Statistica Sinica

When to Refresh a Functional Representation: Online Monitoring of Structural Change in Functional Time Series

Jiajing Sun[1], Meiting Zhu[2], Wolfgang Karl Härdle[3],

Oliver Linton[4] and Zhuo Lin[1]

[1] *University of Chinese Academy of Sciences,* [2] *Xiamen University,*

[3] *Bucharest University of Economic Studies,* [4] *University of Cambridge*

Abstract: Function-valued observations such as yield curves, volatility profiles, and intraday return curves are often summarized by representations estimated from a stable training sample. We study the online decision of when such a representation should be re-estimated. After the training-sample functional principal component analysis (FPCA) basis is fixed, each incoming curve is projected onto the same score coordinates, and cumulative movements in the retained scores are monitored sequentially. We compare kernel-based heteroskedasticity-and-autocorrelation- consistent (HAC) long-run-variance standardization with Shao-type and adjusted-range self-normalized Kolmogorov–Smirnov and weighted Cramér–von Mises monitoring statistics. The adjusted-range self-normalized Kolmogorov–Smirnov statistic is our main recommendation: it avoids kernel and bandwidth choices for long-run-variance estimation and mitigates the denominator-inflation problem that can make direct quadratic self-normalized

Kolmogorov–Smirnov alarms conservative. We establish finite-horizon null limits, consistency, local-power limits, and robustness of the self-normalized null theory to mild training-sample contamination. For one-minute S&P 500 return curves around the COVID-19 market disruption, the adjusted-range alarm occurs well before the comparable HAC alarm, during a period when intraday return magnitudes were already far above their training-sample level. The procedure turns functional instability into a real-time model-maintenance signal for economic and financial applications.

Key words and phrases: Functional time series, online monitoring, self-normalization.

1. Introduction

Economists and financial practitioners increasingly work with objects richer than a single number observed each day. High-frequency recording and modern data methods make it routine to observe within-day or cross-sectional profiles: yield curves, option-implied surfaces, intraday volatility profiles, and high-frequency return curves are leading examples. The economically relevant signal may lie in shape, slope, curvature, or intraday timing rather than in a scalar summary. At the same time, repeated re-estimation is costly, can destabilize the system, and may overreact to short-run noise. The central real-time question is therefore

when a practitioner should stop trusting a functional representation estimated from a training sample.

The issue arises directly in macro-financial monitoring. Central banks, asset managers, and empirical economists often carry a functional representation from a training sample into a monitoring sample that may have changed. A yield-curve representation may become unreliable when term premia are repriced rapidly, and an intraday return-curve representation may miss changes in within-day return dispersion. The useful output is therefore not merely an ex post instability statement, but a real-time monitoring statistic for deciding when the old representation has become too stale to trust.

Functional time series carry more information than scalar time series: depending on the object being monitored, level, slope, curvature, timing, dispersion, and other shape features may all be economically relevant. We use FPCA to represent each function by a finite vector of projection scores estimated from the training sample. A change in the mean of the retained score vector is therefore not merely a scalar level shift: because the eigenfunctions vary over intraday time, maturity, strike, or the relevant functional domain, such as intraday time, maturity, strike, or another economic index, such a change can correspond to movements in the level, slope, cur-

vature, timing, or other shape of the underlying functional object. During online monitoring, the training-sample eigenfunctions are held fixed and incoming observations are projected onto them, giving training and monitoring observations common score coordinates. If the training-sample representation remains adequate, monitoring score increments and cumulative paths should be probabilistically similar to their training-sample counterparts. A monitoring alarm therefore signals that the representation should be refreshed before inference or decisions drift too far from the training-sample environment.

Operationally, the main KS statistics are cumulative sums of centered retained scores after the training sample. HAC versions estimate the LRV directly from the training-sample scores, whereas the self-normalized versions use path functionals from the training sample and therefore avoid choosing a kernel and bandwidth. The alarm asks whether the boundary-scaled monitoring score path is too large relative to training-sample dependence and variation, and the weighted-CvM statistics integrate the corresponding standardized path over the monitoring horizon.

Serial dependence in the score stream makes calibration nontrivial. HAC estimates the LRV directly and therefore requires kernel and bandwidth choices. Building on the fixed-b and sample-path normalization

ideas of Kiefer et al. (2000) and Lobato (2001), Shao (2010) proposed self-normalization (SN) as a way to avoid user-chosen smoothing parameters. The self-normalizer is built from the observed partial-sum path itself. It is not a consistent estimator of the LRV; rather, it is stochastically proportional to the true LRV, so that nuisance scale cancels in the self-normalized statistic. Here Shao's (2010) self-normalizer measures variation in cumulative sums of the retained scores over the training sample, while the adjusted-range self-normalizer is constructed from coordinate-wise adjusted ranges after lag-0 whitening. The adjusted-range KS statistic is motivated most directly by Hong et al. (2024). The finite-sample reason for considering it is the denominator-inflation problem identified in Shao and Zhang (2010): direct quadratic self-normalized KS statistics can lose power when the self-normalizer grows faster than the numerator, and their G_n statistic addresses this issue by constructing the self-normalizer from forward and backward partial sums around each putative break date. We compare these standardizations on the common score process, using a KS monitoring statistic as the default alarm and weighted-CvM as a complementary diagnostic.

To keep the main comparison interpretable, we focus on KS and weighted-CvM statistics, which isolates kernel-based LRV estimation from SN. Other detector shapes can be built from the same FPCA score process,

including Page cumulative-sum (Page-CUSUM), weighted cumulative-sum (weighted-CUSUM), moving-sum (MOSUM), and multiscale moving-sum (multiscale MOSUM) versions. Due to space constraints, Supplement Section S4 gives the full comparison and the corresponding references. The results point to one main recommendation. The KS version of the adjusted-range self-normalized monitoring statistic (RSMS) is tuning-free in the relevant sense: unlike the HAC LRV estimation, it does not require choosing a kernel or a bandwidth. In the simulation studies, it outperforms Shao's sequential monitoring statistic (SSMS) in the settings most relevant for online monitoring and remains informative under mild training-sample contamination. The weighted-CvM family is useful as a secondary diagnostic, but the main empirical recommendation is the RSMS KS monitoring statistic.

We illustrate the framework with 1-minute S&P 500 return curves around the COVID-19 market break. Conditional on a pre-2020 training sample, RSMS KS signals on February 27, 2020 at the main $T = 2$ horizon, whereas the comparable kernel-based HAC benchmark waits until April 1–2. The gap is economically meaningful rather than merely mechanical: over the trading days between the two alarms, average absolute 1-minute returns remain more than ten times their training sample level. This timing

difference is exactly the kind of margin that matters for deciding when a functional forecasting or risk representation should be refreshed while the return-curve environment is still shifting.

The paper relates to several strands of work. On sequential monitoring and change-point detection, our starting points are Chu et al. (1996) and the recent overview by Aue and Kirch (2024). Within functional data analysis, one important line of work studies monitoring after functional compression or within explicitly functional time-series models. Aue et al. (2014), for example, connect dependent functional regression models to structural-change monitoring. Kutta et al. (2026) develop monitoring methods for sparse functional panels observed irregularly and with measurement noise, which is especially useful when the researcher does not observe a clean daily curve on a common grid. Kutta and Kokoszka (2025) provide a broader monitoring framework for functional time series, while Kutta and Dörnemann (2025) emphasize procedures designed to keep detection delay short. These papers clarify that online monitoring of functional objects is feasible, but they address rather different practical bottlenecks: sparse designs, general functional-process monitoring, or short-delay calibration.

A second line of work studies change-point detection more broadly in functional or object-valued settings. Bastian et al. (2024) analyze multiple

change points in functional data, Boniece et al. (2025) propose change-point detection using empirical energy distance, and Zhang et al. (2026) extend the literature beyond classical functions to object-valued time series. These papers show how broadly change-point ideas can be formulated beyond scalar time series. The present paper addresses a different real-time question: after a stable training sample has produced a fixed FPCA score representation, when do incoming function-valued observations provide enough evidence to refresh that representation?

This positioning also separates the present paper from three related functional-data literatures. Fully functional change-point methods can avoid dimension reduction and test directly in the ambient function space (Berkes et al., 2009; Aue et al., 2018); dependent functional-data change-point methods explicitly account for serial dependence in the function-valued process (Aston and Kirch, 2012; Horváth et al., 2010; Sharipov et al., 2016); and dynamic FPCA offers another route under serial dependence through frequency-domain components (Hörmann et al., 2015). We do not claim to dominate these alternatives. Instead, our contribution is complementary and operational: tuning-light real-time monitoring statistics can indicate when an existing functional data model should be re-estimated.

The procedure is not an ex post FPCA change-point test. The FPCA basis is fixed after the training period, incoming curves are projected into that fixed coordinate system, and an alarm is interpreted as a prospective maintenance signal for the functional representation. Our contribution is threefold. First, we formulate online functional time-series monitoring as a model-maintenance problem: after a training-sample FPCA representation has been fixed, the retained score stream is monitored to decide when that representation should be re-estimated. Second, we compare HAC, Shao-type self-normalized, and adjusted-range self-normalized statistics on this common score stream, giving tuning-light monitoring statistics that avoid separate kernel and bandwidth choices. Third, we document their finite-sample and empirical behavior in simulations and in 1-minute S&P 500 return curves, with Supplement Section S4 comparing alternative detector shapes on the same score stream.

The rest of the paper is organized as follows. Section 2 defines the fixed FPCA score process and monitoring statistics, Section 3 gives the finite-horizon theory, and Sections 4–5 report the simulations and S&P 500 illustration. The online Supplement contains proofs, open-end calibration, additional simulations, alternative detectors, and empirical robustness checks.

2. Real-Time Monitoring of Functional Representations

Let $\mathbb{H} = L^2(\mathcal{I})$, where $\mathcal{I} \subset \mathbb{R}^d$ is compact, with inner product $\langle f, g \rangle = \int_{\mathcal{I}} f(t)g(t)\,dt$ and norm $\|f\|_{\mathbb{H}} = \langle f, f \rangle^{1/2}$. We observe a functional time series $\{X_t : t \geq 1\} \subset \mathbb{H}$ sequentially. The first m observations form the training sample, and the subsequent observations form the monitoring sample. For a fixed finite horizon $T \in (0, \infty)$, the monitoring period is $\{X_{m+1}, \ldots, X_{m+\lfloor mT \rfloor}\}$. In the main mean-monitoring formulation, the hypotheses are

$$\begin{aligned} H_0 : \quad & \mathsf{E}\, X_t = \mu, \qquad 1 \leq t \leq m + \lfloor mT \rfloor, \\ H_1 : \quad & \mathsf{E}\, X_t = \mu, \qquad 1 \leq t \leq m + k_m^\dagger, \\ & \mathsf{E}\, X_t = \mu + \Delta_m, \qquad m + k_m^\dagger < t \leq m + \lfloor mT \rfloor, \end{aligned} \tag{2.1}$$

for some $1 \leq k_m^\dagger < \lfloor mT \rfloor$, $\mu, \Delta_m \in \mathbb{H}$, and $\Delta_m \neq 0$.

This section formalizes the sequential monitoring statistics. We estimate FPCA scores from the leading empirical eigenfunctions obtained on a training sample, keep those eigenfunctions fixed when monitoring begins, and project each newly observed function-valued observation onto the same training-sample eigenfunctions. The resulting score vectors are then used to form the KS and weighted-CvM statistics. Using one fixed score coordinate system makes an alarm a re-estimation signal rather than an artifact

of re-estimating the FPCA basis.

Mean constancy is not restrictive: as noted above, a shift in the retained FPCA score mean can represent level, slope, curvature, timing, or other shape changes, because the scores load on empirical eigenfunctions over the function's domain. More general targets can be handled by monitoring suitable score processes; for example, covariance or serial-dependence changes can be studied through lagged products, and functional-regression stability through residual-based score processes that are centered at zero under stability (Carlstein and Siegmund, 1994; Hampel et al., 1986; Aue et al., 2014; Zhang et al., 2011).

2.1 FPCA scores from the training sample

Under H_0 in (2.1), write $X_t = \mu + Y_t$ with $\mathsf{E}(Y_t) = 0$ and $\mu \in \mathbb{H}$ time-invariant. Let $\mathcal{C}$ be the covariance operator of Y_t, with kernel $c(u, v) = \mathrm{Cov}\{Y_t(u), Y_t(v)\}$. Under the usual square-integrability conditions, $\mathcal{C}$ is self-adjoint, nonnegative, and trace class, and hence admits the spectral decomposition $\mathcal{C}\phi_j = \lambda_j \phi_j$, with orthonormal eigenfunctions $\{\phi_j\}$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. The covariance-kernel expansion $c(u, v) = \sum_{j \geq 1} \lambda_j \phi_j(u) \phi_j(v)$ is understood in $L^2(\mathcal{I} \times \mathcal{I})$. If c is continuous, Mercer's theorem additionally gives the usual pointwise expansion.

We use the leading training-sample FPCA eigenfunctions, by which we mean the leading empirical eigenfunctions of the covariance operator estimated from the training sample. Let

$$\widehat{\mu}_m = \frac{1}{m}\sum_{t=1}^{m} X_t, \qquad \widehat{\mathcal{C}}_m = \frac{1}{m}\sum_{t=1}^{m}(X_t - \widehat{\mu}_m) \otimes (X_t - \widehat{\mu}_m),$$

and let $(\widehat{\lambda}_j, \widehat{\phi}_j)$ be the leading empirical eigenpairs of $\widehat{\mathcal{C}}_m$. Fix a score dimension K. For both the training and monitoring observations, we project the function-valued observations onto the same training-sample eigenfunctions:

$$a_t = \big(\langle X_t, \widehat{\phi}_1\rangle, \ldots, \langle X_t, \widehat{\phi}_K\rangle\big)^\top, \qquad t \geq 1.$$

In applications K may be chosen by a fraction-of-variance-explained (FVE) criterion or an information criterion (Li et al., 2013); for the asymptotic theory below, K is fixed.

Define the training score mean $\bar{a}_m = m^{-1}\sum_{t=1}^{m} a_t$ and the centered retained score stream $z_t = a_t - \bar{a}_m$, $t \geq 1$. Equivalently, componentwise, $z_{t,j} = \langle X_t - \widehat{\mu}_m, \widehat{\phi}_j\rangle$, $j = 1, \ldots, K$.

The monitoring partial sums are $S_m(k) = \sum_{t=m+1}^{m+k} z_t$, $k \geq 1$.

Throughout we use the standard monitoring boundary function $g_\gamma(s) = (1+s)\{s/(1+s)\}^\gamma$, $\gamma \in [0, 1/2)$, $s > 0$; its reciprocal is the corresponding time-varying weight, and larger γ gives more emphasis to early monitoring observations.

2.2 HAC-based KS benchmark

To separate the form of the monitoring statistic from the choice of self-normalizer, we also record plug-in HAC benchmarks built from the same training-sample FPCA score stream $\{z_t\}$. Let

$$\widehat{\Gamma}_m = \sum_{\ell=-(m-1)}^{m-1} \mathcal{K}_{\mathrm{HAC}}(\ell/h)\,\widehat{\kappa}_{\ell,m}, \qquad \widehat{\kappa}_{\ell,m} = \begin{cases} m^{-1}\sum_{t=\ell+1}^{m} z_t z_{t-\ell}^{\top}, & \ell \geq 0, \\ \widehat{\kappa}_{-\ell,m}^{\top}, & \ell < 0, \end{cases} \tag{2.2}$$

where $\mathcal{K}_{\mathrm{HAC}}$ is a symmetric kernel and h is a bandwidth. The KS-type HAC benchmark is

$$\mathcal{M}_m^{\mathrm{H}}(k) = \frac{S_m(k)^{\top}\widehat{\Gamma}_m^{-1} S_m(k)}{m g_\gamma(k/m)^2}, \tag{2.3}$$

with stopping time $\mathcal{T}_m^{\mathrm{H}} = \inf\{1 \leq k \leq \lfloor mT \rfloor : \mathcal{M}_m^{\mathrm{H}}(k) > c_\alpha^{\mathrm{H}}(T, K, \gamma)\}$.

The HAC benchmark is familiar and has a clear plug-in interpretation, but it brings kernel and bandwidth choices into the alarm statistic; one aim below is to ask whether a self-normalized statistic can match or improve on it without that tuning.

The stopping time is an alarm time for model maintenance, not an estimator of the break date: the boundary is crossed only after enough post-training evidence accumulates that incoming observations no longer resemble fluctuations around the training sample.

2.3 Shao's KS monitoring statistic

We begin with KS/CUSUM-type monitoring under Shao's quadratic SN. Define Shao's self-normalizer D_m based on the training sample by

$$D_m = \frac{1}{m^2} \sum_{t=1}^{m} \left(\sum_{j=1}^{t} z_j \right) \left(\sum_{j=1}^{t} z_j \right)^\top . \tag{2.4}$$

The corresponding KS-type monitoring statistic is

$$\mathcal{M}_m^{\mathrm{S}}(k) = \frac{S_m(k)^\top D_m^{-1} S_m(k)}{m g_\gamma(k/m)^2}, \qquad \gamma \in [0, 1/2), \tag{2.5}$$

and $\mathcal{T}_m^{\mathrm{S}} = \inf\{1 \le k \le \lfloor mT \rfloor : \mathcal{M}_m^{\mathrm{S}}(k) > c_\alpha^{\mathrm{S}}(T, K, \gamma)\}$.

2.4 Adjusted-range self-normalized KS monitoring statistic (RSMS)

We next consider the adjusted-range self-normalized KS statistic. First we prewhiten the retained score coordinates using the training sample. Let $\widehat{\Sigma}_{z,0} = m^{-1} \sum_{t=1}^{m} z_t z_t^\top$ and write its $LDL^\top$ (square-root-free Cholesky) decomposition $\widehat{\Sigma}_{z,0} = \widehat{L}_m \widehat{\Lambda}_m \widehat{L}_m^\top$, with $\widehat{L}_m$ unit lower triangular and $\widehat{\Lambda}_m$ diagonal. This lag-0 covariance defines a training-sample whitening transformation before the coordinate-wise adjusted ranges are formed.

With a small ridge $\rho > 0$, define

$$\widehat{W}_m = (\widehat{\Lambda}_m + \rho I_K)^{-1/2} \widehat{L}_m^{-1}, \qquad \widetilde{z}_t = \widehat{W}_m z_t, \qquad \widetilde{S}_m(k) = \sum_{t=m+1}^{m+k} \widetilde{z}_t. \tag{2.6}$$

For each coordinate ℓ, let $\widetilde{B}_{m,\ell}(t) = \sum_{j=1}^{t} \widetilde{z}_{j,\ell} - \frac{t}{m}\sum_{j=1}^{m} \widetilde{z}_{j,\ell}, \qquad t = 1, \ldots, m$, and define the adjusted-range self-normalizer $\widetilde{R}_m$ through the componentwise adjusted ranges

$$\widetilde{R}_{m,\ell} = m^{-1/2}\Big(\max_{1\le t\le m} \widetilde{B}_{m,\ell}(t) - \min_{1\le t\le m} \widetilde{B}_{m,\ell}(t)\Big), \qquad \widetilde{R}_m = \mathrm{diag}(\widetilde{R}_{m,1}, \ldots, \widetilde{R}_{m,K}). \tag{2.7}$$

The adjusted-range KS statistic is

$$\mathcal{M}_m^{\mathrm{R}}(k) = \frac{\widetilde{S}_m(k)^\top \widetilde{R}_m^{-2} \widetilde{S}_m(k)}{m g_\gamma(k/m)^2}, \tag{2.8}$$

with $\mathcal{T}_m^{\mathrm{R}} = \inf\{1 \le k \le \lfloor mT \rfloor : \mathcal{M}_m^{\mathrm{R}}(k) > c_\alpha^{\mathrm{R}}(T, K, \gamma)\}$. The boundary exponent γ stabilizes the statistic at very early monitoring times. Setting $\gamma = 0$ gives the standard unweighted KS statistic.

Operationally, RSMS uses the $LDL^\top$ step as a lag-0 training-sample coordinate transformation. Because $\widetilde{S}_m(k)$ and $\widetilde{R}_m$ are constructed after the same transformation, coordinate-wise scale factors cancel from the ratio in $\mathcal{M}_m^{\mathrm{R}}(k)$. Cross-component dependence can still remain, because lag-0 whitening does not remove lagged cross-covariances; Assumption 6 below rules out this case asymptotically, and the diagnostic below checks its finite-sample relevance.

There is no need to use a HAC covariance estimator for $\widehat{W}_m$; doing so would reintroduce separate choices of kernel and bandwidth. Moreover, the unwhitened case $\widehat{W}_m = I_K$ is appropriate when diagnostics or prior structure already suggest weak cross-coordinate dependence.

2.5 Weighted-CvM monitoring statistics

For weighted-CvM monitoring, instead of using $g_\gamma(s)$, we introduce a non-negative deterministic weight function w. For finite-horizon monitoring it is natural to work with weights on $[0,T]$ such as $w_{\mathrm{U}}(s) = 1$, $w_{\mathrm{E}}(s) = 2(1-s/T)$, $w_{\mathrm{M}}(s) = 6(s/T)(1-s/T)$, and $w_{\mathrm{L}}(s) = 2s/T$, which emphasize, respectively, uniform, early, middle, and late changes.

The weighted-CvM versions integrate the standardized monitoring path over the horizon. For the HAC, SSMS, and RSMS standardizations,

$$\mathcal{I}_m^\bullet(k) = \frac{1}{m}\sum_{j=1}^{k} w(j/m)\,\mathcal{M}_m^\bullet(j)\big|_{\gamma=0}, \qquad \bullet \in \{\mathrm{H},\mathrm{S},\mathrm{R}\}, \tag{2.9}$$

with first-crossing stopping times $\mathcal{T}_{m,\mathrm{CvM}}^\bullet = \inf\{1 \le k \le \lfloor mT \rfloor : \mathcal{I}_m^\bullet(k) > c_{\alpha,\mathrm{CvM}}^\bullet(T,K,w)\}$.

The main text works with finite monitoring horizons, which are the horizons used in the simulations and empirical application. Open-end KS and weighted-CvM extensions are proved in Supplement Section S1, and the corresponding $T=\infty$ critical values are reported in Supplement Section S2.

3. Statistical Properties and Implementation

This section gives the probabilistic justification for the alarm thresholds used below. HAC validity comes from estimating the retained-score LRV.

The self-normalized statistics use training-sample path functionals instead, so nuisance scale cancels without a kernel or bandwidth. We also record conditions under which mild training-sample contamination does not change the asymptotic limits of the SSMS and RSMS training-sample self-normalizers. Proofs and extended theorem details are collected in Supplement Section S1, critical values in Supplement Section S2, supplementary simulations in Supplement Section S3, and supplementary monitoring statistics in Supplement Section S4.

The monitoring statistics are built from an FPCA score path whose eigenfunctions are estimated on the training sample and then held fixed. The assumptions below are reduced-form conditions delivering reliable estimation of the leading FPCA eigenfunctions, a weak-dependence partial-sum approximation, and nonsingularity of the covariance and self-normalizing matrices used for standardization.

We first adopt the standard weak-dependence framework of Hörmann and Kokoszka (2010) for functional time series.

Definition 1 (L^p-ℓ-Approximability)**.** *Suppose that* $\{X_t\} \subset L^p_{\mathbb{H}}$ *for some* $p > 0$ *and admits the representation* $X_t = f(\varepsilon_t, \varepsilon_{t-1}, \ldots), \qquad t \geq 1$, *Here* $L^p_{\mathbb{H}}$ *denotes the space of* $\mathbb{H}$*-valued random elements* X *with* $\mathsf{E}\,\|X\|^p_{\mathbb{H}} < \infty$. *The innovations* $\{\varepsilon_t\}$ *are independent and identically distributed (IID) and*

take values in a measurable space, and f is measurable. Let $X_t^{(\ell)}$ be obtained by replacing innovations more than ℓ lags away by an independent copy (here ℓ is the approximation-lag index, distinct from the training-sample size m). The sequence $\{X_t\}$ is called L^p-ℓ-approximable if

$$\sum_{\ell=1}^{\infty} \big(\mathsf{E}\,\|X_\ell - X_\ell^{(\ell)}\|_{\mathbb{H}}^p\big)^{1/p} < \infty.$$

Assumption 1. *Let $Y_t = X_t - \mu$. The centered functional time series $\{Y_t\}$ belongs to $L^p_{\mathbb{H}}$ for some $p > 4$ and is L^p-ℓ-approximable, and the covariance eigenvalues satisfy $\lambda_1 > \lambda_2 > \cdots > \lambda_K > \lambda_{K+1} > 0$.*

This is the reduced-form training-sample stability condition. The L^p-ℓ-approximability restriction allows serial dependence and rich within-function dynamics, but excludes persistence so strong that a training sample cannot support a meaningful warning threshold.

Assumption 2. *The training estimators are well behaved: $\|\widehat{\mu}_m - \mu\|_{\mathbb{H}} = \mathscr{O}_p(m^{-1/2})$ and $\max_{1\le j\le K} \|\widehat{\phi}_j - s_j\phi_j\|_{\mathbb{H}} = o_p(1)$ for suitable signs $s_j \in \{\pm 1\}$.*

Assumptions 1–2 are standard in functional time-series work and guarantee that estimation error from the training-sample mean and eigenfunctions is negligible for the retained score process; see Aue et al. (2014) and Kutta and Kokoszka (2025).

Let $a_t^{(0)} = \big(\langle X_t - \mu, \phi_1\rangle, \ldots, \langle X_t - \mu, \phi_K\rangle\big)^\top$ denote the infeasible population-centered raw score vector under the null. The next condition gives the strong approximation for the population score partial sums and states that the feasible training-centered monitoring sum is uniformly close to its population-score counterpart.

Assumption 3. *For some $\delta > 0$, there exists a positive-definite matrix Σ_K and a standard K-dimensional Brownian motion $\{B_K(r) : r \geq 0\}$ such that, for each fixed finite $T > 0$,*

$$\sup_{0\leq r\leq 1+T} \left\| \sum_{t=1}^{\lfloor mr\rfloor} a_t^{(0)} - \sqrt{m}\,\Sigma_K^{1/2} B_K(r) \right\| = o_p(m^{1/2-\delta}). \tag{3.1}$$

Moreover,

$$\sup_{0\leq r\leq 1} \left\| \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor mr\rfloor} z_t - \frac{1}{\sqrt{m}} \left\{ \sum_{t=1}^{\lfloor mr\rfloor} a_t^{(0)} - \frac{\lfloor mr\rfloor}{m} \sum_{t=1}^{m} a_t^{(0)} \right\} \right\| = o_p(1), \tag{3.2}$$

and

$$\sup_{0\leq s\leq T} \left\| \frac{1}{\sqrt{m}} S_m(\lfloor ms\rfloor) - \frac{1}{\sqrt{m}} \left\{ \sum_{t=m+1}^{m+\lfloor ms\rfloor} a_t^{(0)} - \frac{\lfloor ms\rfloor}{m} \sum_{t=1}^{m} a_t^{(0)} \right\} \right\| = o_p(1). \tag{3.3}$$

Assumption 4 (Small-time tightness)**.** *For every fixed $T < \infty$, every $\eta > 0$, and every $\gamma \in [0, 1/2)$, $\lim_{\varepsilon\downarrow 0} \limsup_{m\to\infty} \mathbb{P}\left(\sup_{1\leq k\leq \lfloor m\varepsilon\rfloor} \frac{S_m(k)^\top A_m S_m(k)}{m g_\gamma(k/m)^2} > \eta \right) = 0$ for $A_m = \widehat{\Gamma}_m^{-1}$ and $A_m = D_m^{-1}$, whenever the corresponding inverse exists. Under Assumption 6, the analogous condition holds with $S_m(k)$ replaced by*

$\widetilde{S}_m(k)$ *and* $A_m = \widetilde{R}_m^{-2}$. *The same conditions are assumed under the local alternatives considered below.*

These conditions are the functional analogue of the Gaussian approximation underlying recursive monitoring schemes and deliver a stable false-alarm benchmark for the training-centered score partial sums. They require more than weak dependence of X_t: the leading K eigenfunctions must be estimable from the training sample, and the training-sample mean and eigenfunctions must converge fast enough that replacing the infeasible population scores $a_t^{(0)}$ by the feasible scores z_t leaves the normalized monitoring partial sums unaffected uniformly over $0 \le s \le T$. The training-mean correction is not negligible estimation error; it is the leading term that produces the Brownian-bridge component $B_K(1+s) - (1+s)B_K(1)$ in the monitoring limit.

Write $B_K^0(r) = B_K(r) - rB_K(1)$, $0 \le r \le 1$, for the training-sample Brownian bridge and, for $s \ge 0$, the monitoring bridge $U_K(s) = B_K(1+s) - (1+s)B_K(1)$. Define $V_K = \int_0^1 B_K^0(r)B_K^0(r)^\top \, dr$ and $R_K = \mathrm{diag}(R_{K,1}, \ldots, R_{K,K})$ with $R_{K,\ell} = \sup_{0\le r\le 1} B_{K,\ell}^0(r) - \inf_{0\le r\le 1} B_{K,\ell}^0(r)$.

3.1 HAC-based KS benchmark

Assumption 5. *The HAC LRV estimator* $\widehat{\Gamma}_m$ *in* (2.2) *is positive definite with probability approaching one and satisfies* $\widehat{\Gamma}_m \xrightarrow{p} \Sigma_K$.

Assumption 5 is the standard consistency condition for the HAC estimator of the retained-score LRV matrix; the kernel and bandwidth enter only through this plug-in estimate of Σ_K.

Theorem 1 (HAC KS-type null limit)**.** *Suppose Assumptions 1–3, 4, and 5 hold. Then, for every fixed* $T < \infty$,

$$\sup_{1\le k\le \lfloor mT\rfloor} \mathcal{M}_m^{\mathrm{H}}(k) \Rightarrow \sup_{0<s\le T} \frac{U_K(s)^\top U_K(s)}{g_\gamma(s)^2}. \tag{3.4}$$

Here $s = k/m$ is the continuous-time scale. Under the null, the retained-score partial sum $m^{-1/2}S_m(\lfloor ms\rfloor)$ converges to the training-centered Brownian limit $\Sigma_K^{1/2}U_K(s)$, and standardizing by the consistent $\widehat{\Gamma}_m^{-1}$ gives the pivotal limit above. The full proof is in Supplement Section S1.3.

3.2 Shao's KS monitoring statistic

Theorem 2 (SSMS KS-type null limit)**.** *Suppose Assumptions 1–3 and 4 hold. Then* $D_m \Rightarrow \Sigma_K^{1/2} V_K \Sigma_K^{1/2}$ *as* $m \to \infty$, *where* $\Sigma_K^{1/2}$ *is the symmetric positive-definite square root of* Σ_K. *Moreover, for every fixed* $T < \infty$,

$$\sup_{1\le k\le \lfloor mT\rfloor} \mathcal{M}_m^{\mathrm{S}}(k) \Rightarrow \sup_{0<s\le T} \frac{U_K(s)^\top V_K^{-1} U_K(s)}{g_\gamma(s)^2}. \tag{3.5}$$

Shao's self-normalizer is built from the training-sample bridge, so the nuisance covariance cancels rather than being estimated; the limit is pivotal. The full proof of Theorem 2 is provided in Supplement Section S1.4.

3.3 Adjusted-range KS monitoring statistic

Assumption 6. *There exists a deterministic nonsingular* W_0 *with* $\widehat{W}_m = W_0 + o_p(1)$ *(the lag-0 whitening matrix in* (2.6)*) such that the transformed-stream LRV is diagonal,* $W_0 \Sigma_K W_0^\top = \Sigma_{K,\mathrm{diag}}$ *for a positive definite diagonal* $\Sigma_{K,\mathrm{diag}}$*. In finite samples this is an approximate-diagonalization requirement, assessed by the off-diagonal diagnostic below.*

The adjusted-range self-normalizer is built coordinate by coordinate. The $LDL^\top$ step removes contemporaneous scale and correlation, so coordinate-wise scale factors cancel from the statistic. The remaining issue is dependence across the transformed monitoring partial sums, whose limiting covariance is governed by $W_0 \Sigma_K W_0^\top$; since Σ_K contains lagged autocovariances, lag-0 whitening alone need not make this diagonal, which is precisely what Assumption 6 requires, exactly in the limit or approximately in finite samples.

This condition holds when serial dependence is mainly coordinate by coordinate after lag-0 whitening (in particular when the transformed scores

are IID). A useful finite-sample diagnostic is

$$\Delta_m^{\mathrm{off}} = \frac{\|\operatorname{offdiag}(\widehat{W}_m \widehat{\Gamma}_m \widehat{W}_m^\top)\|_F}{\|\widehat{W}_m \widehat{\Gamma}_m \widehat{W}_m^\top\|_F}.$$

Small values indicate that the HAC estimate $\widehat{W}_m \widehat{\Gamma}_m \widehat{W}_m^\top$ of the transformed LRV has little off-diagonal mass. The diagnostic is a check on the RSMS approximation, not a claim that lag-0 whitening removes serial dependence. Supplement Section S3.3, Table S3.3 reports Δ_m^{off} for the dependent fMA(1) setting and the empirical S&P 500 score dimensions; that section also discusses HAC-whitened RSMS as a sensitivity check, which we avoid as the default because it reintroduces kernel and bandwidth choices.

Theorem 3 (RSMS KS-type null limit)**.** *Suppose Assumptions 1–3, 4, and 6 hold. Then* $\widetilde{R}_m \Rightarrow \Sigma_{K,\mathrm{diag}}^{1/2} R_K$, *where* $\Sigma_{K,\mathrm{diag}}$ *is the diagonal LRV of the whitened score stream, and for fixed* $T < \infty$,

$$\sup_{1 \le k \le \lfloor mT \rfloor} \mathcal{M}_m^{\mathrm{R}}(k) \Rightarrow \sup_{0 < s \le T} \frac{U_K(s)^\top R_K^{-2} U_K(s)}{g_\gamma(s)^2}. \tag{3.6}$$

Hence, choosing the decision boundary as the corresponding $(1-\alpha)$*-quantile yields asymptotic size* α.

Under approximate diagonalization, the whitening and coordinate-wise adjusted ranges are stochastically proportional to the marginal long-run scales in the whitened coordinates, so the statistic is normalized without choosing a kernel or bandwidth. The proof is in Supplement Section S1.5.

Supplement Section S2 reports simulated critical values for the HAC benchmark statistics and the SSMS and RSMS KS and weighted-CvM statistics, together with the Monte Carlo settings used to obtain those quantiles.

For weighted-CvM monitoring, the only extra ingredient is a deterministic weight function that determines which parts of the monitoring path receive more mass.

Assumption 7 (Weighted-CvM weight function)**.** *For fixed $T < \infty$, $w : [0, T] \to \mathbb{R}_+$ is bounded and Riemann integrable.*

Theorem 4 (Weighted-CvM null limits)**.** *Under Assumption 7 and the assumptions of Theorems 1–3, respectively, the HAC, SSMS, and RSMS weighted-CvM statistics satisfy, for every fixed $T < \infty$, $\sup_{1\le k\le \lfloor mT\rfloor} \mathcal{I}_m^{\bullet}(k) \Rightarrow \sup_{0<s\le T} \int_0^s w(u)\, \frac{U_K(u)^\top A_\bullet U_K(u)}{(1+u)^2}\, du$, with $A_{\mathrm{H}} = I_K$ for HAC, $A_{\mathrm{S}} = V_K^{-1}$ for SSMS, and $A_{\mathrm{R}} = R_K^{-2}$ for RSMS.*

The weighted-CvM limits follow from the KS uniform approximations and the Riemann integrability of w; details are in Supplement Section S1.6.

3.4 Other Theoretical Extensions

We next record the behavior under alternatives and contamination; because weighted-CvM is a secondary diagnostic, the main text states only limiting

results, with proofs and critical values in Supplement Sections S1.6 and S2.2.

Assumption 8 (Fixed alternatives)**.** *There exist a monitoring break location* $k_m^\dagger > 0$*, a retained-score shift* $\delta_{m,K}$*, and a baseline no-change monitoring-score process* $\{z_{m+j}^{(0)}\}$ *such that* $z_{m+j} = z_{m+j}^{(0)} + \mathbf{1}\{j > k_m^\dagger\}\{\delta_{m,K} + \rho_{j,m}\}$, $1 \le j \le \lfloor mT \rfloor$*, where* $\sup_{j > k_m^\dagger} \|\rho_{j,m}\| = o(m^{-1/2})$*, and the baseline process satisfies* $\sup_{0 \le s \le T} \left\| \frac{1}{\sqrt{m}} \sum_{j=1}^{\lfloor ms \rfloor} z_{m+j}^{(0)} - \Sigma_K^{1/2} U_K(s) \right\| = o_p(1)$.

Assumption 9 (Break timing and weighted-CvM exposure)**.** *For fixed-horizon monitoring,* $k_m^\dagger / m \to \kappa \in [0, T)$*. For weighted-CvM monitoring, some post-change interval receives positive weight.*

Assumption 10 (Visible retained-score change)**.** *The change is visible in the retained score space: under fixed alternatives* $\delta_{m,K} \to \delta_K \neq 0$*, and under local alternatives* $\sqrt{m}\,\delta_{m,K} \to \delta_K \neq 0$*.*

Theorem 5 (Consistency under fixed alternatives)**.** *Suppose Assumptions 1–3 hold for the baseline no-change process, and Assumptions 8–10 hold. SSMS KS stops by the horizon with probability tending to one under a fixed visible change. Under Assumption 6, the same conclusion holds for RSMS KS. If Assumption 7 also holds and a post-change interval receives positive weight, the same fixed-alternative consistency conclusion holds for the SSMS and RSMS weighted-CvM versions.*

The proof writes the post-break monitoring sum as a stochastic null component plus the cumulative retained-score mean shift, which grows at rate $\sqrt{m}$ after the break while the null fluctuation stays $O_p(1)$; details are in Supplement Section S1.8.

Theorem 6 (Local-power limits)**.** *Suppose Assumptions 1–3 hold for the baseline no-change process, and Assumptions 4, 8– 10 hold, with* $\frac{k_m^\dagger}{m} \to \kappa \in [0, T)$, $\sqrt{m}\,\delta_{m,K} \to \delta_K \neq 0$. *The SSMS KS local-alternative limit is obtained from Theorem 2 by replacing* $U_K(s)$ *with* $U_K(s) + (s-\kappa)_+\eta_{\mathrm{S}}$, $\eta_{\mathrm{S}} := \Sigma_K^{-1/2}\delta_K$. *Under Assumption 6, the RSMS KS local-alternative limit is obtained, in distribution, from Theorem 3 by replacing* $U_K(s)$ *with* $U_K(s) + (s-\kappa)_+\eta_{\mathrm{R}}$, $\eta_{\mathrm{R}} := \Sigma_{K,\mathrm{diag}}^{-1/2} W_0 \delta_K$. *The corresponding drift replacements also give the SSMS and RSMS weighted-CvM local-alternative limits under Assumption 7.*

Under local alternatives the monitoring sums have the null limit plus a deterministic drift; Supplement Section S1.8 derives the KS and weighted-CvM limits.

Proposition 7 (Mild contamination of the training sample)**.** *Suppose* $z_t = z_t^{(0)} + d_{t,m}$, *where* $z_t^{(0)}$ *satisfies the null assumptions and* $\sup_{0\le r\le 1} \|m^{-1/2}\sum_{t=1}^{\lfloor mr\rfloor} d_{t,m}\| \to 0$ *and* $m^{-1}\sum_{t=1}^{m}\|d_{t,m}\|^2 \to 0$. *Then the*

SSMS and RSMS training-sample self-normalizers D_m and $\widetilde{R}_m$ have the same asymptotic limits as under the clean-training null.

The proof is in Supplement Section S1.8. Consistent with Shao and Zhang (2010), contamination already present in the training sample can inflate the SSMS self-normalizer and make SSMS conservative when a later break occurs, which is why RSMS KS is kept as the main self-normalized statistic, with weighted-CvM and the supplementary detectors as diagnostic companions.

4. Simulation Studies

The simulations ask whether the practical reasons for RSMS–no HAC tuning and less denominator inflation than SSMS–show up in finite samples. They compare HAC, SSMS, and RSMS under independent and serially dependent functional designs with breaks in the retained score representation, separating false-alarm control, post-break responsiveness, and sensitivity to tuning or training-sample contamination. We report empirical size, raw and size-adjusted power (SAP), and average detection delay (ADD); full BB, fIID, and fMA(1) panels, HAC tuning and diagonalization diagnostics, and alternative-detector comparisons are in Supplement Sections S3.1–S4.

4.1 Monte Carlo design and evaluation criteria

Null curves are $X_t(u) = \mu(u) + \varepsilon_t(u)$ on $[0, 1]$ with $\mu(u) = \sin(2\pi u)$. We use three DGPs: Brownian bridge (BB), functional IID B-spline noise (fIID), and functional MA(1) B-spline noise (fMA(1)), which adds short-range dependence; the spline construction is given in Supplement Section S3.

Each replication uses a 301-point grid, estimates FPCA on a clean training sample of length $m = 500$, selects K by 95% FVE, and applies HAC, SSMS, and RSMS to the same selected score stream. HAC uses a Bartlett estimator with bandwidth $L = \lfloor 4(m/100)^{2/9} \rfloor$ ($= 5$ at $m = 500$), with tuning sensitivity in Supplement Section S3.2. The break is at $t^\star = m + s^\star$, $s^\star \in \{50, 200\}$, with horizons $T \in \{1, 2, 5, 10\}$, nominal level 0.05, 1000 replications, and three break designs: abrupt level shift, smooth drift, and abrupt localized change; exact magnitude grids are in Supplement Section S3.

Size is null rejection at 5%; raw power is rejection by the horizon; SAP recomputes power after adjusting by empirical null size; and ADD is the mean post-break lag $\tau_m - s^\star$ among runs stopping after the break, so smaller values mean earlier intervention.

4.2 Empirical size and size-adjusted power

Empirical sizes (Supplement Section S3) show that SSMS is closest to nominal, with KS rejection rates of 3.6–6.6% across BB, fIID, and fMA(1), while RSMS and HAC often reject 7–10%; in fMA(1), varying only the HAC kernel and bandwidth raises HAC size to 8–14% (Supplement Section S3.2). Because the null sizes are not identical, SAP is the more informative comparison of detection ability; MAD summarizes null-size distortion.

The default KS and late-emphasis weighted-CvM summary is in Supplement Table S3.1. Averaged over DGPs, horizons, and break locations in the level-shift and smooth-change designs, RSMS has higher SAP than SSMS (71.0% versus 64.7%) while keeping MAD below HAC; HAC has the highest SAP (75.7%) but the largest MAD. The weighted-CvM rows give the same ordering, so RSMS KS remains the main self-normalized statistic.

4.3 Size-adjusted power and detection delay

The size-adjusted power (SAP) profiles for the dependent fMA(1) setting, in which serial dependence makes the choice of normalization most consequential, are shown in Supplement Section S3.1 (the aggregate tables above already cover BB and fIID). RSMS has higher size-adjusted power than SSMS in the reported fMA(1) panels across horizons, break magnitudes,

and both break designs. The difference is largest in the smooth-change design: when the break accumulates gradually, SSMS is slow to reject because its self-normalizer absorbs more of the post-break movement, whereas RSMS retains more power without moving to the more size-distorted HAC benchmark.

4.4 Detection delay, weighted-CvM, and robustness

The raw rejection and ADD profiles for fMA(1) in Supplement Sections S3.1 and S3.4 reinforce the same ranking. At $\Delta = 0$, HAC is most aggressive and SSMS closest to nominal size; after a break, RSMS usually stops earlier than SSMS, especially under smooth change, while HAC's occasional speed comes with higher false-alarm risk.

For weighted-CvM diagnostics, the late-emphasis weight w_{L} has the highest average SAP within each standardization family, and the full weight-by-weight and ADD comparisons are in Supplement Sections S3.1 and S3.4. RSMS remains above SSMS within every weight profile. Robustness checks (Supplement Sections S3 and S3.6) give the same ordering under localized changes and mild training-sample contamination; once contamination becomes visible, around $b_{\mathrm{train}} = 0.01$, the training sample is no longer a credible reference.

Taken together, the simulation evidence favors RSMS KS: relative to SSMS it is less conservative and detects more after the break, and relative to HAC it avoids kernel and bandwidth choices and stops less aggressively before the break under mild contamination. The weighted-CvM and supplementary detectors are useful checks but do not change the recommendation.

5. Empirical Analysis

We use processed 1-minute S&P 500 prices from TickData to monitor the mean intraday return-curve representation around the COVID-19 market disruption. The training sample is the last 50 trading days of 2019, from October 21 to December 31, and monitoring starts on January 2, 2020. A boundary crossing is the first day on which the fixed training-sample FPCA representation no longer adequately describes incoming return curves. Because $m = 50$ leaves at most 250 monitoring days, we report $T \in \{1, 2, 5\}$.

For each day we compute adjacent 1-minute log returns, map intraday time to $[0, 1]$, smooth with cubic B-splines, and evaluate on a 301-point grid. FPCA is estimated from the training sample only; 80% FVE retains $K = 4$, and HAC, SSMS, and RSMS use this common score process for KS and weighted-CvM monitoring. Supplement Section S5 reports alternative detector shapes, retained-dimension checks, alternative training

lengths, and the 3D return-surface view. The retained-dimension checks preserve the RSMS-before-HAC ordering at $T = 2$ and $T = 5$; with longer training samples, RSMS and HAC move closer to the March stress window, while SSMS remains below threshold in the reported checks. Supplement Section S3.3 reports the RSMS off-diagonal diagnostic.

Although the episode is studied after the fact, each statistic is computed as it would have been in real time, using only the fixed 2019 training sample and observations available up to each day. Relative to the HAC KS benchmark, RSMS KS signals about 24–25 trading days earlier at $T = 2$ and about 14 days earlier at $T = 5$, while the return-curve environment is still changing rather than after stress is already visible in aggregate daily measures. Supplement Section S5 reports daily market background.

Figure 1 shows that the transition is distributed over the trading day–a sharply more negative opening, a more uneven middle, and a more pronounced late-day profile–which is why we monitor the return curve itself rather than only scalar aggregates.

Table 1 shows the key KS pattern. RSMS crosses at all three horizons and both γ values, with first signals on February 26 for $T = 1$, February 27 for $T = 2$, and March 13 for $T = 5$. SSMS never crosses. HAC crosses at $T = 2$ and at $T = 5$ only for $\gamma = 0.15$, first signaling on April 1–2.

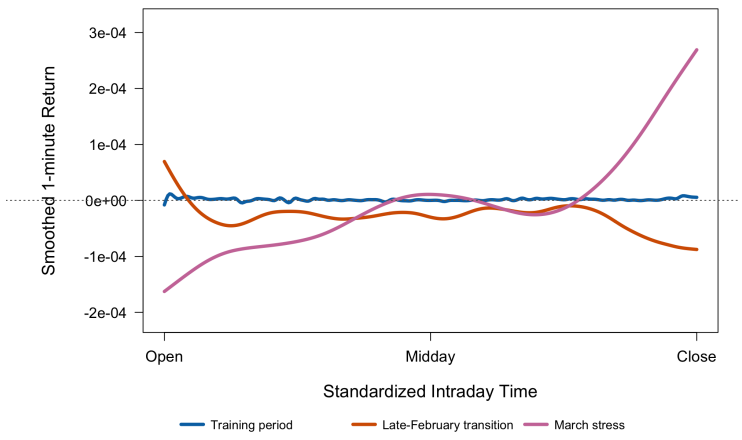

Figure 1: Average intraday return curves across three market states. The training curve is nearly flat, whereas the transition and stress curves show pronounced opening and late-day changes and a more uneven middle segment.

Relative to the March 16 stress reference date, the RSMS signals at $T = 2$ and $T = 5$ occur 12 and 1 trading days earlier, while the corresponding HAC signals occur 12–13 trading days later. In this application, RSMS is the only procedure that signals during the transition from pre-crisis trading to the high-volatility COVID regime.

The weighted-CvM evidence in Supplement Section S5 is more horizon-dependent: only RSMS versions reject, HAC and SSMS remain below threshold, and the earliest weighted-CvM signal is March 13 at $T = 1$.

Table 1: Empirical KS monitoring results for the S&P 500 intraday return curves.

Standardizer	γ	Stat./CV	Reject?	First signal
Panel A: $T = 1$				
RSMS	0.00	1.98	Yes	Feb. 26, 2020
RSMS	0.15	1.90	Yes	Feb. 26, 2020
SSMS	0.00	0.05	No	No signal
SSMS	0.15	0.05	No	No signal
HAC	0.00	0.65	No	No signal
HAC	0.15	0.63	No	No signal
Panel B: $T = 2$				
RSMS	0.00	3.34	Yes	Feb. 27, 2020
RSMS	0.15	3.36	Yes	Feb. 27, 2020
SSMS	0.00	0.10	No	No signal
SSMS	0.15	0.10	No	No signal
HAC	0.00	1.22	Yes	Apr. 2, 2020
HAC	0.15	1.25	Yes	Apr. 1, 2020
Panel C: $T = 5$				
RSMS	0.00	2.57	Yes	Mar. 13, 2020
RSMS	0.15	2.75	Yes	Mar. 13, 2020
SSMS	0.00	0.08	No	No signal
SSMS	0.15	0.09	No	No signal
HAC	0.00	0.96	No	No signal
HAC	0.15	1.05	Yes	Apr. 2, 2020

Note: Stat./CV is the maximum KS path over $1 \leq k \leq \lfloor mT \rfloor$, divided by its 5% finite-horizon critical value. First signal is the first calendar day exceeding the critical value; "No signal" means no crossing within the horizon.

In the days between the RSMS and HAC KS alarms, the raw data were already far from the training regime: average absolute 1-minute returns were about 10–12 times the training-sample level, and daily realized variation was about 95–110 times as large. This is the economic content of the earlier alarm. It would have prompted a review of the return-curve representation during the transition, not only after the stress episode was fully

visible.

6. Conclusion

After a functional representation has been estimated from a training sample, when should it be refreshed? We propose tuning-lean self-normalized monitoring statistics that fix the training-sample FPCA score coordinates and monitor the subsequent score process, keeping HAC, SSMS, and RSMS on common coordinates so that performance differences come from the normalization rather than from different representations.

The evidence points to a clear practical recommendation: the adjusted-range self-normalized KS statistic circumvents the kernel and bandwidth choices of HAC LRV estimation while being less conservative than SSMS in the simulations, and the same pattern appears in the S&P 500 application, where RSMS signals during the COVID return-curve transition, before HAC and well before SSMS. We therefore recommend the adjusted-range self-normalized KS statistic as the main online monitoring statistic.

A natural next step concerns what should happen after an alarm. In the present framework the FPCA coordinates are kept fixed at their training-sample estimates, which makes the monitoring statistic transparent; after a rejection, one must decide which observations define a new stable train-

ing period, whether the basis should be re-estimated, and how monitoring should restart without masking subsequent breaks. Further directions include monitoring functionals with more direct economic content (volatility, covariance, tail risk, or functional-regression stability), localizing which part of the domain drives a rejection, and studying panels of functional objects and finite-sample critical values under approximate diagonalization.

Supplementary Material

The online Supplement contains the proofs of the theoretical results, the simulated critical values for the monitoring statistics, additional simulation and contamination studies, alternative detector definitions and comparisons, and supplementary empirical robustness checks. Throughout the main text, cross-references of the form "Supplement Section Sx" refer to the numbered sections of this online Supplement, which is provided as a single PDF.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (NSFC Grant No. 72173120); the Ministry of Education of Taiwan (Yushan Fellowship, Yushan Scholar); the project "IDA Institute for Digital

Assets" (CF166/15.11.2022; Contract No. CN760046/23.05.2023), financed under Romania's National Recovery and Resilience Plan (Apel nr. PNRR-III-C9-2022-25I8); the Marie Skłodowska-Curie Actions under the European Union's Horizon Europe research and innovation program (Industrial Doctoral Network on Digital Finance, DIGITAL, Project No. 101119635); and COST Action CA19130 and COST Action CA21163, funded by COST (European Cooperation in Science and Technology).

References

Aston, J. A. D. and C. Kirch (2012). Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis 109*, 204–220.

Aue, A., S. Hörmann, L. Horváth, and M. Hušková (2014). Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica 24*(3), 1043–1073.

Aue, A. and C. Kirch (2024). The state of cumulative sum sequential changepoint testing 70 years after page. *Biometrika 111*(2), 367–391.

Aue, A., G. Rice, and O. Sönmez (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society Series B: Statistical Methodology 80*(3), 509–529.

Bastian, P., R. Basu, and H. Dette (2024). Multiple change point detection in functional data with applications to biomechanical fatigue data. *The Annals of Applied Statistics 18*(4),

3109–3129.

Berkes, I., R. Gabrys, L. Horváth, and P. Kokoszka (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society Series B: Statistical Methodology 71*(5), 927–946.

Boniece, B. C., L. Horváth, and L. Trapani (2025). On changepoint detection in functional data using empirical energy distance. *Journal of Econometrics 250*, 106023.

Carlstein, E. G. and D. Siegmund (1994). *Change-point Problems*. Institute of Mathematical Statistics.

Chu, C.-S. J., M. Stinchcombe, and H. White (1996). Monitoring structural change. *Econometrica 64*(5), 1045–1065.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley. MR0829458.

Hong, Y., O. Linton, B. McCabe, J. Sun, and S. Wang (2024). Kolmogorov–smirnov type testing for structural breaks: A new adjusted-range based self-normalization approach. *Journal of Econometrics 238*(2), 105603.

Hörmann, S., Ł. Kidziński, and M. Hallin (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*(2), 319–348.

Hörmann, S. and P. Kokoszka (2010). Weakly dependent functional data. *The Annals of Statistics 38*(3), 1845–1884.

Horváth, L., M. Hušková, and P. Kokoszka (2010). Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis 101*(2), 352–367.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000). Simple robust testing of regression hypotheses. *Econometrica 68*(3), 695–714.

Kutta, T. and N. Dörnemann (2025). Monitoring time series with short detection delay. *Electronic Journal of Statistics 19*(1), 2239–2275.

Kutta, T., A. Jach, and P. Kokoszka (2026). Monitoring panels of sparse functional data. *Journal of Time Series Analysis 47*(3), 660–674.

Kutta, T. and P. Kokoszka (2025). Monitoring of functional time series. *Bernoulli 31*(4), 3356–3381.

Li, Y., N. Wang, and R. J. Carroll (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association 108*(504), 1284–1294.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association 96*(455), 1066–1076.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society Series B: Statistical Methodology 72*(3), 343–366.

Shao, X. and X. Zhang (2010). Testing for change points in time series. *Journal of the American Statistical Association 105*(491), 1228–1240.

Sharipov, O., J. Tewes, and M. Wendler (2016). Sequential block bootstrap in a hilbert space

with application to change point analysis. *Canadian Journal of Statistics 44*(3), 300–322.

Zhang, X., X. Shao, K. Hayhoe, and D. J. Wuebbles (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics 5*, 1765–1796.

Zhang, Y., C. Zhu, and X. Shao (2026). Change-point detection for object-valued time series. *Journal of Business & Economic Statistics 44*(1), 255–269.

Jiajing Sun, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China.
E-mail: jiajing.sun@gmail.com

Meiting Zhu (Corresponding author), School of Economics, Xiamen University, Xiamen, Fujian 361005, China.
E-mail: zzz005297@gmail.com

Wolfgang Karl Härdle, IDA Institute of Digital Assets, Bucharest University of Economic Studies, Bucharest 010374, Romania; National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan; University of Edinburgh, Edinburgh EH8 9AB, United Kingdom; Humboldt-Universität zu Berlin, 10099 Berlin, Germany.
E-mail: haerdle@hu-berlin.de

Oliver Linton, Faculty of Economics, University of Cambridge, Cambridge, United Kingdom.
E-mail: obl20@cam.ac.uk

Zhuo Lin, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; and University of Chinese Academy of Sciences, Beijing 100190, China.
E-mail: linzhuo2021@amss.ac.cn