\{ x : \mathrm{P}^*(T_n^* \le x) \ge 1 - \alpha \right\}$$

satisfies $q_{1-\alpha}^* \to q_{1-\alpha}(\mathbb{W})$ in probability.

Proof. Write $\widetilde{X}_j = X_j - \bar{X}_n$ and note that, by construction, $X_j^* - \bar{X}_n = \widetilde{X}_j^*$ and $\mathsf{E}^*(X_j^*) = \bar{X}_n$. Define the bootstrap "influence" process

$$U_n^*(\lambda) \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor n\lambda \rfloor} \left(X_j^* - \bar{X}_n \right) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor n\lambda \rfloor} \widetilde{X}_j^*, \qquad \lambda \in [0, 1],$$

which takes values in $L^2(T)$.

Under Assumptions 1–4 and the block-length conditions, the (circular) moving-block bootstrap con-

sistently reproduces the weak invariance principle for functional partial sums (see, e.g., Paparoditis and Sapatinas (2016)). In particular, in $C([0,1], L^2(T))$,

$$\big(U_n^*(\lambda)\big)_{\lambda\in[0,1]} \xrightarrow{\mathcal{L}^*} \big(\Gamma(\cdot,\lambda)\big)_{\lambda\in[0,1]} \qquad \text{in probability,} \tag{B.20}$$

where $\Gamma(\cdot,\lambda)$ is the centered Gaussian limit from the functional CLT with covariance $\mathrm{Cov}\{\Gamma(t,\lambda),\Gamma(s,\lambda')\} = (\lambda\wedge\lambda')\,C(s,t)$.

Let $\lambda_n \overset{\text{def}}{=} \lfloor n\lambda\rfloor/n$, so $\sup_{\lambda\in[0,1]}|\lambda_n-\lambda|\le n^{-1}$. Since $S_n^*(\cdot,\lambda) = \lambda_n\bar X_n + n^{-1/2}U_n^*(\lambda)$, we have

$$\begin{aligned}\widehat{\mathbb{I}}_n^*(\lambda) &= \big\|S_n^*(\cdot,\lambda)\big\|^2 = \big\|\lambda_n\bar X_n + n^{-1/2}U_n^*(\lambda)\big\|^2\\ &= \lambda_n^2\|\bar X_n\|^2 + 2\lambda_n\, n^{-1/2}\langle \bar X_n, U_n^*(\lambda)\rangle + n^{-1}\big\|U_n^*(\lambda)\big\|^2.\end{aligned}$$

Recalling $\widehat{\mathbb{I}}_n = \|\bar X_n\|^2$ and multiplying by $\sqrt{n}$ yields the bootstrap analogue of Z_n:

$$Z_n^*(\lambda) \overset{\text{def}}{=} \sqrt{n}\big(\widehat{\mathbb{I}}_n^*(\lambda) - \lambda^2\widehat{\mathbb{I}}_n\big) = 2\lambda_n\,\langle \bar X_n, U_n^*(\lambda)\rangle + R_n^*(\lambda), \tag{B.21}$$

where the remainder is

$$R_n^*(\lambda) = \sqrt{n}(\lambda_n^2-\lambda^2)\|\bar X_n\|^2 + \frac{1}{\sqrt{n}}\|U_n^*(\lambda)\|^2.$$

By (B.20), $\sup_{\lambda\in[0,1]}\|U_n^*(\lambda)\| = O_{\mathrm{p}}^*(1)$ in probability, and hence $\sup_\lambda \frac{1}{\sqrt{n}}\|U_n^*(\lambda)\|^2 = o_{\mathrm{p}}^*(1)$ in probability. Moreover, $\sup_\lambda|\lambda_n^2-\lambda^2|\le 2n^{-1}$ and $\|\bar X_n\| = O_{\mathrm{p}}(1)$, so the first remainder term is $o_{\mathrm{p}}^*(1)$ in probability as well. Therefore,

$$\sup_{\lambda\in[0,1]}\big|Z_n^*(\lambda) - 2\lambda\,\langle \bar X_n, U_n^*(\lambda)\rangle\big| = o_{\mathrm{p}}^*(1) \qquad \text{in probability.} \tag{B.22}$$

Since $\|\bar X_n-\mu\| = o_{\mathrm{p}}(1)$ and $\sup_\lambda\|U_n^*(\lambda)\| = O_{\mathrm{p}}^*(1)$ in probability,

$$\sup_{\lambda\in[0,1]}\big|\langle \bar X_n-\mu, U_n^*(\lambda)\rangle\big| \le \|\bar X_n-\mu\|\sup_\lambda\|U_n^*(\lambda)\| = o_{\mathrm{p}}^*(1) \quad \text{in probability.}$$

Combining with (B.22) gives

$$\big(Z_n^*(\lambda)\big)_{\lambda\in[0,1]} = \big(2\lambda\,\langle \mu, U_n^*(\lambda)\rangle\big)_{\lambda\in[0,1]} + o_{\mathrm{p}}^*(1) \quad \text{in } \ell^\infty([0,1]) \text{ in probability.}$$

The map $x\mapsto\langle\mu,x\rangle$ is a continuous linear functional on $L^2(T)$, so applying it to (B.20) yields

$$\big(\langle \mu, U_n^*(\lambda)\rangle\big)_{\lambda\in[0,1]} \xrightarrow{\mathcal{L}^*} \big(\langle \mu, \Gamma(\cdot,\lambda)\rangle\big)_{\lambda\in[0,1]} \qquad \text{in probability.}$$

The scalar Gaussian process $\langle \mu, \Gamma(\cdot,\lambda)\rangle$ has covariance $(\lambda \wedge \lambda')\,\langle \mu, C\mu\rangle = (\lambda \wedge \lambda')\,\tau^2/4$, hence $\langle \mu, \Gamma(\cdot,\lambda)\rangle \overset{d}{=} \frac{\tau}{2}\mathbb{B}(\lambda)$. Therefore,

$$\big(Z_n^*(\lambda)\big)_{\lambda\in[0,1]} \xrightarrow{\mathcal{L}^*} \big(\lambda\,\tau\,\mathbb{B}(\lambda)\big)_{\lambda\in[0,1]} \qquad \text{in probability.} \tag{B.23}$$

Taking $\lambda = 1$ gives $\sqrt{n}(\widehat{\mathbb{I}}_n^* - \widehat{\mathbb{I}}_n) = Z_n^*(1) \xrightarrow{\mathcal{L}^*} \tau\mathbb{B}(1)$ in probability.

Next, observe that

$$\sqrt{n}G_n^*(\lambda) = \sqrt{n}\big(\widehat{\mathbb{I}}_n^*(\lambda) - \lambda^2\widehat{\mathbb{I}}_n^*(1)\big) = Z_n^*(\lambda) - \lambda^2 Z_n^*(1),$$

and hence, by (B.23) and the continuous mapping theorem in $\ell^\infty([0,1])$,

$$\big(\sqrt{n}G_n^*(\lambda)\big)_{\lambda\in[0,1]} \xrightarrow{\mathcal{L}^*} \big(\tau\,U(\lambda)\big)_{\lambda\in[0,1]} \qquad \text{in probability,}$$

where $U(\lambda) \overset{\text{def}}{=} \lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)$. Since $f \mapsto \big[\int_0^1 f^2(\lambda)\,\nu(d\lambda)\big]^{1/2}$ is continuous on $\ell^\infty([0,1])$, another application of the continuous mapping theorem yields

$$\sqrt{n}\,\widehat{\mathbb{V}}_n^* \xrightarrow{\mathcal{L}^*} \tau\bigg[\int_0^1 U^2(\lambda)\,\nu(d\lambda)\bigg]^{1/2} \qquad \text{in probability.}$$

Combining numerator and denominator limits and using $U^2(\lambda) = \lambda^2\{\mathbb{B}(\lambda) - \lambda\mathbb{B}(1)\}^2$ gives (B.19).

Finally, the distribution function of $\mathbb{W}$ is continuous (the denominator is strictly positive a.s.), and hence the convergence of $q_{1-\alpha}^*$ follows from standard quantile-consistency arguments for bootstrap weak convergence. □

Theorem B.3 (Asymptotic validity of the block bootstrap for the adjusted-range self-normalized pivot)**.** *In addition to the bootstrap quantities defined above, set*

$$\widehat{\mathbb{H}}_n^* \overset{\text{def}}{=} \max_{\lambda\in[0,1]} G_n^*(\lambda) - \min_{\lambda\in[0,1]} G_n^*(\lambda), \qquad T_{n,\mathrm{AR}}^* \overset{\text{def}}{=} \frac{\widehat{\mathbb{I}}_n^* - \widehat{\mathbb{I}}_n}{\widehat{\mathbb{H}}_n^*}.$$

Under the conditions of Theorem B.2,

$$T_{n,\mathrm{AR}}^* \xrightarrow{\mathcal{L}^*} \mathbb{S} \qquad \textit{in probability,}$$

where $\mathbb{S}$ *is defined in* (8)*. Consequently, the bootstrap* $(1-\alpha)$*-quantile*

$$q_{1-\alpha}^{\mathrm{boot}}(\mathbb{S}) \overset{\text{def}}{=} \inf\big\{x : \mathrm{P}^*(T_{n,\mathrm{AR}}^* \le x) \ge 1-\alpha\big\}$$

satisfies $q_{1-\alpha}^{\mathrm{boot}}(\mathbb{S}) \to q_{1-\alpha}^*(\mathbb{S})$ *in probability.*

Proof. From the proof of Theorem B.2, we already have $\big(\sqrt{n}G_n^*(\lambda)\big)_{\lambda\in[0,1]} \xrightarrow{\mathcal{L}^*} \big(\tau U(\lambda)\big)_{\lambda\in[0,1]}$ in probability, where $U(\lambda) = \lambda\mathbb{B}(\lambda) - \lambda^2\mathbb{B}(1)$. Since the range functional $R(f) = \sup_{\lambda\in[0,1]} f(\lambda) - \inf_{\lambda\in[0,1]} f(\lambda)$ is continuous on $\ell^\infty([0,1])$, the continuous mapping theorem yields

$$\sqrt{n}\,\widehat{\mathbb{H}}_n^* = R\big(\sqrt{n}G_n^*\big) \xrightarrow{\mathcal{L}^*} \tau\Big\{\max_{\lambda\in[0,1]} U(\lambda) - \min_{\lambda\in[0,1]} U(\lambda)\Big\} \quad \text{in probability.}$$

Moreover, $\sqrt{n}(\widehat{\mathbb{I}}_n^* - \widehat{\mathbb{I}}_n) = Z_n^*(1) \xrightarrow{\mathcal{L}^*} \tau\mathbb{B}(1)$ in probability. Combining numerator and denominator limits gives $T_{n,\mathrm{AR}}^* \xrightarrow{\mathcal{L}^*} \mathbb{S}$, and quantile consistency follows since the denominator is strictly positive a.s. □

C Complete rolling-window monitor output (BTC IV)

The main text reports only the rolling windows flagged as relevant breaks (Table 2). For completeness, this appendix reproduces the full rolling-window output.

Table C.1: Rolling-window relevant breaks ($W = 30$, $\varepsilon = 0.10$, $\alpha = 0.10$, $\Delta_{\text{rms}} = 1$ vol point). "RMS shift" is $100\sqrt{\widehat{\mathbb{D}}_t}$ (vol points). "RMS boundary$_j$" is $100\sqrt{\widehat{\Delta}^{\star}_{t,j}}$ (vol points), for $j \in \{\text{AR}, Q\}$.

Window	Start	End	$\widehat{t}_{cp}$	Dir.	RMS shift	RMS bnd.$_{AR}$	RMS bnd.$_{Q}$	Reject$_{AR}$	Reject$_{Q}$	Decision
1	20250801	20251002	20250903	Down	1.46	0.00	0.00	0	0	No relevant break
2	20250802	20251003	20250903	Down	1.33	0.00	0.00	0	0	No relevant break
3	20250803	20251004	20250824	Up	1.34	0.00	0.00	0	0	No relevant break
4	20250804	20251005	20250824	Up	1.63	0.00	0.00	0	0	No relevant break
5	20250805	20251006	20250824	Up	1.84	0.00	0.00	0	0	No relevant break
6	20250806	20251007	20250824	Up	2.08	0.00	0.00	0	0	No relevant break
7	20250820	20251008	20250824	Up	2.26	0.00	0.00	0	0	No relevant break
8	20250821	20251121	20251007	Up	8.49	0.00	0.00	0	0	No relevant break
9	20250822	20251122	20251008	Up	15.30	0.00	0.00	0	0	No relevant break
10	20250823	20251123	20251008	Up	15.43	0.00	0.00	0	0	No relevant break
11	20250824	20251124	20251008	Up	14.94	0.00	0.00	0	0	No relevant break
12	20250825	20251127	20251008	Up	13.73	0.00	0.00	0	0	No relevant break
13	20250826	20251128	20251008	Up	12.98	0.00	0.00	0	0	No relevant break
14	20250827	20251129	20251008	Up	12.51	0.00	0.00	0	0	No relevant break
15	20250828	20251130	20251008	Up	12.16	0.00	0.00	0	0	No relevant break
16	20250829	20251201	20251008	Up	12.25	0.00	2.82	0	1	Relevant break
17	20250830	20251202	20251008	Up	12.17	0.00	0.00	0	0	No relevant break
18	20250831	20251203	20251008	Up	11.98	0.00	0.00	0	0	No relevant break
19	20250901	20251224	20251008	Up	11.36	0.00	0.00	0	0	No relevant break
20	20250902	20251225	20251008	Up	10.83	0.00	0.00	0	0	No relevant break
21	20250903	20251226	20251008	Up	10.37	0.00	0.00	0	0	No relevant break
22	20250924	20251227	20251008	Up	9.95	2.14	0.00	1	0	Relevant break
23	20250925	20251228	20251008	Up	9.40	0.00	0.00	0	0	No relevant break
24	20250926	20251229	20251008	Up	9.05	0.00	0.00	0	0	No relevant break
25	20250927	20251230	20251008	Up	8.64	0.00	0.00	0	0	No relevant break
26	20250928	20251231	20251008	Up	8.31	0.00	0.00	0	0	No relevant break
27	20250929	20260101	20251008	Up	7.87	0.00	0.00	0	0	No relevant break
28	20250930	20260102	20251008	Up	7.52	0.00	0.00	0	0	No relevant break
29	20251001	20260103	20251008	Up	7.19	0.00	0.00	0	0	No relevant break
30	20251002	20260104	20251008	Up	6.86	0.00	0.00	0	0	No relevant break
31	20251003	20260105	20251008	Up	6.65	0.00	0.00	0	0	No relevant break
32	20251004	20260106	20251203	Down	4.90	0.00	0.00	0	0	No relevant break
33	20251005	20260107	20251203	Down	5.27	4.28	4.35	1	1	Relevant break
34	20251006	20260121	20251203	Down	5.93	0.00	0.00	0	0	No relevant break
35	20251007	20260122	20251203	Down	6.82	0.00	0.00	0	0	No relevant break
36	20251008	20260123	20251203	Down	7.79	0.00	0.00	0	0	No relevant break
37	20251121	20260124	20251203	Down	8.84	0.00	0.00	0	0	No relevant break
38	20251122	20260125	20251203	Down	8.71	0.00	0.00	0	0	No relevant break
39	20251123	20260126	20251203	Down	8.43	0.00	0.00	0	0	No relevant break
40	20251124	20260127	20251203	Down	8.05	0.00	0.00	0	0	No relevant break
41	20251127	20260128	20251203	Down	7.76	0.00	0.00	0	0	No relevant break
42	20251128	20260129	20251203	Down	7.95	0.00	0.00	0	0	No relevant break
43	20251129	20260130	20251203	Down	8.08	0.00	0.00	0	0	No relevant break

Notes. Dates are in YYYYMMDD format. "Dir." is the direction of the estimated volatility shift (Up/Down). "RMS shift" equals $100\sqrt{\widehat{\mathbb{D}}_t}$ (vol points). "RMS bnd.$_j$" denotes $100\sqrt{\widehat{\Delta}^{\star}_{t,j}}$ (vol points), where $j \in \{AR, Q\}$ indexes the two monitor variants; bounds are reported as 0.00 when no positive relevance threshold is rejected in that window. "Reject" indicators are reported as 1/0.

D Additional empirical analysis

We revisit the two datasets analyzed in Dette et al. (2020). This enables an apples-to-apples comparison: any difference in the inferred relevant change size is attributable to the choice of method rather than to data handling.

D.1 Two-sample test

The first dataset comprises annual temperature curves from Cape Otway (1865–2011) in southern Australia and Sydney (1859–2011) on the eastern coast. These stations are approximately 1000 km apart, so differences in temperature profiles are expected. The two-sample test quantifies the extent of these differences.

Following Fremdt et al. (2014) and Dette et al. (2020), we project the daily minimum temperature values of each year onto a Fourier basis consisting of 49 basis functions, resulting in smooth annual temperature curves for each location under study. This choice balances the bias–variance trade-off in functional data smoothing. Too few basis functions may miss key seasonal patterns, increasing bias, while too many risk overfitting to noise, raising variance. A dimension of 49 provides enough flexibility to capture seasonal trends and local anomalies while ensuring stable functional representations.

We first apply the quadratic SN procedure considered in Dette et al. (2020); see Test (2.17) therein for the testing procedure, and Eqs. (2.18) and (2.19) therein for the definitions of $\hat{\mathbb{D}}_{m,n}$ and the self-normalizer $\hat{\mathbb{V}}_{m,n}$. They report $\hat{\mathbb{D}}_{m,n} = 14.115$ and $\hat{\mathbb{V}}_{m,n} = 0.315$. Choosing values of Δ between 9.1 and 10.7 leads to rejection of the null hypothesis at the $\alpha \geqslant 5\%$ significance level. For $\Delta \in [10.8, 11.7]$, the null hypothesis is rejected only at the $\alpha \geqslant 10\%$ level, indicating weaker evidence against the null.

Next, we apply our proposed adjusted-range-based SN method. The corresponding test statistic and self-normalizer are defined in (15). When applying the adjusted-range-based approach, we still have $\hat{\mathbb{D}}_{m,n} = 14.115$, and the adjusted-range-based self-normalizer is $\hat{\mathbb{H}}_{m,n} = 1.005$. This method yields comparable results: for Δ between 9.1 and 10.5, the null hypothesis is rejected at the $\alpha \geqslant 5\%$ level but not at the 1% level. For Δ values between 10.6 and 11.4, rejection occurs only at the $\alpha \geqslant 10\%$ level.

Notably, the locations of the cut-off points are comparable, while the lengths of the intervals over which rejection occurs are slightly shorter than those reported in Dette et al. (2020). This is consistent with the fact that the quadratic and adjusted-range procedures use different SN functionals and hence different critical values (cf. Table 1 in Dette et al. (2020)). Therefore, once the adjusted-range-based statistic detects a deviation large enough to reject the null hypothesis, the sharpness of the detection boundary ensures that the rejection zone is tightly concentrated. Overall, our method prioritizes robustness and precision, which is reflected in the structure of the rejection intervals.

D.2 Change-point test

Here, we revisit the daily river flow data (in m^3/s) from the Chemnitz at Göritzhain (Germany), spanning hydrological years 1910–2014. Treating each year as a functional observation yields $N = 105$ curves. Using the same trimming choice as in Dette et al. (2020) ($\varepsilon = 0.1$), the estimated change-point is 1964; see Figure 10(b) therein for a visual comparison of the pre- and post-change mean curves.

For the quadratic SN approach in Dette et al. (2020), Table 3 therein implies that the null hypothesis of no relevant change is rejected at the 10% level for $\Delta = 0.72$ but not for $\Delta \geq 0.73$ (and not at the 5% level at $\Delta = 0.72$). We reproduce the corresponding decision boundary in Table D.1 to provide a direct comparison.

Applying the adjusted-range-based SN on the same split yields a different relevance boundary: at the 10% level the null is rejected up to $\Delta = 0.45$ and not rejected for $\Delta \geq 0.46$ (Table D.1). This side-by-side comparison shows that the choice of normalizer can materially affect the practical magnitude of change supported by the data. In particular, the adjusted-range scaling produces a more conservative relevance bound in this example, which is attractive when the goal is to avoid overstating effect sizes in dependent functional settings.

Table D.1: Comparison of change-point test decisions for varying Δ values based on the quadratic SN method in Dette et al. (2020) and the adjusted-range-based SN method, applied to daily river flow data (in m^3/s) from the Chemnitz at Göritzhain, Germany.

	Quadratic SN method (Dette et al. 2020)			Adjusted-range-based SN method		
Δ	99%	95%	90%	99%	95%	90%
0.45	—	—	—	0	0	1
0.46	—	—	—	0	0	0
0.72	0	0	1	—	—	—
0.73	0	0	0	—	—	—

Note: "1" indicates rejection of the null hypothesis; "0" indicates failure to reject it. To keep the comparison focused, we report only Δ values around the decision boundary for each method and significance level.