Bernoulli **31**(4), 2025, 3356–3381
https://doi.org/10.3150/24-BEJ1850

Monitoring of functional time series

TIM KUTTA[a] and PIOTR KOKOSZKA[b]

Department of Statistics, Colorado State University, 851 Oval Dr, Fort Collins, CO 80523, United States,
[a]*tim.kutta@colostate.edu,* [b]*piotr.kokoszka@colostate.edu*

Detecting changes in dependent time series is an important topic in statistics. Tests for change points have been developed for different data structures, including multivariate, high-dimensional and even functional data. In functional data analysis, most literature focuses on retrospective settings, where the full dataset is collected before searching for changes. In this work, we follow the different paradigm of sequential monitoring. Here, data arrive successively, and the user tries to determine on an ongoing basis whether a change has already occurred - for an open-ended time period. This problem has not been investigated before for infinite dimensional data, where the standard tools to validate monitoring schemes are not available. In this work, we propose a different approach, combining strong approximations for Banach space valued data, with bounds from empirical process theory. Therewith, we validate monitoring schemes on a range of function spaces, including L^p-functions and spaces of continuous functions. In the latter case, we also propose theory for data with an unbounded domain - a problem rarely considered for functional data. An application of our theory to monitoring sequences of estimated densities is given. We investigate practical aspects of our approach in a simulation study and a data example on Exchange Traded Funds.

Keywords: Banach space; change point; functional data; monitoring

1. Introduction

This paper is concerned with monitoring time series with values in infinite-dimensional Banach spaces. As new functional observations arrive, we want to determine in real time if the mean of a sequence has changed at some point before the current time. Methods of this type have been proposed and justified in finite dimensional settings. The main probabilistic tools, if the dimension is finite, are adaptations of the Komlós, Major, Tusnády (KMT) approximation (Komlós, Major and Tusnády (1975), Komlós, Major and Tusnády (1976)). These results permit close approximation of partial sums of model errors by a Brownian motion. Another approach frequently used to validate finite dimensional monitoring schemes involves applying Hájek–Rényi bounds (see, e.g. Aue and Kirch (2024)). These bounds are then combined with weak invariance principles to prove convergence of the change point detector. However, up to this point, neither KMT approximations nor the required Hájek–Rényi bounds have been established to justify monitoring schemes in infinite dimensional function spaces. In this paper, we present a different approach to the development and asymptotic justification of monitoring procedures in several important Banach spaces. We also show how our general theory can be applied to monitoring for a mean change in a sequence of random probability density functions that are accessible only indirectly through small samples. We now provide a detailed background and explanation of our contribution.

Consider a time series $\{X_i, i \in \mathbb{Z}\}$ of Banach space valued random elements with corresponding parameters θ_i that are functionals of the distributions $\mathcal{L}(X_i)$. These parameters need not be finite dimensional. Suppose we observe a realization $X_1, \dots, X_N$. The *retrospective* change point problem is the testing problem

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_N, \quad \text{vs.} \quad H_1 : \exists\, k^* \;\; \theta_1 = \theta_2 = \cdots = \theta_{k^*} \neq \theta_{k^*+1} = \cdots = \theta_N. \tag{1.1}$$

1350-7265

In the literature there exist many generalizations of (1.1), such as the inclusion of multiple change points, but for sake of clarity we stick with this simple formulation. A key characteristic of (1.1) is the idea of "looking back" at the past N observations. Problem (1.1) has been extensively studied for X_i taking values in various function spaces and abstract metric spaces as well as other complex data structures, e.g. high-dimensional or network data. The problem we study in this paper is different, and can be formulated as testing

$$H_0 : \theta_1 = \ldots = \theta_M = \theta_{M+1} = \theta_{M+2} = \cdots \tag{1.2}$$

against the alternative of a parameter change at $M + k^*$ for some $k^* \geq 1$, i.e.

$$H_1 : \exists\, k^* \; \theta_1 = \cdots = \theta_{M+k^*} \neq \theta_{M+k^*+1} = \theta_{M+k^*+1} = \cdots \tag{1.3}$$

The first M observations are assumed not to contain a change point, and we want to test if there is a change sometime after time M. In this setting, we are looking ahead at the sequentially arriving new data. Problem (1.2)–(1.3) has been studied only in finite dimensional settings related to multiple regression and parametric time series.

Problem (1.2)–(1.3) is related to classical sequential testing. The chief differences are: 1) The time M corresponds to the initial stable period in which parameters are estimated. In classical sequential analysis, the parameters under the null, and often under the alternative, are specified a priori. There is no initial estimation period. 2) Sequential analysis is based on likelihood ratio tests, typically under the assumption of multivariate normality. Likelihoods are impossible to define in most functional settings. 3) Sequential analysis focuses on expected detection delay. We focus on the probability of a false alarm and power of detection, which are analogous to the well-known concepts in the Neyman–Pearson paradigm.

Our main contribution to statistical theory is showing that strong approximations on $(0, \infty)$, available in certain Banach spaces, can be used to construct monitoring schemes for functional time series. In our proofs, we combine these approximations with advanced bounds on covering numbers from empirical process theory and moment inequalities for mixing time series. As a result, our new theory is applicable to a broad array of different function spaces. First, it applies to the classical L^2-functions with compact support, a separable Hilbert space that is typically considered in FDA. Second, it extends to L^p-spaces with $p \in [2, \infty)$ and even to the space of continuous functions equipped with the supremum-norm. These latter spaces are infinite-dimensional Banach spaces, and hence proving central limit theorems and invariance principles requires careful arguments to exploit their specific, geometric structure. Developing theory for random elements on the space of continuous functions is theoretically challenging, but seems particularly natural: In applications, random functions are typically created using non-parametric estimators that are inherently smooth. It is on account of their continuity, that these functions are embedded in an L^2-space - essentially because the supremum norm is stronger than the L^2-norm. The fact that the supremum norm is so strong also implies that some relevant functionals are continuous w.r.t. the sup-norm, but not w.r.t. the L^2-norm. Finally, we present theory for the space of continuous functions with infinite support - a space rarely considered in FDA literature, but needed to study important statistical objects, including densities. Here we use arguments from the theory of stochastic processes on σ-compact domains to obtain weak convergence of change point detectors. As an application of our theory, we propose a new method to monitor for a change in a sequence of estimated random densities.

The remainder of the paper is organized as follows. In Section 2, we put our work in a broader perspective by discussing some closely related research. Before formulating the problem and main asymptotic results in Section 4, we collect useful prerequisites in Section 3. In Section 5, we apply the theory of Section 4 to the problem of detecting a change in a sequence of density estimators.

Section 6 is dedicated to the examination of finite sample properties of our monitoring procedures and an application to densities of intraday returns on selected Exchange Traded Funds. A summary of our contributions is given in Section 7. The proofs of our results and extensive additional information are placed in the online Supplementary Material Kutta and Kokoszka (2025).

2. Related research

Classical sequential testing is explained in Siegmund (1985). A concise account of the change point problem is given by Horváth and Rice (2014) and a recent review of the monitoring literature is provided by Aue and Kirch (2024). There are many monographs and dozens of thousands of papers on sequential and change point analyses, so the background we provide is limited and targeted at the problem we address.

Change point research in the framework of FDA. In FDA, observations are random elements in some function space, e.g. the space L^2 or C (continuous functions). Berkes et al. (2009) proposed a test for a change in the mean function in an L^2 setting; extensions were considered by Aston and Kirch (2012), Horváth, Kokoszka and Rice (2014), Dette and Wied (2016), Gromenko, Kokoszka and Reimherr (2017) and Aue, Rice and Sönmez (2018), among others. An invariance principle for L^2-functions -the typical building block of a retrospective change point detector- is derived in Berkes, Horváth and Rice (2013). Structural breaks of time series in the space $C([0,1])$ were studied in Dette, Kokot and Aue (2020). Within the L^2 framework, changes in the covariance of a functional time series were considered in Stoehr, Aston and Kirch (2021) and Horváth, Rice and Zhao (2022), and in the cross-covariance operator in Rice and Shum (2019). Dette and Kokot (2022) considered inference for the covariance kernel in C. Tests for the stability of eigenvalues and principal components were presented in Aue, Rice and Sönmez (2020) and Dette and Kutta (2021). A self-normalization approach, see Shao and Zhang (2010) and Shao (2015), was used in the context of change point detection in functional time series by Dette, Kokot and Volgushev (2020). Dubey and Müller (2020) present change point detection methodology and theory in general metric spaces. These papers, and many others, study some form of the retrospective problem (1.1). *Within the framework of FDA, we are not aware of any work on the monitoring problem* (1.2)–(1.3).

Monitoring research. The monitoring problem (1.2)–(1.3) was introduced by Chu, Stinchcombe and White (1996) in the context of a scalar response regression. They used two specific boundary crossing functions for the Wiener process to compute probabilities of false alarms. Their approach was advanced by Horváth et al. (2004) who introduced a parametric family of boundary functions and two types of regression residuals. The boundary functions we consider are related to those proposed by Horváth et al. (2004). Extensions to various scalar time series models were developed by Berkes et al. (2004), Aue et al. (2006) and Aue et al. (2012). Gösmann, Kley and Dette (2021) modified the monitoring paradigm and developed methodology suitable for multivariate time series. Horváth, Kokoszka and Wang (2021) worked out monitoring of quantile functions in various tail behavior scenarios. Their work has a functional aspect, but it does not involve a sequence of random distributions; the randomness is only in the scalar observations that come from deterministic distributions (one before a change, another after it). Aue et al. (2014) monitor for changes in a function-on-function regression, using projections on a finite (and fixed) number of functional principal components. Wu et al. (2022) consider monitoring of independent high-dimensional observations. Chen, Wang and Samworth (2022) study online change point detection for high-dimensional data in a classical sequential paradigm. For an informative review of recent developments in change point monitoring, we refer to Gösmann (2020). *While monitoring* *methodology is well developed for scalar observations, there are at present no methods to monitor data in function spaces without dimension reduction.*

3. Preliminaries

For ease of reference, we provide in this section several definitions and properties that will be used throughout the paper.

Random continuous functions. For $K > 0$, we denote by $C([-K,K])$ the space of continuous functions on the interval $[-K,K]$. It is a separable Banach space equipped with the supremum norm $\|f\|_\infty := \sup_{t\in[-K,K]} |f(t)|$. Next, if $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, we call a Borel map $X : \Omega \to C([-K,K])$ a random function in $C([-K,K])$. Recall that the measurability of X is equivalent to the measurability of all evaluations $X(t)$ for $t \in [-K,K]$, see e.g. Billingsley (1968), p. 84. Supposing that X satisfies the first moment condition $\mathbb{E}\|X\|_\infty < \infty$, we can define its mean function $\mu \in C([-K,K])$ by the pointwise identity $\mathbb{E}X(t) = \mu(t)$ for all t. Moreover, if X satisfies the second moment condition $\mathbb{E}\|X\|_\infty^2 < \infty$, we can define its continuous covariance kernel $c : [-K,K] \times [-K,K] \to \mathbb{R}$ by the pointwise identity

$$c(s,t) := \mathbb{E}[(X(t) - \mu(t))(X(s) - \mu(s))], \qquad \forall s,t \in [-K,K].$$

We say that a random function X is normally distributed, if for any $N \in \mathbb{N}$ and $t_1 < \cdots < t_N$, the vector $(X(t_1), ..., X(t_N))$ follows a multivariate normal distribution. Additional aspects of continuous random functions are discussed in the online Supplement (Kutta and Kokoszka (2025)), and for further details, we refer the reader to Janson and Kaijser (2015).

Polygon chains. An important subspace of $C([-K,K])$ consists of the polygon chains with $N \geq 2$ grid-points $\mathcal{PC}([-K,K],N)$. Any function $q \in \mathcal{PC}([-K,K],N)$ is characterized by values v_n at the grid points $x_n := (2nK/N - K)$, $n = 0,\dots,N$, i.e.

$$q(x) = \frac{v_{n-1}x_n - v_n x_{n-1}}{x_n - x_{n-1}} + \frac{v_n - v_{n-1}}{x_n - x_{n-1}}x, \ \text{ for } x \in [x_{n-1}, x_n]. \tag{3.1}$$

The canonical projection $\Pi_N : C([-K,K]) \to \mathcal{PC}([-K,K],N)$ is defined by setting $v_n = f(x_n)$ in (3.1). By definition, q is a piecewise linear function, interpolating between the grid points. The space of N polygon chains is an $(N+1)$-dimensional subspace of $C([-K,K])$ and can be used to assess the complexity of a function $f \in C([-K,K])$ by the decay rate of $\|f - \Pi_N(f)\|_\infty$.

Continuous functions on an unbounded domain. Denote by $C_b(\mathbb{R})$ the space of bounded continuous functions on $\mathbb{R}$. Equipped with the (suitably adapted) supremum norm it is a separable Banach space. Furthermore, we define the closed subspace $C_0(\mathbb{R})$ of functions vanishing at $\pm\infty$, i.e. the functions $f \in C_b(\mathbb{R})$ that satisfy $\lim_{t\to\infty} |f(t)| = \lim_{t\to-\infty} |f(t)| = 0$. We call a Borel map $X : \Omega \to C_b(\mathbb{R})$ a random variable in $C_b(\mathbb{R})$ and say that it is in $C_0(\mathbb{R})$, if $\mathbb{P}(X \in C_0(\mathbb{R})) = 1$. As before, measurability of X is equivalent to the measurability of all evaluations. Furthermore, the notions of expectation, covariance function and normality in $C_b(\mathbb{R})$ can be defined in analogy to the case of continuous functions on a compact interval (again, see the online Supplement).

Random L^p-functions. For any $p \in [1, \infty)$, we refer by $L^p([-K, K])$ (or short L^p) to the Lebesgue space of measurable functions $f : [-K, K] \to \mathbb{R}$, with finite norm

$$\|f\|_p := \left\{ \int |f(x)|^p dx \right\}^{1/p}.$$

Here, as common, we identify functions that are norm-equivalent. Equipped with this norm, L^p is a separable Banach space for all $p \geq 1$ and, similarly as before, we call a Borel map $X : (\Omega, \mathcal{A}, \mathbb{P}) \to L^p([-K, K])$ a random function in L^p. If the first moment of X exists, i.e. if $\mathbb{E}\|X\|_p < \infty$, we can define its mean function $\mu \in L^p([-K, K])$ by the identity

$$\int \mu(t) f(t) dt = \mathbb{E} \int X(t) f(t) dt, \qquad \forall f \in L^q([-K, K]),$$

where $1/q + 1/p = 1$. Here, for $p = 1$ we obtain $q = \infty$, and we denote by $L^\infty([-K, K])$ the space of essentially bounded functions. If $\mathbb{E}\|X\|_p^2 < \infty$, the covariance can be characterized by the bilinear operator

$$C[f, g] := \mathbb{E}\left[\left(\int X(s) f(s) dt \right) \left(\int X(t) g(t) ds \right) \right], \qquad f, g \in L^q([-K, K]).$$

We say that a random function X is normally distributed, if the product $\int X(t) f(t) dt$ is a real valued normal, for any $f \in L^q$. For details on random variables on Banach spaces in general and L^p-spaces in particular, we refer to the monograph of Ledoux and Talagrand (1991).

Brownian motion in a Banach space. In order to describe the limiting distribution of our test statistics, we employ the Brownian motion $\{W(x) : x \geq 0\}$ in the separable Banach space $\mathcal{B}$. It is is a stochastic process with continuous sample paths in $\mathcal{B}$ that has stationary, independent and normally distributed increments. It is characterized by its mean $\mathbb{E}W(1) \in \mathcal{B}$ (in our case, always the constant 0-function) and the covariance of $W(1)$. The existence and uniqueness of this process are established similarly as for the finite dimensional Brownian motion, see Kuelbs (1973) and references therein.

ϕ-mixing. To specify temporal dependence, we use the notion of ϕ-mixing. Consider random elements X_n in a separable Banach space. We say that the time series $(X_n)_{n \in \mathbb{N}}$ is ϕ-mixing, if there exists a sequence of non-negative numbers $\phi(k) \downarrow 0$ such that

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq \phi(k)\mathbb{P}(A)$$

for any two events $A \in \sigma(X_i : 1 \leq i \leq m)$ and $B \in \sigma(X_i : i \geq m + k)$ and any $m, k \geq 1$. Here, as always, σ denotes the sigma-algebra generated by the respective random variables. It is a direct consequence of the definition of ϕ-mixing that it is stable against measurable transforms. More precisely, if X_i takes values in the Banach space $\mathcal{X}$ and $g : \mathcal{X} \to \mathcal{Y}$ is a measurable map into another Banach space, the time series $(g(X_n))_{n \in \mathbb{N}}$ is also ϕ-mixing with coefficients $(\phi_g(n))_{n \in \mathbb{N}}$ satisfying $\phi_g(n) \leq \phi(n)$ for all $n \geq 1$. This property makes mixing more convenient for our purposes than measures of dependence that involve moments. For details on ϕ-mixing, as well as helpful covariance inequalities, we refer to the monograph of Dehling, Mikosch and Sørensen (2002) and for a survey on different mixing types to Bradley (1986).

4. Monitoring for a mean change in Banach spaces

In this section, we investigate the problem of monitoring a functional time series $(X_n)_{n\in\mathbb{N}}$ for a mean change. We first define a general setting on a separable Banach space $(\mathcal{B}, \|\cdot\|)$, and then split up our analysis along different function spaces.

The statistical model. Consider two different mean functions $\mu^{(1)}, \mu^{(2)} \in \mathcal{B}$ and a strictly stationary time series of model errors $(\varepsilon_n^{(1)}, \varepsilon_n^{(2)})_{n\in\mathbb{N}}$ in $\mathcal{B}\times\mathcal{B}$. We assume that these errors are centered, i.e. $\mathbb{E}[\varepsilon_1^{(1)}] = \mathbb{E}[\varepsilon_1^{(2)}] = 0$. We assume that the observations X_n satisfy

$$X_n := \begin{cases} \mu^{(1)} + \varepsilon_n^{(1)}, & n \le M + k^*, \\ \mu^{(2)} + \varepsilon_n^{(2)}, & n > M + k^*. \end{cases} \tag{4.1}$$

Here, as before, M denotes the length of an initial, stable set and $k^* \in \mathbb{N}\cup\{\infty\}$ the location of the change, where $k^* = \infty$ means that no change occurs. We are interested in testing the hypotheses

$$H_0 : \mu^{(1)} = \mu_1 = \cdots = \mu_M = \mu_{M+1} = \mu_{M+2} = \cdots, \tag{4.2}$$

$$\text{versus} \qquad H_1 : \mu^{(1)} = \mu_1 = \cdots = \mu_{M+k^*} \neq \mu^{(2)} = \mu_{M+k^*+1} = \mu_{M+k^*+2} = \cdots$$

or, $H_0 : k^* = \infty$ versus $H_1 : k^* < \infty$.

A change point detector. Let $\gamma \in [0, 1/2)$, $\zeta \in (0,1)$ and define the notation $(x)_\zeta := \max(x, \zeta)$. Then we can define the weight function,

$$g(M,k) := M^{1/2}\Big(1 + \frac{k}{M}\Big)\Big(\frac{k}{M+k}\Big)_\zeta^\gamma. \tag{4.3}$$

The cutoff $\zeta > 0$ is needed in infinite-dimensional spaces, but can be made arbitrarily small in practice, see Remark 4.1. We provide an intuitive explanation for the choice of the weight function (4.3) in the online Supplement (Kutta and Kokoszka (2025)). With this notation in hand, we define the detector

$$\Gamma_M(k) := \left\| \frac{\frac{k}{M}\sum_{n=1}^M X_n - \sum_{n=M+1}^{M+k} X_n}{g(M,k)} \right\|. \tag{4.4}$$

Level-α and consistency. We now discuss the behavior of the detector under H_0 and H_1. We will show that under H_0,

$$\sup_{k\ge1} \Gamma_M(k) \xrightarrow{d} \sup_{x\in[0,1]} \left\| \frac{W(x)}{(x)_\zeta^\gamma} \right\|, \qquad \text{as } M \to \infty, \tag{4.5}$$

where W is a centered, $\mathcal{B}$-valued Brownian motion, with the covariance structure of the time series $(\varepsilon_n^{(1)})_{n\in\mathbb{N}}$ (see Theorems 4.1–4.3 for explicit expressions). If the covariance structure of W were known, we could, for any $\alpha \in (0,1)$, approximate the upper α-quantile of the limiting distribution, say $q_{1-\alpha}$, and propose the test decision to reject as soon as

$$\Gamma_M(k) > q_{1-\alpha}. \tag{4.6}$$

In practice, the covariance structure of W is unknown, but it can be approximated using the stable, initial sample, see Section 6.4. If (4.5) holds, it follows that the test decision (4.6) has asymptotic level

α, as $M \to \infty$. Besides the level of the test decision, we also want to analyze its consistency. Therefore, we consider local alternatives with mean functions $\mu^{(1)}, \mu^{(2)}$ and a change point location k^* that may depend on M. We call our approach $\sqrt{M}$-consistent if

$$\sqrt{M}\|\mu^{(1)} - \mu^{(2)}\| \to \infty, \tag{4.7}$$

implies

$$\mathbb{P}\Big(\sup_{k\geq 1} \Gamma_M(k) > q_{1-\alpha}\Big) \to 1. \tag{4.8}$$

We now formulate the corresponding results in specific function spaces.

Monitoring in $C([-K,K])$. We now analyze the change point detector for continuous data on a compact interval. We first specify the assumptions.

Assumption 4.1.

i) (Dependence) The time series $(\varepsilon_n^{(1)}, \varepsilon_n^{(2)})_{n\in\mathbb{N}}$ in $C([-K,K]) \times C([-K,K])$ is strictly stationary and ϕ-mixing. For some integer $J \geq 1$ and $\nu > 2J^2$, the mixing coefficients satisfy

$$\phi(k) = \mathcal{O}(k^{-\nu}).$$

ii) (Moments) For some $\delta > 0$ and $i = 1,2$ it holds that $\mathbb{E}\|\varepsilon_1^{(i)}\|_\infty^{2J+\delta} < \infty$.
iii) (Centering) For $i = 1,2$, it holds that $\mathbb{E}[\varepsilon_1^{(i)}] = 0$.
iv) (Smoothness) For some $\rho > 1/(2J)$ and $i = 1,2$, it holds that

$$\sup_{s\neq t} \mathbb{E}\left[\left(\frac{|\varepsilon_1^{(i)}(t) - \varepsilon_1^{(i)}(s)|}{|t-s|^\rho}\right)^{2J}\right] < \infty.$$

Moreover, each mean function $\mu^{(i)}$ is ρ-Hölder continuous.

Condition i) means that the noise terms in (4.1) are weakly dependent and stationary, before and after the change. Allowing a break in the noise distribution at the change point location is often necessary. Conditions ii) and iii) are standard in the change point literature and usually required to prove invariance principles, see e.g. Dehling (1983), Berkes, Horváth and Rice (2013). Finally, condition iv) ensures smoothness of the observations, which translates into smoothness of the sample average. If dependence of the random errors $\varepsilon_n^{(i)}$ is weaker (see Condition i)) and if more moments exist, we allow for less smoothness of the sample paths. For $\rho < 1/2$, Condition iv) is satisfied, e.g., by the Brownian motion or by general classes of Itô processes, as discussed in Theorem 1 of Fischer and Nappo (2009). In the space $C([-K,K])$, there exists a close connection between smoothness of stochastic processes and their tightness, which can be thought of as a narrow concentration around a finite dimensional vector space - in our case a space of polygon chains. A similar, but stronger condition was used in Dette, Kokot and Aue (2020) for the study of functional time series in a retrospective change point setting. Now, we state our main mathematical result, the weak convergence of the random variable $\sup_{k\geq 1} \Gamma_M(k)$.

Theorem 4.1. *Suppose that Assumption 4.1 is satisfied. Then, under H_0, the weak convergence (4.5) holds and the covariance kernel of the Brownian motion W is given by $c_W^{(1)}$, where*

$$c_W^{(i)}(s,t) := \mathbb{E}\big[\varepsilon_1^{(i)}(s)\varepsilon_1^{(i)}(t)\big] + 2\sum_{n\geq 2}\mathbb{E}\big[\varepsilon_1^{(i)}(s)\varepsilon_n^{(i)}(t)\big], \qquad \forall s,t \in [-K,K]. \tag{4.9}$$

Under H_1 and (4.7), the convergence (4.8) holds ($\sqrt{M}$-consistency).

The test decision (4.6) based on Theorem 4.1 presents a new approach to the monitoring of functional time series. First, to the best of our knowledge, it is the first monitoring scheme derived in an infinite-dimensional function space. Previous work (Aue et al. (2014)) relied on projections on a fixed number of principal components combined with strong approximations of KMT-type for multivariate data that do not extend to infinite-dimensional data. Second, in contrast to most FDA work, it considers functional data as elements in the Banach space of continuous functions $C([-K,K])$ rather than the Hilbert space L^2. In the context of continuous functions, we employ strong approximations for Banach space valued data, established by Dehling (1983), to replace sums of random functions by a functional Brownian motion. A critical condition for these results is the tightness of the partial sum process in $C([-K,K])$. We show that it holds because the process concentrates closely around a finite dimensional subspace of polygon chains. The tightness proof relies on maximum inequalities from the theory of stochastic processes.

Remark 4.1. We provide a few more details on Theorem 4.1:

i) The proof of this theorem can be adapted to include other types of dependence conditions, such as absolutely regular sequences of β-mixing random functions. In this case, different results of Dehling (1983) would need to be used.

ii) The constant $\zeta > 0$ can be chosen arbitrarily small in practice (see, Wied and Galeano (2013) and Pape, Wied and Galeano (2016)). An investigation of the proof of Theorem 4.1 together with the results in Dehling (1983), shows that a choice of $\zeta = 0$ is possible, for sufficiently small γ. With this parameter choice, g is identical to the popular weight function from Horváth et al. (2004). Regarding the relation of ζ and γ, our results for Banach valued data differ from those in the multivariate setting, where the KMT theorems imply that $\zeta = 0$ is permissible for any $\gamma \in [0, 1/2)$ if sufficiently many moments exist. A potential strategy to generalize our results for $\zeta = 0$ and $\gamma > 0$ might be the application of Hölderian invariance principles to control the sum $\sum_{n=M+1}^{M+k} X_n$ for small values of k. For Hilbert space valued data such results have been established e.g. by Račkauskas and Suquet (2009) and Theorem 2.1 of that work indicates that extensions to other separable Banach spaces such as $C([-K,K])$ might be possible.

iii) In this paper, we consider functions on the compact interval $[-K,K]$. It is however clear that by shifting and rescaling the functions the same procedures can be used for any compact subinterval of the real line.

Monitoring in $C_0(\mathbb{R})$. Most literature on FDA is focused on random functions living on a compact domain, typically a compact interval. Such assumptions are motivated by the clear-cut, bounded support that functions in many problems exhibit, say the study of yearly temperature profiles, where the temperature is measured over the compact time interval of one year, see, e.g., Aue, Rice and Sönmez (2018), among hundreds of references. Yet, the choice of the supporting interval is not always that simple: Consider the case of a sequence of random densities $(f_n)_{n\in\mathbb{N}}$ discussed in Section 6.3. Many important families of parametric densities have unbounded support and considering them on a compact subinterval of $\mathbb{R}$ seems artificial. Even if all densities in the time series were supported on one joint, compact interval I, it might not be clear what I is, specifically when we are interested in monitoring schemes and not all data is available at the outset. For these reasons, we now present a version of our theory for data in $C_0(\mathbb{R})$, the space of continuous functions that vanish at $\pm\infty$, equipped with the sup-norm. We require essentially the same assumptions as in the compactly supported case.

Assumption 4.2. Assumption 4.1 holds with $C([-K,K])$ replaced everywhere by $C_0(\mathbb{R})$.

In contrast to the space of continuous functions on a compact interval, the space $C_0(\mathbb{R})$ poses additional challenges in terms of tightness of stochastic processes. More precisely, tightness of a random element in an infinite-dimensional vector space means concentration around a low-dimensional subspace, also known as *flatness*. In the case of $C([-K,K])$, we use the space of polygon chains with N edges. This construction does not generalize to $C_0(\mathbb{R})$, where a stochastic process may fluctuate for large arguments outside the range $[-K,K]$ of any finite polygon chain. Therefore, we modify our approach and endow $C_0(\mathbb{R})$ for the study of our change point detector with a weaker topology than that induced by the sup-norm. For a constant $B>0$ and a positive weight function $\psi \in C_0(\mathbb{R})$, we define

$$\|f\|_{B,\psi} := \sup_{t\in\mathbb{R}} \big|\big(f(t)\wedge B\big)\cdot\psi(t)\big|, \qquad f\in C_0(\mathbb{R}). \tag{4.10}$$

Here "$x \wedge y$" denotes the minimum of $x, y \in \mathbb{R}$. Notice that we can assess the distance of two functions $f_1, f_2 \in C_0(\mathbb{R})$ by $\|f_1 - f_2\|_{B,\psi}$, which defines a metric on $C_0(\mathbb{R})$. To make it unambiguously clear which norm (and topology) we are considering, we will in the following always write $(C_0(\mathbb{R}), \|\cdot\|_{B,\psi})$ whenever referring to the continuous functions equipped with $\|\cdot\|_{B,\psi}$ and $(C_0(\mathbb{R}), \|\cdot\|_\infty)$ if we consider the ordinary sup-norm. Also notice that in Assumption 4.2 we are referring to $(C_0(\mathbb{R}), \|\cdot\|_\infty)$. While $\|\cdot\|_{B,\psi}$ is not a norm (it lacks homogeneity), it has certain norm-like properties, making it a *G-norm*, compatible with the additive group structure of the vector space of functions, see Chapter 22 in Schechter (1997). We provide further details on $\|\cdot\|_{B,\psi}$ and the choice of B and ψ in Remark 4.2.

Redefining the detector Γ_M in (4.4) using the G-norm (4.10), allows us to detect changes in $(C_0(\mathbb{R}), \|\cdot\|_{B,\psi})$. For the formulation of local alternatives, we assume additionally to (4.7) (change of size larger than $1/\sqrt{M}$), a concentration of the difference $\mu^{(1)} - \mu^{(2)}$, i.e. that for some large enough (but unknown) $K>0$

$$\sqrt{M} \sup_{t\in[-K,K]} |\mu^{(1)}(t) - \mu^{(2)}(t)| \to \infty. \tag{4.11}$$

This is in particular true for any fixed pair of functions $\mu^{(1)}, \mu^{(2)} \in C_0(\mathbb{R})$. Within the above setting, we can define and prove the validity of the monitoring scheme under basically the same assumptions as in the compactly supported case.

Theorem 4.2. *Suppose that Assumption 4.2 is satisfied. Then, under H_0, convergence (4.5) holds with the norm $\|\cdot\|$ replaced by the G-norm $\|\cdot\|_{B,\psi}$. The covariance kernel of the Brownian motion W in $(C_0(\mathbb{R}), \|\cdot\|_\infty)$ is given by*

$$c_W^{(i)}(s,t) := \mathbb{E}\big[\varepsilon_1^{(i)}(s)\varepsilon_1^{(i)}(t)\big] + 2\sum_{n\geq 2} \mathbb{E}\big[\varepsilon_1^{(i)}(s)\varepsilon_n^{(i)}(t)\big], \qquad s,t\in\mathbb{R}. \tag{4.12}$$

Under H_1, (4.7) and (4.11), convergence (4.8) holds for any large enough $B>0$.

We note that W as stochastic process on $(C_0(\mathbb{R}), \|\cdot\|_\infty)$ is also a stochastic process on $(C_0(\mathbb{R}), \|\cdot\|_{B,\psi})$. Moreover, the covariance function (4.12) in Theorem 4.2 is the same as the covariance function (4.9) in Theorem 4.1, except that the ranges for s and t are different.

Remark 4.2. We provide additional comments on the G-norm and its application.

i) The G-norm (4.10) metricizes the *topology of compact convergence* on $C_0(\mathbb{R})$. In this topology, a sequence of functions $(f_n)_{n\in\mathbb{N}}$ converges to f, if for any compact set $\mathcal{E}$

$$\sup_{t\in\mathcal{E}} |f_n(t) - f(t)| \to 0.$$

The resulting topology is strictly weaker than that induced by the sup-norm. One consequence of this is the fact that while $(C_0(\mathbb{R}), \|\cdot\|_\infty)$ is a closed metric space, the space $(C_0(\mathbb{R}), \|\cdot\|_{B,\psi})$ is not closed.

ii) Even though from a topological viewpoint it does not matter which pair of (B,ψ) is selected to define the G-norm, it matters for the statistical application. While the specific choice may depend on the application at hand, we recommend to generically use large values of B and a slowly decaying weight function ψ, for high power against a broad spectrum of alternatives.

Monitoring in L^p. Up to this point, we have considered the detector Γ_M for continuous functions. Now, we will consider an analogue of Theorem 4.1 that is specifically tailored to L^p-spaces, $p \in [2,\infty)$. The space L^2 is commonly used in FDA, but there are relatively few results for other L^p-spaces. For the next result, we require weaker dependence, moment and regularity conditions than before. Most importantly, we do not require our observations to be continuous.

Assumption 4.3.

i) (Dependence) The time series $(\varepsilon_n^{(1)}, \varepsilon_n^{(2)})_{n\in\mathbb{N}}$ in $L^p([-K,K]) \times L^p([-K,K])$ is strictly stationary and ϕ-mixing, such that for some $\nu > 2$

$$\phi(k) = \mathcal{O}(k^{-\nu}).$$

ii) (Moments) For some $\delta > 0$ and $i = 1,2$, $\mathbb{E}\|\varepsilon_1^{(i)}\|_p^{2+\delta} < \infty$.

iii) (Centering) For $i = 1,2$, $\mathbb{E}[\varepsilon_1^{(i)}] = 0 \in L^p([-K,K])$.

iv) (Regularity) For $i = 1,2$, there exist sequences of N-dimensional subspaces $V_N^{(i)} \subset L^p([-K,K])$ with projections $\Lambda_N^{(i)} : L^p([-K,K]) \to V_N^{(i)}$, such that

$$\mathbb{E}\big[\|\{\mathrm{Id} - \Lambda_N^{(i)}\}[\varepsilon_1^{(i)}]\|_p^{2+\delta}\big] = \mathcal{O}(N^{-\xi}), \tag{4.13}$$

for some $\xi > 0$. Moreover, the mean functions satisfy $\mu^{(1)}, \mu^{(2)} \in L^p([-K,K])$.

Conditions i)-iii) are analogues of those in Assumption 4.1 for the (least restrictive) case of a Hölder index $\rho > 1/2$. The final condition ensures that the observations are concentrated around a low-dimensional subspace, thus guaranteeing tightness of the partial sum processes in the definition of Γ_M. An appropriate choice of the function space $V_N^{(i)}$ depends on the distribution of the data. It can, for instance, make sense to choose the space of polygon chains, defined in Section 3. If the functions $\varepsilon_n^{(i)}$ are piecewise Hölder continuous, it can be shown that they satisfy condition iv). Another choice for $V_N^{(i)}$ might be a space of step functions, where $\varepsilon_n^{(i)}$ is essentially approximated by its Riemann sum.

In the case of L^2 data, Condition iv) can be further weakened by replacing moments of order $2+\delta$ by second moments.

Assumption 4.3 (Continued).

iv)' (Regularity) For $i = 1,2$, there exist sequences of N-dimensional subspaces $V_N^{(i)} \subset L^2([-K,K])$ with projections $\Lambda_N^{(i)} : L^2([-K,K]) \to V_N^{(i)}$, such that

$$\mathbb{E}\big[\|\{\mathrm{Id} - \Lambda_N^{(i)}\}[\varepsilon_1^{(i)}]\|_2^2\big] = O(N^{-\xi}), \tag{4.14}$$

for some $\xi > 0$. Moreover, the mean functions satisfy $\mu^{(1)}, \mu^{(2)} \in L^2([-K,K])$.

The advantage of Condition iv)' compared to iv) for L^2-spaces is that it can be checked by a simple, sufficient criterion, in terms of the ordered eigenvalues $\lambda_1^{(i)} \geq \lambda_2^{(i)} \geq \ldots$ of the covariance of $\varepsilon_1^{(i)}$.

Lemma 4.1. *Suppose that Assumption 4.3, Conditions i)-iii) hold for $p = 2$. Furthermore, suppose that there exist constants $C, \xi > 0$ such that $\lambda_n^{(i)} \leq Cn^{-(1+\xi)}$. Then (4.14) holds with $\Lambda_N^{(i)}$ being the projection on the subspace spanned by the leading N functional principal components of $\varepsilon_1^{(i)}$.*

We can now formulate our results for L^p-spaces.

Theorem 4.3. *Suppose that $p \in [2,\infty)$ and that Assumption 4.3 is satisfied with Condition iv)' instead of iv) if $p = 2$. Then, under H_0, the weak convergence (4.5) holds and the covariance of the Brownian motion W is given by $C_W^{(1)}$, where*

$$\begin{aligned} C_W^{(i)}(f,g) :=& \mathbb{E}\Big[\int \varepsilon_1^{(i)}(s)f(s)ds \int \varepsilon_1^{(i)}(t)g(s)ds\Big] \\ &+ 2\sum_{n\geq 2}\mathbb{E}\Big[\int \varepsilon_1^{(i)}(s)f(s)ds \int \varepsilon_n^{(i)}(t)g(s)ds\Big], \quad f,g \in L^q[-K,K], \end{aligned} \tag{4.15}$$

for $1/p + 1/q = 1$. Under H_1 and (4.7), the convergence (4.8) holds ($\sqrt{M}$-consistency).

Remark 4.3. The results formulated in this section rely on strong approximations for sums of Banach space valued random variables. It is important to notice that such approximations depend on the geometric structure of the Banach space in question. The case of continuous functions is particularly challenging, since $C([-K,K])$ is well-known to be a Banach space with (trivial) type 1 and cotype ∞. In contrast, L_p spaces with $p \geq 2$ are known to have type 2 and cotype p, see e.g. Chapter 9.3 in Ledoux and Talagrand (1991), meaning that many tools from probability theory extend to these spaces, such as the Marcinkiewicz–Zygmund inequality with index 2, Utev (1991). Another tool that extends to certain spaces of type 2, including all L^p-spaces for $p \geq 2$, are Rosenthal inequalities with index > 2, which entail CLTs for Banach spaces (see Chapters 9 and 10 in Ledoux and Talagrand (1991)). The role of these inequalities is to transfer tightness from individual random variables to partial sums. For instance, we exploit in our proofs for L^p-spaces the bounds

$$\mathbb{E}\Big\|\{\mathrm{Id} - \Lambda_N^{(i)}\}\Big[\frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j^{(i)}\Big]\Big\|_p^2 \leq \frac{C}{n}\sum_{i=1}^n \mathbb{E}\|\{\mathrm{Id} - \Lambda_N^{(i)}\}[\varepsilon_j^{(i)}]\|_p^2 = O(N^{-\xi}),$$

where we have used Assumption 4.3, iv) for the second equality. Such inequalities are generally not available for Banach spaces of type $p < 2$ and in particular potential extensions of Theorem 4.3 to L^p-spaces with $p \in [1,2)$ will have to rely on different proof strategies than those pursued in this paper. For further details on the relation between probabilistic properties of Banach spaces and their geometry, we refer the reader to Hoffmann-Jørgensen (1974), Diestel and Uhl (1976) and Ledoux and Talagrand (1991).

5. Monitoring for a change in a sequence of estimated random densities

In this section, we consider monitoring a time series of random density functions. In contrast to the previous section, we assume that the true functions of interest, the densities, are not directly observable, and we only have access to estimators based on a (small) data sample. We focus on densities to provide more specific results in an important case, but it is clear that our results could be extended to other time series of estimated functions.

Statistical model. We now specify our model: Let $(f_n^{(1)})_{n\in\mathbb{N}}$ and $(f_n^{(2)})_{n\in\mathbb{N}}$ denote two stationary time series of continuous probability densities on $\mathbb{R}$ and define

$$f_n := \begin{cases} f_n^{(1)}, & n \le M + k^*, \\ f_n^{(2)}, & n > M + k^*. \end{cases} \tag{5.1}$$

Here, as before, $k^* \in \mathbb{N} \cup \{\infty\}$ specifies the time of a change, with $k^* = \infty$ corresponding to no change. As in the previous section, we define the mean functions $\mu_n := \mathbb{E} f_n$ and $\mu^{(i)} := \mathbb{E} f_n^{(i)}$ for $i = 1, 2$, and consider the hypothesis pair (4.2). Notice, that if the mean function of a random probability density is well-defined, it is again a probability density. One could clearly work with densities transformed in some way, using LQD, CLR or other transforms, see Petersen, Zhang and Kokoszka (2022) for a review, but we want to focus here on the estimation aspect, so we do not include any of those transformations. We now assume that we cannot directly observe the functions f_n. Rather, we base our inference on a sample of L i.i.d. observations sampled from the densities, i.e.

$$x_{1,n}^{(i)}, \dots, x_{L,n}^{(i)} \sim f_n^{(i)}, \qquad i = 1, 2. \tag{5.2}$$

We assume that conditionally on the time series $(f_n^{(1)}, f_n^{(2)})_{n\in\mathbb{N}}$ the random variables $\{x_{\ell,n}^{(i)} : 1 \le \ell \le L, n \in \mathbb{N}, i = 1, 2\}$ are independent. Based on a sample $\{x_{1,n}^{(i)}, \dots, x_{L,n}^{(i)}\}$, we construct the corresponding kernel density estimator $\hat{f}_n^{(i)}$. Denoting by $\mathscr{K}$ a non-negative kernel function with $\int \mathscr{K}(t)dt = 1$ and by h_L a positive bandwidth, $\hat{f}_n^{(i)}$ is then defined as

$$\hat{f}_n^{(i)}(t) := \frac{1}{Lh_L} \sum_{\ell=1}^{L} \mathscr{K}\left(\frac{t - x_{\ell,n}^{(i)}}{h_L}\right). \tag{5.3}$$

In reality, we do not know the value of i, which indicates whether we are before or after the change point. To reflect this in our notation, we define the ℓth observation in period n as $x_{\ell,n} := x_{\ell,n}^{(i)}$ if $f_n = f_n^{(i)}$ and the corresponding kernel density estimator as $\hat{f}_n := \hat{f}_n^{(i)}$. So, in practice, for some time $k \ge 1$, we observe the panel of points

$$\{x_{\ell,n} : 1 \le \ell \le L, n = 1, ..., M + k\}, \tag{5.4}$$

from which construct the functional observations $\hat{f}_1, ..., \hat{f}_{M+k}$. We regard these density estimates as random functions in the space $C_0(\mathbb{R})$ and apply the sequential change point based on Theorem 4.2. (In the online Supplement, we establish that these estimators are indeed measurable random functions in $C_0(\mathbb{R})$.) A key issue is whether the assumptions imposed on the random densities f_n translate to suitable conditions on the $\hat{f}_n$, so that Theorem 4.2 can be applied.

Remark 5.1. We emphasize that the sampling scheme proposed in this section introduces two sources of randomness: First, the densities f_n are themselves random functions, i.e. functional data objects. Second, the act of sampling from them to calculate the estimates $\hat{f}_n$ introduces additional randomness. This framework differs from non-functional models in which the distribution remains constant before and after a change point - see, Matteson and James (2014), Kojadinovic and Verdier (2021) and Horváth, Kokoszka and Wang (2021) for some examples. While these methodologies offer robust tools to investigate distributional changes, they presuppose a fixed data-generating process between adjacent time periods. This assumption is unrealistic in typical functional settings where the underlying functions are random objects. A scenario of this type is explored in our data analysis in Section 6.3, where we examine the distributions of intraday log-returns that vary from day to day.

We begin by specifying the assumptions on the unobservable densities, the kernels and the mean functions before and after the change.

Assumption 5.1.

i) (Dependence) The time series $(f_n^{(1)}, f_n^{(2)})_{n\in\mathbb{N}}$ in $C_0(\mathbb{R}) \times C_0(\mathbb{R})$ is strictly stationary and ϕ-mixing, with mixing coefficients satisfying for some $\delta > 0$

$$\phi(k) = O(k^{-2(1+\delta)}).$$

ii) (Kernel & Bandwidth) The kernel function $\mathscr{K}$ is Lipschitz continuous and bounded, and the bandwidth h_L is positive.

iii) (Moments) It holds that $\mathbb{E}\|f_1^{(1)}\|_\infty, \mathbb{E}\|f_1^{(2)}\|_\infty < \infty$. Furthermore, the expectations $\mu^{(1)}, \mu^{(2)}$ satisfy for some $t \in [-K, K]$

$$\tilde{\mu}^{(1)}(t) := \int \mathscr{K}(x)\mu^{(1)}(xh_L - t)dx \neq \int \mathscr{K}(x)\mu^{(2)}(xh_L - t)dx =: \tilde{\mu}^{(2)}(t).$$

The last condition guarantees that a change in mean density translates into a change in the mean function of the kernel density estimator, which can be shown to equal $\mathbb{E}\hat{f}_n^{(i)} = \tilde{\mu}^{(i)}$, see Lemma 5.1 below. The assumption holds for any fixed $\mu^{(1)} \neq \mu^{(2)}$ and any kernel function $\mathscr{K}$ for sufficiently small h_L, see Proposition 4.1.1. in Giné and Nickel (2016).

Next, we demonstrate that the time series of density estimators is weakly dependent and strictly stationary before and after the potential change point.

Lemma 5.1. *Suppose that Assumption 5.1 holds.*

i) *Then* $\tilde{\mu}^{(i)} = \mathbb{E}\hat{f}_n^{(i)}$ *for* $i = 1, 2$.
ii) *The sequence* $(\hat{f}_n^{(1)}, \hat{f}_n^{(2)})_{n\in\mathbb{N}}$ *is strictly stationary in* $C_0(\mathbb{R}) \times C_0(\mathbb{R})$ *and* ϕ*-mixing, with coefficients* $\hat{\phi}(k) \le \phi(k)$.

Lemma 5.1 demonstrates that the assumptions of Theorem 4.2 are met (in this formulation for fixed mean functions). We can now redefine the change point detector for the estimated functions as

$$\widehat{\Gamma}_M(k) := \left\| \frac{\frac{k}{M}\sum_{n=1}^{M} \hat{f}_n - \sum_{n=M+1}^{M+k} \hat{f}_n}{g(M,k)} \right\|_{B,\psi}. \tag{5.5}$$

For ease of notation, we also define the residual $\hat{\varepsilon}_n^{(i)} := \hat{f}_n^{(i)} - \tilde{\mu}^{(i)}$.

Theorem 5.1. *Suppose that Assumption 5.1 and the null hypothesis* H_0 *hold* ($k^* = \infty$)*. Then, the weak convergence*

$$\sup_{k\ge 1} \widehat{\Gamma}_M(k) \xrightarrow{d} \sup_{x\in[0,1]} \left\| \frac{W(x)}{(x)_\zeta^\gamma} \right\|_{B,\psi}, \qquad \text{as } M \to \infty, \tag{5.6}$$

follows, where where W *is a centered, functional Brownian motion on* $C_0(\mathbb{R})$*. The covariance of* W *is* $c_W^{(1)}$*, where*

$$c_W^{(i)}(s,t) := \mathbb{E}\big[\hat{\varepsilon}_1^{(i)}(s)\hat{\varepsilon}_1^{(i)}(t)\big] + 2\sum_{n\ge 2} \mathbb{E}\big[\hat{\varepsilon}_1^{(i)}(s)\hat{\varepsilon}_n^{(i)}(t)\big] \quad s,t \in \mathbb{R}.$$

If H_1 *holds for* $k^* = k^*(M) < \infty$ *and* $B > 0$ *is sufficiently large, then*

$$\mathbb{P}\Big(\sup_{k\ge 1} \widehat{\Gamma}_M(k) > q_{1-\alpha}\Big) \to 1. \tag{5.7}$$

Here $q_{1-\alpha}$ *is the upper* α*-quantile of the limiting distribution in* (5.6).

Remark 5.2. For simplicity of presentation, we have in this section assumed that the bandwidth h_L is fixed and deterministic for all densities f_n. However, our theory allows a straightforward extension to random bandwidths that depend on n. For this purpose, define a bandwidth selection rule H that can be understood as a measurable function $H : \mathbb{R}^L \to [a, b]$, where $a < b$ are positive numbers. Then, defining $h_{L,n}^{(i)} := H(x_{1,n}^{(i)}, ..., x_{L,n}^{(i)})$ and redefining

$$\hat{f}_n^{(i)}(t) := \frac{1}{Lh_L^{(i)}} \sum_{\ell=1}^{L} \mathscr{K}\left(\frac{t - x_{\ell,n}^{(i)}}{h_{L,n}^{(i)}}\right), \tag{5.8}$$

all results from this section remain true. One could also modify our arguments to accommodate random sample sizes L, assuming they are uniformly bounded from below and above.

6. Finite sample properties

In this section, we study the finite sample properties the proposed monitoring schemes. We begin with two simulation studies, first for fully observed functional data, in Section 6.1, and second for estimated random densities, in Section 6.2. In Section 6.3, we consider a data example of return distributions during the Great Recession of 2008. Finally, Sections 6.4 and 6.5 provide additional details on parameter choices and covariance estimation.

6.1. Monitoring functional time series

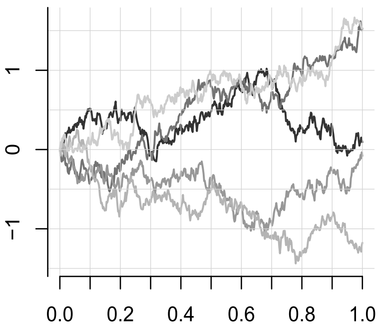
We consider a functional time series of the form (4.1), where the mean functions before and after the change point are given by

$$\mu^{(1)}(t) = 0 \qquad \text{and} \qquad \mu^{(2)}(t) = a\cos(2\pi t), \qquad t \in [0,1],$$

respectively. Here, $a \geq 0$ is a constant that quantifies the distance to the null hypothesis H_0, with $a = 0$ corresponding to no change.

For the noise terms, we set $\varepsilon_n^{(1)} = \varepsilon_n^{(2)} =: \varepsilon_n$ and consider different scenarios. Defining by $(\tilde{\varepsilon}_n)_{n\in\mathbb{N}}$ a sequence of i.i.d., centered noise functions, we consider 1) the independent case $\varepsilon_n = \tilde{\varepsilon}_n$ and 2) the dependent case $\varepsilon_n = (\tilde{\varepsilon}_n + \tilde{\varepsilon}_{n-1})/\sqrt{2}$. For the distribution of the $(\tilde{\varepsilon}_n)_{n\in\mathbb{N}}$, we consider a) the standard Brownian motion and b) a diffusion process, defined as the stochastic integral $\tilde{\varepsilon}_n(t) := \int_0^t \sigma(s)dw_n(s)$, where $w_1, w_2, \ldots$ are i.i.d. standard Brownian motions and $\sigma(t) := 10(t-0.5)^2$ is a U-shaped volatility function. In Figure 1, we display some observations $X_n = \mu^{(i)} + \varepsilon_n$ before and after a mean change, with errors generated by the Brownian motions. All functions are discretized by evaluating them on an equidistant grid on $[0,1]$ with 50 grid-points. Notice that for convenience, the random functions X_n are indexed in the unit interval, rather than some interval symmetric around 0, which is permissible according to Remark 4.1, part iii).

In order to approximate the quantiles of the limiting distribution, we estimate in each simulation run the long-run variance kernel c_W, based on the first M (stable) observations. The estimator $\hat{c}_W$ is based on autocovariance functions up to order 2 and the precise definition is given in Section 6.5, below. We consider $a \in \{0, 1/10, 2/10, \ldots, 8/10\}$ to study power, with $a = 0$ corresponding to H_0. We fix the parameters $\zeta = 0.05$ and $\gamma = 0.3$ (their impact is studied in additional simulations in Section 6.4). The nominal level is fixed at $\alpha = 5\%$ and is indicated in Figure 2 by a light gray horizontal line. We consider



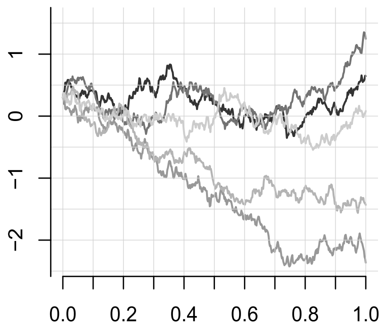


Figure 1. Realizations of functions X_n before (left) and after (right) the change point, for $a = 0.3$ and independent Brownian motions as errors.

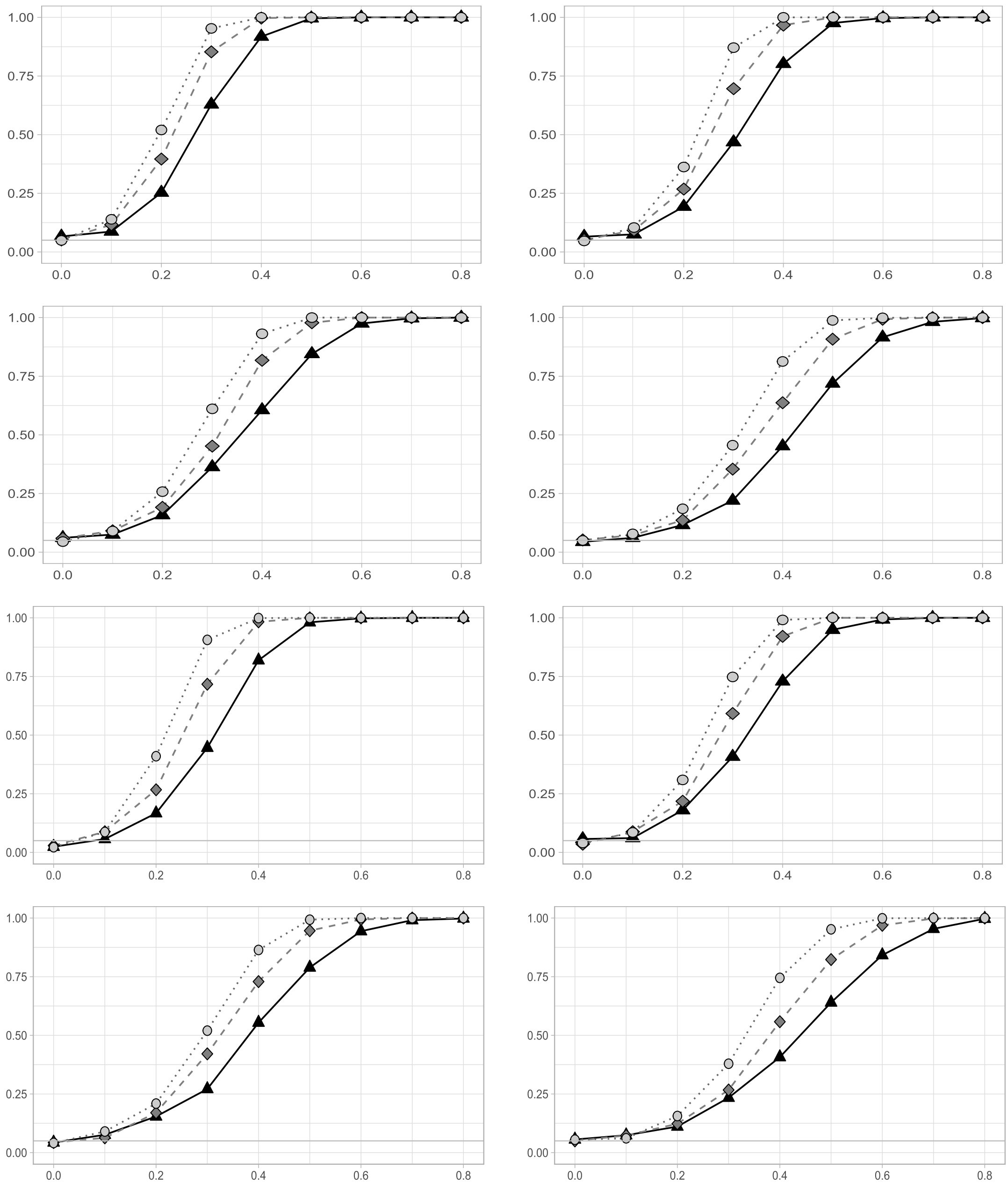


Figure 2. Power of the test (4.6) for randomfunctions (y-axis) against alternatives of different sizes a (x-axis), with $a = 0$ standing for the null hypothesis. The curves correspond to different sizes of the stable data set with $M = 50$ (solid), $M = 75$ (dashed) and $M = 100$ (dotted). The model errors are i.i.d. diffusion processes (first row), dependent diffusion processes (second row), i.i.d. Brownian motions (third row) and dependent Brownian motions (fourth row). On the left side, results for $k^* = 25$ and on the right side for $k^* = 75$ are displayed.

	Brownian motion		Diffusion process	
M	Independent	Dependent	Independent	Dependent
50	5.30%	4.95%	6.55%	5.25%
75	4.55%	4.65%	5.15%	5.55%
100	4.95%	4.80%	4.8%	4.75%

Table 1. Empirical sizes of the test (4.6) for a theoretical nominal level of $\alpha = 5\%$.

different values for the stable data set, $M \in \{50, 75, 100\}$. The potential change point location is $M + k^*$ with $k^* \in \{25, 75\}$.

In Table 1 (above), we report empirical sizes based on 2,000 simulation runs. In Figure 2, we display the power curve based on 1,000 simulation runs. In all different scenarios, we observe a close approximation of the nominal level if $a = 0$ where larger sample sizes generally lead to better approximations. As $a > 0$ grows, we see a rapid increase of power, where a larger size of the stable dataset M implies higher power. Moreover, power is higher for the earlier change at $k^* = 25$ (left) than for the larger value $k^* = 75$ (right). The introduction of dependence predictably reduces power, with the effect being equally strong for both diffusion processes and Brownian motions.

6.2. Monitoring time series of density estimators

We now focus on random densities considered in Section 5. Denote by $\varphi^{(\nu,\sigma)}$ the density of the normal distribution $\mathcal{N}(\nu, \sigma^2)$. Let $(\tilde{\nu}_n)_{n\in\mathbb{N}}$ be a sequence of independent uniformly distributed random variables on the interval $[0, 1]$. Therewith, we define the random mean parameters $\nu_n := \tilde{\nu}_n$ for independent data and $\nu_n := (\tilde{\nu}_n + \tilde{\nu}_{n-1})/2$ for dependent data. Moreover, independent of these means, we define the sequence of i.i.d. standard deviations $(\sigma_n)_{n\in\mathbb{N}}$ that are uniformly distributed on the interval $[0.01, 1]$. Therewith, we define the time series of random densities considered in (5.1):

$$f_n := \begin{cases} \varphi^{(\nu_n,\sigma_n)}, & n \le M + k^*, \\ \varphi^{(\nu_n+a,\sigma_n)}, & n > M + k^*. \end{cases}$$

The constant $a \ge 0$ quantifies the distance to H_0, and $a = 0$ corresponds to the null hypothesis of no change. For each n, we generate a sample of L i.i.d. observations, sampled from the density f_n. This sample is used to build the kernel density estimator $\hat{f}_n$.

In Figure 3, we display a few density estimates before and after the change point for $a = 0.5$. Throughout our simulations, we use $L = 30$ and choose the kernel function $\mathscr{K}$ as the Gaussian kernel. To calculate the density estimates, we employ the `density` function in `R` that is approximated on a grid of 50 evenly spaced points on $[-5, 5]$. Bandwidths are selected using the `SJ` method proposed by Sheather and Jones (1991). Otherwise, the setup is analogous as in the previous simulation section: We consider sizes of the stable data set $M \in \{50, 75, 100\}$. We fix the parameters $\zeta = 0.05$, $\gamma = 0.3$ and maintain a nominal level of $\alpha = 5\%$. Changes are located at $M + k^*$ with $k^* \in \{25, 75\}$. To calculate the detector $\widehat{\Gamma}_M$, we also specify the G-norm, where we use $B = 30$ and for $A = 2$ the function

$$\psi(x) := \begin{cases} \exp(x + A), & x < -A, \\ 1, & x \in [-A, A] \\ \exp(A - x), & x > A. \end{cases}$$

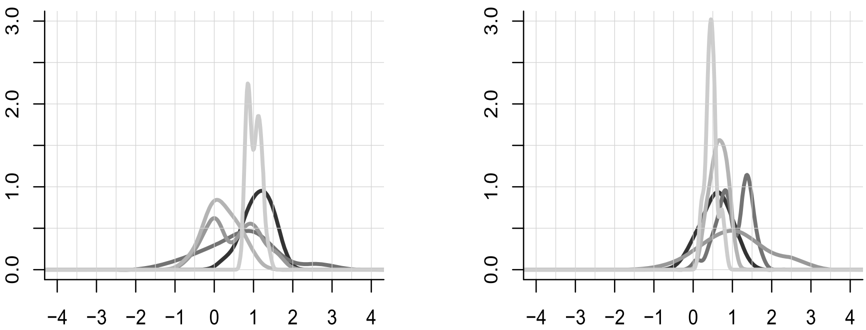


Figure 3. Realizations of five kernel density estimators $\hat{f}_n$ before (left) and after (right) the change point, for $a = 0.5$ in the independent setting. The sample size for each estimator is $L = 30$.

In Figure 4, we display the power curevs based on 1000 simulation run for each empirical probability. Our findings are similar to those in the previous section: larger M values are associated with higher power. Dependence has only a minor influence on the power for the test, while later changes lead to a more noticeable loss in power.

Table 2, based on 2,000 replications, shows that the monitoring scheme controls size very well.

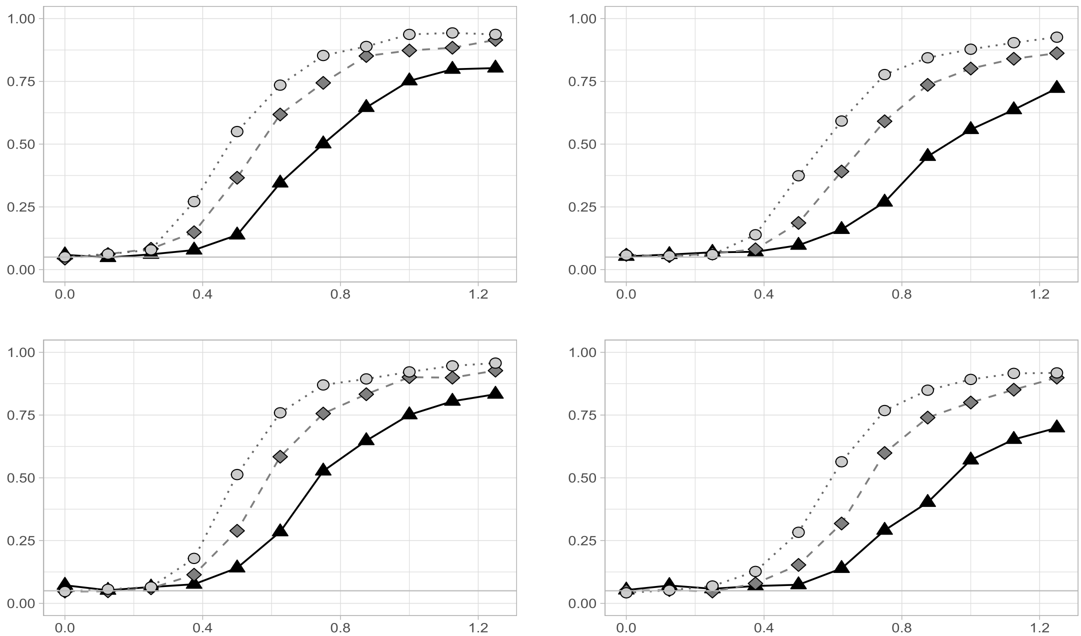


Figure 4. Power of the test (4.6) applied to random densities (y-axis) against alternatives of different sizes a (x-axis), with $a = 0$ standing for the null hypothesis. The curves correspond to different sizes of the stable data set with $M = 50$ (solid), $M = 75$ (dashed) and $M = 100$ (dotted). In the first row we consider the case of independent densities and in the second one the dependent case. On the left side, results for $k^* = 25$ and on the right side for $k^* = 75$ are displayed.

M	Independent	Dependent
50	6.55%	5.30%
75	4.55%	5.45%
100	4.90%	5.00%

Table 2. Empirical sizes of the test (4.6) for density estimators and a theoretical nominal level of $\alpha = 5\%$.

6.3. Application to exchange traded funds

Exchange Traded Funds (ETFs) are popular investment vehicles because they allow investors to make bets on entire sectors of the U.S. economy rather than choosing specific publicly traded companies. Researching individual companies is beyond the means and skills of most investors. To illustrate the statistical methodology developed in this paper, we consider four ETFs from the SPDR (Standard & Poor's Depositary Receipt) family. The sectors under consideration are finance (XLF), energy (XLE), industry (XLI) and consumer staples (XLP).

Denoting the logarithmic price of an ETF at time t by p_t, and dividing a trading day into $L+1$ equidistant time points $t_0, t_2, \ldots, t_L$, the intraday log-returns are defined as

$$x_{t_\ell} := p_{t_\ell} - p_{t_{\ell-1}}, \qquad \ell = 1, \ldots, L.$$

By construction, these returns are a measure of relative price change of the ETF. Repeating this process for every day in our observation period provides samples of the form $\{x_{t_\ell, n} : 1 \le \ell \le L\}$ that can be used to approximate the density of log-returns for the ETF on day n. (Forecasting of densities of intraday returns was studied by Horta and Ziegelmann (2018) and Kokoszka et al. (2019).) We consider log-returns over 5 minute intervals and only include regular trading days (9:30–16:00 EST). We have employed exactly the same kernel density estimators, bandwidth selection methods, definition of the G-norm and long-run variance estimator as in the previous section. As before, we have discretized the random functions on a grid with 50 equally spaced points, but this time on the interval $[-3, 3]$. This is a very conservative range because the five minute returns on funds that contain hundreds of companies basically never exceed $\pm 3\%$, see Figure 5 for an illustration. As before, we fix the nominal level at

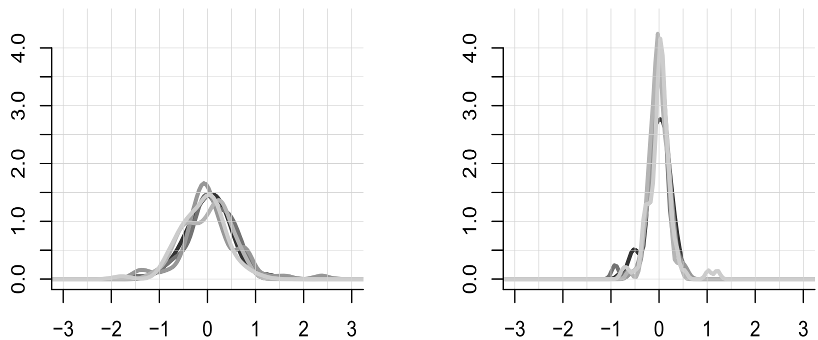


Figure 5. Realizations of five kernel density estimators $\hat{f}_n$ for the energy sector at the very beginning of the monitoring period (left) and after the first 100 days of the monitoring period (right).

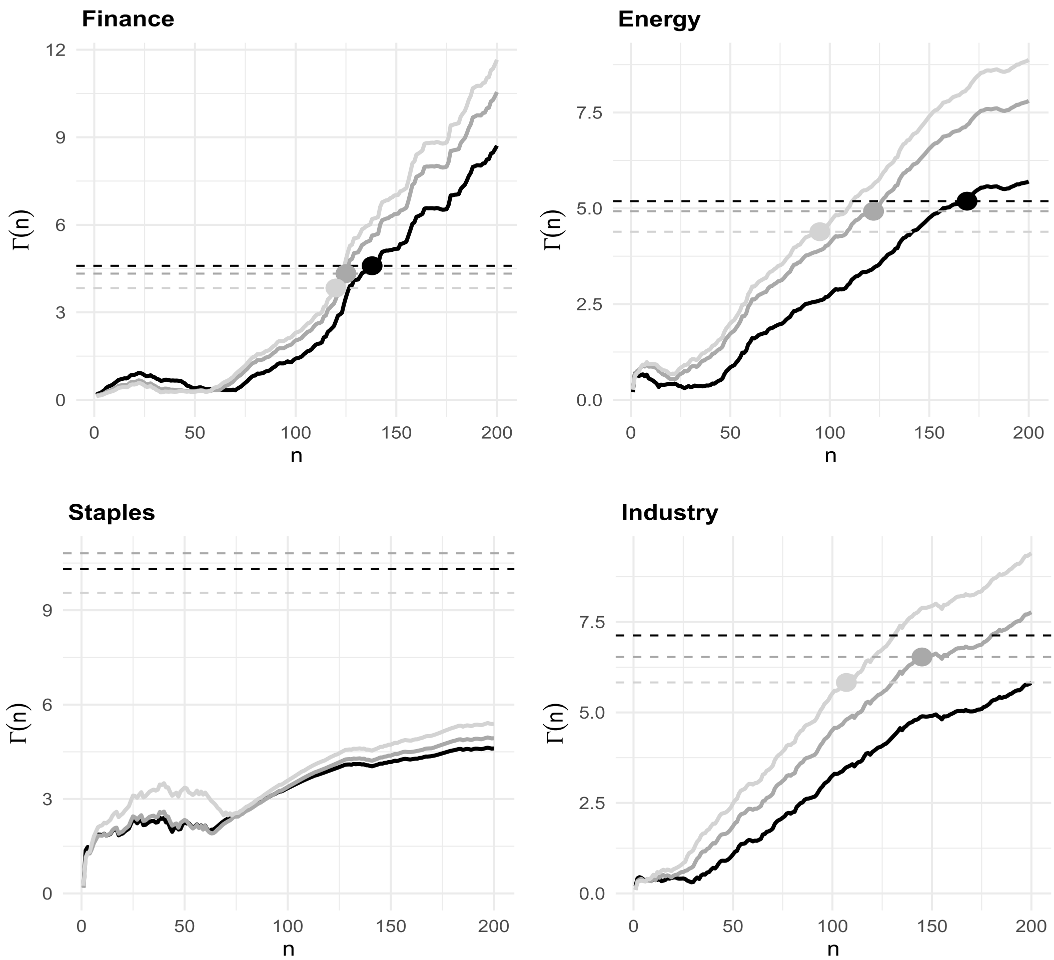


Figure 6. Statistic $\widehat{\Gamma}_M(n)$ (solid lines), for trading days $n = 1, 2, \ldots, 200$ starting on 02/01/2009, for $M = 50$ (black), $M = 75$ (dark grey) and $M = 100$ (light grey). Dashed lines indicate the respective upper 5% quantiles and large dots mark the first day n, where $\widehat{\Gamma}_M(n)$ surpasses this threshold.

$\alpha = 5\%$ and consider sizes of the initial period of $M = 50, 75, 100$ trading days, where the monitoring period begins with the first regular trading day in 2009. In Figure 5 we display examples of densities at the very beginning of and in mid 2009. We observe that at later dates the densities are less spread out, indicating a decline in the variance of log-returns. This shape change is visible in other sectors as well and might be associated with the less volatile environment of the fading recession.

However, it is difficult to determine by eye if and when, and for which sectors, a statistically significant change occurs, and this is even more difficult to determine in real time, after the close of each trading day. We applied our monitoring scheme to these densities. In Figure 6, we display the value of the test statistic $\widehat{\Gamma}_M(n)$ at each trading day n in the monitoring period for the different sectors. Changes are detected in the financial, energy and industry sector, while even for an initial period of $M = 100$ no changes are detected in consumer staples. The fact that no change in consumer staples is detected, is likely due to the higher stability of this sector during crises compared to other parts of the economy.

We emphasize, that the domains of the curves like those in Figure 6 would be growing each day. If a curve crosses a critical value, the monitoring would stop. The decision would then be that the density has changed sometime before the crossing point. The whole curves are shown only because we need to produce static graphs.

6.4. Additional simulation results

We provide additional simulation results to study the impact of the parameter choices γ and ζ. For this purpose, we consider the same simulation setup as in Section 6.1, while restricting ourselves to the case of independent errors of Brownian motion type. To study the impact of the parameter γ, we first fix $\zeta = 0.05$, the value from Section 6.1, and consider the parameter choices $\gamma \in \{0, 0.49\}$. Subsequently, we fix $\gamma = 0.3$, the value from Section 6.1, and consider $\zeta \in \{0, 0.1\}$. The resulting power curves are displayed in Figure 7. As we can see by comparing our results for $\gamma = 0$ with $\gamma = 0.49$, a small choice of γ enhances the power of our test procedure, most notably for later changes when $k^* = 75$. In can be shown that for very early changes such as $k^* = 5$ larger values for γ are beneficial, but even for $k^* = 25$, power is usually higher for $\gamma = 0$ compared to $\gamma = 0.49$. Importantly, this higher power can not be traced back to a higher Type I error. Indeed, in our simulations, $\gamma = 0$ leads to a slightly more conservative test than $\gamma = 0.49$. Altering ζ while keeping γ fixed seems to have no discernible effect on the outcomes.

6.5. Estimation of c_W

The estimation of the long-run covariance is an important part of inference for functional time series. For a rigorous investigation, with asymptotic consistency results, we refer the reader to Hörmann and Kokoszka (2010). Here, we only provide a small sketch of the estimation procedure. Recall the definition of the long-run variance kernel $c_W^{(1)}$ defined in (4.9). Defining the n-th lag function $c^{(1)}[n](s,t) := \mathbb{E}[\varepsilon_1^{(1)}(s)\varepsilon_{n+1}^{(1)}(t)]$, a standard estimation strategy for $c^{(1)}$ is to define the truncated series estimator

$$\hat{c}_W(s,t) := w_M(0)\hat{c}[0](s,t) + \sum_{n=1}^{b} 2w_M(n)\hat{c}[n](s,t).$$

Here $w_M : \mathbb{N}_0 \to [0,1]$ is a weight function with finite support, $b \in \mathbb{N}$ a bandwidth parameter and $\hat{c}[n](s,t)$ an estimator of the lag functions. For asymptotic considerations, it is usually assumed that the weights satisfy $w_M(n) \to 1$ as $M \to \infty$ (for any $n \in \mathbb{N}$) and that $b \to \infty$ at a sufficiently slow rate. The lag estimator in the definition of $\hat{c}_W$ is defined as

$$\hat{c}[n](s,t) := \frac{1}{M} \sum_{i=1}^{M-n} [X_i(s) - \bar{X}_M(s)][X_{i+n}(t) - \bar{X}_M(t)],$$

where $\bar{X}_M := \frac{1}{M}\sum_{i=1}^{M} X_i$ is the average of all functions during the stable period. In Section 6, we have used these estimators with the parameter choices $b = 2$ and $w_M(0) = w_M(1) = 1$ and $w_M(2) = 1/2$.

7. Summary

We conclude with a summary of our main contributions. We formulated a general paradigm of monitoring for a change point in the expected value of a sequence of random functions in a Banach space.

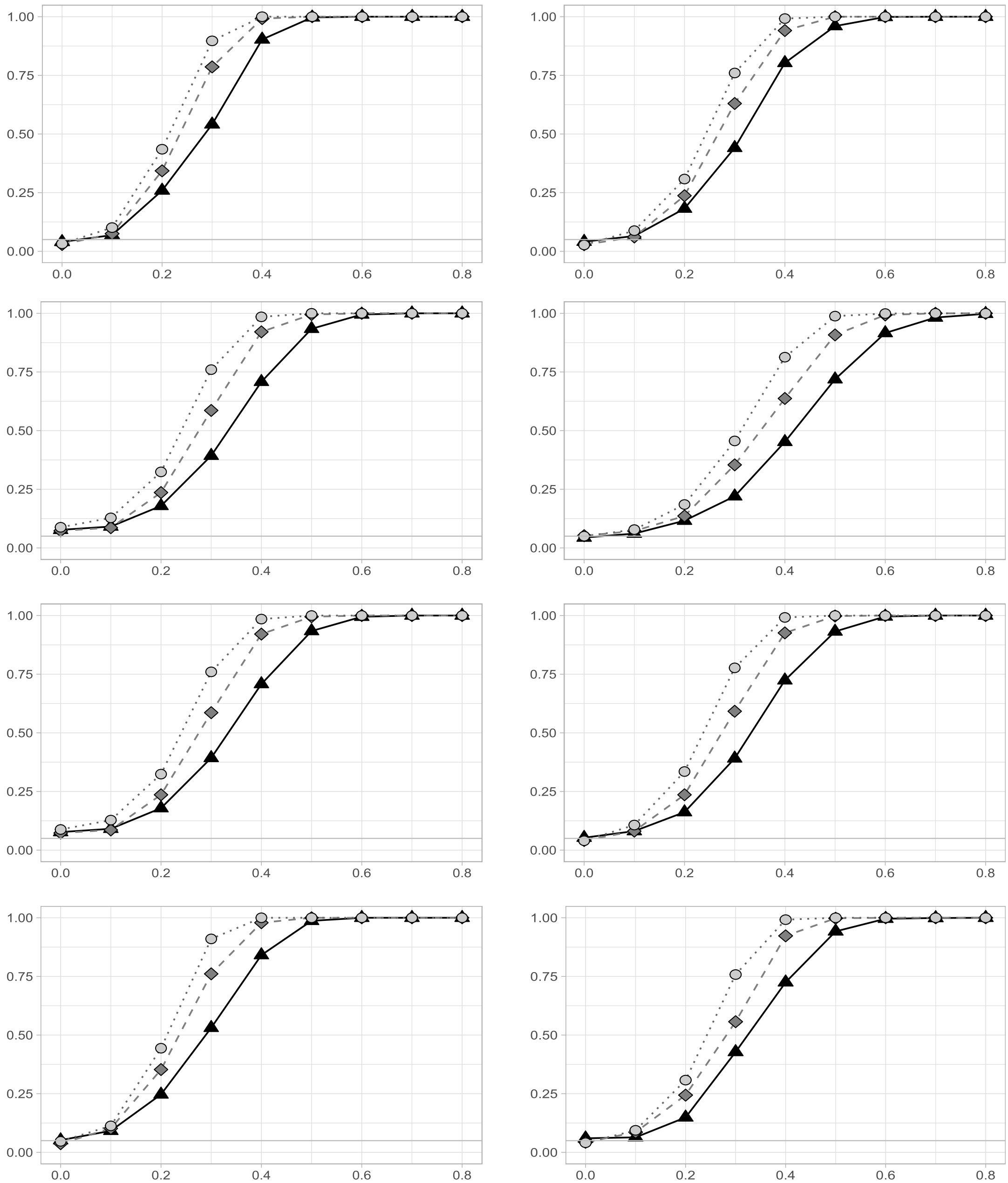


Figure 7. Power of the test (4.6) (y-axis) against alternatives of different sizes a (x-axis), with $a = 0$ denoting the null hypothesis. The curves correspond to different sizes of the stable data set with $M = 50$ (solid), $M = 75$ (dashed) and $M = 100$ (dotted). The model errors are i.i.d. Brownian motions and we display results for $k^* = 25$ (left) and $k^* = 75$ (right). The parameter choices are $(\gamma, \zeta) = (0, 0.05)$, (first row) $(\gamma, \zeta) = (0.49, 0.05)$, (second row) $(\gamma, \zeta) = (0.3, 0)$, (third row) $(\gamma, \zeta) = (0.3, 0.1)\}$, (fourth row).

In contrast to previous research, we do not use projections. Projections (coordinates) obviously arise in multivariate settings and are natural in separable Hilbert spaces due to the presence of countable orthonormal bases. While bases can be defined in separable Banach spaces in several ways, such concepts are much less natural, so our framework does not use them. We explained how detectors can be constructed in general Banach spaces and formulated relevant concepts of detection at a specified type I error, as well as consistent detection similar to the concept of local power.

We then formulated assumptions and established the above noted asymptotic properties in key Banach spaces used in statistical applications: We considered the space $C([-K,K])$ of continuous functions on a compact interval, the space $C_0(\mathbb{R})$ of continuous functions vanishing at $\pm\infty$, and the space $L^p([-K,K])$ with $p \geq 2$. Particularly novel arguments were developed in the case of $C_0(\mathbb{R})$ which has basically not been treated in existing FDA research, but is very important because probability density functions live in $C_0(\mathbb{R})$, in general. Temporal dependence is characterized by power-law decay of ϕ-mixing coefficients.

As a separate contribution, we showed that our approach can be applied to monitoring for a change in the mean of probability density functions. The density functions are not directly observed, but are reconstructed from i.i.d. scalar observations via smoothing. We established conditions that guarantee consistent detection with a controlled type I error in such two stage settings (a sequence of random densities, each observed through a random sample).

After verifying via simulations that our approach works in various settings, including the estimated densities, we applied it to monitoring for a change in the mean density of intraday returns (one density per day). Monitoring a period after the real-estate crisis of 2008, we detected changes in several sectors, but not in Consumer Staples, a sector containing companies focusing on everyday consumer needs.

Acknowledgements

Tim Kutta is the corresponding author.

Funding

This research was partially supported by the United States National Science Foundation grants DMS-2123761 and DMS-2412408.

Supplementary Material

Supplement to "Monitoring of functional time series" (DOI: 10.3150/24-BEJ1850SUPPA; .pdf). The online Supplement (Kutta and Kokoszka (2025)) provides detailed proofs, technical derivations, and additional background material. It includes an intuitive explanation for the choice of the boundary function g and a brief overview of random elements in Banach spaces.

Supplement to "Monitoring of functional time series" (DOI: 10.3150/24-BEJ1850SUPPB; .zip). The R code for our simulations.

References

Aston, J.A.D. and Kirch, C. (2012). Detecting and estimating changes in dependent functional data. *J. Multivariate Anal.* **109** 204–220. MR2922864 https://doi.org/10.1016/j.jmva.2012.03.006

Aue, A. and Kirch, C. (2024). The state of cumulative sum sequential changepoint testing 70 years after Page. *Biometrika* **111** 367–391. MR4745572 https://doi.org/10.1093/biomet/asad079

Aue, A., Rice, G. and Sönmez, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 509–529. MR3798876 https://doi.org/10.1111/rssb.12257

Aue, A., Rice, G. and Sönmez, O. (2020). Structural break analysis for spectrum and trace of covariance operators. *Environmetrics* **31** e2617, 14. MR4061134 https://doi.org/10.1002/env.2617

Aue, A., Horváth, L., Hušková, M. and Kokoszka, P. (2006). Change-point monitoring in linear models. *Econom. J.* **9** 373–403. MR2328569 https://doi.org/10.1111/j.1368-423X.2006.00190.x

Aue, A., Hörmann, S., Horváth, L., Hušková, M. and Steinebach, J.G. (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory* **28** 804–837. MR2959126 https://doi.org/10.1017/S0266466611000673

Aue, A., Hörmann, S., Horváth, L. and Hušková, M. (2014). Dependent functional linear models with applications to monitoring structural change. *Statist. Sinica* **24** 1043–1073. MR3241277

Berkes, I., Horváth, L. and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Process. Appl.* **123** 385–403. MR3003356 https://doi.org/10.1016/j.spa.2012.10.003

Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P. (2004). Sequential change-point detection in GARCH(p,q) models. *Econometric Theory* **20** 1140–1167. MR2101953 https://doi.org/10.1017/S0266466604206041

Berkes, I., Gabrys, R., Horváth, L. and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 927–946. MR2750251 https://doi.org/10.1111/j.1467-9868.2009.00713.x

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396

Bradley, R.C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability and Statistics (Oberwolfach, 1985)* (E. Eberlein and M.S. Taqqu, eds.). *Progr. Probab. Statist.* **11** 165–192. Boston, MA: Birkhäuser. MR0899990

Chen, Y., Wang, T. and Samworth, R.J. (2022). High-dimensional, multiscale online changepoint detection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 234–266. MR4400396 https://doi.org/10.1111/rssb.12447

Chu, C.S.J., Stinchcombe, M. and White, H. (1996). Monitoring structural change. *Econometrica* **64** 1045–1065.

Dehling, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. Verw. Gebiete* **63** 393–432. MR0705631 https://doi.org/10.1007/BF00542537

Dehling, H., Mikosch, T. and Sørensen, M. (2002). *Empirical Process Techniques for Dependent Data*. Boston, MA: Birkhäuser, Inc. MR1958776 https://doi.org/10.1007/978-1-4612-0099-4

Dette, H. and Kokot, K. (2022). Detecting relevant differences in the covariance operators of functional time series: A sup-norm approach. *Ann. Inst. Statist. Math.* **74** 195–231. MR4396673 https://doi.org/10.1007/s10463-021-00795-2

Dette, H., Kokot, K. and Aue, A. (2020). Functional data analysis in the Banach space of continuous functions. *Ann. Statist.* **48** 1168–1192. MR4102692 https://doi.org/10.1214/19-AOS1842

Dette, H., Kokot, K. and Volgushev, S. (2020). Testing relevant hypotheses in functional time series via self-normalization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 629–660. MR4112779 https://doi.org/10.1111/rssb.12370

Dette, H. and Kutta, T. (2021). Detecting structural breaks in eigensystems of functional time series. *Electron. J. Stat.* **15** 944–983. MR4255289 https://doi.org/10.1214/20-ejs1796

Dette, H. and Wied, D. (2016). Detecting relevant changes in time series models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 371–394. MR3454201 https://doi.org/10.1111/rssb.12121

Diestel, J. and Uhl, J.J. Jr. (1976). The Radon-Nikodym theorem for Banach space valued measures. *Rocky Mountain J. Math.* **6** 1–46. MR0399852 https://doi.org/10.1216/RMJ-1976-6-1-1

Dubey, P. and Müller, H.-G. (2020). Fréchet change-point detection. *Ann. Statist.* **48** 3312–3335. MR4185810 https://doi.org/10.1214/19-AOS1930

Fischer, M. and Nappo, G. (2009). On the moments of the modulus of continuity of Itô processes. *Stoch. Anal. Appl.* **28** 103–122. MR2597982 https://doi.org/10.1080/07362990903415825

Giné, E. and Nickel, R. (2016). *Infinite Dimensional Statistics*. Cambridge Univ. Press.

Gösmann, J. (2020). New aspects of sequential change point detection. PhD thesis, Ruhr-University Bochum.

Gösmann, J., Kley, T. and Dette, H. (2021). A new approach for open-end sequential change point monitoring. *J. Time Series Anal.* **42** 63–84. MR4192928 https://doi.org/10.1111/jtsa.12555

Gromenko, O., Kokoszka, P. and Reimherr, M. (2017). Detection of change in the spatiotemporal mean function. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 29–50. MR3597963 https://doi.org/10.1111/rssb.12156

Hoffmann-Jørgensen, J. (1974). Sums of independent Banach space valued random variables. *Studia Math.* **52** 159–186. MR0356155 https://doi.org/10.4064/sm-52-2-159-186

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. MR2662361 https://doi.org/10.1214/09-AOS768

Horta, E. and Ziegelmann, F. (2018). Dynamics of financial returns densities: A functional approach applied to the Bovespa intraday index. *Int. J. Forecast.* **34** 75–88.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *J. Econometrics* **179** 66–82. MR3153649 https://doi.org/10.1016/j.jeconom.2013.11.002

Horváth, L., Kokoszka, P. and Wang, S. (2021). Monitoring for a change point in a sequence of distributions. *Ann. Statist.* **49** 2271–2291. MR4319250 https://doi.org/10.1214/20-aos2036

Horváth, L. and Rice, G. (2014). Extensions of some classical methods in change point analysis. *TEST* **23** 219–255. MR3210268 https://doi.org/10.1007/s11749-014-0368-4

Horváth, L., Rice, G. and Zhao, Y. (2022). Change point analysis of covariance functions: A weighted cumulative sum approach. *J. Multivariate Anal.* **189** Paper No. 104877, 23. MR4384119 https://doi.org/10.1016/j.jmva.2021.104877

Horváth, L., Hušková, M., Kokoszka, P. and Steinebach, J. (2004). Monitoring changes in linear models. *J. Statist. Plann. Inference* **126** 225–251. MR2090695 https://doi.org/10.1016/j.jspi.2003.07.014

Janson, S. and Kaijser, S. (2015). Higher moments of Banach space valued random variables. *Mem. Amer. Math. Soc.* **238** vii+110. MR3402381 https://doi.org/10.1090/memo/1127

Kojadinovic, I. and Verdier, G. (2021). Nonparametric sequential change-point detection for multivariate time series based on empirical distribution functions. *Electron. J. Stat.* **15** 773–829. MR4203345 https://doi.org/10.1214/21-EJS1798

Kokoszka, P., Miao, H., Petersen, A. and Shang, H.L. (2019). Forecasting of density functions with an application to cross-sectional and intraday returns. *Int. J. Forecast.* **35** 1304–1317.

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883 https://doi.org/10.1007/BF00532688

Kuelbs, J. (1973). The invariance principle for Banach space valued random variables. *J. Multivariate Anal.* **3** 161–172. MR0328986 https://doi.org/10.1016/0047-259X(73)90020-1

Kutta, T. and Kokoszka, P. (2025). Supplement to "Monitoring of functional time series." https://doi.org/10.3150/24-BEJ1850SUPPA, https://doi.org/10.3150/24-BEJ1850SUPPB

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]* **23**. Berlin: Springer. MR1102015 https://doi.org/10.1007/978-3-642-20212-4

Matteson, D.S. and James, N.A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *J. Amer. Statist. Assoc.* **109** 334–345. MR3180567 https://doi.org/10.1080/01621459.2013.849605

Pape, K., Wied, D. and Galeano, P. (2016). Monitoring multivariate variance changes. *J. Empir. Finance* **39** 54–68.

Petersen, A., Zhang, C. and Kokoszka, P. (2022). Modeling probability density functions as data objects. *Econom. Stat.* **21** 159–178. MR4366852 https://doi.org/10.1016/j.ecosta.2021.04.004

Račkauskas, A. and Suquet, C. (2009). Hölderian invariance principle for Hilbertian linear processes. *ESAIM Probab. Stat.* **13** 261–275. MR2528083 https://doi.org/10.1051/ps:2008011

Rice, G. and Shum, M. (2019). Inference for the lagged cross-covariance operator between functional time series. *J. Time Series Anal.* **40** 665–692. MR3995663 https://doi.org/10.1111/jtsa.12447

Schechter, E. (1997). *Handbook of Analysis and Its Foundations*. San Diego, CA: Academic Press. MR1417259

Shao, X. (2015). Self-normalization for time series: A review of recent developments. *J. Amer. Statist. Assoc.* **110** 1797–1817. MR3449074 https://doi.org/10.1080/01621459.2015.1050493

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *J. Amer. Statist. Assoc.* **105** 1228–1240. MR2752617 https://doi.org/10.1198/jasa.2010.tm10103

Sheather, S.J. and Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **53** 683–690. MR1125725

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. *Springer Series in Statistics*. New York: Springer. MR0799155 https://doi.org/10.1007/978-1-4757-1862-1

Stoehr, C., Aston, J.A.D. and Kirch, C. (2021). Detecting changes in the covariance structure of functional time series with application to fMRI data. *Econom. Stat.* **18** 44–62. MR4238907 https://doi.org/10.1016/j.ecosta.2020.04.004

Utev, S.A. (1991). Sums of weakly dependent random variables. *Sib. Math. J.* **32** 675–690.

Wied, D. and Galeano, P. (2013). Monitoring correlation change in a sequence of random variables. *J. Statist. Plann. Inference* **143** 186–196. MR2969021 https://doi.org/10.1016/j.jspi.2012.06.007

Wu, T., Wang, R., Yan, H. and Shao, X. (2022). Adaptive change point monitoring for high-dimensional data. *Statist. Sinica* **32** 1583–1610. MR4449907 https://doi.org/10.5705/ss.202020.0438

Received June 2024 and revised December 2024