JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **47**: 660–674 (2026)
Published online 21 November 2024 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12796

SPECIAL ISSUE ARTICLE

MONITORING PANELS OF SPARSE FUNCTIONAL DATA

TIM KUTTA[a] AGNIESZKA JACH[b] AND PIOTR KOKOSZKA[a*]
[a]*Colorado State University, Fort Collins, Colorado, USA*
[b]*Hanken School of Economics, Helsinki, Finland*

Panels of random functions are common in applications of functional data analysis. They often occur when sequences of functions are observed at a number of different locations. We propose a methodology to monitor for structural breaks in such panels and to identify the changing components with statistical certainty. Our approach relies on a Full-CUSUM statistic that has proved to be powerful in finite dimensions but has not been applied to functional data. To account for the practically relevant problem of sparsity, we formulate our results for triangular arrays of nonstationary, sparse estimators. The derivation of our asymptotic theory relies on new Gaussian approximations on the Banach space of continuous functions, which imply new convergence results for the change point detectors. We illustrate our approach with a simulation study and application to intraday returns on exchange traded funds.

Received 26 August 2024; Accepted 25 October 2024

Keywords: change point monitoring; functional time series; sparse functional data.

MSC subject classification: 62L10; 62M10.

1. INTRODUCTION

We propose new methodology to detect and localize changes in panels of functional data. Our approach is formulated in a monitoring setup where a training set is available at the onset of the procedure and afterwards functional observations arrive sequentially. With each incoming observation, the user has to decide whether a change has already occurred, and if so, in which component of the panel it is located. This monitoring process might go on for a fixed, predetermined period (closed-end scenario) or indefinitely (open-end scenario).

To illustrate the problem we solve, we refer to Figure 1 that shows five consecutive observations of a multivariate time series of functions. The data are described in Section 5. We refer to such a data structure as a functional panel. In applications, we observe several hundred functions in each panel components. In our application, we observe a new panel of nine function every day. Each day, we want to make a decision if the mean function of each of the nine components has changed or not. We however do not treat each component separately but want to sequentially identify the set of components that have undergone a change up to the current day. We want to do it is such a way that this set is identified with a prescribed probability of type I error.

Sequential tests have been developed in various contexts, and for a detailed overview of the historical development, we refer to Aue and Kirch (2024). While monitoring procedures exist for different data structures (univariate, multivariate, high dimensional), their extension to infinite dimensional, functional data have proved challenging, partly due to the scarcity of appropriate probabilistic tools on abstract Hilbert and Banach spaces. The only work in this direction that we could find is Aue *et al.* (2014), who consider a change in a multivariate, projected parameter with fixed dimension. Yet, it is of natural interest to extend inference to functional (infinite dimensional) parameters, such as the mean function, as typically considered in retrospective analyses (see, e.g., Horváth *et al.* (2014), Dubey and Müller (2019), Dette *et al.* (2020b)). Such fully functional approaches will

*Correspondence to: Piotr Kokoszka, Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA.
E-mail: Piotr.Kokoszka@colostate.edu

[This article is part of the special issue "Data Segmentation"]

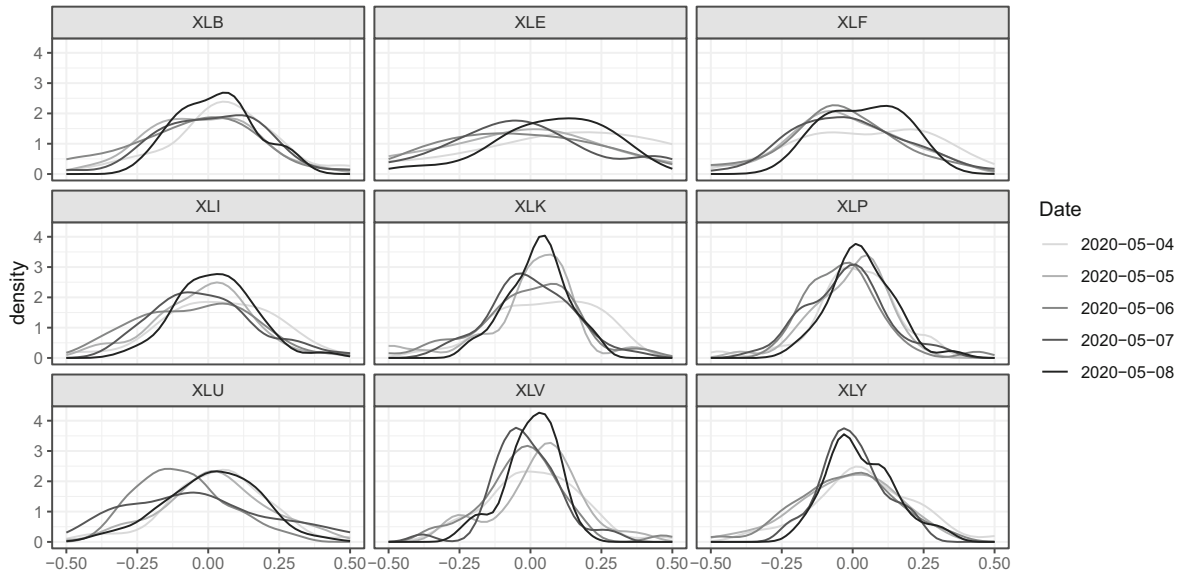

Figure 1. Estimated densities of intraday 10-min returns on five consecutive days for nine Exchange Traded Funds.

be developed in this paper. Applications of change point analyses for functional time series range across disciplines, including finance and economics (Dette and Wied, 2016; Bardsley *et al.*, 2017; Horváth *et al.*, 2021; Horváth *et al.*, 2022), climate research (Berkes *et al.*, 2009; Zhang *et al.*, 2011; Gromenko *et al.*, 2017; Tucker and Yarger, 2023; Chen *et al.*, 2023), medical applications (Aston and Kirch, 2012b; Stöhr *et al.*, 2021; Bastian *et al.*, 2023). To create relevant inference methods, one has to account for two common data features, namely, panel structure and sparsity. Panels of dependent random functions typically occur when sequences of functions are measured at a number of (broadly understood) locations. Examples include annual temperature curves at different places or daily return curves on different assets. Sparsity refers to effects on the statistical analysis introduced by imperfect reconstructions of curves, due to finite resolution and observational noise. It is important to investigate when and how reconstruction effects impinge on statistical inference and when they can be safely ignored.

We develop a procedure that detects and localizes changes in a panel of functions as new data arrive. To the best of our knowledge, this is the first procedure that sequentially locates the components that change. More precisely, suppose we have at each time period $n = 1, 2, \ldots$ a vector $(X_n[\ell])_{\ell=1,\ldots,d}$ of d random functions. We want to find out whether at time n there have (already) occurred changes, and if so in which components ℓ. Those components where a change has been detected at time n are gathered in the set $\widehat{\mathcal{I}}(n)$, that is sequentially updated for each n. We construct $\widehat{\mathcal{I}}(n)$ to be asymptotically consistent (all changes are included in $\widehat{\mathcal{I}}(n)$), while holding a user-determined error rate $\alpha \in (0, 1)$ (locations are only erroneously included in $\widehat{\mathcal{I}}(n)$ at global rate α). Asymptotics are formulated for an increasing training data set. Our analysis is applicable to open- and closed-end scenarios. Open-end scenarios here refer to procedures that can potentially run indefinitely, while still respecting the global error rate. In such scenarios, functional weak invariance principles on compact intervals are insufficient, and our theory rests on a new Gaussian approximation in function spaces, recently derived by Kutta and Kokoszka (2024). Finally, our theory encompasses sparse functional data, with nonstationary reconstructions as they often occur in practice, which may even have infinite support. In the majority of FDA research, it is assumed that functions are defined on a common compact interval, but to accommodate densities, functions defined on an unbounded subset of the line must be considered.

DOI: 10.1111/jtsa.12796 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

The remainder of this paper is organized as follows: In Section 2, we place our contribution in a broader setting and review related literature on change point detection and sequential testing. Section 3 provides the mathematical and notational background for the statistical procedures derived in Section 4. A simulation study and an application to monitoring of intraday return densities are presented in Section 5. Proofs of mathematical results and some details are deferred to the Online Supporting Information.

2. THE SETTING AND RELATED CONTRIBUTIONS

The literature on change point detection is so extensive that we can only review a small fraction of representative works, most closely related to our own. We will split our review into two parts, first focusing on change point detection in FDA and then on monitoring of nonfunctional sequences. The recent monograph of Horváth and Rice (2024) provides a broad and in-depth overview of recent advances in change point research.

Change point detection in FDA. Retrospective change point detection has been extensively explored in the framework of FDA. To keep our discussion succinct, we want to focus here on statistical tests for changes and omit related, but distinct research topics, such as change point estimation or segmentation (such as found in Hariz *et al.* (2007), Chiou *et al.* (2019), Wang *et al.* (2023)). Consider therefore a time series $(X_n)_{n\in\mathbb{N}}$ with corresponding parameters $(\theta_n)_{n\in\mathbb{N}}$, where θ_n depends on the marginal distribution $\mathcal{L}(X_n)$ (say a mean). Inference is based on a sample $X_1, \dots, X_N$ of size N, and we want to test the hypotheses pair of θ_n being constant over time, the null hypothesis, against a change at some point n^*, the alternative, that is,

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_N, \qquad \text{vs.} \qquad H_1 : \theta_1 = \cdots = \theta_{n^*} \neq \theta_{n^*+1} = \theta_{n^*+2} = \cdots = \theta_N.$$

In FDA, the X_n take values in a function space, typically the separable Hilbert space $L^2([0,1])$ or sometimes in the Banach space $\mathcal{C}([0,1])$ of continuous functions equipped with the sup-norm. The space $L^2([0,1])$ has many favourable analytic properties, for example, there exists a straightforward extension of PCA to this space. Nevertheless, considering functions on $\mathcal{C}([0,1])$ can be statistically advantageous, given the strength and simple geometric interpretation of the norm. Change point tests for L^2-valued data were first developed by Berkes *et al.* (2009) who combined projections on finite dimensional subspaces with multivariate invariance principles. Various extensions have been proposed. Aston and Kirch (2012a) considered epidemic changes. Horváth *et al.* (2014) developed fully functional tests (without the need of projections) for weakly dependent time series. The corresponding probabilistic result, a weak invariance principle for m-approximable Hilbert space valued time series, was derived by Berkes *et al.* (2013) and has since then been used repeatedly in change point detection research. Similar methods were used in the investigation of multivariate functional time series in Gromenko *et al.* (2017) who studied a similar data structure as considered in this paper. Relevant changes (changes beyond a certain magnitude) were tested in Dette *et al.* (2020b) using a self-normalized test statistic that is asymptotically distribution-free. Structural breaks of time series in the space $\mathcal{C}([0,1])$ were studied in Dette *et al.* (2020a) and in general metric spaces by Dubey and Müller (2019). While most of the above works focus on changes in the mean function of a functional time series, other parameters have been considered, such as covariance operators in Stöhr *et al.* (2021) and Horváth *et al.* (2022), the cross-covariance operator in Rice and Shum (2019), the spectrum of covariance operators in Aue *et al.* (2020) or the eigensystems of the covariance in Dette and Kutta (2021). Madrid Padilla *et al.* (2022) propose an algorithm for the estimation of multiple change points for sparse functions observed on a multivariate spatial domain; the functions are real-valued (no panel). Madrid Padilla *et al.* (2023) study a similar problem, but rather than a single function, a multivariate time series with change points is observed.

Sequential monitoring. The problem we study is related to the following general testing problem:

$$H_0 : \theta_1 = \cdots = \theta_M = \theta_{M+1} = \theta_{M+2} = \cdots \tag{2.1}$$

versus

$$H_1 : \exists\, k^* \ \theta_1 = \cdots = \theta_{M+k^*} \neq \theta_{M+k^*+1} = \theta_{M+k^*+2} = \cdots \tag{2.2}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12796

The first M observations are assumed not to contain a change point, and we want to test, as new data arrive, if there is a change sometime after time M. Problem (2.1) to (2.2) has been studied only in finite dimensional settings related to multiple regression and parametric time series. Chu *et al.* (1996) introduced it 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.* (2003) who considered a large family of boundary functions and two types of regression residuals. Extensions to various scalar time series models were developed by Berkes *et al.* (2004) and Aue *et al.* (2006). Gösmann *et al.* (2021) modified the monitoring paradigm and developed methodology suitable for multivariate time series. Horváth *et al.* (2021) worked out monitoring of quantile functions in various tail behaviour 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). Wu *et al.* (2022) consider monitoring of independent high-dimensional observations. 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 infinite dimensional spaces*. Problem (2.1) to (2.2) is related to classical sequential testing. The chief differences are as follows: (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 impractical to define in most functional settings. (3) Sequential analysis focuses on expected detection delay. We focus on probabilities of correct decisions. Moreover, we are chiefly concerned with the sequential identification of the components of a multivariate functional time series where changes have occurred.

3. PRELIMINARIES

3.1. Mathematical Preliminaries

Vector space notation. We often consider the d-dimensional real vector space, equipped with the maximum norm, which is denoted by $(\mathbb{R}^d, |\cdot|)$. For a non-empty index set $\mathcal{I} \subset \{1, \dots, d\}$ and a vector $v \in \mathbb{R}^d$, we define the subvector with components in $\mathcal{I}$ as $v^{(\mathcal{I})} := (v[\ell])_{\ell \in \mathcal{I}}$.

Space of continuous functions. Throughout this work, we denote by I a non-empty subinterval of the real line that is either compact or equal to $\mathbb{R}$. By $\mathcal{C}(I, \mathbb{R}^d)$ we denote the space of bounded continuous functions $f : I \to \mathbb{R}^d$. Any such function can be interpreted as a vector of continuous functions $f = (f[1], \dots, f[d])$ with $f[\ell] \in \mathcal{C}(I, \mathbb{R})$. A natural distance on $\mathcal{C}(I, \mathbb{R}^d)$ is given by the sup-norm

$$\|f\| := \sup_{u \in I} |f(u)| = \sup_{u \in I} \max_{1 \le \ell \le d} |f[\ell](u)|.$$

Notice that in both the max-norm (for vectors) and sup-norm (for vectors of functions), we do not make dependence on the dimension explicit, for the sake of simplicity. For the special case of $I = \mathbb{R}$, we define the closed subspace $\mathcal{C}_0(I, \mathbb{R}^d)$ of $\mathcal{C}(I, \mathbb{R}^d)$, as the set of functions that vanish at $\pm\infty$, that is,

$$\lim_{y \to \infty} \sup_{|u| > y} |f(u)| = 0.$$

We subsequently present unified results for the space of continuous functions on a compact interval I and for the space of functions that vanish at infinity with $I = \mathbb{R}$ and hence generically write $\mathcal{C}_0(I, \mathbb{R}^d)$. We recall that, equipped with the sup-norm, both of these spaces are Banach spaces.

Random functions in $\mathcal{C}_0$. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We call a measurable map $X : \Omega \to \mathcal{C}_0(I, \mathbb{R}^d)$ a random function in $\mathcal{C}_0(I, \mathbb{R}^d)$. Recall that the measurability of X is equivalent to the measurability of all evaluations $X(u)$ for $u \in I$, see p. 84 in Billingsley (1968). If $\mathbb{E}\|X\| < \infty$, we can define the mean function $\mu \in \mathcal{C}_0(I, \mathbb{R}^d)$ by

the point-wise identity

$$\mathbb{E}\,[X[\ell](u)] = \mu[\ell](u),$$

for all $u \in I$ and $1 \le \ell \le d$. Moreover, if $\mathbb{E}\|X\|^2 < \infty$, we can define its continuous covariance kernel c by the pointwise identity

$$c[\ell, k](u, v) := \mathbb{E}[(X[\ell](u) - \mu[\ell](u))(X[k](v) - \mu[k](v))], \quad 1 \le \ell, k \le d,\ u, v \in I.$$

For details on continuous random functions, we refer the reader to Janson and Kaijser (2015).

The Banach space valued Brownian motion. To describe the limiting distribution of our test statistics, we employ the Brownian motion in $C_0(I, \mathbb{R}^d)$. This Brownian motion $\{W(x) : x \in [0, 1]\}$ is a stochastic process with continuous sample paths in $C_0(I, \mathbb{R}^d)$ that has stationary, independent, and normally distributed increments. It is characterized by its mean $\mathbb{E}W(1) \in C_0(I, \mathbb{R}^d)$ (in our case, always the constant zero function) and the covariance of $W(1)$. Since $W(x)$ is a random function, we sometimes write its evaluation at $u \in I$ as $W(x, u) := W(x)(u)$.

3.2. Sparse Functional Data

Latent functions and sparse estimators. The random functions of interest need not be directly observable and must be reconstructed from a set of scalar observations. We denote the underlying functions by $X_1, X_2, \ldots \in C(I, \mathbb{R}^d)$ and call them *latent functions*. Even though we have no direct access to the latent functions, we want to make statistical inference about their distribution. To approximate the unavailable latent functions, we will rely on *sparse estimators* denoted by $X_{1,M}, X_{2,M}, \ldots$ that are available to the user. Roughly speaking, we assume that the quality of the approximation of X_i by $X_{i,M}$ increases with the length M of the training period. Related asymptotics have been considered in monitoring of high-dimensional time-series, where the dimension of the data increases with M (see Horváth *et al.*, 2021; Gösmann *et al.*, 2022). Precise assumptions are formulated in Section 4.

Distortions. When talking about sparse estimators of latent functions, it is helpful to think about randomness on two levels. First, there is randomness involved in drawing the function X_i from a distribution on $C(I, \mathbb{R}^d)$. Second, there is randomness in the process of reconstruction. Say, we have $d = 1$ and observe the function X_i on a grid of points $g_1, \ldots, g_M$

$$y_{i,m} := X_i(g_m) + \epsilon_{i,m}, \qquad m = 1, \ldots, M,$$

with real valued model errors $\epsilon_{i,m}$. We use a nonparametric estimator $X_{i,M}$ to reconstruct X_i from the sample $\{(y_{i,M}, g_m) : 1 \le m \le M\}$. Then, $X_{i,M}$ not only depends on the randomness of X_i but also on the randomness in the second layer–the model errors $\epsilon_{i,M}$. Abstractly, we express this second source of randomness as a variable $\delta_{i,M}$ that takes values in some metric space $\mathcal{S}_M$. In the previous example, $\delta_{i,M} = (\epsilon_{i,1}, \ldots, \epsilon_{i,M}) \in \mathbb{R}^M$. We can then conceptualize the sparse estimators $X_{i,M}$ as a measurable transform, say

$$\mathscr{A}_M : C_0(I, \mathbb{R}^d) \times \mathcal{S}_M \to C_0(I, \mathbb{R}^d), \tag{3.1}$$

with

$$X_{i,M} := \mathscr{A}_M(X_i, \delta_{i,M}). \tag{3.2}$$

We call the random variables $\delta_{i,M}$ the *distortions*, since they only allow us to obtain an imperfect reconstruction of the X_i.

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12796

Remark 3.1.

1) *The meaning of sparsity*: There exist different notions of sparsity in the context of functional data. Some authors refer to sparsity as a setup where the estimator $X_{i,M}$ for X_i does not improve (in our above regression example, the sample size M would be fixed). In this work, we present asymptotics for the case of improving estimators $X_{i,M} \to X_i$ and call this a sparse setup to contrast it with scenarios where the entire latent function X_i is available.
2) *A general setup*: Most works on sparse functional data (in one of the above senses) investigate a specific sampling regime and reconstruction method for $X_{i,M}$. To make our results more generally applicable, we formulate them abstractly for sparse estimators $X_{i,M}$ of the form (3.2) and only impose some high-level conditions on them. Specifically, we will formulate in Assumption 4.1 certain uniform consistency and smoothness properties of the estimators, which can be validated in many different cases and are often intuitively plausible. Many common estimators can be fit into the form (3.2). Examples include, for nonparametric regression, the Nadarya-Watson estimator or the local polynomial estimator. In the case of nonparametric density estimation, the kernel density estimator fits the above setup.
3) *Fully observed functions*: The case where the functions X_i are observed without any error is also covered by our above formulation if we set $\mathscr{A}_M(f, \delta_{i,M}) = \mathscr{A}(f) = f$.
4) *Temporal dependence*: It is intuitively clear that if both the latent functions X_i and the distortions $\delta_{i,M}$ are weakly dependent, the sparse estimators $X_{i,M}$ are also weakly dependent. It often makes sense to treat the distortions $\delta_{i,M}$ (microstructure) as independent of the latent functions $(X_i)_{i\in\mathbb{N}}$ (macrostructure) and to assume that distortions are independent across i while the latent functions are time-dependent. We will model temporal dependence by strong mixing, see Assumption 4.1, part i).

4. MONITORING PANELS OF FUNCTIONAL DATA

In this section, we present our main contributions. We develop monitoring schemes (open-end and closed-end) for multivariate time series of sparsely observed random functions. Our approach is based on a Full-CUSUM statistic that has not been employed in the monitoring of functional data before. As a main probabilistic result, we establish the convergence of the change point detector's cdf at polynomial speed with increasing size of the training set. Such results have not been available before, even for one-dimensional data. We use this convergence to create a sequential set estimator that, with a prescribed failure rate, determines those components where a change has occurred.

4.1. The Setup

Monitoring and the change point model. We consider a time series of latent functions $X_1, X_2, X_3, \ldots \in \mathcal{C}_0(I, \mathbb{R}^d)$ with mean $\mu_n := \mathbb{E}[X_n]$. Recall that X_n is a multivariate function $X_n := (X_n[1], \ldots, X_n[d])$ with $X_n[\ell] \in \mathcal{C}_0(I, \mathbb{R})$ and μ_n as a component-wise mean $\mu_n[\ell] = \mathbb{E}X_n[\ell]$ (for details see Section 3). We are interested in monitoring for changes in the mean function and identifying in which components $\ell = 1, \ldots, d$ those changes were located. Here, monitoring refers to a statistical setup, where the mean function is constant on an initial training data set of known length M. After this initial stable period, the monitoring period begins, where new observations are sequentially made and the user has to decide at each point whether a change has already occurred or not. The procedure ends after $\lfloor TM \rfloor$ observations, where $T > 1$ determines the relative length of the monitoring period compared to the stable period. For $T < \infty$ the setup is referred to as a *closed-end* regime, and for $T = \infty$ as an *open-end* regime, where monitoring continues for a (theoretically) infinite time span. Statistical results are formulated for $M \to \infty$, i.e. for more training data. Mathematically, we can express our mean change model as follows: There exist mean functions $\mu^{(1)}, \mu^{(2)} \in \mathcal{C}_0(I, \mathbb{R}^d)$ and parameters $\tau_1, \tau_2, \ldots, \tau_d \in (1, T]$ such that for $n = 1, \ldots, \lfloor TM \rfloor$ and $\ell = 1, \ldots, d$

$$\mu_n[\ell] = \begin{cases} \mu^{(1)}[\ell], & \text{if } 1 \le n \le \lfloor M\tau_\ell \rfloor \\ \mu^{(2)}[\ell], & \text{if } \lfloor M\tau_\ell \rfloor < n \le \lfloor TM \rfloor. \end{cases} \tag{4.1}$$

DOI: 10.1111/jtsa.12796 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

We assume that $\mu^{(1)}[\ell] \neq \mu^{(2)}[\ell]$ for any $\ell = 1, \dots, d$ and thus a change in component ℓ occurs if and only if $\tau_\ell < T$. Changes may occur simultaneously or at different times, and some components may experience changes, whereas others may not. We are interested in mean changes of the latent functions X_n, which are only accessible through sparse estimators $X_{n,M}$ defined in Section 3.2 with means $\mu_{n,M} := \mathbb{E}X_{n,M}$.

The inference problem. Our objective is to find a suitable set estimator for the components of the latent functions that experience a mean change. We can express this set mathematically as

$$\mathcal{I}^* := \{\ell : 1 < \tau_\ell < T\}. \tag{4.2}$$

Since we work in a monitoring setup, we will construct a set estimator $\widehat{\mathcal{I}}$ for $\mathcal{I}^*$ that is sequential–i.e., at some time point $M + n$ in the monitoring period, it will be based only on the observations $\{X_{1,M}, \dots, X_{M+n,M}\}$ and will be updated when the next observation $X_{M+n+1,M}$ arrives, and so on. Statistically the estimator $\widehat{\mathcal{I}}(n)$ then tells us at time $M + n$ in which components a change has already been identified. Before we construct a set estimator, we specify three desirable properties it should have. The set estimator at the end of the monitoring period is defined as $\widehat{\mathcal{I}}(\lfloor TM \rfloor - M)$, and if there is no end to the monitoring ($T = \infty$), we simply define it as $\widehat{\mathcal{I}}(\lfloor TM \rfloor - M) = \widehat{\mathcal{I}}(\infty) := \bigcup_{n=1}^{\infty} \widehat{\mathcal{I}}(n)$.

Desiderata 4.1.

i) Monotonicity: $\widehat{\mathcal{I}}(n) \subset \widehat{\mathcal{I}}(n+1)$ for all $n = 1, \dots, \lfloor TM \rfloor - M - 1$.
ii) Consistency: $\lim_{M\to\infty} \mathbb{P}(\mathcal{I}^* \subset \widehat{\mathcal{I}}(\lfloor TM \rfloor - M)) = 1$.
iii) Asymptotic level α: $\lim_{M\to\infty} \mathbb{P}(\widehat{\mathcal{I}}(\lfloor TM \rfloor - M) = \mathcal{I}^*) \geq 1 - \alpha,\ \alpha \in (0, 1)$.

Condition (i) implies that the discovery of a change in a component at time $M + n$ will not be revoked at a later time. Condition (ii) implies that, as $M \to \infty$, all regions where a change occurs are correctly identified, while Condition (iii) ensures that the probability of including additional components where no change occurs can be made arbitrarily small.

Assumptions. We impose some formal assumptions for our subsequent mathematical analysis. Therefore, we introduce the error functions $\varepsilon_n^{(j)} \in C_0(I, \mathbb{R}^d)$, where

$$X_n - \mu_n = \begin{cases} \varepsilon_n^{(1)}, & \text{if } 1 \leq n \leq \lfloor M\tau_\ell \rfloor \\ \varepsilon_n^{(2)}, & \text{if } \lfloor M\tau_\ell \rfloor < n \leq \lfloor TM \rfloor. \end{cases} \tag{4.3}$$

Hence, $\varepsilon_n^{(j)}$ captures the random fluctuations of the latent function X_n. The fact that the distribution of $\varepsilon_n^{(j)}$ can change before and after a structural break (from $j = 1$ to $j = 2$) is often required in functional models.

Assumption 4.1.

i) (Dependence) The time series $(\varepsilon_n^{(1)}, \varepsilon_n^{(2)})_{n\in\mathbb{N}}$ is weakly stationary. For any $M \in \mathbb{N}$, the time series $(\varepsilon_n^{(1)}, \varepsilon_n^{(2)}, \delta_{n,M})_{n\in\mathbb{N}}$ is α-mixing, with mixing coefficients denoted by $\alpha_{\varepsilon,\delta,M}$. There exist two numbers $C_1, \nu > 0$, such that for all $n \in \mathbb{N}$

$$\sup_M \alpha_{\varepsilon,\delta,M}(n) \leq C_1 n^{-\nu}.$$

ii) (Smoothness & Moments) Let $\xi > 1/2$ and assume that the functions $\mu^{(1)}$ and $\mu^{(2)}$ are ξ-Hölder continuous. Let $J > 4$ be a constant that satisfies

$$3 + \frac{9}{J-4} < \nu.$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12796

For $j = 1, 2$ and some $C_2 > 0$ it holds that

$$\sup_n \mathbb{E}\|\varepsilon_n^{(j)}\|^J \leq C_2, \quad \sup_n \sup_{u \neq v} \mathbb{E}\left(\frac{|\varepsilon_n^{(j)}(u) - \varepsilon_n^{(j)}(v)|}{|u-v|^\xi} \right)^J \leq C_2.$$

iii) (Approximation) There exist $C_3 > 0$, $\gamma > 0$ such that

$$\sup_n \mathbb{E}\|X_n - X_{n,M}\|^J \leq C_3 M^{-\gamma}, \quad \sup_{n,M} \sup_{u \neq v} \mathbb{E}\left(\frac{|X_{n,M}(u) - X_{n,M}(v)|}{|u-v|^\xi} \right)^J \leq C_3.$$

Moreover, for some $\eta > 0$

$$\sup_n \|\mu_n - \mu_{n,M}\| \leq C_4 M^{-1/2-\eta}.$$

iv) (Decaying tails) This assumption is only required if $I = \mathbb{R}$. There exist constants $\kappa, C_5 > 0$ such that for any $y > 0$ and $M \in \mathbb{N}$

$$\sup_n \mathbb{E}\left[\sup_{|u|>y} |X_n(u)|^J \right] \leq C_5 \exp(-y^\kappa), \qquad \sup_n \mathbb{E}\left[\sup_{|u|>y} |X_{n,M}(u)|^J \right] \leq C_5 \exp(-y^\kappa).$$

Condition i) of Assumption 4.1 implies that the latent functions before and after a change are weakly stationary (in each component). Dependence is quantified by α-mixing coefficients, and since the sparse estimators are measurable transforms of the latent functions X_n and the distortions $\delta_{n,M}$, they are α-mixing, with the same coefficients (for further details on our dependence assumption, see Remark 4.1). Condition ii) implies an interplay between the existence of moments and dependence, where higher moments (larger J) permit stronger dependence (smaller ν). Condition iii) implies that the sparse estimators $X_{n,M}$ are uniformly consistent for the latent functions X_n and are in their own right sufficiently smooth. Evidently, these conditions depend on the specific construction of the sparse estimators $X_{n,M}$. For instance, in the case of kernel-type estimators (kernel density, Nadaraya-Watson), concentration results on the uniform distance between X_n and $X_{n,M}$, as well as their derivatives, can be used. We refer to the classical results in Schuster (1969) and Schuster and Yakowitz (1979) as relevant examples. Moreover, we assume that the mean functions μ_n and $\mu_{n,M}$ are sufficiently close to each other so that random fluctuations due to varying $\mu_{n,M}$ do not confound true changes of the μ_n. Finally, Condition (iv) ensures that the random functions have fast decaying tails if the indexing interval I is infinitely large. Typical examples include density functions with exponentially decaying tails. Notice that the vast majority of works in FDA are confined to data on a compact interval, which can however be inappropriate to describe random densities with unbounded support.

4.2. Statistical Inference

A change point detector. We now develop statistical inference methods in the monitoring regime of the previous section. For this purpose, we devise a *detector* $\widehat{\Gamma}_M^{(\mathcal{I})}$ that assumes large values in the presence of change points and is small otherwise. We will later on use the detector $\widehat{\Gamma}_M^{(\mathcal{I})}$ as a building block of the set estimator $\widehat{\mathcal{I}}$ described in the Desiderata 4.1. Yet, our convergence results concerning $\widehat{\Gamma}_M^{(\mathcal{I})}$, collected in Theorem 4.1 below, are of independent interest.

The detector $\widehat{\Gamma}_M^{(\mathcal{I})}$ has two inputs: First, a non-empty index set $\mathcal{I} \subset \{1, \dots, d\}$. In the case of the complete index set $\mathcal{I} = \{1, 2, \dots, d\}$, we define $\widehat{\Gamma}_M := \widehat{\Gamma}_M^{(\mathcal{I})}$. Second, the detector $\widehat{\Gamma}_M^{(\mathcal{I})} = \widehat{\Gamma}_M^{(\mathcal{I})}(n)$ depends on the data collected up to time $M + n$. At time point $M + n$, only the functions $\{X_{1,M}, \dots, X_{M+n,M}\}$ are available to the user.

We want to base the detector $\widehat{\Gamma}_M^{(\mathcal{I})}(n)$ on the CUSUM principle. There are many variants of the CUSUM statistic, such as standard CUSUM, mCUSUM, MOSUM, and Full-CUSUM statistics. These statistics have been used in the context of univariate and multivariate data and are reviewed in detail in Aue and Kirch (2024). While all of these variants are based on similar intuitions, they differ somewhat in their power to detect alternatives. In this work, we adapt a Full-CUSUM statistic that, in finite dimensions, is motivated by the likelihood ratio test and yields particularly powerful change point tests (see Gösmann *et al.* (2021)). We define for a non-empty index set $\mathcal{I} \subset \{1, \dots, d\}$ and a weight function $w(M, n)$ the detector

$$\begin{aligned}\widehat{\Gamma}_M^{(\mathcal{I})}(n) &:= \max_{0\le \ell < n} w(M,n) \cdot \left\| \frac{n-\ell}{M+\ell} \sum_{i=1}^{M+\ell} X_{i,M}^{(\mathcal{I})} - \sum_{i=M+\ell+1}^{M+n} X_{i,M}^{(\mathcal{I})} \right\| \\ &= \max_{j\in\mathcal{I}} \max_{0\le \ell < n} w(M,n) \cdot \left\| \frac{n-\ell}{M+\ell} \sum_{i=1}^{M+\ell} X_{i,M}[j] - \sum_{i=M+\ell+1}^{M+n} X_{i,M}[j] \right\|. \end{aligned} \tag{4.4}$$

The intuition behind the detector is as follows: Suppose there is a change in a component $j^* \in \mathcal{I}$ at time $k^* = \lfloor M\tau_{j^*} \rfloor$. Then, an ideal statistic to consider would be (for $n > k^*$)

$$\left\| \frac{n-k^*}{M+k^*} \sum_{i=1}^{M+k^*} X_{i,M}^{(\mathcal{I})}[j^*] - \sum_{i=M+k^*+1}^{M+n} X_{i,M}^{(\mathcal{I})}[j^*] \right\|,$$

that is, a comparison of the sample mean before and after k^*. Since the location of the change point is in practice unknown, the Full-CUSUM detector in (4.4) maximizes over all potential points $\ell = 0, \dots, n-1$ and over all components $j \in \mathcal{I}$. The weight function w scales the difference, assuring that it converges weakly as $M \to \infty$. In the following, we will use the weight function

$$w(M,n) := \frac{M^{1/2}}{M+n},$$

but other choices are possible (see Gösmann (2020)). The advantage of w is that it yields a particularly simple limit, described below in Eq. (4.6), which facilitates quantile approximations. For our theoretical analysis, we define the long-run covariance

$$c[l,k](u,v) := \mathbb{E}\left[\varepsilon_1[l](u)\varepsilon_1[k](v)\right] + 2\sum_{n\ge 2} \mathbb{E}\left[\varepsilon_1[l](u)\varepsilon_n[k](v)\right], \tag{4.5}$$

which in turn defines a centered Brownian motion $\{W(x) : x \in [0,T]\}$ on the space of d-dimensional continuous functions. Throughout this work, we will always assume that c is not identically 0, and hence, the limit W is nontrivial. Recall from Section 3, that $W(x)$ is itself a function for any $x \ge 0$ and that we can write its evaluation at point $u \in I$ as $W(x,u)$. Furthermore, for any $u \in I$, we can define the pointwise range of the Brownian motion in components of set $\mathcal{I}$ as

$$\mathcal{R}_T(u,\mathcal{I}) := \max_{\ell\in\mathcal{I}} \left\{ \max_{x\in[0,(T-1)/T]} W[\ell](x,u) - \min_{x\in[0,(T-1)/T]} W[\ell](x,u) \right\}, \tag{4.6}$$

where we interpret $\infty/\infty =: 1$ in the open-end case. With these definitions in place, we state the key approximations for the distribution of $\widehat{\Gamma}_M^{(\mathcal{I})}(n)$.

DOI: 10.1111/jtsa.12796

Theorem 4.1. Suppose that Assumption 4.1 holds and let $\mathcal{I}$ be a non-empty index set with $\mathcal{I} \cap \mathcal{I}^* = \emptyset$ (no changes in $\mathcal{I}$). Then there exists a $\tau > 0$ such that

$$\sup_{y \in \mathbb{R}} \left| \mathbb{P}\left(\sup_{1 \le n \le \lfloor TM \rfloor - M - 1} \widehat{\Gamma}_M^{(\mathcal{I})}(n) \le y \right) - \mathbb{P}\left(\sup_{u \in I} \mathscr{R}_T(u, \mathcal{I}) \le y \right) \right| = \mathcal{O}(M^{-\tau}).$$

On the other hand, if $\mathcal{I} \cap \mathcal{I}^* \neq \emptyset$ holds, then for any fixed $y_0 \in \mathbb{R}$,

$$\mathbb{P}\left(\sup_{1 \le n \le \lfloor TM \rfloor - M - 1} \widehat{\Gamma}_M^{(\mathcal{I})}(n) > y_0 \right) = 1 - \mathcal{O}(M^{-\tau}).$$

The constant $\tau > 0$ in the above Theorem depends only on the constants in Assumption 3.1, namely, $C_i, J, \nu, \xi, \eta, \kappa, d$ and the $\mathcal{O}$-term in the second part of the Theorem depends on the number y_0. The proof of the Theorem relies on a Gaussian approximation of the partial sum process $P_M^{(\mathcal{I})}(x) = \frac{1}{\sqrt{M}} \sum_{i=1}^{\lfloor xM \rfloor} X_{i,M}^{(\mathcal{I})}$ by the Brownian motion $W^{(\mathcal{I})}(x)$ on the indexing interval $[0, T]$ for $T < \infty$ and on a growing interval $[0, M^\zeta]$ for $T = \infty$. In the case of $T = \infty$, additional challenges arise when controlling partial sums with $x \in [M^\zeta, \infty)$. Essentially, these are weighted to be negligible by our choice of the weight function w. Our approximations rely on a coupling result for the pair (P_M, W) on a probability space Ω_M, where we construct a version of the pair that is close with high probability, see Section A of the Supporting Information. The weak convergence in Theorem 4.1 can be used to construct a set estimator for $\mathcal{I}^*$ (defined in Eq. (4.2)), the collection of components where a change occurs.

Set estimation. We denote by $q_{1-\alpha}^{(\mathcal{I})}$ the upper α-quantile of the limiting distribution in (4.6). The covariance $c^{(\mathcal{I})}$ of $W^{(\mathcal{I})}$ is estimated using the training data as explained in Section C.2 of the Supporting Information. This leads to an estimator of $q_{1-\alpha}^{(\mathcal{I})}$, but for simplicity of notation, we use in the following algorithm the true quantiles. We denote the indicator function by $\mathbb{I}\{\cdot\}$.

The algorithm can be informally described as follows. At each time $m \le n$, we use the detector $\widehat{\Gamma}^{(\mathcal{B})}(m)$ to test for a change in the components of $\mathcal{B}$. If no change is detected in the mth step, we move on to the $(m+1)$th step and repeat the procedure. If a change is detected, say in component j, we eliminate j from $\mathcal{B}$ giving us the updated version $\mathcal{B} \to \mathcal{B} \setminus \{j\}$ and continue our search for changes in the next step $m+1$ with the updated set $\mathcal{B}$ and the statistic $\widehat{\Gamma}^{(\mathcal{B})}(m+1)$. Thus, starting out with $\mathcal{B} = \{1, \ldots, d\}$ for $m = 1$ we sequentially weed out those components where changes are detected. Those components that have been identified as containing a change up until n make up the set estimator $\widehat{\mathcal{I}}(n)$.

Theorem 4.2. Suppose that Assumption 4.1 holds and let $\alpha \in (0, 1)$. Then, the set estimator $\widehat{\mathcal{I}}$ from Algorithm 1 satisfies the Desiderata 4.1, with asymptotic level

$$\mathbb{P}\left(\widehat{\mathcal{I}}(\lfloor TM \rfloor - M) \subseteq \mathcal{I}^* \right) \ge 1 - \alpha - \mathcal{O}(M^{-\tau}).$$

Remark 4.1. We give some details on Theorem 4.2:

1) *Local alternatives*: Sometimes inference methods are investigated for small changes. This means that asmyptotics are considered for change point magnitudes that decay as the sample size increases (see, for details, Aue and Kirch (2024)). It is possible to modify our results in this direction. More precisely, suppose that the mean functions $\mu^{(j)}, j = 1, 2$ may depend on M and satisfy for any component $k \in \mathcal{I}^*$

$$0 < \liminf_M \sqrt{\frac{M}{\log(M)}} \cdot \|\mu^{(1)}[k] - \mu^{(2)}[k]\|.$$

   Then, the results of Theorem 4.1 and Theorem 4.2 remain true.
2) *Asymptotic error rate*: The asymptotic error rate of $\alpha \in (0, 1)$ can be made arbitrarily small. It follows from our proofs that it is exactly approximated if there are no changes in any component.

DOI: 10.1111/jtsa.12796 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Algorithm 1. Set estimator of $\mathcal{I}^*$

Require: length of stable period M, number of steps in the monitoring period n, quantiles $q_{1-\alpha}^{(\mathcal{I})}$ for $\mathcal{I} \subset \{1, \dots, d\}$, sparse estimators $\mathbf{X} = (X_{1,M}, \dots, X_{M+n,M})$

function IEST($M, n, \mathbf{X}, (q_{1-\alpha}^{(\mathcal{I})})_{\mathcal{I}}$)
$\vec{I} = (0, 0, \dots, 0) \in \mathbb{R}^d$
$\mathcal{B} = \{k : \vec{I}[k] = 0\}$
  for $m = 1, 2, \dots, n$ **do**
    if $\mathcal{B} \neq \emptyset$ **then**
      for $k \in \mathcal{B}$ **do**
        $\vec{I}[k] = \mathbb{I}\{\widehat{\Gamma}_M^{(k)}(m) > q_{1-\alpha}^{(\mathcal{B})}\}$
      end for
      $\mathcal{B} = \{k : \vec{I}[k] = 0\}$
    end if
  end for
$\widehat{\mathcal{I}}(n) = \{k : \vec{I}[k] = 1\}$
return $\widehat{\mathcal{I}}(n)$
end function

3) *Dependence assumption*: In this work, we have assumed that the functional noise ε_n and the distortions δ_n are α-mixing. Since the reconstructions $X_{n,M}$ are measurable transforms of these two components, they inherit this mixing assumption, making them weakly dependent as well. This property of being stable under measurable transforms is not generally shared by other dependence concepts, such as m-approximable time series (see, e.g., Hörmann and Kokoszka (2010)), which makes mixing uniquely suitable for the study of sparse FDA. However, if we neglect the problem of sparsity for the moment and consider the fully observed case ($X_{n,M} = X_n$), other dependence concepts, such as m-approximability, could be used instead of mixing. For this purpose, it would be necessary to derive an approximation of the partial sum process on the function space by a functional Brownian motion, at a polynomial rate in M. We do not know of such results, but they can most likely be derived (a closely related result was shown in Berkes *et al.* (2013)).

5. A SIMULATION STUDY AND APPLICATION TO INTRADAY RETURNS ON EXCHANGE TRADED FUNDS

The objective of this section is to show how the methodology proposed and asymptotically justified in previous sections works in practice. We do not aim at a comprehensive simulation study and extensive applications but merely want to illustrate the approach. We do so by considering sparse density estimators corresponding to truncated normal observations in a simulation, and non-normal densities estimated form intraday financial data.

We applied Algorithm 1 to a panel of simulated densities with d components, $d \in \{5, 10\}$ and temporal index $n = 1, 2, \dots, TM$, with $T \in \{3, 4, 5\}$. We considered stable/training periods of length $M \in \{50, 60, 70\}$. Denote by ϕ_{a,b^2} the density of the normal distribution with mean a and variance b^2. In our simulations, a is random and drawn from the uniform distribution, $a \sim U(\underline{a}, \overline{a})$, while b is fixed at 1. The limits $\underline{a}$ and $\overline{a}$ vary in simulations. The count of observations that follow ϕ_{a,b^2}, from which the density is estimated, is $M/2$. We estimate the densities based on observations that fall into the interval $I = [-1, 1]$, so their number is random and equal to about 68% of all generated data points. We use the standard kernel density estimator (`R` function `density` with the default bandwidth selection) as a sparse estimator, discretizing the interval $I = [-1, 1]$ with equidistant grid $u_1 < u_2, \dots < u_J, J = 50$. If changes occur, we place them in about half of the components with $\tau_\ell = 1 + 1/\ell, \ell = 1, 2, \dots, \lfloor d/2 \rfloor$ and switch the distribution of the random variable a from $a^{(1)} \sim U(\underline{a}^{(1)}, \overline{a}^{(1)})$ to $a^{(2)} \sim U(\underline{a}^{(2)}, \overline{a}^{(2)})$. To simulate the

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12796

Table I. Proportion of correct detection of components with a change point (the set $\mathcal{I}^*$) with $T = 4$, $M = 60$. The terms "big" and "small" change, explained in the text, refer to the typical size of the mean change (it is random in the simulations). The theoretical proportion, as $M \to \infty$, should be 0.95. The proportion of correct detections for our method is under the "Proposed Method" header. Under the header "Naïve Method", we show the performance of the method that counts changes in components one by one.

d	$\lvert\mathcal{I}^*\rvert$	Distance	Scenario $(\underline{a}^{(2)}, \overline{a}^{(2)})$	Naïve Method	Proposed Method
5	2	Small	Fully overlapping $(0.3, 0.4), (0.3, 0.4)$	0.846	0.967
5	2	Small	Partially overlapping $(0.3, 0.4), (0.35, 0.45)$	0.846	0.955
5	2	Small	Nonoverlapping $(0.3, 0.4), (0.4, 0.5)$	0.849	0.955
5	2	Big	Fully overlapping $(0.4, 0.5), (0.4, 0.5)$	0.842	0.963
5	2	Big	Partially overlapping $(0.4, 0.5), (0.45, 0.55)$	0.834	0.951
5	2	Big	Nonoverlapping $(0.4, 0.5), (0.5, 0.6)$	0.825	0.956
10	5	Small	Fully overlapping $(0.3, 0.4), (0.3, 0.4), \ldots, (1.5, 2)$	0.723	0.963
10	5	Small	Partially overlapping $(0.3, 0.4), (0.35, 0.45), \ldots, (0.5, 0.6)$	0.762	0.963
10	5	Small	Nonoverlapping $(0.3, 0.4), (0.4, 0.5), \ldots, (0.7, 0.8)$	0.693	0.945
10	5	Big	Fully overlapping $(0.4, 0.5), (0.4, 0.5), \ldots, (0.4, 0.5)$	0.745	0.965
10	5	Big	Partially overlapping $(0.4, 0.5), (0.45, 0.55), \ldots, (0.6, 0.7)$	0.736	0.963
10	5	Big	Nonoverlapping $(0.4, 0.5), (0.5, 0.6), \ldots, (0.8, 0.9)$	0.740	0.968

asymptotic distribution, we use covariance estimator of Section C.2 with $b = 2$ and Bartlett weights and generate 500 vectors of i.i.d. standard normal random variables. We repeated the computations for selected cases with $b = 3, 4, 5$ and obtained similar rates. The frequencies reported in Table I are based on 1000 replications. In the components with the change, we switch the distribution of the random mean from $(\underline{a}^{(1)}, \overline{a}^{(1)}) = (0.0, 0.1)$ to various configurations for $(\underline{a}^{(2)}, \overline{a}^{(2)})$, which can be divided into two distance groups, small and big, in terms of how far the interval $(\underline{a}^{(2)}, \overline{a}^{(2)})$ is from the baseline interval. Then, within each group, we consider three types of scenarios in which the intervals $(\underline{a}^{(2)}, \overline{a}^{(2)})$ are fully overlapping, partially overlapping and nonoverlapping.

Table I reports proportions of correct identifications of the set $\mathcal{I}^*$ at the significance level $\alpha = 0.05$. The proposed method, Algorithm 1, is compared to a naïve method of just applying the univariate change point test to each component separately, and sequentially increasing the number of components with a detected change point. We see that the proposed method is much closer to the target of 0.95 with only a slight overestimation of the correct proportion. Similar results hold for other choices of M and T.

To illustrate how our method works in practice, we apply Algorithm 1 to intraday 10-min simple returns (in percent) on nine exchange traded funds (ETFs) covering the trading days between May 1, 2020 and October 31, 2020, hence, overlapping with the initial period of the Covid-19 pandemic. The ETFs cover nine sectors of the US economy, see Table II. For each of the $d = 9$ components, at each day $n = 1, 2, \ldots, 128$, we estimate the density of the 39 intraday returns on the interval $[-0.5, 0.5]$ with a grid of 50 points. Then, we apply Algorithm 1 to these densities with $M = 40$ (this leads to $T = 3$) and $\alpha = 0.05$.

DOI: 10.1111/jtsa.12796 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Table II. Results of the monitoring for a mean change in the density of intraday returns.

Ticker	Sector	Change point
XLP	Consumer staples	July 27
XLI	Industrials	August 07
XLV	Health Care	August 13
XLB	Materials	August 18
XLF	Financials	August 20
XLU	Utilities	August 21
XLY	Consumer discretionary	August 21
XLE	Energy	No change point
XLK	Technology	No change point

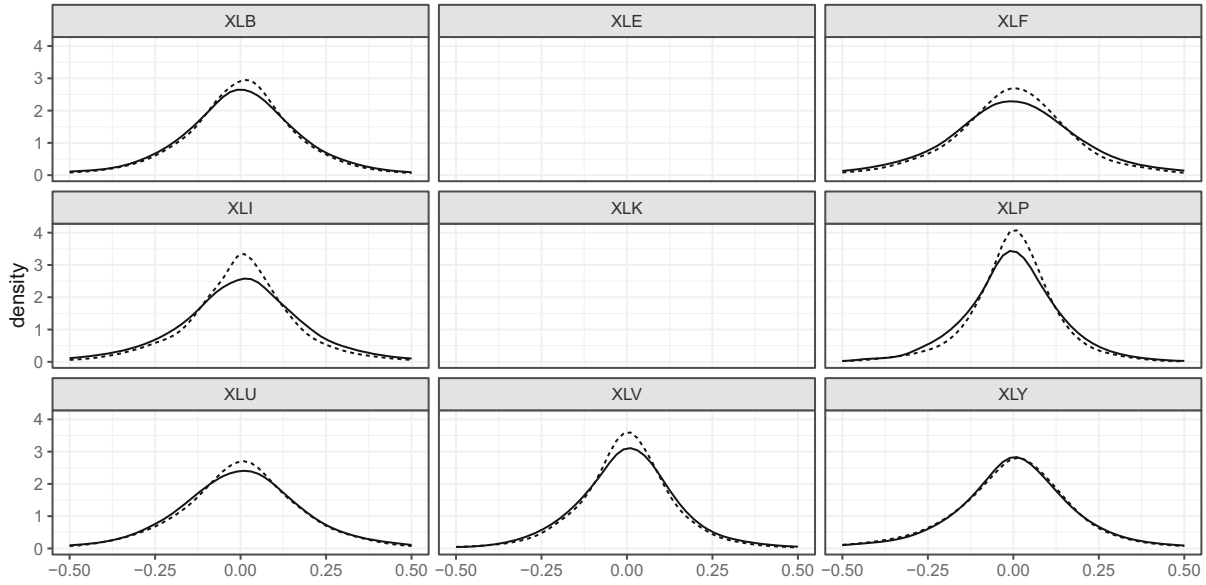

Figure 2. Mean estimated densities of intraday 10-min returns on $[-0.5, 0.5]$ before (solid line) and after (dashed line) the estimated change points for the 9 ETFs.

The results of the application of Algorithm 1 are shown in Table II. Energy (XLE) and Technology (XLK) are the only two sectors where a change is not detected. Almost all sectors where a significant change has been detected exhibit a density with a sharper peak (Figure 2), that is, a higher proportion of very small returns than before the change.

ACKNOWLEDGEMENTS

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

DATA AVAILABILITY STATEMENT

The intraday price data can be obtained from Bloomberg as part of the subscription to "Bloomberg Terminal". The data that support the findings will be available in Bloomberg at https://www.bloomberg.com/professional/products/bloomberg-terminal/ following an embargo from the date of publication to allow for commercialization of research findings.

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12796

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Aston M, Kirch C. 2012a. Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis* **109**:204–220.

Aston M, Kirch C. 2012b. Evaluating stationarity via change-point alternatives with applications to fMRI data. *Annals of Applied Statistics* **6**:1906–1949.

Aue A, Hörmann S, Horváth L, Hušková M. 2014. Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica* **24**:1043–1073.

Aue A, Horváth L, Hušková M, Kokoszka P. 2006. Change–point monitoring in linear models with conditionally heteroskedastic errors. *Econometrics Journal* **9**:373–403.

Aue A, Kirch C. 2024. The state of cumulative sum sequential change point testing seventy years after page. *Biometrika* **111**:367–391.

Aue A, Rice G, Sonmez G. 2020. Structural break analysis for spectrum and trace of covariance operators. *Environmetrics* **31**:e2617.

Bardsley P, Horváth L, Kokoszka P, Young G. 2017. Change point tests in functional factor models with application to yield curves. *The Econometrics Journal* **20**:86–117.

Bastian P, Basu R, Dette H. 2023. Multiple change point detection in functional data with applications to biomechanical fatigue data. arXiv preprint, arXiv:2312.11108.

Berkes I, Gabrys R, Horváth L, Kokoszka P. 2009. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society (B)* **71**:927–946.

Berkes I, Gombay E, Horváth L, Kokoszka P. 2004. Sequential change-point detection in GARCH(p, q) models. *Econometric Theory* **20**:1140–1167.

Berkes I, Horváth L, Rice G. 2013. Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications* **123**:385–403.

Billingsley P. 1968. *Convergence of Probability Measures* Wiley, New York.

Chen Y-T, Chiou J-M, Huang T-Z. 2023. Greedy segmentation for a functional data sequence. *Journal of the American Statistical Association* **118**:959–971.

Chiou J-M, Chen Y-T, Hsing T. 2019. Identifying multiple changes for a functional data sequence with application to freeway traffic segmentation. *The Annals of Applied Statistics* **13**(3):1430–1463.

Chu C-SJ, Stinchcombe M, White H. 1996. Monitoring structural change. *Econometrica* **64**:1045–1065.

Dette H, Kokot K, Aue A. 2020a. Functional data analysis in the Banach space of continuous functions. *The Annals of Statistics* **48**:1168–1192.

Dette H, Kokot K, Volgushev S. 2020b. Testing relevant hypotheses in functional time series via self-normalization. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82**:629–660.

Dette H, Kutta T. 2021. Detecting structural breaks in eigensystems of functional time series. *Electronic Journal of Statistics* **15**:944–983.

Dette H, Wied D. 2016. Detecting relevant changes in time series models. *Journal of the Royal Statistical Society (B)* **78**:371–396.

Dubey P, Müller H-G. 2019. Fréchet change-point detection. *The Annals of Statistics* **48**:3312–3335.

Gösmann J. 2020. *New Aspects of Sequential Change Point Detection*. Ph.D. thesis, Ruhr-University Bochum, Bochum.

Gösmann J, Kley T, Dette H. 2021. A new approach for open-end sequential change point monitoring. *Journal of Time Series Analysis* **42**:63–84.

Gösmann J, Stoehr C, Heiny J, Dette H. 2022. Sequential change point detection in high dimensional time series. *Electronic Journal of Statistics* **16**:3608–3671.

Gromenko O, Kokoszka P, Reimherr M. 2017. Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society (B)* **79**:29–50.

Hariz S, Wylie J, Zhang Q. 2007. Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. *The Annals of Statistics* **35**:1802–1826.

Hörmann S, Kokoszka P. 2010. Weakly dependent functional data. *The Annals of Statistics* **38**:1845–1884.

Horváth L, Hušková M, Kokoszka P, Steinebach J. 2003. Monitoring changes in linear models. *Journal of Statistical Planning and Inference* **126**:225–251.

Horváth L, Kokoszka P, Rice G. 2014. Testing stationarity of functional time series. *Journal of Econometrics* **179**:66–82.

Horváth L, Kokoszka P, Wang S. 2021. Monitoring for a change point in a sequence of distributions. *The Annals of Statistics* **49**:2271–2291.

Horváth L, Rice G. 2024. *Change Point Analysis for Time Series*. Springer, New York.

DOI: 10.1111/jtsa.12796 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Horváth L, Rice G, Zhao L. 2022. Change point analysis of covariance functions: a weighted cumulative sum approach. *Journal of Multivariate Analysis* **189**:104877.
Janson S, Kaijser S. 2015. *Higher moments of Banach space valued random variables*. *Memoirs of the American Mathematical Society*, V **238**. American Mathematical Society, Providence, RI, 110 pp.
Kutta T, Kokoszka P. 2024. *Prokhorov Metric Convergence of the Partial Sum Process for Reconstructed Functional Data*. Technical report. Colorado State University, Fort Collins, CO.
Madrid Padilla CM, Wang D, Zhao Z, Yu Y. 2022. Change-point detection for sparse and dense functional data in general dimensions. *Advances in Neural Information Processing Systems* **35**:37121–37133.
Madrid Padilla CM, Xu H, Wang D, Madrid Padilla OH, Yu Y. 2023. Change point detection and inference in multivariate non-parametric models under mixing conditions. *Advances in Neural Information Processing Systems* **36**:21081–21134.
Rice G, Shum M. 2019. Inference for the lagged cross-covariance operator between functional time series. *Journal of Time Series Analysis* **40**:665–692.
Schuster E. 1969. Estimation of a probability density function and its derivatives. *The Annals of Mathematical Statistics* **40**:1187–1195.
Schuster E, Yakowitz S. 1979. Contributions to the theory of nonparametric regression, with application to system identification. *The Annals of Statistics* **7**:139–149.
Stöhr C, Aston J, Kirch C. 2021. Detecting changes in the covariance structure of functional time series with application to fMRI data. *Econometrics and Statistics* **18**:44–62.
Tucker D, Yarger D. 2023. Elastic functional changepoint detection of climate impacts from localized sources. *Environmetrics* **35**:e2826.
Wang M, Harris T, Li B. 2023. Asynchronous changepoint estimation for spatially correlated functional time series. *Journal of Agricultural, Biological, and Environmental Statistics* **28**(1):157–176.
Wu T, Wang R, Yan H, Shao X. 2022. Adaptive change point monitoring for high-dimensional data. *Statistica Sinica* **32**:1583–1610.
Zhang X, Shao X, Hayhoe K, Wuebbles D. 2011. Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics* **5**:1765–1796.

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12796