Online Monitoring of Structural Change with Adjusted-Range Self-Normalization

Meiting Zhu[1]
[1]School of Economics, Xiamen University, Xiamen, China
and
Yongmiao Hong[2]
[2] Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing, China
and
Jiajing Sun[3]*
[3]School of Economics and Management,
University of Chinese Academy of Sciences, Beijing, China
and
Oliver Linton[4]
[4]Faculty of Economics, University of Cambridge, UK

December 25, 2025

Abstract

We study online detection of structural change in time series and propose a sequential monitoring scheme based on adjusted-range self-normalization. Unlike CUSUM procedures that require long-run variance estimation—and thus bandwidth, kernel, or block choices—our method is fully tuning-free. Relative to classic self-normalization, which can erode power for Kolmogorov–Smirnov–type detectors as shifts grow, the adjusted-range normalizer curbs over-inflation and preserves sensitivity in real time. We derive the limiting null distribution of the monitor and establish consistency under weak conditions. Simulations benchmarked against kernel-HAC CUSUM and standard self-normalized monitors show accurate size, materially higher

*Corresponding author: jiajing.sun@gmail.com. We thank participants at Tianjin University's Fifth Ph.D. Academic Forum (2023), the Asian Meeting of the Econometric Society (2024), and the Doctoral Thesis Presentation Workshop at Xiamen University's Paula and Gregory Chow Institute for valuable comments. We are grateful to the London Mathematical Society and the University of Birmingham for supporting the workshop *Advanced Techniques in Time Series Analysis*. We also thank Cheng Hsiao, Whitney Newey, Lorenzo Trapani, Shouyang Wang, Cuihong Yang, and Wolfgang Karl Härdle for helpful insights. This work was supported by the National Natural Science Foundation of China (NSFC) under grants 71988101 and 72173120. The authors declare no competing interests. R code is available at Zenodo: 10.5281/zenodo.17460569.

power, and shorter average run lengths, especially when the training window is mildly contaminated by undetected breaks. An application to high-frequency USD/GBP around the 2016 UK referendum and the 2020 COVID-19 turmoil demonstrates timely alarms and robust performance without tuning.

Keywords: sequential monitoring; change-point detection; self-normalization; adjusted-range-based; average run length.

1 Introduction

Structural change is a pervasive feature of modern data—arising with shifts in monetary policy, market microstructure, public-health crises, and environmental shocks. Ignoring such breaks can induce model misspecification and misleading inference. Early work established powerful retrospective tools for testing and estimating break dates (*e.g.* Andrews 1993, Bai & Perron 1998, 2003, Chen & Hong 2012, 2016, Perron & Yamamoto 2015, Perron 2006), while surveys emphasize their centrality for economic time series (Aue & Horváth 2013).

Increasingly, however, the relevant question is sequential and practical: *Is yesterday's model still adequate for today's data?* (Chu et al. 1996). Online monitoring schemes address this by using an initial training window of length m and repeatedly testing incoming observations against the baseline at a controlled false-alarm rate. CUSUM-type procedures are popular (Gut & Steinebach 2002, Berkes et al. 2004, Zeileis et al. 2005, Gombay & Horváth 2009, Gombay & Serban 2009, Horváth et al. 2022), and related *sample-normalized* and *ratio-type* monitors have a long history in the sequential-change literature (see, *e.g.*, Horváth et al. 2008, Horváth & Rice 2024). These approaches typically normalize by quantities estimated from the training sample, and their justification rests on partial-sum approximations for estimators and empirical-process functionals. In practice, however, procedures that rely on heteroskedasticity and autocorrelation consistent (HAC) long-run variance (LRV) estimation introduce tuning choices (kernel, bandwidth, block length) and sensitivity to dependence and heteroskedasticity, which matter in finite samples and can distort size or erode power (Newey & West 1987, Andrews 1991, Newey & West 1994, Müller 2007). From a methodological standpoint, our setup follows the classic linearization/partial-sum route—tracing back to Durbin's early ideas on recursive/score-based expansions and sequential diagnostics (*e.g.*, Durbin 1969, Brown et al. 1975) and developed systematically in change-point monographs and surveys (Csörgő & Horváth 1997, Horváth & Rice 2024)—which makes transparent the

link between observations, influence functions (IFs), and the resulting monitoring statistics.

These considerations motivate stabilization strategies that retain sequential responsiveness while reducing tuning and robustness burdens. A compelling remedy is self–normalization (SN) (Lobato 2001, Kiefer & Vogelsang 2005, Shao 2010), which replaces the unknown LRV with a data–dependent, nuisance–free normalizer constructed from partial sums. In sequential monitoring, SN is attractive because it eliminates delicate tuning choices (no bandwidths, kernels, or block sizes), yields pivotal or near–pivotal asymptotic critical values under weak conditions that accommodate serial correlation and heteroskedasticity, maintains stable size in the presence of realistic dependence, and imposes a far lower computational burden than resampling schemes. At the same time, the classical SN normalizer can grow too quickly in the face of sizable structural shifts, producing "good size but weak power" for KS–type statistics (Shao & Zhang 2010); this shortcoming motivates power–enhanced self–normalization designs tailored to preserve sensitivity under pronounced changes.

We build on recent progress showing that an *adjusted-range* self-normalizer—based on the range of partial sums rather than their quadratic variation—ameliorates this power defect for KS-type tests (Hong, Linton, McCabe, Sun & Wang 2024, Hong et al. 2025). Intuitively, the adjusted-range grows more slowly under structural change than the quadratic SN normalizer, so the resulting test statistic increases monotonically with break magnitude and is less sensitive to mild contamination of the training window (a realistic risk as m grows). This improved behavior is particularly relevant for online monitoring, where sequential accumulation makes occasional Type II "leakage" of borderline breaks into the training set plausible, and where the self-normalizer is computed once and reused throughout monitoring.

Building on these insights, we extend the adjusted-range self-normalization framework to sequential change-point monitoring and develop its accompanying asymptotic theory. In the online setting, the self-normalizer is recursively reused over time, which raises analytical

challenges related to dependence between training and monitoring periods and joint weak convergence under sequential stopping rules. While our theoretical analysis assumes a stationary training sample, the adjusted-range design offers practical robustness to mild contamination that may arise in real-time applications. We provide a unified asymptotic treatment for both finite and open-end horizons, establish the validity of simulation-based critical values, and demonstrate that adjusted-range–based self-normalization can be effectively operationalized for sequential change-point detection.

In this paper, we propose an adjusted–range–based self–normalized sequential monitoring scheme (RSMS) for online change–point detection. This is, to our knowledge, the first online monitoring scheme to deploy adjusted-range self-normalization, relative to a standard CUSUM monitor with HAC LRV (CSMS) and the self–normalized scheme of Chan et al. (2021) (SSMS). RSMS preserves the tuning–free, nuisance–free advantages of self–normalization—eschewing bandwidths and block choices while admitting simple simulation–based critical values; its adjusted–range normalizer retains power under sizable shifts and exhibits robustness to mild contamination in the training sample, thereby mitigating the monotonicity and power losses documented for quadratic SN in KS settings (Shao & Zhang 2010, Hong, Linton, McCabe, Sun & Wang 2024, Hong et al. 2025); and it accommodates a standard early–signal weight $\gamma \in [0, 0.5)$ to emphasize incipient changes, with simulations showing that small positive γ substantially shortens average run length (ARL) for RSMS at little size cost, whereas the same adjustment is less beneficial for SSMS (see also Homm & Breitung 2012, Horváth et al. 2022, Kirch & Stoehr 2022).

We study functionals of empirical distributions, covering approximately linear statistics and M-estimators, and derive the null asymptotics with simulation-based critical values for finite and open-end horizons. Under the alternative hypothesis, we establish the consistency of the proposed procedures. Monte Carlo results indicate that RSMS achieves accurate size

(relative to SSMS, which tends to be over-conservative, and CSMS, which is tuning-sensitive), *higher power*, and *shorter ARL*. An empirical analysis of the USD/GBP exchange rate around the 2016 Brexit referendum and the 2020 COVID-19 shock corroborates these findings: RSMS flags breaks promptly and robustly, often earlier than SSMS and competitively with (or better than) well-tuned CSMS. Overall, RSMS offers a practical, computation-light, and tuning-free online monitor with strong finite-sample performance and theoretical guarantees, making it well-suited for high-frequency and real-time decision environments.

2 Adjusted-Range-Based Self-Normalized Online Change-Point Detection

2.1 Quantities of Interest: Empirical Functionals and Asymptotics

We work with a broad class of targets—functionals of empirical distributions—covering approximately linear statistics (e.g., marginal mean, variance, covariance) (Shao 2010, Dette & Gösmann 2020, Hong, Linton, McCabe, Sun & Wang 2024, Hong et al. 2025) and standard estimators defined by estimating equations. The latter includes M-estimators, encompassing maximum likelihood (ML), generalized method of moments (GMM), least squares (LS), and robust or quantile regression (Kunsch 1984, 1989).[1]

Let $\{X_t\}_{t=1}^n$ be a d-dimensional stationary time series. The empirical d-variate distribution based on $\{X_t\}_{t=i}^j$ is

$$\widehat{\mathbf{F}}_{i,j}^d = \frac{1}{j-i+1}\sum_{t=i}^{j}\delta_{X_t},$$

where δ_x denotes the Dirac measure placing unit mass at $x \in \mathbb{R}^d$. For the full sample we write $\widehat{\mathbf{F}}_{1,n}^d \equiv \widehat{\mathbf{F}}_n^d$.

[1] The framework is deliberately general to accommodate these estimator classes. Because M-estimators are widely used and robust to outliers through flexible loss functions, we give a dedicated treatment in Appendix A.3.

Let $\theta = \mathcal{G}(\mathbf{F}^d) \in \mathbb{R}^q$ be the target, where $\mathcal{G}$ is a functional on the d-variate distribution $\mathbf{F}^d$, and define the plug-in estimator $\widehat{\theta}_n = \mathcal{G}(\widehat{\mathbf{F}}_n^d)$.

Following Kunsch (1984), consider the class of plug-in estimators

$$\widehat{\theta}_n = \mathcal{G}\Big(\widehat{\mathbf{F}}_n^d\Big), \qquad \mathcal{G} : \mathscr{M}_{\text{stat}}^d \to \Theta, \tag{1}$$

where $\widehat{\theta}_n \in \mathbb{R}^q$ is computed from $\{X_t\}_{t=1}^n$, $\Theta \subset \mathbb{R}^q$ is compact, and $\mathscr{M}_{\text{stat}}^d$ denotes the set of d-variate distributions of stationary processes.

The representation in (1) accommodates estimators defined implicitly as solutions to

$$\sum_{t=1}^{n} \psi\Big(X_t; \widehat{\theta}_n\Big) = 0, \tag{2}$$

for a measurable map $\psi : \mathbb{R}^d \times \Theta \to \mathbb{R}^q$. The population counterpart $\theta = \mathcal{G}(\mathbf{F}^d)$ is characterized by

$$\int \psi(x; \theta)\, \mathbf{F}^d(\mathrm{d}x) = 0, \tag{3}$$

linking the underlying distribution $\mathbf{F}^d$ to the estimating function $\psi(\cdot)$. Illustrative choices of ψ include: (i) *Descriptive statistics.* For univariate X_t and mean μ, $\psi(X_t; \mu) = X_t - \mu$, so $\sum_{t=1}^n \psi(X_t; \widehat{\mu}_n) = 0 \Rightarrow \widehat{\mu}_n = n^{-1} \sum_{t=1}^n X_t$; for variance with $\theta = (\mu, \sigma^2)'$, take $\psi_1(X_t; \mu, \sigma^2) = X_t - \mu$ and $\psi_2(X_t; \mu, \sigma^2) = (X_t - \mu)^2 - \sigma^2$. (ii) *Least squares.* In the linear model $Y_t = X_t'\beta + u_t$ with $E(u_t \mid X_t) = 0$, the estimator solves $\sum_{t=1}^n \psi((X_t, Y_t); \beta) = 0$ with $\psi((X_t, Y_t); \beta) = X_t\Big(Y_t - X_t'\beta\Big)$, the first-order condition from minimizing $\sum_{t=1}^n (Y_t - X_t'\beta)^2$ (Kiefer & Vogelsang 2002). (iii) *Maximum likelihood.* With log-likelihood contribution $\ell_t(\theta)$, the score $\psi(Z_t; \theta) = \nabla_\theta \ell_t(\theta)$ yields $\sum_{t=1}^n \psi(Z_t; \widehat{\theta}_n) = 0$. See also Hong et al. (2025) and, for additional time-series examples (including AR(p)), Kunsch (1984).

We define the IF as

$$\mathbf{IF}\Big(x; \mathbf{F}^d\Big) = \lim_{\varepsilon \searrow 0} \frac{\mathcal{G}\Big((1-\varepsilon)\mathbf{F}^d + \varepsilon \delta_x\Big) - \mathcal{G}(\mathbf{F}^d)}{\varepsilon},$$

where $\delta_x(z) = \mathbf{1}\{x \leq z\}$ denotes the distribution function of the Dirac measure at $x \in \mathbb{R}^d$, with the inequality interpreted component-wise. The standing conditions below parallel those in Shao (2010), Shao & Zhang (2010), Hong, Linton, McCabe, Sun & Wang (2024); see these references for discussion and motivation.

Denote $D([0, T+1], \mathbb{R}^q)$ as the Skorokhod space of càdlàg (right-continuous with left limits) functions, equipped with the J_1 topology; "$\Longrightarrow$" stands for weak convergence in this space.

Assumption 1. *Let $\{X_t\}_{t\in\mathbb{N}}$ be strictly stationary with $E\{\mathbf{IF}(X_t; \mathbf{F}^d)\} = 0$. As $m \to \infty$ and for any $T \in (0, \infty]$,*

$$m^{-1/2} \sum_{t=1}^{\lfloor rm \rfloor} \mathbf{IF}(X_t; \mathbf{F}^d) \implies \sqrt{\mathbf{\Sigma}(\mathbf{F}^d)}\, \mathbf{B}_q(r) \quad \text{in } D\Big([0, T+1], \mathbb{R}^q\Big) \text{ under the Skorokhod } J_1 \text{ topology,}$$

where $\mathbf{B}_q(\cdot)$ is a q-dimensional standard Brownian motion and $\sqrt{\mathbf{\Sigma}(\mathbf{F}^d)}$ is a $q \times q$ lower–triangular matrix with nonnegative diagonal entries. The long–run variance–covariance matrix is

$$\mathbf{\Sigma}(\mathbf{F}^d) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\Big\{\mathbf{IF}(X_0; \mathbf{F}^d), \mathbf{IF}(X_k; \mathbf{F}^d)\Big\},$$

and is positive definite.

The functional CLT above holds under standard primitives (e.g., summable α- or β-mixing with a $2+\delta$ moment, near-epoch dependence, or finite-q physical dependence) that ensure weak convergence in $D([0, \infty), \mathbb{R}^q)$ under J_1; see, for example, Billingsley (1968), Whitt (2002). This covers both fixed $T < \infty$ and open-end $T = \infty$.

Assumption 1 aligns with Assumption 1 in Shao (2010) and Assumption 3.1 in Shao & Zhang (2010). As those papers emphasize, it is a high-level (functional CLT) requirement rather than a primitive one; see, for example, Assumption 2.1 in Phillips (1987) for primitive conditions under which it holds.

Assumption 2. *The remainder term* $\mathbf{Re}_{i,j}$ *is defined via*

$$\mathcal{G}\left(\widehat{\mathbf{F}}_{i,j}^{d}\right)=\mathcal{G}\left(\mathbf{F}^{d}\right)+\frac{1}{j-i+1}\sum_{t=i}^{j}\mathbf{IF}\left(X_{t},\mathbf{F}^{d}\right)+\mathbf{Re}_{i,j},$$

and satisfies $\sup_{1\leq i<j\leq n}(j-i+1)\cdot|\mathbf{Re}_{i,j}|=o_{p}\left(n^{1/2}\right)$.

In view of (3), we set $\mathbf{\Sigma}_{\psi}=M\mathbf{\Sigma}\left(\mathbf{F}^{d}\right)M'$, where M is a normalization matrix that aligns the scaling of the IF with $\psi(\cdot)$. We use the notation $\mathbf{\Sigma}_{\psi}$ when emphasizing a parametric specification. See (A.21) and (A.22) in Appendix A.3 for the mapping between the IF and $\psi(\cdot)$, and for the explicit expression of M.

2.2 Sequential Monitoring Framework and Hypothesis Formulation

The IF together with Assumptions 1 and 2 provide a baseline framework for analyzing estimators under stationarity. In practice, however, time series frequently exhibit structural breaks that violate these conditions. To accommodate such departures, we extend attention to a dynamic environment in which change-points can shift the parameters of interest. The aim is to detect these shifts sequentially, allowing the procedure to adapt promptly to evolving data-generating mechanisms.

We observe the sequence $\{X_t\}$ in real time. A change-point occurs at time t^* if the parameter equals θ_0 for $\{X_t\}_{t=1}^{t^*}$ and switches to $\theta_1\neq\theta_0$ for $\{X_t\}_{t=t^*}^{n}$. The objective is to identify this change in θ with minimal delay.

Let the joint likelihood of $\{X_t\}_{t=1}^{n}$ be f_θ, and consider $\{f_\theta:\theta\in\Theta\}$ as a parametric family indexed by θ. The parameter is estimated by (2): $\sum_{t=1}^{n}\psi(X_t,\widehat{\theta}_n)=0$.

The first m observations serve as the training sample and, in practice, are typically verified by retrospective tests to be free of structural breaks. The subsequent observations, $\{X_t\}_{t=m+1,m+2,\ldots,m+mT}$, arrive sequentially and form the testing sample, where $T\in(0,\infty]$

is a user-specified (possibly infinite) ratio of the monitoring horizon to the training-sample size.

Online change-point detection can be posed as a hypothesis test with null

$$H_0 : \theta = \theta_0, \quad \text{for } t = 1, 2, \ldots, m + mT,$$

against the alternative

$$H_1 : \theta = \begin{cases} \theta_0, & \text{for } t = 1, 2, \ldots, t^* - 1, \\ \theta_1, & \text{for } t = t^*, t^* + 1, \ldots, m + mT, \end{cases}$$

where $t^* > m$ is the unknown change-point and $t^* = m + k^*$ for some $k^* > 0$. Provided $\theta = \theta_0$ in the training sample, standard regularity conditions yield $\widehat{\theta}_m \xrightarrow{p} \theta_0$, where $\widehat{\theta}_m$ is obtained from the training sample by solving (2).

We refer to $\{\psi(X_t, \widehat{\theta}_m)\}_{t=1}^{m}$ as the "generalized residuals" and to $\{\psi(X_t, \widehat{\theta}_m)\}_{t=m+1}^{m+mT}$ as the "generalized forecast errors." Intuitively, in the absence of change-points, the sequence $\{\psi(X_t, \widehat{\theta}_m)\}_{t=m+1,m+2,\ldots}$ should be similar in distribution to $\{\psi(X_t, \widehat{\theta}_m)\}_{t=1}^{m}$.

More precisely, consider the CUSUM statistic

$$S_m\left(k, \widehat{\theta}_m\right) = \sum_{t=m+1}^{m+k} \psi\left(X_t, \widehat{\theta}_m\right). \tag{4}$$

Under no change, the sequence $\{\psi(X_t, \widehat{\theta}_m)\}_{t=m+1}^{mT}$ has expectation close to zero, so $S_m(k, \widehat{\theta}_m)$ behaves like a Brownian motion. Note that the "generalized residuals," and hence the "generalized forecast errors," typically exhibit temporal dependence; independence arises only in special cases—for example, when $\widehat{\theta}_m$ comprises the mean and variance for serially i.i.d. Gaussian data.

Accordingly, the CUSUM process (4) should be standardized by an LRV estimator, such as the HAC estimator (Newey & West 1987, 1994, Andrews 1991). However, Müller (2007) shows that HAC-robust tests can suffer from size distortions in finite samples with realistic

dependence. Alternatives—fixed-b asymptotics, block bootstrap, and subsampling—require tuning choices (Hong, Linton, McCabe, Sun & Wang 2024, Hong et al. 2025). See Appendix A.1 for the detailed specification of the **C**USUM-based **s**equential change-point **m**onitoring **s**cheme (CSMS), which employs the standard CUSUM process normalized by the HAC LRV estimator.

To avoid tuning-parameter choices, Chan et al. (2021) adopt Shao's (2010) self-normalization and propose the SSMS, denoted $\mathbb{M}_m^S(k)$. See (A.2) in Appendix A.1 for the exact expression of $\mathbb{M}_m^S(k)$, and Chan et al. (2021) for further discussion. However, (A.2) is a KS-type statistic. Shao & Zhang (2010) show via simulations that the self-normalized KS statistic—constructed using Shao's (2010) self-normalization—exhibits declining power as the magnitude of the structural shift increases; see their Figure 1. This weakness does not affect (A.2) when the sequence $\{X_t\}_{t=1}^m$ is free of change-points; the issue arises when "mild" breakpoints inadvertently enter the training sample, inflating the self-normalizer and deflating $\mathbb{M}_m^S(k)$, thereby reducing the sensitivity of SSMS. This is particularly problematic for Chan et al.'s (2021) framework, which fixes T and lets $m \to \infty$, so $mT \to \infty$. At some point, a type II error is bound to occur. This limits the use of Shao's (2010) self-normalization in online change-point detection. See Appendix A.5 for simulation evidence under contaminated training, where RSMS outperforms SSMS in both detection and ARL.

We propose using Hong, Linton, McCabe, Sun & Wang's (2024) adjusted-range-based self-normalization, as the adjusted-range KS test exhibits strong power under the alternative. Owing to the robustness of range statistics (cf. Hong, Linton, McCabe, Sun & Wang (2024), Hong et al. (2025) for a literature review and discussion), the presence of "mild" structural changes—potentially included in the training sample due to type II errors—may attenuate detection sensitivity far less than in Chan et al.'s (2022) approach.

Following Hong, Linton, McCabe, Sun & Wang (2024) and Hong et al. (2025), we ap-

ply partial prewhitening by linearly transforming $\psi(\cdot)$. Specifically, define $\psi^{\dagger}\big(X_t, \widehat{\theta}_m\big) = \widehat{\mathbf{C}}_m^{-1}\psi\big(X_t, \widehat{\theta}_m\big)$, where $\widehat{\mathbf{C}}_m$ equals the identity when $\{\psi\big(X_t, \widehat{\theta}_m\big)\}_{t=1}^{m}$ exhibits no cross-dependence. In general, $\widehat{\mathbf{C}}_m$ is obtained from the square-root-free Cholesky (LDL) decomposition of the sample variance $\widehat{\mathbf{\Sigma}}_m$ of $\{\psi\big(X_t, \widehat{\theta}_m\big)\}_{t=1}^{m}$. Let $\mathbf{\Sigma}_m = \mathbf{C}_m\mathbf{D}_m\mathbf{C}_m'$ be the LDL decomposition of the variance–covariance matrix of $\{\psi\big(X_t, \widehat{\theta}_m\big)\}_{t=1}^{m}$, where $\mathbf{D}_m$ is diagonal and $\mathbf{C}_m$ is the unique unit lower triangular factor. The sample counterpart satisfies $\widehat{\mathbf{\Sigma}}_m = \widehat{\mathbf{C}}_m\widehat{\mathbf{D}}_m\widehat{\mathbf{C}}_m'$. Since the sample variance–covariance matrix consistently estimates the population counterpart, and the LDL decomposition is unique for any positive definite matrix, we have $\widehat{\mathbf{C}}_m \xrightarrow{p} \mathbf{C}$ as $m \to \infty$.

We define the CUSUM processes based on the training and testing samples as follows:

$$\widetilde{S}_m^{\dagger}\left(k, \widehat{\theta}_m\right) = \left\{\widetilde{s}_m^{\dagger(1)}\left(k, \widehat{\theta}_m\right), \ldots, \widetilde{s}_m^{\dagger(d)}\left(k, \widehat{\theta}_m\right)\right\}' = \sum_{t=1}^{k} \psi^{\dagger}\left(X_t, \widehat{\theta}_m\right) \in \mathbb{R}^d, \quad k = 1, 2, \ldots, m, \tag{5}$$

and

$$S_m^{\dagger}\left(k, \widehat{\theta}_m\right) = \left\{s_m^{\dagger(1)}\left(k, \widehat{\theta}_m\right), \ldots, s_m^{\dagger(d)}\left(k, \widehat{\theta}_m\right)\right\}' = \sum_{t=m+1}^{m+k} \psi^{\dagger}\left(X_t, \widehat{\theta}_m\right) \in \mathbb{R}^d, \quad k = 1, 2, \ldots, mT. \tag{6}$$

respectively.

The adjusted-range self-normalizer is computed from the training sample:

$$\mathbf{R}_m\left(\widehat{\theta}_m\right) = m^{-1/2}\,\mathrm{diag}\begin{pmatrix} \max_{1\le k\le m}\widetilde{s}_m^{\dagger(1)}\big(k, \widehat{\theta}_m\big) - \min_{1\le k\le m}\widetilde{s}_m^{\dagger(1)}\big(k, \widehat{\theta}_m\big) \\ \vdots \\ \max_{1\le k\le m}\widetilde{s}_m^{\dagger(d)}\big(k, \widehat{\theta}_m\big) - \min_{1\le k\le m}\widetilde{s}_m^{\dagger(d)}\big(k, \widehat{\theta}_m\big) \end{pmatrix} \in \mathbb{R}^d \times \mathbb{R}^d, \tag{7}$$

where $\mathrm{diag}(\cdot)$ returns the square matrix whose diagonal entries are given by the argument vector.

The adjusted-range monitoring statistic is defined by

$$\mathbb{M}_m^R(k) = \frac{S_m^{\dagger}\left(k, \widehat{\theta}_m\right)' \mathbf{R}_m\left(\widehat{\theta}_m\right)^{-2} S_m^{\dagger}\left(k, \widehat{\theta}_m\right)}{m\left(1 + \frac{k}{m}\right)^2\left(\frac{k}{k+m}\right)^{2\gamma}}, \tag{8}$$

where the factor $\left(\frac{k}{k+m}\right)^{2\gamma}$ increases the weight on smaller k/m ratios, thereby enhancing sensitivity to early changes.

This setup aligns with Chu et al. (1996), who argue that monitoring schemes with $\gamma \in [0, 0.5)$ are effective for detecting early structural breaks; see also Horváth et al. (2004), Horvath et al. (2007), Homm & Breitung (2012), Aue & Kirch (2024), Horváth et al. (2022), and Kirch & Stoehr (2022).

Setting $\gamma > 0$, however, can distort the type I error under the null. To mitigate this, Horváth et al. (2022) propose a correction term that converges in probability to one. While it leaves the asymptotic distribution unchanged, it relies on a consistent LRV estimator—undermining the key benefit of self-normalization, which is to avoid LRV estimation altogether.

Our simulations further indicate that $\gamma > 0$ does not improve early detection for SSMS, which may explain why Chan et al. (2021) do not incorporate a γ term. By contrast, for RSMS, a modest choice (e.g., $\gamma = 0.15$ as in our experiments) yields only minor size distortion while materially improving early detection through lower ARL. We therefore recommend a positive γ when early detection is paramount or delays are costly, such as in financial risk management or fault detection in safety-critical settings.

The proposed monitoring scheme employs the following stopping-time criterion:

Algorithm 1 Adjusted-Range-Based Online Change-Point Detection Scheme

Input: Monitoring horizon T, asymptotic decision boundaries c_R for significance level α
for $k = 1$ to mT **do**
  if $\mathbb{M}_m^R(k) > c_R$ **then**
    Stop at time k; change detected.
    return
  end if
end for
No change detected; set stopping time to $mT + 1$

The stopping rule balances sensitivity to structural change against false alarms via the choice of threshold c_R and monitoring length mT. If early termination occurs, the change-point location can be estimated ex post by casting detection as a model estimation or selection problem; see, e.g., Bai & Perron (1998, 2003).

3 Asymptotic Properties of the RSMS

3.1 Asymptotic Properties Under H_0

We impose the following identifiability conditions.

Assumption 3. *The true parameter θ_0 lies in the interior of Θ.*

Assumption 4. $\mathrm{E}[\sup_{\theta\in\Theta} \|\psi(X_t,\theta)\|] < \infty$, *and θ_0 is the unique solution to* $\mathrm{E}[\psi(X_t,\theta)] = 0$. *Specifically, for any $\epsilon > 0$ there exists $\kappa > 0$ such that $\|\theta - \theta_0\| > \epsilon$ implies $\mathrm{E}[\psi(X_t,\theta)] > \kappa$.*

Assumption 5. $\mathrm{E}\left[\sup_{\theta\in\Theta} \|\psi(X_t,\theta_0)\|^{2+\delta}\right] < \infty$ *for some $\delta > 0$, and $\{X_t\}$ is strongly mixing with coefficients α_k satisfying $\sum_{k=1}^{\infty} \alpha_k^{\delta/(2+\delta)} < \infty$.*

Assumption 6. *$\psi(X_t,\theta)$ is continuously differentiable in θ on a neighborhood V_{θ_0} of θ_0, and $\mathrm{E}\big(\sup_{\theta\in V_{\theta_0}}\|\partial\psi(X_t,\theta)/\partial\theta\|\big)<\infty$.*

The next result establishes the weak convergence of $\mathbf{R}_m\big(\widehat{\theta}_m\big)$, which in turn delivers the decision boundary $c_R=c_R(\alpha,d,T)$ for asymptotic size α to control the type I error.

Theorem 1. *Under Assumptions 1–6, we have $\widehat{\theta}_m=\theta_0+o_p(m^{-1/2})$. Moreover:*

(a). $\mathbf{R}_m^2\big(\widehat{\theta}_m\big)\Longrightarrow\big(\mathbf{\Sigma}_\psi^{1/2}\big)\,\mathbf{R}\,\big(\mathbf{\Sigma}_\psi^{1/2}\big)'$ *as $m\to\infty$, where*

$$\mathbf{R}=\left[\mathrm{diag}\left\{\sup_{r\in(0,1]}\left(\mathbb{B}_d(r)-r\mathbb{B}_d(1)\right)-\inf_{r\in(0,1]}\left(\mathbb{B}_d(r)-r\mathbb{B}_d(1)\right)\right\}\right]^2,$$

$\mathbb{B}_d(r)$ is a standard d-dimensional Brownian motion, $\mathbf{R}_m\big(\widehat{\theta}_m\big)$ is defined in (7), and $\mathbf{\Sigma}_\psi=\mathbb{V}\mathrm{ar}(\psi(X_t,\theta_0))=\sum_{k=-\infty}^{\infty}\mathrm{E}[\psi(X_t,\theta_0)\,\psi(X_{t+k},\theta_0)']$.

(b). The asymptotic size of RSMS with boundary c_R for $T<\infty$ is

$$\lim_{m\to\infty}P\Big(\mathbb{T}_m^R\le mT\Big|\,H_0\Big)=P\left(\sup_{1\le s\le T}\frac{\mathbb{U}_d(s)'\mathbf{R}^{-1}\mathbb{U}_d(s)}{(1+s)^2\left(\frac{s}{s+1}\right)^{2\gamma}}>c_R\right),\tag{9}$$

where $\mathbb{U}_d(s)=\mathbb{B}_d(1+s)-(1+s)\,\mathbb{B}_d(1)$.

(c). For $T=\infty$, if $\{X_t\}$ is geometrically ρ-mixing,[2] then

$$\begin{aligned}\lim_{m\to\infty}P\Big(\mathbb{T}_m^R<\infty\Big|\,H_0\Big)&=P\left(\sup_{1\le s<\infty}\frac{\mathbb{U}_d(s)'\mathbf{R}^{-1}\mathbb{U}_d(s)}{(1+s)^2\left(\frac{s}{s+1}\right)^{2\gamma}}>c_R\right)\\&=P\left(\sup_{0<u\le 1}\frac{\mathbb{B}_d^\dagger(u)'\mathbf{R}^{-1}\mathbb{B}_d^\dagger(u)}{u^{2\gamma}}>c_R\right),\end{aligned}\tag{10}$$

where $\mathbb{U}_d(s)=\mathbb{B}_d(1+s)-(1+s)\,\mathbb{B}_d(1)$, and $\mathbb{B}_d^\dagger(u)$ is a standard d-dimensional Brownian motion independent of $\mathbf{R}$.

The decision boundary c_R is chosen to achieve a prespecified asymptotic size α, i.e.,

$$P\left(\sup_{1\le s\le T}\frac{[\mathbb{B}_d(1+s)-(1+s)\mathbb{B}_d(1)]'\,\mathbf{R}^{-1}[\mathbb{B}_d(1+s)-(1+s)\mathbb{B}_d(1)]}{(1+s)^2\left(\frac{s}{s+1}\right)^{2\gamma}}>c_R\right)=\alpha,\tag{11}$$

[2] That is, $\rho(k):=\sup_{f\in\mathcal{L}^2(\mathcal{A}_0),\,g\in\mathcal{L}^2(\mathcal{B}_k)}|\mathrm{Corr}(f,g)|=O(a^k)$ for $a\in(0,1)$, with $\mathcal{A}_0$ and $\mathcal{B}_k$ the σ-fields generated by $\{X_t:t\le 0\}$ and $\{X_t:t\ge k\}$, respectively, and $\mathcal{L}^2(\mathcal{F})$ the space of square-integrable $\mathcal{F}$-measurable variables.

or equivalently,

$$P\left(\sup_{0<u\leq 1}\frac{\mathbb{B}_d^{\dagger}(u)'\,\mathbf{R}^{-1}\mathbb{B}_d^{\dagger}(u)}{u^{2\gamma}} > c_R\right) = \alpha. \tag{12}$$

The probabilities in (9)–(10), and hence the asymptotic boundaries c_R, are approximated via Monte Carlo simulation. We consider $T \in \{1, 2, 5, 10\}$, where $T = 1, 2$ denote short monitoring horizons, $T = 5$ a medium horizon, and $T = 10$ a long horizon. Simulations are conducted for $d = 1, \ldots, 5$ under $\gamma = 0$ and $\gamma = 0.15$. Results are reported in Table 1.

Table 1: Simulated asymptotic decision boundaries for the RSMS with $T = 1, 2, 5, 10$ and $\gamma = 0$ or $\gamma = 0.15$

	$\gamma = 0$								$\gamma = 0.15$							
	$T = 1$		$T = 2$		$T = 5$		$T = 10$		$T = 1$		$T = 2$		$T = 5$		$T = 10$	
$d\backslash\alpha$	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	2.1	1.5	2.7	2	3.4	2.5	3.9	2.8	2.7	2	3.3	2.5	3.9	2.9	4.3	3.2
2	3.2	2.4	5.2	4.1	10.9	8.3	21.1	15.9	4.8	3.7	8	6.2	19.2	14.7	44.6	33.1
3	4	3.3	6.5	5.1	13.8	10.8	26.6	21.1	5.8	4.7	9.6	7.9	23.8	18.4	53.8	41.7
4	4.8	3.9	7.6	6.2	16.2	12.8	30.7	24	6.9	5.7	11.5	9.2	28.1	22.2	63.4	49.7
5	5.5	4.5	8.9	7.3	18.3	14.8	35.3	28.6	7.8	6.5	12.7	10.4	29.9	24.5	71.3	55.3

Notes: Asymptotic decision boundaries c_R are computed by Monte Carlo with 10,000 replications. Brownian motions are simulated using $10{,}000 \times (m/T + 1)$ i.i.d. $N(0, 1)$ draws.

3.2 Asymptotic Properties Under H_1

This section establishes the consistency of the proposed RSMS. We begin by imposing mild regularity conditions on the post-change process.

Assumption 7.

(a) Let $\{X_t^*\}_{t\geq t^*}$ *denote the process after the change-point. Assume that* $\mathrm{E}[\psi(X_t^*, \theta_0)] = c \neq 0$ *for some constant* $c \in \mathbb{R}^d$.

(b) Assume that $\mathrm{E}\big(\sup_{\theta\in U_{\theta_0}} \|\psi(X_t^*, \theta)\|\big) < \infty$ *for some neighborhood* U_{θ_0} *of* θ_0.

Assumption 7 ensures that the behavior of $\psi(X_t, \theta_0)$ changes after the break, so that $\mathrm{E}[\psi(X_t^*, \theta_0)]$ becomes nonzero but remains finite in expectation.

Assumption 8. *Under the alternative* $(H_1 : \theta_1 \neq \theta_0)$*, the change occurs at* $m + k^* = \lfloor mc \rfloor$ *for some* $c \in (1, T+1)$*. Moreover, there exist two Brownian bridges,* $\big(\mathbb{B}_d(s) - s\mathbb{B}_d(1)\big)$ *and* $\big(\mathbb{B}_d(s) - \mathbb{B}_d(c) - (s-c)\mathbb{B}_d(1)\big)$*, such that the joint weak convergence*

$$\begin{pmatrix} \left\{ \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} \mathbf{IF}(X_t, \theta_0) \right\}_{s \in (0,c]} \\ \left\{ \frac{1}{\sqrt{m}} \sum_{t=\lfloor mc \rfloor + 1}^{\lfloor ms \rfloor} \mathbf{IF}(X_t^*, \theta_1) \right\}_{s \in (c, T+1]} \end{pmatrix} \xRightarrow{\mathcal{D}} \begin{pmatrix} \left\{ \sqrt{\mathbf{\Sigma}^{(1)}} \big(\mathbb{B}_d(s) - s\mathbb{B}_d(1)\big) \right\}_{s \in (0,c]} \\ \left\{ \sqrt{\mathbf{\Sigma}^{(2)}} \big(\mathbb{B}_d(s) - \mathbb{B}_d(c) - (s-c)\mathbb{B}_d(1)\big) \right\}_{s \in (c, T+1]} \end{pmatrix}$$

holds, where $\mathbf{\Sigma}^{(1)}$ *and* $\mathbf{\Sigma}^{(2)}$ *are positive definite matrices defined by*

$$\mathbf{\Sigma}^{(\ell)} = M^{-1} \mathbf{\Sigma}_{\psi}^{(\ell)} \big(M^{-1}\big)' = M^{-1} \sum_{k=-\infty}^{\infty} E\Big[\psi^{(\ell)}(X_t, \theta_0)\, \psi^{(\ell)}(X_{t+k}, \theta_0)'\Big] \big(M^{-1}\big)', \quad \ell = 1, 2.$$

Assumption 8 fixes the change-point location to ensure a non-negligible post-change segment for monitoring. A similar condition is used by Dette & Gösmann (2020).

Theorem 2 establishes the asymptotic consistency of RSMS.

Theorem 2. *For the RSMS with decision boundary* c_R *chosen to satisfy* (11) *or* (12) *at significance level* $\alpha \in (0, 1)$*, and under Assumptions 7 and 8, the asymptotic power equals one, i.e.,*

$$\lim_{m \to \infty} P\Big(T_m^R \leq mT \,\Big|\, H_1\Big) = 1 \quad \textit{for } T < \infty,$$

and

$$\lim_{m \to \infty} P\Big(T_m^R \leq mT \,\Big|\, H_1\Big) = 1 \quad \textit{for } T = \infty.$$

4 Simulation Studies

Modeling the conditional mean and variance is central to statistical inference and underlies forecasting in economics and finance. The conditional mean captures systematic dynamics over time, whereas the conditional variance quantifies forecast uncertainty.

We study structural breaks in the conditional mean of continuous time-series processes and evaluate the ability of the RSMS, SSMS, and CSMS statistics to detect such changes. Performance is assessed by (i) simulated type-I error (empirical size) under the null, (ii) rejection probability (power) under the alternative, and (iii) ARL. All results are based on 1,000 Monte Carlo replications. Empirical size is reported at the 5% and 10% nominal levels; power and ARL are reported at the 5

Appendix A.4 considers parameter constancy in a Poisson autoregressive (PAR) model, widely used for count-valued series such as transaction or default counts. In this setting the conditional mean depends on past realizations, while the conditional variance follows from the Poisson law. Because the process is equidispersed—its marginal mean equals its marginal variance—tests of PAR parameter constancy coincide with tests for constancy of both the unconditional mean and variance.

Finally, motivated by the possibility that online procedures may inadvertently admit post-shift observations into the training window, Appendix A.5 studies training-sample contamination. RSMS remains markedly more robust than SSMS in this case, delivering higher detection rates and shorter ARLs.

We consider four data-generating processes (DGPs) under the null hypothesis.

DGP1: Simple Homoskedastic Errors. Let X_t follow a bivariate VAR(1) model $X_t = \Psi X_{t-1} + \varepsilon_t$ with

$$\Psi = \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{pmatrix}, \qquad \{\varepsilon_t\} \text{ i.i.d. } \mathrm{MN}(0, I_2),$$

where MN denotes the multivariate normal distribution and I_2 is the 2×2 identity.

DGP2: VAR with Homoskedastic Errors. As in DGP1, but allowing cross-dependence

in X_t and ε_t:

$$\Psi = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}, \qquad \{\varepsilon_t\} \sim \mathrm{MN}(0, \Sigma_\varepsilon)\,, \ \Sigma_\varepsilon = \begin{pmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{pmatrix}.$$

DGP3: VAR with Conditional Heteroskedastic Errors. As in DGP2, but with conditional heteroskedasticity in the errors. The bivariate process $\{\varepsilon_t\}$ has components $\varepsilon_{i,t}$, $i = 1, 2$, each following GARCH(1,1):

$$\varepsilon_t = \Sigma_t^{1/2} e_t, \qquad \sigma_{i,t}^2 = (1 - \alpha_1 - \beta_1) + \alpha_1 \varepsilon_{i,t-1}^2 + \beta_1 \sigma_{i,t-1}^2, \ i = 1, 2,$$

where $\Sigma_t = \mathrm{diag}(\sigma_{1,t}^2, \sigma_{2,t}^2)$, $(\alpha_1, \beta_1) = (0.1, 0.2)$, and $\{e_t\}$ is i.i.d. with $e_t \sim \mathrm{MN}(0, I_2)$.

DGP4: VAR with Unconditional Heteroskedastic Errors. Introduce a volatility shift:

$$\sigma_{i,t}^2 = \sigma_0^2 \big[1 + \delta\, I(t > m/2)\big], \qquad \sigma_0 = 1, \ \delta = 0.2, \ i = 1, 2,$$

where $I(\cdot)$ is the indicator function. All other elements are as in DGP2. The series X_t is nonstationary due to δ. Note that the null remains valid (the conditional mean is unchanged), but Assumption 1 is violated.

The goal is to test constancy of the mean level. The null is $H_0 : \mu = \mu_0$, and the alternative H_1 is the negation of H_0. Accordingly, set $\psi(X_t, \widehat{\mu}_m) = X_t - \widehat{\mu}_m$ with $\widehat{\mu}_m = m^{-1} \sum_{t=1}^m X_t$. This choice of ψ is used to construct RSMS, SSMS, and CSMS. In particular, the RSMS for the mean employs the stopping time

$$C_m = \left\{ \min \left\{ k : \ \frac{\left| \sum_{t=m+1}^{m+k} (X_t - \widehat{\mu}_m) \right|}{\mathbf{R}_m(\widehat{\mu}_m)} > m^{1/2} \left(1 + \frac{k}{m}\right) c_R \right\}, \ mT + 1 \right\}.$$

The SSMS and CSMS stopping rules are defined analogously, replacing $\mathbf{R}_m(\widehat{\mu}_m)$ with the square root of either Shao's (2010) self-normalizer or the long-run variance (LRV) estimator based on the training sample, respectively. For CSMS, the LRV bandwidth is selected by Silverman's rule of thumb.

Monitoring horizons are set to mT with $T \in \{1, 2, 5, 10\}$ under the null and $T \in \{1, 2, 5\}$ under the alternative. Note that $T = 10$ approximates the open-end scheme ($T = \infty$), which is typically infeasible in practice.

We deliberately do not fit ARMA–GARCH models in order to assess how temporal dependence and heteroskedasticity affect monitoring performance. Likewise, we omit the partial prewhitening from Section 2.2 when evaluating RSMS to preserve cross-dependence effects.

Under the null (Table 2), RSMS type I error rates generally lie close to the nominal 5% and 10% levels across most scenarios, especially for moderate and larger training sizes ($m = 500, 1000$) and smaller T. For example, with $m = 500$ or 1000, realized sizes are typically within a few percentage points of the targets for all T. SSMS tends to be undersized, often below the nominal levels; for $m = 1000$ and larger T, the 5% size can approach or fall below 1%. CSMS is the most conservative, with realized type I errors frequently near zero for moderate to large T and m.

Increasing γ from 0 to 0.15 makes RSMS more liberal, with type-I error rates rising appreciably above nominal. By contrast, SSMS becomes more conservative and frequently under-rejects. While Shao–style self-normalization often delivers good size in fixed-sample settings, SSMS combines it with the horizon weight γ; in our experiments, setting $\gamma = 0.15$ makes SSMS increasingly conservative as either m or T grows. For context, Chan et al. (2021) set $\gamma = 0$ and do not report results for $\gamma > 0$. CSMS is comparatively insensitive to γ, but its size is already too low at $T = 5$ and $T = 10$, indicating overly conservative behavior regardless of γ. Overall, RSMS provides the most reliable size control among the three procedures, particularly for moderate or large training samples ($m = 500, 1000$) and for the longer monitoring horizon ($T = 10$).

Second, we evaluate empirical rejection rates under the alternative H_1. Starting from the null DGPs, we introduce a change-point at $t^* = m + k^*$. The training sample is fixed at

Table 2: Simulated type I error rates for the RSMS (R), SSMS (S), and CSMS (C) for a change in the mean ($\gamma = 0$ and 0.15).

		T=1						T=2						T=5						T=10					
		R		S		C		R		S		C		R		S		C		R		S		C	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$\gamma = 0$																									
DGP1	$m = 100$	9.0	14.4	2.6	5.1	12.4	18.8	10.5	16.2	3.0	5.8	9.0	13.1	8.2	12.6	3.6	5.9	2.2	3.0	7.4	14.0	2.6	6.4	0.0	0.3
	$m = 500$	4.3	7.3	1.5	2.5	5.0	9.0	3.3	6.8	1.1	2.0	1.8	3.7	4.2	8.6	1.4	2.9	0.1	0.3	3.4	5.3	1.4	2.6	0.0	0.0
	$m = 1000$	3.9	6.2	1.5	3.2	3.3	5.9	3.1	5.4	0.7	2.0	1.9	3.3	2.9	7.8	2.0	3.7	0.1	0.3	2.8	5.1	0.9	2.4	0.0	0.0
DGP2	$m = 100$	10.8	16.0	2.8	4.4	11.9	18.1	11.4	16.9	3.0	6.2	8.0	12.9	10.9	16.7	4.0	6.3	1.7	3.0	10.1	15.5	2.8	5.6	0.2	0.3
	$m = 500$	5.0	8.4	1.4	2.6	4.6	8.5	5.4	10.0	1.1	2.0	1.8	4.0	5.8	8.8	1.4	2.9	0.1	0.3	3.6	7.2	1.6	2.4	0.0	0.0
	$m = 1000$	4.4	7.0	1.5	3.2	3.0	5.4	4.7	8.0	0.7	2.0	2.0	3.4	4.2	7.1	2.0	3.8	0.1	0.3	3.8	6.8	0.9	2.3	0.0	0.0
DGP3	$m = 100$	10.9	16.3	2.5	4.9	11.8	17.8	11.2	17.4	2.9	6.3	7.9	12.7	10.5	16.1	3.9	5.8	1.7	3.0	10.7	15.2	2.9	5.8	0.1	0.3
	$m = 500$	4.8	8.6	1.2	2.8	5.5	8.4	5.3	10.0	1.0	1.9	1.8	4.0	5.8	9.3	1.4	2.7	0.1	0.4	4.3	7.1	1.7	2.2	0.0	0.0
	$m = 1000$	4.5	6.9	1.6	3.0	2.9	5.9	4.7	8.1	1.0	2.5	1.8	3.6	4.5	6.8	1.9	3.9	0.1	0.2	3.9	7.3	1.1	2.3	0.0	0.0
DGP4	$m = 100$	13.6	19.2	3.7	5.5	15.5	21.8	13.0	19.9	3.5	7.1	10.1	15.5	12.5	17.7	4.1	7.1	1.5	4.0	11.3	16.7	3.1	6.7	0.1	0.2
	$m = 500$	6.8	11.0	2.0	4.0	6.5	11.4	6.6	12.2	1.3	2.5	2.4	5.3	6.1	10.8	1.7	3.0	0.1	0.4	4.2	7.9	1.5	2.7	0.0	0.0
	$m = 1000$	5.4	8.9	1.8	3.7	4.4	7.6	5.2	9.6	1.2	2.8	2.6	4.2	4.4	8.1	2.2	4.6	0.2	0.4	4.3	7.6	1.3	2.9	0.0	0.0
$\gamma = 0.15$																									
DGP1	$m = 100$	14.1	20.5	0.8	1.5	14.7	20.9	14.9	20.5	1.1	1.9	8.2	13.5	11.4	18.1	0.9	2.2	1.7	2.6	12.5	19.1	1.0	1.8	0.0	0.0
	$m = 500$	6.6	11.8	0.1	0.8	6.5	10.3	6.4	12.0	0.2	0.9	1.7	3.5	7.7	11.5	0.3	0.8	0.1	0.3	6.1	11.4	0.4	1.0	0.0	0.0
	$m = 1000$	6.4	9.8	0.5	0.8	4.0	6.6	5.1	8.6	0.3	0.5	1.7	3.1	6.4	10.6	0.7	1.4	0.1	0.2	5.8	10.0	0.0	0.5	0.0	0.0
DGP2	$m = 100$	15.7	21.7	1.0	2.1	13.7	20.0	15.6	22.2	1.3	1.9	8.3	13.0	15.3	20.5	1.0	2.5	1.4	2.3	14.6	21.4	1.1	2.2	0.0	0.3
	$m = 500$	8.5	14.0	0.1	0.7	6.6	9.9	9.1	14.8	0.2	0.9	1.2	3.3	9.4	15.2	0.4	0.9	0.1	0.3	7.6	12.6	0.5	1.1	0.0	0.0
	$m = 1000$	7.1	11.7	0.5	0.8	3.7	6.6	7.5	11.9	0.3	0.5	1.7	3.2	6.3	11.4	0.7	1.3	0.1	0.2	6.8	12.6	0.0	0.7	0.0	0.0
DGP3	$m = 100$	16.0	21.6	0.8	1.5	14.1	19.9	15.9	22.5	1.3	2.0	7.8	12.0	14.8	20.1	1.2	2.4	1.1	2.2	14.1	20.6	1.2	2.7	0.1	0.3
	$m = 500$	8.4	13.9	0.1	0.8	6.5	9.2	9.4	14.8	0.3	0.8	1.5	3.3	9.4	14.9	0.5	1.1	0.1	0.3	7.4	12.6	0.5	0.9	0.0	0.0
	$m = 1000$	7.0	11.5	0.5	0.8	3.7	6.8	7.7	12.6	0.4	0.7	1.6	3.2	6.8	11.6	0.6	1.0	0.1	0.2	6.9	11.8	0.1	0.5	0.0	0.0
DGP4	$m = 100$	18.8	26.6	1.3	2.6	18.2	24.9	18.5	25.8	1.5	2.6	10.1	16.0	16.4	22.4	1.5	2.6	1.3	2.7	15.9	22.4	1.1	2.4	0.1	0.1
	$m = 500$	11.3	17.6	0.3	1.1	8.5	13.6	11.8	18.5	0.2	1.0	2.2	5.7	11.7	18.1	0.3	0.8	0.2	0.3	8.0	14.2	0.6	1.4	0.0	0.0
	$m = 1000$	9.1	14.7	0.6	0.8	5.8	9.9	9.5	14.8	0.3	1.0	2.3	4.1	8.3	14.5	0.6	1.3	0.1	0.2	8.2	13.4	0.1	1.0	0.0	0.0

$m = 500$ to balance size control and computational cost. The magnitude of the structural shift is $\Delta \in \{0.25, 0.50, 0.75, 1.00\}$. Given the deflated type I errors observed for CSMS at longer horizons (Table 2), we focus on $T \in \{1, 2, 5\}$.

We consider two forms of change:

(i) Abrupt change:

$$Y_t = \begin{cases} X_t, & 1 \leq t \leq t^*, \\ \Delta + X_t, & t^* \leq t \leq m + mT. \end{cases}$$

(ii) Smooth change:

$$Y_t = \begin{cases} X_t, & 1 \leq t \leq t^*, \\ \Delta \times \dfrac{t}{m + mT} + X_t, & t^* \leq t \leq m + mT. \end{cases}$$

Tables 3 and 4 summarize the results under H_1. As expected, power increases with the magnitude of the shift (Δ). Holding other factors fixed, abrupt changes deliver higher power than smooth changes at the same Δ, reflecting a stronger signal. Earlier breaks (e.g., $k^* = 50$) are also easier to detect than later ones (e.g., $k^* = 200$), as the monitoring scheme has more post-break observations to accumulate evidence.

Extending the monitoring horizon T initially raises power by giving the statistic more opportunity to capture the break. Beyond a point, however, this effect reverses: as the window becomes long, noise accumulation dominates and the signal-to-noise ratio deteriorates, reducing power—particularly for small or smooth changes.

Across a broad range of designs, RSMS attains the highest or near-highest rejection rates, especially as Δ grows. SSMS generally trails RSMS, with more noticeable gaps for smaller breaks or under more demanding DGPs (e.g., DGPs 2 and 3). CSMS exhibits mixed behavior: it tends to be least powerful under DGPs 1, 2, and 4, but can outperform under DGP 3.

Introducing $\gamma = 0.15$ typically boosts power for all three procedures, particularly for small-to-moderate shifts and early horizons. For example, at $k^* = 50$, $T = 1$, and $\Delta = 0.25$, RSMS rises from 45.9% to 48.2%, SSMS from 40.0% to 42.5%, and CSMS from 40.5% to 42.6%. Larger shifts ($\Delta \geq 0.5$) and $T = 2$ show even greater gains. For longer horizons ($T = 5$) and smooth changes, improvements are more muted and SSMS or CSMS may experience slight declines relative to $\gamma = 0$. Overall, $\gamma = 0.15$ enhances sensitivity to abrupt breaks, with the largest absolute gains for RSMS.

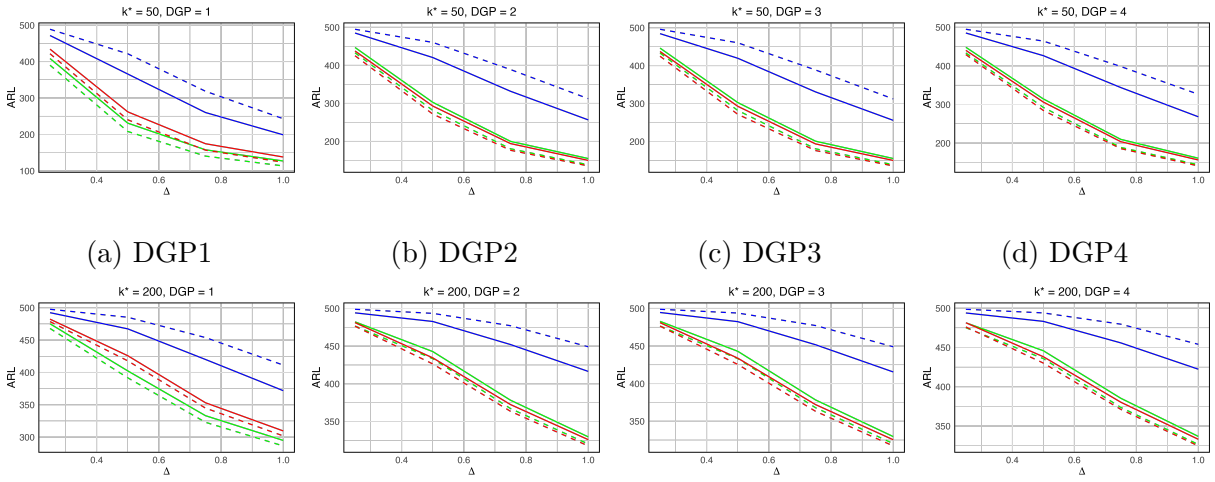
Table 3: Rejection rates (%) for detecting a mean shift using RSMS (R), SSMS (S), and CSMS (C). Training size $m = 500$; break locations $k^* \in \{50, 200\}$; $\gamma = 0$. Results shown for change types (i) abrupt and (ii) smooth.

		$k^* = 50$												$k^* = 200$									
		DGP1			DGP2			DGP3			DGP4			DGP1			DGP2			DGP3			DGP4		
T	Δ	R	S	C	R	S	C	R	S	C	R	S	C	R	S	C	R	S	C	R	S	C	R	S	C
Type i - abrupt change																									
$T=1$	0.25	45.9	40.0	40.5	35.1	21.9	12.4	12.8	11.8	59.4	37.0	36.9	33.8	20.1	18.9	19.0	18.3	9.2	6.7	6.3	6.3	29.1	18.5	18.6	19.0
$T=1$	0.5	97.6	91.9	92.1	84.4	73.0	49.7	50.0	42.9	99.4	92.4	92.5	85.6	71.9	60.2	60.0	50.1	37.4	22.9	23.1	19.2	83.0	57.2	56.7	48.2
$T=1$	0.75	100.0	99.9	99.9	99.0	93.6	82.4	82.6	74.4	100.0	100.0	99.9	99.5	97.2	91.6	91.8	83.8	72.0	48.8	48.6	41.6	99.2	91.8	91.6	84.4
$T=1$	1	100.0	100.0	100.0	100.0	98.2	94.9	95.0	90.7	100.0	100.0	100.0	100.0	99.9	99.1	99.2	97.0	90.5	73.6	73.4	63.0	100.0	99.6	99.6	97.9
$T=2$	0.25	54.1	46.0	46.1	39.6	32.5	18.7	18.6	16.1	66.7	41.5	41.9	35.5	37.0	32.6	32.6	28.6	21.1	12.5	12.3	10.9	47.6	28.7	29.2	24.9
$T=2$	0.5	99.4	96.9	96.9	92.4	86.6	66.5	66.8	58.4	100.0	97.7	97.8	92.4	95.7	89.3	89.4	81.2	74.0	51.5	51.6	43.2	98.7	87.4	87.7	79.0
$T=2$	0.75	100.0	100.0	100.0	99.8	98.4	90.8	90.4	86.2	100.0	100.0	100.0	100.0	100.0	99.5	99.4	98.8	95.2	82.3	82.1	75.0	100.0	99.9	99.9	99.1
$T=2$	1	100.0	100.0	100.0	100.0	99.9	98.0	98.0	95.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.0	94.2	94.2	91.2	100.0	100.0	100.0	100.0
$T=5$	0.25	36.4	32.6	32.4	25.4	43.9	26.5	26.9	22.3	49.1	25.0	25.0	19.4	30.2	26.2	26.5	20.7	38.5	22.3	23.1	18.6	40.6	19.8	19.9	16.2
$T=5$	0.5	98.6	94.3	94.2	87.7	91.0	77.0	77.1	69.4	100.0	95.3	95.2	87.6	97.3	90.6	90.6	82.9	87.7	72.3	72.2	63.6	99.9	91.2	90.8	80.3
$T=5$	0.75	100.0	100.0	100.0	99.7	99.4	95.5	95.8	92.1	100.0	100.0	100.0	100.0	100.0	99.9	99.9	99.4	99.1	93.7	93.6	89.5	100.0	100.0	100.0	99.8
$T=5$	1	100.0	100.0	100.0	100.0	99.9	99.6	99.6	98.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	99.1	99.1	97.7	100.0	100.0	100.0	100.0
Type ii - smooth change																									
$T=1$	0.25	26.6	23.8	24.6	22.0	12.1	8.0	8.0	7.7	36.8	23.2	23.3	22.7	14.7	15.0	14.9	14.4	6.7	5.4	4.9	5.4	21.4	14.6	14.9	16.2
$T=1$	0.5	84.2	71.5	71.9	62.1	49.1	31.8	31.8	26.7	91.4	71.2	70.9	61.3	52.6	44.5	44.8	37.6	26.3	14.6	14.8	12.6	66.2	40.9	40.8	36.5
$T=1$	0.75	99.7	96.4	96.4	91.8	83.0	60.1	59.8	51.2	100.0	97.2	97.1	92.6	90.3	78.5	79.2	69.4	56.7	36.8	37.1	31.1	94.9	78.6	78.6	68.3
$T=1$	1	100.0	100.0	99.9	99.0	93.9	83.7	83.5	75.0	100.0	100.0	99.9	99.6	99.5	95.4	95.4	90.1	80.8	56.8	57.5	48.1	100.0	96.4	96.2	90.9
$T=2$	0.25	22.3	21.1	21.6	19.0	11.7	7.6	6.9	7.1	29.2	19.0	19.0	16.9	17.2	17.6	17.8	15.8	9.5	5.8	5.4	5.8	23.5	15.2	15.2	14.0
$T=2$	0.5	83.4	71.4	71.2	61.3	54.6	33.3	33.3	27.4	91.1	67.5	67.1	56.5	73.3	61.6	61.5	51.3	44.7	26.5	26.5	21.8	83.7	56.1	56.1	46.4
$T=2$	0.75	99.4	97.0	97.0	92.0	86.7	66.4	66.5	57.7	100.0	97.7	97.6	91.7	97.7	93.3	93.4	86.5	80.4	57.5	57.4	49.4	99.6	93.3	92.9	84.7
$T=2$	1	100.0	99.9	99.8	99.2	96.5	86.3	85.9	79.0	100.0	100.0	100.0	99.6	100.0	99.4	99.3	98.1	94.1	79.9	79.6	71.9	100.0	99.7	99.6	98.8
$T=5$	0.25	6.6	6.9	7.0	5.7	13.1	8.4	8.3	8.0	9.4	4.5	4.5	3.6	6.0	6.5	6.6	5.5	12.6	8.2	8.2	7.9	9.1	4.2	4.4	3.4
$T=5$	0.5	53.0	44.1	43.7	33.6	54.9	33.2	33.6	26.8	66.4	35.4	35.6	26.0	50.0	42.0	41.7	32.1	53.1	31.4	32.3	26.2	63.4	33.0	33.2	24.4
$T=5$	0.75	95.2	85.4	85.6	75.1	84.0	66.9	66.5	56.8	99.4	84.4	84.0	70.5	94.3	83.8	83.8	72.6	82.6	64.6	64.7	54.9	99.2	81.7	81.5	68.0
$T=5$	1	99.9	99.0	99.0	95.7	96.3	86.2	85.3	79.0	100.0	99.4	99.4	96.5	99.9	98.7	98.6	95.2	96.0	85.0	84.4	77.8	100.0	99.1	99.0	95.8

Table 4: Rejection rates (%) for detecting a mean shift using RSMS (R), SSMS (S), and CSMS (C). Training size $m = 500$; break locations $k^* \in \{50, 200\}$; $\gamma = 0.15$. Results shown for change types (i) abrupt and (ii) smooth.

		$k^* = 50$												$k^* = 200$									
		DGP1			DGP2			DGP3			DGP4			DGP1			DGP2			DGP3			DGP4	
T	Δ	R	S	C	R	S	C	R	S	C	R	S	C	R	S	C	R	S	C	R	S	C	R	S	C
Type i - abrupt change																									
$T=1$	0.25	48.2	42.5	42.6	37.6	10.3	5.8	5.4	5.2	61.3	39.3	39.1	36.7	22.4	20.5	20.3	20.4	3.5	2.0	2.0	2.2	30.8	20.2	20.5	21.1
$T=1$	0.5	97.9	92.3	92.4	85.6	53.7	33.3	33.2	27.5	99.4	92.9	92.8	86.5	74.0	61.9	61.7	52.5	21.4	10.1	9.8	7.9	84.3	58.5	58.9	49.8
$T=1$	0.75	100.0	100.0	99.9	99.1	87.0	65.5	65.8	55.9	100.0	100.0	100.0	99.6	97.7	92.3	92.4	85.1	52.9	32.3	32.1	26.2	99.3	92.3	92.1	85.5
$T=1$	1	100.0	100.0	100.0	100.0	95.6	87.5	87.7	79.6	100.0	100.0	100.0	100.0	100.0	99.3	99.3	97.2	78.9	53.9	54.0	45.1	100.0	99.6	99.6	98.0
$T=2$	0.25	51.8	44.8	44.8	38.1	17.1	8.6	8.5	7.1	64.1	39.2	39.6	33.5	34.2	31.1	31.2	27.1	10.2	4.9	4.5	4.5	44.1	26.4	26.4	23.2
$T=2$	0.5	99.1	96.4	96.5	91.7	73.7	50.3	49.7	40.9	100.0	97.4	97.2	91.0	95.0	88.2	88.4	79.6	56.8	34.1	33.9	26.6	98.4	85.6	85.6	76.6
$T=2$	0.75	100.0	100.0	100.0	99.8	95.1	82.4	81.9	74.9	100.0	100.0	100.0	100.0	100.0	99.4	99.4	98.8	88.6	69.6	69.4	60.6	100.0	99.8	99.8	99.1
$T=2$	1	100.0	100.0	100.0	100.0	99.0	94.2	94.2	90.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.3	88.0	87.9	82.1	100.0	100.0	100.0	100.0
$T=5$	0.25	30.8	27.4	27.5	21.4	27.1	13.3	14.1	10.9	41.2	19.2	19.5	15.3	24.1	21.4	21.7	16.8	22.2	11.4	11.5	9.2	32.3	15.2	15.3	11.3
$T=5$	0.5	97.8	92.2	92.3	84.7	81.2	62.2	62.5	53.7	99.9	92.8	92.8	82.2	96.1	87.4	87.4	77.7	76.5	57.1	56.4	46.9	99.8	86.7	86.5	73.3
$T=5$	0.75	100.0	99.9	99.9	99.6	97.9	89.7	89.3	83.6	100.0	100.0	100.0	99.9	100.0	99.7	99.7	99.1	96.4	86.8	85.7	79.8	100.0	100.0	100.0	99.7
$T=5$	1	100.0	100.0	100.0	100.0	99.6	97.7	97.9	95.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.6	96.2	96.4	93.3	100.0	100.0	100.0	100.0
Type ii - smooth change																									
$T=1$	0.25	28.7	26.0	26.3	24.2	5.2	3.0	2.6	3.0	38.7	25.1	25.1	25.1	16.5	16.2	16.3	16.4	2.8	1.5	1.5	1.7	22.9	16.0	16.4	18.0
$T=1$	0.5	85.1	72.9	73.1	64.3	30.0	15.3	15.7	11.8	91.7	72.3	72.3	62.9	55.2	46.3	47.0	40.1	12.7	6.2	6.5	5.5	67.8	42.1	42.2	38.1
$T=1$	0.75	99.7	96.5	96.7	92.2	65.6	42.0	42.0	34.5	100.0	97.2	97.3	93.2	91.0	79.8	80.4	71.0	36.7	20.4	20.1	15.0	95.3	79.3	79.8	69.9
$T=1$	1	100.0	100.0	99.9	99.1	87.4	67.0	67.0	56.5	100.0	100.0	99.9	99.6	99.6	95.9	95.8	91.0	62.7	39.6	40.3	32.6	100.0	96.6	96.5	91.7
$T=2$	0.25	19.9	19.6	20.1	17.9	5.8	3.0	2.5	2.8	27.1	17.1	16.7	15.3	15.8	16.4	16.6	15.3	4.3	2.2	2.2	2.3	21.4	13.2	13.2	12.6
$T=2$	0.5	81.7	69.2	68.8	58.6	36.6	18.5	18.8	14.7	89.8	63.9	64.3	53.6	70.1	59.4	59.4	49.3	28.8	14.1	13.8	10.6	81.4	53.0	53.1	43.5
$T=2$	0.75	99.2	96.4	96.5	91.3	73.8	50.3	49.9	40.4	100.0	97.0	96.9	90.3	97.3	92.6	92.3	85.1	65.2	40.1	40.2	32.0	99.6	91.4	91.4	82.4
$T=2$	1	100.0	99.8	99.8	99.1	91.5	74.8	74.4	66.2	100.0	100.0	100.0	99.4	100.0	99.4	99.2	98.0	86.7	66.7	66.7	57.1	100.0	99.6	99.6	98.4
$T=5$	0.25	4.8	5.1	5.3	4.3	6.1	3.5	3.6	3.1	6.2	2.8	3.1	2.6	4.5	4.8	4.9	3.9	5.6	3.3	3.2	2.9	5.6	2.6	3.0	2.5
$T=5$	0.5	44.7	37.2	37.5	27.4	37.3	18.8	18.9	14.3	58.2	28.1	27.8	20.2	42.1	35.1	35.0	25.9	35.2	17.5	17.9	13.6	54.7	25.9	25.6	18.5
$T=5$	0.75	93.1	81.7	81.7	69.7	72.0	49.1	49.5	39.9	98.8	78.0	77.8	63.7	91.8	79.1	79.8	67.1	70.4	47.2	47.5	37.9	97.8	75.0	74.7	60.1
$T=5$	1	99.8	98.2	98.1	94.3	89.6	74.8	74.4	66.0	100.0	99.0	98.8	94.9	99.6	97.8	97.7	93.1	88.6	73.2	72.6	64.0	100.0	98.3	98.5	93.2

While power measures a scheme's ability to detect a change by the end of the monitoring window, the ARL captures the expected detection delay—i.e., the average number of observations until a change is flagged. In many real-time settings (operational risk, systems monitoring), minimizing delay is more critical than maximizing end-of-period power. For example, timely detection of shifts in inflation lets central banks adjust policy proactively; early identification of structural changes in asset volatility enables risk managers to adapt exposures; and rapid detection of deviations in pollution or temperature can mitigate health and ecological risks. We therefore assess ARL performance for all three schemes; see Figure 1. Because ARL is most relevant in dynamic environments—where changes are anticipated and prompt detection averts adverse outcomes—we focus on $T = 1$ in the simulations. We report results for Type (i) (abrupt changes) only; Type (ii) (smooth changes) exhibits the same qualitative pattern with uniformly longer ARLs, reflecting the greater difficulty of detecting gradual shifts.



(a) DGP1 (b) DGP2 (c) DGP3 (d) DGP4

(e) DGP1 (f) DGP2 (g) DGP3 (h) DGP4

Figure 1: ARL for detecting a mean shift using RSMS, SSMS, and CSMS. Results shown for $T = 1$ under the Type (i) (abrupt) alternative, with break locations $k^* = 50$ (top row) and $k^* = 200$ (bottom row). For each method, solid lines correspond to $\gamma = 0$ and dashed lines to $\gamma = 0.15$.

The simulation results exhibit several clear patterns. First, ARL decreases as Δ increases, indicating faster detection for larger shifts. For CSMS and RSMS across all DGPs, setting $\gamma = 0.15$ further reduces ARL, consistent with greater emphasis on early signals. By contrast, $\gamma = 0.15$ increases ARL for SSMS, which may explain why Chan et al. (2021) do not introduce this tuning parameter.

Second, among the three schemes, CSMS attains the shortest ARLs under DGP1, whereas RSMS delivers shorter ARLs for DGPs 2–4, suggesting superior detection efficiency in more challenging environments. SSMS consistently exhibits the longest ARLs, reflecting slower detection regardless of the DGP.

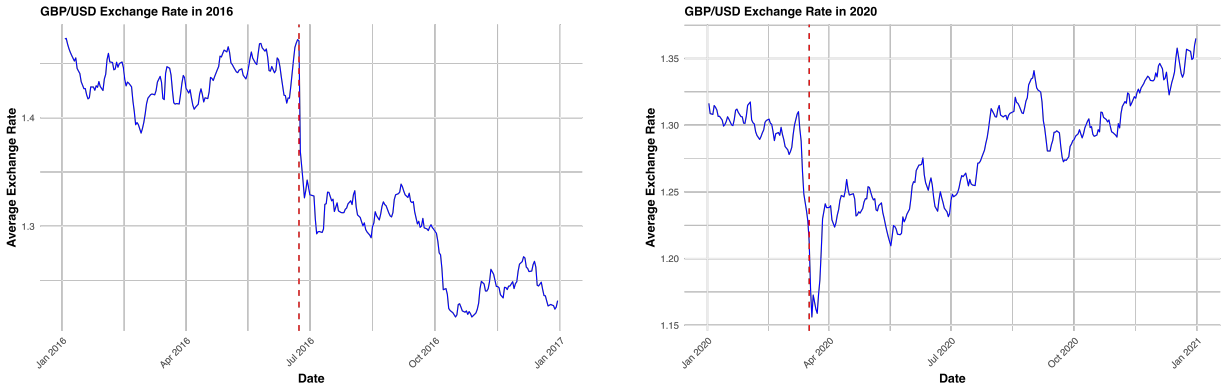
Overall, RSMS combines reliable size control with strong detection performance across a wide range of designs. This balance supports both the credibility of alarms and timely responsiveness, making RSMS especially suitable for high-stakes settings in which false alarms or delayed responses are costly.

5 Empirical Analysis

We examine the dynamics of the USD/GBP exchange rate around major events in 2016, 2020, and 2022, with particular focus on the 2016 EU referendum and the onset of COVID-19. Data are sourced from Refinitiv Eikon and consist of 5-minute intraday USD/GBP quotes for 2016. To improve numerical stability—given the small scale of exchange-rate movements in basis points (0.0001)—observations are rescaled by a factor of 100 prior to analysis. Intraday timestamps are standardized to the unit interval $[0, 1]$, and daily curves are obtained by smoothing with cubic B-spline bases. The number of basis functions is set to $\min\{21,\ 0.8 \times (\text{minimum number of intraday observations across all trading days})\}$. The final sample comprises 312 valid daily curves for both 2019 and 2020.

For subsequent inference, we employ functional principal component analysis (FPCA) to extract dominant modes of variation from the smoothed curves, enabling online detection of structural change in a functional time-series setting. This approach provides a natural bridge from functional data to time-series monitoring via FPCA scores; see, e.g., Sun et al. (2025) for an adjusted-range-based self-normalized KS test for functional time series. Principal component scores for the testing sample are computed by projecting each testing-day curve onto the eigenfunctions estimated from the training sample, ensuring alignment between the training and monitoring phases.

Figure 2a depicts the daily USD/GBP series in 2016 and shows a sharp, persistent depreciation immediately following the Brexit referendum, consistent with acute market stress and sustained uncertainty. Figure 2b presents the 2020 series, with a marked decline around 17 March 2020 coinciding with the initial rollout of social-distancing policies and heightened pandemic-driven uncertainty—highlighting market sensitivity to large macro shocks.



(a) 2016 Brexit referendum; dashed line marks 23 June 2016.

(b) 2020 COVID-19 shock; dashed line marks 17 March 2020.

Figure 2: Daily USD/GBP exchange rate in 2016 and 2020 (5-minute quotes aggregated to daily means). Dashed lines indicate event dates.

We apply RSMS, SSMS, and CSMS to the USD/GBP data for 2016 and 2020, varying the initial window size (m), monitoring horizon (T), and sensitivity parameter (γ).

For 2016, Table 5 reports, for configurations yielding rejections only, the test statistic, the rejection decision, and the calendar date of the first detected break. The (m, T) choices ensure the post-referendum period lies within the monitoring window while respecting the sample of 312 trading days. Overall, RSMS delivers earlier detection—especially for larger T—and, when H_0 is rejected, typically identifies the break before SSMS and often before CSMS. For example, under ($m = 50$, $T = 5$, $\gamma = 0.00$), RSMS flags a break on July 12, 2016, whereas CSMS and SSMS signal on July 15, 2016 and August 16, 2016, respectively. Likewise, for ($m = 75$, $T = 2$, $\gamma = 0.15$), RSMS detects on July 22, 2016, essentially in line with CSMS (July 21, 2016) and clearly ahead of SSMS (September 1, 2016).

Second, Table 6 summarizes RSMS, SSMS, and CSMS results for structural breaks in the USD/GBP rate during 2019. From January 1 to March 17 there are 55 trading days; we therefore set $m = 50$. At ($m = 50$, $T = 5$, $\gamma = 0.00$), RSMS signals a break on August 5, closely matching CSMS (August 4), while SSMS triggers later on October 10.

Setting $\gamma = 0.15$ does not improve early detection; it either delays detection or prevents rejection altogether. For example, RSMS does not signal for $T = 1$ under $\gamma = 0.15$. Notably, SSMS fails to detect any change when $m = 50$ and $T = 1$ for both γ values, and also fails under $m = 5$ with $\gamma = 0.15$. This pattern likely reflects the limitations of SSMS when the training window is short and may be contaminated by mild early-pandemic shifts, which inflate the self-normalizer and hinder detection.

In sum, RSMS exhibits consistently earlier detection across configurations, implying shorter ARLs when a true break occurs. These empirical results align with the simulation evidence and underscore the responsiveness and robustness of RSMS in dynamic settings (e.g., post-Brexit and COVID-19), making it a practical choice for timely anomaly detection.

Regarding the sensitivity parameter, $\gamma = 0.15$ does not improve detection overall. In the empirical analysis it advances detection only for CSMS, and in the simulations it benefits

Table 5: USD/GBP (2016): first detected break dates for RSMS (R), SSMS (S), and CSMS (C)

m	T	γ	Scheme	Statistic	Date of first detection
50	5	0.00	R	213.92	07-12-2016
50	5	0.00	S	459.22	08-16-2016
50	5	0.00	C	186.74	07-15-2016
50	5	0.15	R	225.94	07-26-2016
50	5	0.15	S	459.22	09-20-2016
50	5	0.15	C	197.23	08-01-2016
75	2	0.00	R	41.83	07-18-2016
75	2	0.00	S	132.23	08-10-2016
75	2	0.00	C	57.21	07-17-2016
75	2	0.15	R	47.24	07-22-2016
75	2	0.15	S	132.23	09-01-2016
75	2	0.15	C	64.61	07-21-2016
100	1	0.00	R	22.28	07-17-2016
100	1	0.00	S	133.54	07-27-2016
100	1	0.00	C	39.88	07-14-2016
100	1	0.15	R	27.43	07-18-2016
100	1	0.15	S	133.54	08-08-2016
100	1	0.15	C	49.10	07-15-2016
100	2	0.00	R	188.70	07-19-2016
100	2	0.00	S	1131.14	07-31-2016
100	2	0.00	C	337.83	07-15-2016
100	2	0.15	R	213.11	07-22-2016
100	2	0.15	S	1131.14	08-14-2016
100	2	0.15	C	381.53	07-19-2016

Notes: Entries are shown only for configurations that reject H_0. "Scheme" abbreviations: R = RSMS, S = SSMS, C = CSMS. Dates are in month–day–year format.

Table 6: USD/GBP (2019): first detected break dates for RSMS (R), SSMS (S), and CSMS (C)

m	T	γ	Scheme	Statistic	Date of first detection
50	1	0.00	R	2.38	04-04-2019
50	1	0.00	C	3.09	04-02-2019
50	1	0.15	C	3.92	04-15-2019
50	5	0.00	R	21.48	08-05-2019
50	5	0.00	S	56.59	10-10-2019
50	5	0.00	C	27.88	08-04-2019
50	5	0.15	R	23.00	08-27-2019
50	5	0.15	C	29.86	08-23-2019

Notes: Entries shown only for configurations that reject H_0. Scheme abbreviations: R = RSMS, S = SSMS, C = CSMS. Dates are month–day–year.

RSMS and CSMS but has negligible or slightly adverse effects for SSMS. Thus, while γ is intended to tune sensitivity, its effectiveness is scheme- and context-dependent.

6 Conclusion

In this paper, we develop RSMS—an online change–point detector built on the CUSUM framework with adjusted–range self–normalization (Hong, Linton, McCabe, Sun & Wang 2024, Hong et al. 2025). RSMS avoids LRV estimation and tuning choices (e.g., HAC bandwidths/kernels or block lengths). It achieves correct size under the null and consistency under the alternative. Relative to KS–type self–normalization (Shao 2010), the adjusted range yields a tuning–free implementation that is more robust to mild contamination in the training window. We also derive finite– and open–horizon asymptotics with ready–to–use critical values $c_R(\alpha, d, T)$ and provide practical guidance on a simple early–detection weight

γ, clarifying when modest $\gamma > 0$ improves timeliness without reintroducing LRV estimation.

Simulations across diverse DGPs and break types in both the main text and the appendices show that RSMS outperforms the self–normalized scheme of Chan et al. (2021) (based on Shao (2010)) and the standard CUSUM–based approach (Horváth et al. 2022), delivering higher power and shorter ARLs. Notably, Appendix A.5 demonstrates that RSMS remains power-stable and continues to reduce ARLs even when the training sample is contaminated—a situation that is difficult to avoid in practice because type II errors and rolling re-estimation naturally push recent (potentially shifted) observations into the training window.

An empirical application to functional USD/GBP series around the 2016 EU referendum and the COVID–19 period confirms timely alarms for major breaks, with RSMS often detecting earlier than SSMS and, in some cases, achieving the shortest ARL.

Future work includes extending RSMS to additional domains (e.g., climate or social–media analytics) and exploring integrations with deep learning to improve adaptability and real–time efficiency (Hong, Linton, Sun & Zhu 2024).

References

Al-Osh, M. A. & Alzaid, A. A. (1987), 'First-order integer-valued autoregressive (INAR (1)) process', *Journal of Time Series Analysis* **8**(3), 261–275.

Andrews, D. W. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica: Journal of the Econometric Society* pp. 817–858.

Andrews, D. W. (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica* **61**(4), 821–856.

Aue, A. & Horváth, L. (2013), 'Structural breaks in time series', *Journal of Time Series Analysis* **34**(1), 1–16.

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

Bai, J. & Perron, P. (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* **66**(1), 47–78.

Bai, J. & Perron, P. (2003), 'Computation and analysis of multiple structural change models', *Journal of applied econometrics* **18**(1), 1–22.

Berkes, I., Gombay, E., Horváth, L. & Kokoszka, P. (2004), 'Sequential change-point detection in GARCH (p, q) models', *Econometric Theory* **20**(6), 1140–1167.

Billingsley, P. (1968), *Convergence of Probability Measures*, 1st edn, Wiley, New York.

Brown, R. L., Durbin, J. & Evans, J. M. (1975), 'Techniques for testing the constancy of regression relationships over time', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **37**(2), 149–163.

Chan, N. H., Ng, W. L. & Yau, C. Y. (2021), 'A self-normalized approach to sequential change-point detection for time series', *Statistica Sinica* **31**(1), 491–517.

Chen, B. & Hong, Y. (2012), 'Testing for smooth structural changes in time series models via nonparametric regression', *Econometrica* **80**(3), 1157–1183.

Chen, B. & Hong, Y. (2016), 'Detecting for smooth structural changes in GARCH models', *Econometric Theory* **32**(3), 740–791.

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

Csörgő, M. & Horváth, L. (1997), *Limit Theorems in Change-Point Analysis*, John Wiley & Sons, Chichester.

Dette, H. & Gösmann, J. (2020), 'A likelihood ratio approach to sequential change point

detection for a general class of parameters', *Journal of the American Statistical Association* **115**(531), 1361–1377.

Durbin, J. (1969), 'Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals', *Biometrika* **56**(1), 1–15.

Gombay, E. & Horváth, L. (2009), 'Sequential tests and change detection in the covariance structure of weakly stationary time series', *Communications in Statistics - Theory and Methods* **38**(16-17), 2872–2883.

Gombay, E. & Serban, D. (2009), 'Monitoring parameter change in AR(p) time series models', *Journal of Multivariate Analysis* **100**(4), 715–725.

Gut, A. & Steinebach, J. (2002), 'Truncated sequential change-point detection based on renewal counting processes', *Scandinavian Journal of Statistics* **29**(4), 693–719.

Homm, U. & Breitung, J. (2012), 'Testing for speculative bubbles in stock markets: a comparison of alternative methods', *Journal of Financial Econometrics* **10**(1), 198–231.

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

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

Hong, Y., Linton, O., Sun, J. & Zhu, M. (2024), 'Yongmiao Hong, Oliver Linton, Jiajing Sun, and Meiting Zhu's contribution to the discussion of 'the discussion meeting on probabilistic and statistical aspects of machine learning", *Journal of the Royal Statistical Society Series B: Statistical Methodology* **86**(2), 320–321.

Horváth, L., Horváth, Z. & Hušková, M. (2008), Ratio tests for change point detection, *in* N. Balakrishnan, E. A. Peña & M. J. Silvapulle, eds, 'Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen', Vol. 1 of *Institute of Mathematical Statistics Collections*, Institute of Mathematical Statistics, pp. 293–304.

Horváth, L., Hušková, M., Kokoszka, P. & Steinebach, J. (2004), 'Monitoring changes in linear models', *Journal of Statistical Planning and Inference* **126**(1), 225–251.

Horvath, L., Kokoszka, P. & Steinebach, J. (2007), 'On sequential detection of parameter changes in linear regression', *Statistics & Probability Letters* **77**(9), 885–895.

Horváth, L., Liu, Z. & Lu, S. (2022), 'Sequential monitoring of changes in dynamic linear models, applied to the us housing market', *Econometric Theory* **38**(2), 209–272.

Horváth, L. & Rice, G. (2024), *Change Point Analysis for Time Series*, Springer.

Kiefer, N. M. & Vogelsang, T. J. (2002), 'Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size', *Econometric Theory* **18**(6), 1350–1366.

Kiefer, N. M. & Vogelsang, T. J. (2005), 'A new asymptotic theory for heteroskedasticity-autocorrelation robust tests', *Econometric Theory* **21**(6), 1130–1164.

Kirch, C. & Stoehr, C. (2022), 'Asymptotic delay times of sequential tests based on u-statistics for early and late change points', *Journal of Statistical Planning and Inference* **221**, 114–135.

Kunsch, H. (1984), 'Infinitesimal robustness for autoregressive processes', *The Annals of Statistics* pp. 843–863.

URL: *http://www.jstor.org/stable/2240964*

Kunsch, H. R. (1989), 'The jackknife and the bootstrap for general stationary observations', *The Annals of Statistics* **17**(3), 1217–1241.

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

Müller, U. K. (2007), 'A theory of robust long-run variance estimation', *Journal of Econometrics* **141**(2), 1331–1352.

Newey, W. K. & West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

Newey, W. K. & West, K. D. (1994), 'Automatic lag selection in covariance matrix estimation', *The review of economic studies* **61**(4), 631–653.

Perron, P. (2006), Dealing with structural breaks, *in* T. C. Mills & K. Patterson, eds, 'Palgrave Handbook of Econometrics: Volume 1, Econometric Theory', Palgrave Macmillan, Basingstoke, pp. 278–352.

Perron, P. & Yamamoto, Y. (2015), 'Using OLS to estimate and test for structural changes in models with endogenous regressors', *Journal of Applied Econometrics* **30**(1), 119–144.

Phillips, P. C. (1987), 'Time series regression with a unit root', *Econometrica* pp. 277–301.

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

Sun, J., Hong, Y., Lin, Z. & Xu, W. (2025), 'Structural stability of functional data-a new adjusted-range-based self-normalization approach', *Economics Letters* p. 112350.

Wang, R., Zhu, C., Volgushev, S. & Shao, X. (2022), 'Inference for change points in high-dimensional data via selfnormalization', *The Annals of Statistics* **50**(2), 781–806.

Whitt, W. (2002), *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*, Springer Series in Operations Research, Springer, New York.

Wu, W. B. (2005), 'Nonlinear system theory: Another look at dependence', *Proceedings of the National Academy of Sciences* **102**(40), 14150–14154.

Zeileis, A., Leisch, F., Kleiber, C. & Hornik, K. (2005), 'Monitoring structural change in dynamic econometric models', *Journal of Applied Econometrics* **20**(1), 99–121.

A.1 The CSMS and SSMS

We first detail the CSMS, which standardizes the statistic using a LRV estimator, following Horváth et al. (2022). In this paper, bandwidth selection for the LRV is implemented via Silverman's rule of thumb, which performs better in our applications than the default settings in Horváth et al. (2022). More broadly, alternative tuning choices (bandwidths, kernels) can yield materially different outcomes, underscoring the drawbacks of LRV-based normalization.

The CSMS statistic is

$$\mathbb{M}_m^C(k) = \frac{S_m\left(k, \widehat{\boldsymbol{\theta}}_m\right)' \boldsymbol{\Gamma}_m\left(\widehat{\boldsymbol{\theta}}_m\right)^{-1} S_m\left(k, \widehat{\boldsymbol{\theta}}_m\right)}{m\left(1+\frac{k}{m}\right)^2\left(\frac{k}{k+m}\right)^{2\gamma}},$$

where

$$\boldsymbol{\Gamma}_m(\widehat{\theta}_m) = \sum_{l=-(m-1)}^{m-1} K\left(\frac{\ell}{h}\right)\widehat{\kappa}_{\ell,m},$$

$$\widehat{\kappa}_{\ell,m} = \begin{cases} \frac{1}{m-\ell}\sum_{j=1}^{m-l}\left(\sum_{j=1}^{t}\psi\left(X_j,\widehat{\theta}_m\right)\right)\left(\sum_{j=1}^{t}\psi\left(X_j,\widehat{\theta}_m\right)\right)', & \text{if } 0 \le \ell < m, \\ \frac{1}{M-|\ell|}\sum_{j=-(\ell-1)}^{M}\left(\sum_{j=1}^{t}\psi\left(X_j,\widehat{\theta}_m\right)\right)\left(\sum_{j=1}^{t}\psi\left(X_j,\widehat{\theta}_m\right)\right)', & \text{if } -m < \ell < 0. \end{cases}$$

The stopping time is defined by

$$\mathbb{T}_m^C = \begin{cases} \min\left\{k : \mathbb{M}_m^C(k) > c_C,\ 1 \le k \le mT\right\}, & \text{if such } k \text{ exists;} \\ mT+1, & \text{if } \mathbb{M}_m^C(k) \le c_C \text{ for all } 1 \le k \le mT. \end{cases}$$

Table A.1 reports simulated asymptotic decision boundaries; Monte Carlo settings match those used for Table 1.

Table A.1: Simulated asymptotic critical values for the CSMS under various horizons ($T = 1, 2, 5, 10$) and sensitivity parameters ($\gamma = 0$ or 0.15).

	$\gamma = 0$								$\gamma = 0.15$							
	$T = 1$		$T = 2$		$T = 5$		$T = 10$		$T = 1$		$T = 2$		$T = 5$		$T = 10$	
$d\backslash\alpha$	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	2.5	1.9	3.2	2.5	4.1	3.2	4.5	3.5	3.3	2.5	4.0	3.1	4.6	3.5	4.9	3.8
2	3.9	3.1	6.2	4.9	12.6	10.1	23.5	18.9	5.4	4.4	8.6	6.9	17.9	14.3	33.6	26.7
3	4.7	3.9	7.6	6.3	16.0	13.1	29.4	24.3	6.5	5.5	10.5	8.6	22.2	18.2	41.1	33.4
4	5.6	4.8	9.1	7.6	18.9	15.9	34.9	29.2	7.7	6.5	12.5	10.6	25.6	21.6	48.8	40.4
5	6.6	5.6	10.3	8.8	21.2	18.2	40.0	33.9	8.9	7.5	14.1	11.9	29.3	24.8	55.6	46.7

Second, the SSMS replaces the LRV estimator with the self-normalizer

$$\mathbf{D}_m\left(\widehat{\theta}_m\right) = \frac{1}{m^2}\sum_{t=1}^{m}\left\{\left(\sum_{j=1}^{t}\psi\left(X_j, \widehat{\theta}_m\right)\right)\left(\sum_{j=1}^{t}\psi\left(X_j, \widehat{\theta}_m\right)\right)'\right\} \tag{A.1}$$

and defines the statistic

$$\mathbb{M}_m^S(k) = \frac{S_m\left(k, \widehat{\boldsymbol{\theta}}_m\right)' \mathbf{D}_m\left(\widehat{\boldsymbol{\theta}}_m\right)^{-1} S_m\left(k, \widehat{\boldsymbol{\theta}}_m\right)}{m\left(1+\frac{k}{m}\right)^2\left(\frac{k}{k+m}\right)^{2\gamma}}. \tag{A.2}$$

The stopping time is

$$\mathbb{T}_m^S = \begin{cases} \min\left\{k : \mathbb{M}_m^S(k) > c_S,\ 1 \le k \le mT\right\}, & \text{if such } k \text{ exists;} \\ mT+1, & \text{if } \mathbb{M}_m^S(k) \le c_S \text{ for all } 1 \le k \le mT. \end{cases}$$

Table A.2 reports simulated asymptotic decision boundaries; Monte Carlo settings match those used in Tables 1 and A.1.

Since the CSMS and SSMS stopping rules mirror the structure of Algorithm 1, we omit them for brevity.

Table A.2: Simulated asymptotic critical values for the SSMS under monitoring horizons $T \in \{1, 2, 5, 10\}$ and sensitivity parameters $\gamma \in \{0, 0.15\}$.

	$\gamma = 0$								$\gamma = 0.15$							
	$T = 1$		$T = 2$		$T = 5$		$T = 10$		$T = 1$		$T = 2$		$T = 5$		$T = 10$	
$d\backslash\alpha$	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	34	22.8	44.1	30	55.5	38.1	59	40.8	43.6	30.7	53.1	36	60.9	42.6	67.8	47.4
2	68.7	50.4	93.3	68.2	113.4	84.3	126.8	93.1	89.7	67.5	111	83.5	128.7	96.8	140.5	103.4
3	108.4	83.7	153.4	115	184.7	140.5	206.1	155.9	143.9	112.3	185.9	140.5	204.8	158.2	219.6	171.1
4	165	127.9	219.2	171.3	265.5	210.4	293.8	229.3	217.9	171.7	259.3	207.9	297.3	234.7	319.5	252.3
5	218.5	175	284.4	226.9	363.4	287.1	394.3	319.1	290	230.4	345.9	273.6	404.5	317.4	425.7	336.7

A.2 Proof of the Main Results

A.2.1 Proof of Theorem 1(a)

For brevity, write $\psi_j(\theta) = \psi(X_j, \theta)$ and $\psi_j^*(\theta) = \psi(X_j^*, \theta)$. Let $\|\cdot\|$ denote the max norm, i.e., for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $\|x\| = \max_{i \in \{1,\ldots,d\}} |x_i|$.

Under the conditions of Theorem 1 and by the continuous mapping theorem, it suffices to show

$$\sup_{1 \le s \le m} \left\| \frac{\sum_{j=1}^{s} \psi_j\left(\widehat{\theta}_m\right) - \left(\sum_{j=1}^{s} \psi_j\left(\theta_0\right) - \frac{s}{m} \sum_{j=1}^{m} \psi_j\left(\theta_0\right)\right)}{m^{1/2}} \right\| \xrightarrow{p} 0. \tag{A.3}$$

Consider

$$\sum_{j=1}^{s} \psi_j\left(\widehat{\theta}_m\right) = \left(\sum_{j=1}^{s} \psi_{j1}\left(\widehat{\theta}_m\right), \sum_{j=1}^{s} \psi_{j2}\left(\widehat{\theta}_m\right), \ldots, \sum_{j=1}^{s} \psi_{jd}\left(\widehat{\theta}_m\right) \right),$$

and apply the mean value theorem coordinate-wise. For each $i = 1, \ldots, d$ and all $s = 1, \ldots, m$,

$$\left| \frac{\sum_{j=1}^{s} \psi_{ji}\left(\widehat{\theta}_m\right) - \left[\sum_{j=1}^{s} \psi_{ji}\left(\theta_0\right) + \sum_{j=1}^{s} \psi'_{ji}\left(\theta^*_{msi}\right) \left(\widehat{\theta}_m - \theta_0\right) \right]}{m^{1/2}} \right| = 0, \tag{A.4}$$

where θ^*_{msi} lies on the line segment between $\widehat{\theta}_m$ and θ_0.

By Theorem 1 and a uniform law of large numbers (ULLN), for all $i = 1, \ldots, d$,

$$\begin{aligned}
&\sup_{1\le s\le m}\left|\frac{\left[\sum_{j=1}^{s}\psi'_{ji}(\theta^*_{msi}) - s\,\mathrm{E}\{\psi'_{ji}(\theta_0)\}\right]\left(\widehat{\theta}_m-\theta_0\right)}{m^{1/2}}\right| \\
&\le \sup_{1\le s\le m}\left|\left[\frac{\sum_{j=1}^{s}\psi'_{ji}(\theta^*_{msi}) - s\,\mathrm{E}\{\psi'_{ji}(\theta_0)\}}{m}\right]\sqrt{m}\left(\widehat{\theta}_m-\theta_0\right)\right| \\
&= \left|\sqrt{m}\left(\widehat{\theta}_m-\theta_0\right)\right| \sup_{1\le s\le m}\left|\frac{\sum_{j=1}^{s}\psi'_{ji}(\theta^*_{msi}) - s\,\mathrm{E}\{\psi'_{ji}(\theta_0)\}}{m}\right| \\
&= |O_p(1)| \sup_{1\le s\le m}\left|\frac{\sum_{j=1}^{s}\psi'_{ji}(\theta^*_{msi}) - s\,\mathrm{E}\{\psi'_{ji}(\theta_0)\}}{m}\right| \xrightarrow{p} 0.
\end{aligned} \tag{A.5}$$

Applying the mean value theorem coordinate-wise for $i = 1, \ldots, d$ with $s = m$,

$$\left|\frac{\sum_{j=1}^{m}\psi_{ji}(\widehat{\theta}_m) - \left[\sum_{j=1}^{m}\psi_{ji}(\theta_0) + \sum_{j=1}^{m}\psi'_{ji}(\theta^*_{mmi})(\widehat{\theta}_m-\theta_0)\right]}{m}\right| = 0,$$

where $\theta^*_{mmi} \in (\widehat{\theta}_m, \theta_0)$, and $\sum_{j=1}^{m}\psi_{ji}(\widehat{\theta}_m) = 0$ by definition. By the ULLN and positive definiteness of $\mathrm{E}\{\psi'_j(\theta_0)\}$, solving the linear system yields

$$\mathrm{E}\{\psi'_j(\theta_0)\}\,(1+o_p(1))(\widehat{\theta}_m-\theta_0) = -\frac{1}{m}\sum_{j=1}^{m}\psi_j(\theta_0).$$

Hence,

$$\widehat{\theta}_m-\theta_0 = \left[-\frac{1}{m}\,\mathrm{E}\left\{\psi'_j(\theta_0)\right\}^{-1}\sum_{j=1}^{m}\psi_j(\theta_0)\right](1+o_p(1)). \tag{A.6}$$

Combining (A.4), (A.5), and (A.6), we obtain

$$\sup_{1\le s\le m}\left\|\frac{\sum_{j=1}^{s}\psi_j\left(\widehat{\theta}_m\right) - \left(\sum_{j=1}^{s}\psi_j(\theta_0) - \frac{s}{m}\sum_{j=1}^{m}\psi_j(\theta_0)\right) + o_p(1)\left(\frac{s}{m}\sum_{j=1}^{m}\psi_j(\theta_0)\right)}{m^{1/2}}\right\| \xrightarrow{p} 0. \tag{A.7}$$

Since

$$\sup_{1\le s\le m}\left\|\frac{\frac{s}{m}\sum_{j=1}^{m}\psi_j(\theta_0)}{m^{1/2}}\right\| = \sup_{1\le s\le m}\left|\frac{s}{m}\right|\left\|\frac{\sum_{j=1}^{m}\psi_j(\theta_0)}{m^{1/2}}\right\| = O_p(1),$$

(A.7) implies (A.3).

Under Assumptions 1–2, for any $r \in (0, 1]$,

$$\frac{\sum_{j=1}^{\lfloor mr\rfloor}\psi_j(\widehat{\theta}_m)}{\sqrt{m}} \Longrightarrow \mathbf{\Sigma}_\psi^{1/2}\Big(\mathbb{B}_d(r) - r\,\mathbb{B}_d(1)\Big),$$

and by the continuous mapping theorem,

$$\max_{1\le\lfloor mr\rfloor\le m}\frac{\sum_{j=1}^{\lfloor mr\rfloor}\psi_j(\widehat{\theta}_m)}{\sqrt{m}}-\min_{1\le\lfloor mr\rfloor\le m}\frac{\sum_{j=1}^{\lfloor mr\rfloor}\psi_j(\widehat{\theta}_m)}{\sqrt{m}}$$
$$\xrightarrow{\mathcal{D}(0,1]}\boldsymbol{\Sigma}_\psi^{1/2}\left[\sup_{r\in(0,1]}\Big(\mathbb{B}_d(r)-r\,\mathbb{B}_d(1)\Big)-\inf_{r\in(0,1]}\Big(\mathbb{B}_d(r)-r\,\mathbb{B}_d(1)\Big)\right].$$

Hence,

$$\mathbf{R}_m^2(\widehat{\theta}_m)\Longrightarrow\boldsymbol{\Sigma}_\psi^{\frac{1}{2}}\mathbf{R}(\boldsymbol{\Sigma}_\psi^{\frac{1}{2}})'\quad\text{as}\quad m\to\infty.$$

□

A.2.2 Proof of Theorem 1(b)

For $T<\infty$, proceeding as in the proof of Theorem 1(a) and applying the mean value theorem coordinate-wise (for each $i=1,\ldots,d$ and $k=1,\ldots,mT$),

$$\left|\frac{\sum_{j=m+1}^{m+k}\psi_{ji}(\widehat{\theta}_m)-\left[\sum_{j=m+1}^{m+k}\psi_{ji}(\theta_0)+\sum_{j=m+1}^{m+k}\psi'_{ji}(\theta^*_{mki})(\widehat{\theta}_m-\theta_0)\right]}{m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^\gamma}\right|=0, \tag{A.8}$$

where $\theta^*_{mki}\in(\widehat{\theta}_m,\theta_0)$.

By the conditions of Theorem 1 and the ULLN, for all $i=1,\ldots,d$,

$$\begin{aligned}
&\sup_{1\le k\le mT}\left|\frac{\left[\sum_{j=m+1}^{m+k}\psi'_{ji}(\theta^*_{mki})-k\,\mathrm{E}\{\psi'_{ji}(\theta_0)\}\right](\widehat{\theta}_m-\theta_0)}{m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^\gamma}\right|\\
&\le\sup_{1\le k\le mT}\left|\left[\frac{\sum_{j=m+1}^{m+k}\psi'_{ji}(\theta^*_{mki})-k\,\mathrm{E}\{\psi'_{ji}(\theta_0)\}}{(m+k)\left(\frac{k}{k+m}\right)^\gamma}\right]\sqrt{m}(\widehat{\theta}_m-\theta_0)\right|\\
&=\left|\sqrt{m}(\widehat{\theta}_m-\theta_0)\right|\sup_{1\le k\le mT}\left|\frac{\sum_{j=m+1}^{m+k}\psi'_{ji}(\theta^*_{mki})-k\,\mathrm{E}\{\psi'_{ji}(\theta_0)\}}{(m+k)\left(\frac{k}{k+m}\right)^\gamma}\right|\\
&=|O_p(1)|\sup_{1\le k\le mT}\left|\frac{\sum_{j=m+1}^{m+k}\psi'_{ji}(\theta^*_{mki})-k\,\mathrm{E}\{\psi'_{ji}(\theta_0)\}}{(m+k)\left(\frac{k}{k+m}\right)^\gamma}\right|\xrightarrow{p}0.
\end{aligned} \tag{A.9}$$

Combining (A.6), (A.8), and (A.9), it follows that

$$\sup_{1\le k\le mT}\left\|\frac{\sum_{j=m+1}^{m+k}\psi_j(\widehat{\theta}_m)-\left(\sum_{j=m+1}^{m+k}\psi_j(\theta_0)-\frac{k}{m}\sum_{j=1}^{m}L_j(\theta_0)\right)+o_p(1)\left(\frac{k}{m}\sum_{j=1}^{m}\psi_j(\theta_0)\right)}{m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^\gamma}\right\|\xrightarrow{p}0.$$

Since

$$\sup_{1\le k\le mT}\left\|\frac{\frac{k}{m}\sum_{j=1}^{m}\psi_j(\theta_0)}{m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}}\right\| = \sup_{1\le k\le mT}\left|\frac{\frac{k}{m}}{1+\frac{k}{m}}\right|^{-\gamma}\left\|\frac{\sum_{j=1}^{m}\psi_j(\theta_0)}{m^{1/2}}\right\| = O_p(1),$$

it follows that

$$\sup_{1\le k\le mT}\left\|\frac{\sum_{j=m+1}^{m+k}\psi_j\left(\widehat{\theta}_m\right)-\left(\sum_{j=m+1}^{m+k}\psi_j(\theta_0)-\frac{k}{m}\sum_{j=1}^{m}L_j(\theta_0)\right)}{m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}}\right\| \xrightarrow{p} 0. \tag{A.10}$$

To prove Theorem 1(b), by (A.10) it suffices to establish the weak limit

$$\sup_{1\le k<\infty}\frac{C_m(k,\widehat{\theta}_m)'\,\mathbf{R}_m(\widehat{\theta}_m)^{-2}\,C_m(k,\widehat{\theta}_m)}{m\left(1+\frac{k}{m}\right)^2\left(\frac{k}{k+m}\right)^{2\gamma}} \Longrightarrow \sup_{0<u\le 1}\frac{\mathbb{B}^{\dagger}(u)'\,\mathbf{R}^{-1}\,\mathbb{B}^{\dagger}(u)}{u^{2\gamma}},$$

where $C_m(k,\widehat{\theta}_m)=\sum_{j=m+1}^{m+k}\psi_j(\widehat{\theta}_m)-\frac{k}{m}\sum_{j=1}^{m}\psi_j(\widehat{\theta}_m)$ and $\mathbb{B}^{\dagger}(\cdot)$ is a standard Brownian motion independent of $\mathbf{R}$.

Following the structure of Lemma A.5 in Horváth et al. (2022), for some $\delta>0$ define two independent Wiener processes $\{\mathbb{B}_{m,1}(k):0\le k\le m\}$ and $\{\mathbb{B}_{m,2}(k):0\le k<\infty\}$. Then

$$\begin{aligned}&\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|\frac{k}{m}\left(\sum_{j=1}^{m}\psi_j(\theta_0)-\mathbf{\Sigma}_{\psi}^{1/2}\mathbb{B}_{m,1}(m)\right)\right|\\ &= O_p(m^{1/2-\delta})\sup_{1\le k<\infty}\frac{k/m}{g(m,k)} = O_p(m^{-\delta})\sup_{0<x<\infty}\frac{x}{(1+x)\left(\frac{x}{1+x}\right)^{\gamma}} = o_p(1),\end{aligned}$$

where $g(m,k)=m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}$ and $x=k/m$.

Similarly,

$$\begin{aligned}&\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|\sum_{j=m+1}^{m+k}\psi_j(\theta_0)-\mathbf{\Sigma}_{\psi}^{1/2}\mathbb{B}_{m,2}(k)\right|\\ &= O_p(1)\sup_{1\le k<\infty}\frac{k^{\frac{1}{2}-\delta}}{g(m,k)} = O_p(m^{-\delta})\sup_{0<x<\infty}\frac{x^{1/2-\delta}}{(1+x)\left(\frac{x}{1+x}\right)^{\gamma}} = o_p(1).\end{aligned}$$

Hence,

$$\begin{aligned}
&\sup_{1\le k<\infty} \frac{\left(\boldsymbol{\Sigma}_\psi^{1/2}\mathbb{B}_{m,2}(k) - \frac{k}{m}\boldsymbol{\Sigma}_\psi^{1/2}\mathbb{B}_{m,1}(m)\right)' \left[\boldsymbol{\Sigma}_\psi^{1/2}\mathbf{R}(\boldsymbol{\Sigma}_\psi^{1/2})'\right]^{-1} \left(\boldsymbol{\Sigma}_\psi^{1/2}\mathbb{B}_{m,2}(k) - \frac{k}{m}\boldsymbol{\Sigma}_\psi^{1/2}\mathbb{B}_{m,1}(m)\right)}{m\left(1+\frac{k}{m}\right)^2\left(\frac{k}{k+m}\right)^{2\gamma}} \\
= &\sup_{1\le k<\infty} \frac{\left(\mathbb{B}_{m,2}(k) - \frac{k}{m}\mathbb{B}_{m,1}(m)\right)'\mathbf{R}^{-1}\left(\mathbb{B}_{m,2}(k) - \frac{k}{m}\mathbb{B}_{m,1}(m)\right)}{m\left(1+\frac{k}{m}\right)^2\left(\frac{k}{k+m}\right)^{2\gamma}} \\
= &\sup_{1/m\le x<\infty} \frac{\left(\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)\right)'\mathbf{R}^{-1}\left(\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)\right)}{(1+x)^2\left(\frac{x}{x+1}\right)^{2\gamma}} \\
\xrightarrow{a.s.} &\sup_{0<x<\infty} \frac{\left(\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)\right)'\mathbf{R}^{-1}\left(\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)\right)}{(1+x)^2\left(\frac{x}{x+1}\right)^{2\gamma}},
\end{aligned}$$

where $\mathbb{B}_1$ and $\mathbb{B}_2$ are independent Brownian motions. As shown by Chu et al. (1996) (see also Horváth et al. (2004)),

$$\{\mathbb{B}_2(x) - x\,\mathbb{B}_1(1) : 0 \le x < \infty\} \overset{\mathcal{D}}{\Longrightarrow} \left\{(1+x)\,\mathbb{B}^\dagger\left(\tfrac{x}{1+x}\right) : 0 \le x < \infty\right\}.$$

Therefore,

$$\sup_{0<x<\infty} \frac{\left(\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)\right)'\mathbf{R}^{-1}\left(\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)\right)}{(1+x)^2\left(\frac{x}{x+1}\right)^{2\gamma}} \overset{\mathcal{D}}{\Longrightarrow} \sup_{0<u\le 1} \frac{\mathbb{B}^\dagger(u)'\mathbf{R}^{-1}\mathbb{B}^\dagger(u)}{u^{2\gamma}}.$$

In summary, by (A.10),

$$\sup_{1\le k<\infty} \frac{S_m(k,\widehat{\theta}_m)'\,\mathbf{R}_m(\widehat{\theta}_m)^{-2}\,S_m(k,\widehat{\theta}_m)}{m\left(1+\frac{k}{m}\right)^2\left(\frac{k}{k+m}\right)^{2\gamma}} \overset{\mathcal{D}}{\Longrightarrow} \sup_{0<u\le 1} \frac{\mathbb{B}^\dagger(u)'\,\mathbf{R}^{-1}\,\mathbb{B}^\dagger(u)}{u^{2\gamma}},$$

where $S_m(k,\widehat{\theta}_m) = \sum_{j=m+1}^{m+k}\psi_j(\widehat{\theta}_m)$. □

A.2.3 Proof of Theorem 1(c)

For $T=\infty$, by arguments analogous to those above,

$$\sup_{1\le k<\infty}\left\|\frac{\sum_{j=m+1}^{m+k}\psi_j\left(\widehat{\theta}_m\right) - \left(\sum_{j=m+1}^{m+k}\psi_j(\theta_0) - \frac{k}{m}\sum_{j=1}^{m}\psi_j(\theta_0)\right)}{m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}}\right\| \xrightarrow{p} 0.$$

Thus, it suffices to show that

$$\sup_{1\le k<\infty}\left\|\frac{\sum_{j=m+1}^{m+k}\psi_j(\theta_0) - \frac{k}{m}\sum_{j=1}^{m}\psi_j(\theta_0)}{m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}}\right\| \overset{\mathcal{D}}{\Longrightarrow} \sup_{0<x<\infty}\left\|\frac{\boldsymbol{\Sigma}_\psi^{1/2}[\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)]}{(1+x)\left(\frac{x}{x+1}\right)^{\gamma}}\right\|.$$

It is therefore enough to verify

$$\sup_{mT\le k<\infty}\left\|\frac{\sum_{j=m+1}^{m+k}\psi_j(\theta_0)}{m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}}\right\|\xrightarrow{p}0 \tag{A.11}$$

and

$$\sup_{T\le x<\infty}\left\|\frac{\boldsymbol{\Sigma}_{\psi}^{1/2}\left[\mathbb{B}_2(x)-x\,\mathbb{B}_1(1)\right]}{(1+x)\left(\frac{x}{x+1}\right)^{\gamma}}\right\|\xrightarrow{p}0. \tag{A.12}$$

Hence, it suffices to verify

$$\sup_{mT\le k<\infty}\left\|\frac{\sum_{j=m+1}^{m+k}\psi_j(\theta_0)}{m^{\frac{1}{2}}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}}\right\|\xrightarrow{p}0 \tag{A.13}$$

and

$$\sup_{T\le x<\infty}\left\|\frac{\boldsymbol{\Sigma}_{\psi}^{\frac{1}{2}}\left[\mathbb{B}_2(x)-x\,\mathbb{B}_1(1)\right]}{(1+x)\left(\frac{x}{x+1}\right)^{\gamma}}\right\|\xrightarrow{p}0. \tag{A.14}$$

For (A.13), under the additional ρ-mixing conditions, a ρ-mixing Hájek–Rényi inequality applies (see Theorem 1(c) of Chan et al. (2021)). Thus, for some constant c^*, each $i=1,2,\ldots,d$, and any $\epsilon>0$,

$$\begin{aligned}P\left(\max_{mT\le k\le n}\left|\frac{\sum_{j=m+1}^{m+k}\psi_{ji}(\theta_0)}{m^{\frac{1}{2}}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}}\right|\ge\epsilon\right)\le\frac{c^*}{\epsilon^2}&\left(\sum_{j=m+1}^{m+mT}\frac{\mathbb{V}\mathrm{ar}(\psi_{ji}(\theta_0))}{\left[m^{\frac{1}{2}}(1+T)^{1-\gamma}T^{\gamma}\right]^2}\right)\\&+4\sum_{j=m+mT+1}^{n}\frac{\mathbb{V}\mathrm{ar}(\psi_{ji}(\theta_0))}{\left[m^{\frac{1}{2}}\left(1+\frac{j}{m}\right)\left(\frac{j}{j+m}\right)^{\gamma}\right]^2},\end{aligned} \tag{A.15}$$

where $\gamma\in[0,1/2)$.

Next, taking $\lim_{T\to\infty}\limsup_{m\to\infty}\lim_{n\to\infty}$ on both sides of (A.15), for any $\epsilon>0$ we obtain

$$\lim_{T\to\infty}\limsup_{m\to\infty}\lim_{n\to\infty}P\left(\sup_{mT\le k\le n}\left\|\frac{\sum_{j=m+1}^{m+k}\psi_j(\theta_0)}{m^{\frac{1}{2}}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}}\right\|\ge\epsilon\right)=0,$$

which yields (A.13).

For (A.14), the law of the iterated logarithm implies

$$\sup_{T\le x<\infty}\left\|\frac{\boldsymbol{\Sigma}_{\psi}^{\frac{1}{2}}\left[\mathbb{B}_2(x)-x\,\mathbb{B}_1(1)\right]}{(1+x)\left(\frac{x}{x+1}\right)^{\gamma}}\right\|\xrightarrow[T\to\infty]{\text{a.s.}}0,$$

and hence (A.14) holds. Consequently,

$$\begin{aligned}
&\lim_{m\to\infty} P\Big(\mathbb{T}_m^R < \infty \mid H_0\Big) \\
&= P\left(\sup_{0<x<\infty} \frac{[\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)]'\,\mathbf{R}^{-1}\,[\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)]}{(1+x)^2\Big(\frac{x}{x+1}\Big)^{2\gamma}} > c \right).
\end{aligned}$$

Note that $\mathbf{R}$ is a functional of Brownian motion. By the independent-increments property, $\{\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)\}_{x\in(0,\infty)}$ is independent of $\mathbf{R}$. Therefore,

$$\begin{aligned}
&P\left(\sup_{0<x<\infty} \frac{[\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)]'\,\mathbf{R}^{-1}\,[\mathbb{B}_2(x) - x\,\mathbb{B}_1(1)]}{(1+x)^2\Big(\frac{x}{x+1}\Big)^{2\gamma}} > c \right) \\
&= P\left(\sup_{0<x<\infty} \frac{\Big[(1+x)\,\mathbb{B}\Big(\frac{x}{1+x}\Big)\Big]'\,\mathbf{R}^{-1}\,\Big[(1+x)\,\mathbb{B}\Big(\frac{x}{1+x}\Big)\Big]}{(1+x)^2\Big(\frac{x}{x+1}\Big)^{2\gamma}} > c \right) \\
&= P\left(\sup_{0<u\le 1} \frac{\mathbb{B}^\dagger(u)'\mathbf{R}^{-1}\mathbb{B}^\dagger(u)}{u^{2\gamma}} > c \right),
\end{aligned}$$

which proves Theorem 1(c). □

A.2.4 Proof of Theorem 2

By Assumptions 7 and 8 and the ULLN,

$$\frac{1}{m+mT-t^*+1} \sum_{t=t^*}^{m+mT} \psi_j^*\Big(\widehat{\theta}_m\Big) = \mathrm{E}\Big(\psi_j^*(\theta_0)\Big) + o_p(1).$$

Hence, for any coordinate $i = 1, 2, \ldots, d$ with $\mathrm{E}\Big(\psi_{ji}^*(\theta_0)\Big) \neq 0$,

$$\begin{aligned}
&\frac{S_m\Big(mT, \widehat{\theta}_m\Big)}{m^{\frac{1}{2}}\Big(1+\frac{mT}{m}\Big)^{1-\gamma}\Big(\frac{mT}{m}\Big)^{\gamma}} \\
&= m^{-\frac{1}{2}}(1+T)^{-(1-\gamma)}T^{-\gamma}\left(\sum_{t=m+1}^{t^*-1} \psi_j\Big(\widehat{\theta}_m\Big) + \sum_{t=t^*}^{m+mT} \psi_j^*\Big(\widehat{\theta}_m\Big) \right) \\
&= O_p(1) + \frac{m+mT-t^*+1}{m^{\frac{1}{2}}(1+T)^{1-\gamma}T^{\gamma}} \Big(\mathrm{E}\Big(\psi_j^*(\theta_0)\Big) + o_p(1) \Big) \\
&= O_p(\sqrt{m}).
\end{aligned}$$

Consequently,

$$\sup_{1\le k\le mT} \frac{S_m\big(k,\widehat{\theta}_m\big)' \mathbf{R}_m\big(\widehat{\theta}_m\big)^{-2} S_m\big(k,\widehat{\theta}_m\big)}{m\big(1+\frac{k}{m}\big)^2\big(\frac{k}{k+m}\big)^{2\gamma}}$$
$$\ge \left(\frac{S_m\big(mT,\widehat{\theta}_m\big)}{m^{\frac{1}{2}}\big(1+\frac{mT}{m}\big)^{1-\gamma}\big(\frac{mT}{m}\big)^{\gamma}}\right)' \mathbf{R}_m\big(\widehat{\theta}_m\big)^{-2} \left(\frac{S_m\big(mT,\widehat{\theta}_m\big)}{m^{\frac{1}{2}}\big(1+\frac{mT}{m}\big)^{1-\gamma}\big(\frac{mT}{m}\big)^{\gamma}}\right) \to \infty,$$

as $m \to \infty$. It follows that

$$\lim_{m\to\infty} P\big(\mathbb{T}_m^R \le mT \mid H_1\big)$$
$$= \lim_{m\to\infty} P\left(\sup_{1\le k\le mT} \frac{S_m\big(k,\widehat{\theta}_m\big)' \mathbf{R}_m\big(\widehat{\theta}_m\big)^{-2} S_m\big(k,\widehat{\theta}_m\big)}{m\big(1+\frac{k}{m}\big)^2\big(\frac{k}{k+m}\big)^{2\gamma}} > c \,\middle|\, H_1\right) \to 1.$$

The same reasoning applies to the open-end procedure. □

A.3 RSMS for M-Estimators

This appendix develops the RSMS framework for M-estimators. Here "M" denotes "maximum-likelihood–type." M-estimators optimize a general sample objective, extending likelihood maximization to broader criteria and enabling robust procedures tailored to the data. Our development mirrors the functional approach in Section 2 and delivers a general asymptotic theory under standard regularity conditions.

The assumptions in Section 2 carry over to the M-estimation setting. We consider solutions to the first-order conditions

$$\sum_{t=m+1}^{m+k} \psi(X_t, \widehat{\theta}_m) = 0, \tag{A.16}$$

where $\psi(X_t, \theta)$ is the score (estimating) function. Equation (A.16) is evaluated on the monitoring sample $\{X_{m+1}, \ldots, X_{m+k}\}$ using $\widehat{\theta}_m$ obtained from the training sample $\{X_1, \ldots, X_m\}$. Since an M-estimator maximizes or minimizes a criterion function, the maximum likelihood estimator is a special case in which $\psi(X_t, \theta)$ is the derivative of the log-likelihood.

Remark A.1. *The function $\psi(\cdot)$ encompasses approximately linear statistics as special cases. For $d = 1$, to monitor a mean, set $\psi(X_t, \widehat{\mu}_m) = X_t - \widehat{\mu}_m$; to monitor a variance, set $\psi(X_t, \widehat{\sigma}_m^2) = \frac{k}{k-1}(X_t - \widehat{\mu}_{m+1,m+k})^2 - \widehat{\sigma}_m^2$, where $\widehat{\mu}_m$ and $\widehat{\sigma}_m^2$ are training-sample estimates and $\widehat{\mu}_{m+1,m+k}$ is the mean of $\{X_{m+1}, \ldots, X_{m+k}\}$.*

Consider the process $\{X_t\}_{t\in\mathbb{Z}}$:

$$X_t = \begin{cases} g(\varepsilon_t, \varepsilon_{t-1}, \ldots), & \text{for } t < m + k^*, \\ h(\varepsilon_t, \varepsilon_{t-1}, \ldots), & \text{for } t \geq m + k^*, \end{cases} \tag{A.17}$$

where $g(\cdot)$ and $h(\cdot)$ exhibit temporal decay, $m + k^*$ is the change point, and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are i.i.d. on a measurable space $\mathbb{S}$. The mappings $g, h : \mathbb{S}^{\mathbb{N}} \to \mathbb{R}^d$ describe pre- and post-change dynamics. Under $H_0 : g = h$, the series is stationary; under $H_1 : g \neq h$, the series forms a triangular array.

To quantify temporal dependence, we employ physical dependence measures (Wu 2005, Dette & Gösmann 2020). Let ε_0' be an independent copy of ε_0 and define

$$\rho_{t,q}^{(1)} = \left(\mathrm{E}\left[\left|g(\varepsilon_t, \varepsilon_{t-1}, \ldots) - g(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \ldots)\right|^q\right]\right)^{1/q}, \tag{A.18}$$

$$\rho_{t,q}^{(2)} = \left(\mathrm{E}\left[\left|h(\varepsilon_t, \varepsilon_{t-1}, \ldots) - h(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \ldots)\right|^q\right]\right)^{1/q}. \tag{A.19}$$

Assume $\rho_{t,q}^{(1)} = O(\zeta^t)$ and $\rho_{t,q}^{(2)} = O(\zeta^t)$ for some $\zeta \in (0, 1)$. Define the cumulative measures

$$\theta_q^{(\ell)} = \sum_{t=1}^{\infty} \rho_{t,q}^{(\ell)}, \quad \ell = 1, 2. \tag{A.20}$$

Using the IF defined in Section 2, a Taylor approximation under standard regularity conditions yields

$$\mathbf{IF}\left(\theta, \mathbf{v}^d\right) = M^{-1} \int \psi\left(X_t; \theta\right) \mathbf{v}^d(dX_t)\,, \tag{A.21}$$

where

$$M = \int \psi(x, \theta) \lambda^d(x, \theta)^t \mathbf{F}^d(dX), \tag{A.22}$$

and $\lambda^d(x,\theta) = \frac{\partial}{\partial\theta}\log f^d(x,\theta)$ with f^d the joint likelihood of $X_t \in \mathbb{R}^d$ under parameter θ. Because $\{X_t\}$ are observed at integer times, $\mathbf{v}^d$ is approximated by a functional of Dirac measures and the empirical distribution is $\widehat{\mathbf{F}}_n^d = \frac{1}{n}\sum_{t=1}^n \delta_{X_t}$, so

$$\mathbf{IF}\left(\theta, \mathbf{v}^d\right) = M^{-1}\frac{1}{n}\sum_{t=1}^{n}\psi\left(X_t;\theta\right).$$

We next state sufficient conditions to apply the general theory in Section 3 to M-estimators based on $\psi(X_t,\theta)$. Assumption 1 holds if $\{X_t\}$ is stationary and an invariance principle holds, i.e., for $r \in [0, 1+T]$,

$$\frac{1}{\sqrt{m}}\sum_{t=1}^{\lfloor mr\rfloor}\mathbf{IF}\left(X_t,\theta_0\right) \overset{\mathcal{D}}{\Longrightarrow} \sqrt{\mathbf{\Sigma}}\Big(\mathbb{B}_d(r) - r\mathbb{B}_d(1)\Big),$$

with $\mathbf{\Sigma} = M^{-1}\mathbf{\Sigma}_\psi M^{-1}$ the LRV of the M-estimator. Using

$$\mathbf{IF}(X_t,\theta_0) = M^{-1}\psi(X_t,\theta_0),$$

we obtain

$$\frac{1}{\sqrt{m}}\sum_{t=1}^{\lfloor mr\rfloor}\psi(X_t,\theta_0) \overset{\mathcal{D}}{\Longrightarrow} \sqrt{\mathbf{\Sigma}_\psi}\Big(\mathbb{B}_d(r) - r\mathbb{B}_d(1)\Big).$$

Here $\mathbf{\Sigma}_\psi = \mathbb{V}\mathrm{ar}\Big(\psi(X_t,\theta_0)\Big) = \sum_{k=-\infty}^{\infty}\mathrm{E}\Big[\psi(X_t,\theta_0)\psi(X_{t+k},\theta_0)'\Big]$, as in Wu (2005) under $\Theta_q^{(1)} < \infty$. With a change point, stationarity holds piecewise if $\Theta_q^{(2)} < \infty$, and weak convergence remains valid provided $h = g + \eta$ with $\eta \neq 0$.

A higher-order remainder $\widetilde{\mathbf{Re}}$ arises from the Taylor expansion:

$$\mathcal{G}((1-\varepsilon)\mathbf{F}^d + \varepsilon\mathbf{v}^d) = \mathcal{G}(\mathbf{F}^d) + \varepsilon\int \mathbf{IF}(x,\theta_0)\,\mathbf{v}^d(dX) + \frac{\varepsilon^2}{2}\iint\frac{\partial^2\mathcal{G}}{\partial\mathbf{v}^d\partial\mathbf{v}^d}(x,x';\theta)\,\mathbf{v}^d(dX)\,\mathbf{v}^d(dX') + o(\varepsilon^2), \tag{A.23}$$

with remainder

$$\widetilde{\mathbf{Re}} = \frac{\varepsilon^2}{2}\iint\frac{\partial^2\mathcal{G}}{\partial\mathbf{v}^d\partial\mathbf{v}^d}(x,x';\theta)\,\mathbf{v}^d(dx)\,\mathbf{v}^d(dx').$$

Since $\mathbf{v}^d$ is a Dirac measure at the sample points X_t, i.e., $\mathbf{v}^d(dx) = \frac{1}{j-i+1}\sum_{t=i}^{j}\delta_{X_t}(dx)$, the double integral becomes

$$\iint\frac{\partial^2\mathcal{G}}{\partial\mathbf{v}^d\partial\mathbf{v}^d}(x,x';\theta)\mathbf{v}^d(dx)\mathbf{v}^d(dx') = \frac{1}{(j-i+1)^2}\sum_{t=i}^{j}\sum_{t'=i}^{j}\frac{\partial^2\mathcal{G}}{\partial\mathbf{v}^d\partial\mathbf{v}^d}(X_t,X_{t'};\theta).$$

Thus

$$\widetilde{\mathbf{Re}}_{i,j} = \frac{\varepsilon^2}{2} \cdot \frac{1}{(j-i+1)^2} \sum_{t=i}^{j} \sum_{t'=i}^{j} \frac{\partial^2 \mathcal{G}}{\partial \mathbf{v}^d \partial \mathbf{v}^d}(X_t, X_{t'}; \theta)\,,$$

and

$$(j-i+1)\,\widetilde{\mathbf{Re}}_{i,j} = O_p\Big(n^{1/2}\Big),$$

provided $\varepsilon^2 = o_p\Big(n^{-1/2}\Big)$. Hence the higher-order term satisfies the same order requirement as in Assumption 2.

The validity of this condition depends on the dependence in $\{X_t\}$ via g and h in (A.17). For M-estimators, sufficient conditions can be phrased in terms of moments and dependence of $\psi(X_t, \theta_0)$.

Theorem A.1. *Assume* (A.17) *holds with bounded g and h; Assumptions 3–6 hold under H_0, and Assumptions 7–8 hold under H_1. Suppose the covariance matrices $\Gamma(g)$ and $\Gamma(h)$ are positive definite.*

(a) Under $H_0 : g = h$, the assumptions of Theorem 1 are satisfied, and the RSMS statistic $\mathbb{M}_m^R(k)$ attains the nominal asymptotic size α.

(b) Under $H_1 : g \neq h$, the assumptions of Theorem 2 are satisfied, and $\mathbb{M}_m^R(k)$ is consistent.

Proof of Theorem A.1. For part (a), we first show $\sum_{t=1}^{\infty} \rho_{t,q} < \infty$, where

$$\rho_{t,q} = \Big(\mathrm{E}\big[|\psi(X_t, \theta) - \psi(X_t', \theta)|^q\big]\Big)^{1/q},$$

with $X_t = g(\varepsilon_t, \varepsilon_{t-1}, \dots)$ and $X_t' = g(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_0', \varepsilon_{-1}, \dots)$, the latter obtained by replacing ε_0 with an independent copy ε_0'. The term $\rho_{t,q}$ gauges the impact on $\psi(X_t, \theta)$ of perturbing ε_0 and thus measures temporal dependence.

By Assumption 6, ψ is Lipschitz, so

$$|\psi(X_t, \theta) - \psi(X_t', \theta)| \leq L_\psi \|X_t - X_t'\|.$$

Taking qth moments,

$$\mathrm{E}\Big[|\psi(X_t,\theta)-\psi(X_t',\theta)|^q\Big] \le L_\psi^q\,\mathrm{E}\Big[\|X_t-X_t'\|^q\Big],$$

hence

$$\rho_{t,q} \le L_\psi\Big(\mathrm{E}[\|X_t-X_t'\|^q]\Big)^{1/q}.$$

From (A.17) and temporal decay of g,

$$X_t - X_t' = g(\varepsilon_t,\varepsilon_{t-1},\ldots) - g(\varepsilon_t,\varepsilon_{t-1},\ldots,\varepsilon_1,\varepsilon_0',\varepsilon_{-1},\ldots),$$

and thus

$$\|X_t-X_t'\| \le C\zeta^t\|\varepsilon_0-\varepsilon_0'\|,$$

for some $C>0$ and $\zeta\in(0,1)$. Therefore,

$$\rho_{t,q} \le L_\psi\Big(\mathrm{E}\Big[C^q\zeta^{tq}\|\varepsilon_0-\varepsilon_0'\|^q\Big]\Big)^{1/q}.$$

By independence and finiteness of $\mathrm{E}\|\varepsilon_0\|^q$,

$$\mathrm{E}\|\varepsilon_0-\varepsilon_0'\|^q \le 2^q\mathrm{E}\|\varepsilon_0\|^q,$$

so

$$\rho_{t,q} \le L_\psi C\zeta^t\Big(2^q\mathrm{E}\|\varepsilon_0\|^q\Big)^{1/q} \le K\zeta^t,$$

with $K = L_\psi C(2^q\mathrm{E}\|\varepsilon_0\|^q)^{1/q}$. Summing over t,

$$\sum_{t=1}^{\infty}\rho_{t,q} \le K\sum_{t=1}^{\infty}\zeta^t = K\,\frac{\zeta}{1-\zeta} < \infty.$$

By Theorem 3 in Wu (2005), the weak convergence

$$\frac{1}{\sqrt{m}}\sum_{t=1}^{\lfloor mr\rfloor}\psi(X_t,\theta) \overset{\mathcal{D}}{\Longrightarrow} \sqrt{\boldsymbol{\Sigma}_\psi}\Big(\mathbb{B}_d(r)-r\mathbb{B}_d(1)\Big)$$

holds in $l^\infty([0,T+1])$ as $m\to\infty$.

For the remainder in (A.23), write

$$\widetilde{\mathbf{Re}} = \frac{\varepsilon^2}{2} \iint \frac{\partial^2 \mathcal{G}}{\partial \mathbf{v}^d \, \delta \mathbf{v}^d}(x, x'; \theta) \, \mathbf{v}^d(dx) \, \mathbf{v}^d(dx') + o(\varepsilon^2).$$

On $[i, j]$, the empirical distribution is

$$\widehat{\mathbf{F}}^d_{i,j} = \frac{1}{j - i + 1} \sum_{t=i}^{j} \delta_{X_t},$$

so

$$\widehat{\mathbf{F}}^d_{i,j} - \mathbf{F}^d = \frac{1}{j - i + 1} \sum_{t=i}^{j} (\delta_{X_t} - \mathbf{F}^d).$$

Expanding $\mathcal{G}(\cdot)$,

$$\mathcal{G}(\widehat{\mathbf{F}}^d_{i,j}) = \mathcal{G}(\mathbf{F}^d) + \frac{1}{j - i + 1} \sum_{t=i}^{j} \mathbf{IF}(X_t, \mathbf{F}^d) + \widetilde{\mathbf{Re}}_{i,j},$$

with

$$\widetilde{\mathbf{Re}}_{i,j} = \frac{1}{2} \iint \frac{\partial^2 \mathcal{G}}{\partial \mathbf{v}^d \partial \mathbf{v}^d}(x, x'; \theta) \, (\widehat{\mathbf{F}}^d_{i,j} - \mathbf{F}^d)(dx) \, (\widehat{\mathbf{F}}^d_{i,j} - \mathbf{F}^d)(dx') + o_p(n^{1/2}).$$

Equivalently,

$$\widetilde{\mathbf{Re}}_{i,j} = \frac{1}{2} \iint \frac{\partial^2 \mathcal{G}}{\partial \mathbf{v}^d \partial \mathbf{v}^d}(x, x'; \theta) \cdot \frac{1}{(j - i + 1)^2} \sum_{t=i}^{j} \sum_{t'=i}^{j} (\delta_{X_t} - \mathbf{F}^d)(dx) \, (\delta_{X_{t'}} - \mathbf{F}^d)(dx').$$

If $\sup_{x,x'} \left| \partial^2 \mathcal{G} / (\partial \mathbf{v}^d \partial \mathbf{v}^d)(x, x'; \theta) \right| \leq C$, then

$$|\widetilde{\mathbf{Re}}_{i,j}| \;\leq\; \frac{C}{(j - i + 1)^2} \sum_{t=i}^{j} \sum_{t'=i}^{j} \left| (\delta_{X_t} - \mathbf{F}^d)(dx) \, (\delta_{X_{t'}} - \mathbf{F}^d)(dx') \right|.$$

Since $\mathbb{V}\mathrm{ar}\Big(\sum_{t=i}^{j} (\delta_{X_t} - \mathbf{F}^d) \Big) \propto (j - i + 1)$, it follows that

$$\widetilde{\mathbf{Re}}_{i,j} = O_p\Big((j - i + 1)^{-1} \Big), \quad \text{so} \quad (j - i + 1) \, |\widetilde{\mathbf{Re}}_{i,j}| = o_p(n^{1/2}).$$

Thus,

$$\sup_{1 \leq i < j \leq n} (j - i + 1) \, |\widetilde{\mathbf{Re}}_{i,j}| = o_p(n^{1/2}),$$

under boundedness of $\partial^2 \mathcal{G} / (\partial \mathbf{v}^d \partial \mathbf{v}^d)$ and sufficiently weak dependence of X_t.

For part (b), the consistency of $\mathbb{M}_m^R(k)$ follows from the joint weak convergence of the pre- and post-change IF processes under the alternative, the existence of a structural change point, and the positive definiteness of the long-run covariance matrices. These conditions satisfy Assumption 8, ensuring that the regularity conditions of Theorem 1(b) hold, from which the consistency result follows. □

A.4 Online Change-Point Detection of Both Conditional Mean and Variance Using the PAR Model

The PAR(1) process is a count-valued process $\{X_t\}$ with $X_t \in \mathbb{Z}_+$ evolving via

$$X_t = \alpha X_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{Pois}(\lambda).$$

The operator "$\circ$" denotes binomial thinning, a discrete-time mechanism introduced by Al-Osh & Alzaid (1987) for modeling count time series. Conditional on X_{t-1}, binomial thinning is defined by

$$\alpha \circ X_{t-1} = \sum_{i=1}^{X_{t-1}} Z_i, \qquad Z_i \sim \text{Bernoulli}(\alpha) \text{ independently.}$$

The parameter vector $\theta = (\alpha, \lambda)$ governs the dynamics: $\alpha \in [0, 1)$ is the thinning parameter controlling temporal dependence, and $\lambda > 0$ is the Poisson innovation rate determining the arrival intensity of the independent innovations ε_t.

The choice of the PAR(1) model is motivated by three considerations. First, its marginal distribution is Poisson with equal mean and variance, permitting simultaneous monitoring of changes in both parameters. Second, whereas online change-point detection in linear time series often uses forecast errors, the PAR model typically calls for maximum likelihood (ML) estimation; this yields score functions that naturally underpin CUSUM constructions and

align with our asymptotic framework. Third, the growing demand for robust count-data models—especially since the COVID-19 pandemic—underscores the practical relevance of PAR(1). Unlike linear AR models, PAR preserves key distributional features of counts (non-negativity and discreteness), providing a coherent probabilistic basis for analysis.

Given a count series $\{X_t\}_{t=1}^n$, the conditional (on the initial observation) log-likelihood is

$$l(\alpha, \lambda) = \sum_{t=2}^{n} \log P(X_t \mid X_{t-1}),$$

where

$$P(X_t \mid X_{t-1}) = \sum_{i=0}^{X_t \wedge X_{t-1}} \binom{X_{t-1}}{i} \alpha^i (1-\alpha)^{X_{t-1}-i} \frac{e^{-\lambda} \lambda^{X_t - i}}{(X_t - i)!}.$$

The score vector is

$$\psi(X_t, \theta) = \left(\frac{\partial}{\partial \alpha} l(\alpha, \lambda),\ \frac{\partial}{\partial \lambda} l(\alpha, \lambda) \right)'.$$

The null hypothesis is $H_0 : \theta$ is constant; the alternative is $H_1 : H_0$ is false. We estimate the PAR(1) model by maximum likelihood on the training sample to obtain $\widehat{\theta}_m$. The resulting score functions, evaluated using $\widehat{\theta}_m$ for both training and testing samples, are then used to construct the RSMS, SSMS, and CSMS.

We first examine type I error for the PAR(1) design (Table A.3). For $\gamma = 0$ with $m = 100$, CSMS and RSMS are closest to nominal at small T: at $T = 1$, CSMS yields 11.5% (10% nominal) and 8.1% (5% nominal), while RSMS yields 12.3% and 9.2%, respectively. SSMS is most conservative at $T = 1$ and 2 (e.g., 7.1% and 3.9% at $T = 1$). At $T = 5$ and 10, CSMS becomes the most conservative, whereas RSMS remains reasonably sized even at $T = 10$, which approximates the open-end scheme. As m increases, all methods become more conservative, with realized sizes falling below nominal.

With $\gamma = 0.15$, CSMS and RSMS show inflated rejection at $m = 100$—especially at $T = 1$ and 2—but quickly turn conservative as m grows. For large m and T, all three procedures under-reject. Notably, Chan et al. (2021) implement SSMS with $\gamma = 0$ and do not report

Table A.3: Type I error rates (%) for RSMS (R), SSMS (S), and CSMS (C) under the PAR(1) null. Columns report monitoring horizons $T \in \{1, 2, 5, 10\}$ at nominal levels 5% and 10%. Rows show training sizes $m \in \{100, 500, 1000\}$. Panels correspond to $\gamma = 0$ and $\gamma = 0.15$.

	T=1						T=2						T=5						T=10					
	R		S		C		R		S		C		R		S		C		R		S		C	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$\gamma = 0$																								
$m = 100$	9.2	12.3	3.9	7.1	11.5	15.4	8.7	13.9	3.1	5.8	8.1	11.6	8.3	12.9	3.4	6.3	2.1	3.9	8.5	12.1	3.3	5.5	0.8	1.3
$m = 500$	4.6	8.2	1.1	3.5	2.9	6.3	4.3	8.5	1.4	3.3	1.6	3.8	4.8	7.1	1.8	3.0	0.1	0.3	4.3	6.4	1.7	3.5	0.0	0.0
$m = 1000$	2.8	5.5	1.1	3.0	1.4	3.9	2.7	6.5	1.2	2.5	0.7	1.7	2.9	5.3	1.7	3.0	0.0	0.2	2.3	4.7	1.6	2.8	0.0	0.0
$\gamma = 0.15$																								
$m = 100$	13.4	18.1	1.5	2.7	13.3	17.5	13.4	19.5	1.1	2.4	8.6	12.0	12.1	18.1	1.5	2.8	2.3	3.1	12.8	18.1	1.5	2.8	0.6	1.1
$m = 500$	8.0	13.4	0.3	0.6	4.2	7.8	7.9	11.8	0.3	1.3	1.6	3.5	6.9	11.9	0.3	1.2	0.1	0.1	7.2	12.5	0.2	1.1	0.0	0.0
$m = 1000$	5.6	10.1	0.0	0.6	2.3	5.2	6.3	10.5	0.2	0.7	0.5	1.4	6.0	9.8	0.5	1.0	0.0	0.0	5.8	10.3	0.6	1.2	0.0	0.0

$\gamma > 0$ results, a choice consistent with the finite–sample conservatism we document at $\gamma = 0.15$.

In sum, CSMS and RSMS are well-sized or somewhat liberal in small samples, while all methods—particularly SSMS—become increasingly conservative as the training size or monitoring horizon rises. Accordingly, we report power for $T \in \{1, 2, 5\}$ and restrict ARL analysis to $T = 1$, reflecting the focus on rapid detection under short horizons.

Second, to evaluate empirical rejection under H_1, we simulate series with a structural break at $t^* = m + k^*$, fixing $m = 500$ and considering $k^* \in \{50, 200\}$.

We consider two types of structural change:

(i) Abrupt shift:

$$\alpha_t = \begin{cases} \alpha_1, & t \le m + k^*, \\ \alpha_2, & t > m + k^*, \end{cases} \qquad \lambda_t = \begin{cases} \lambda_1, & t \le m + k^*, \\ \lambda_2, & t > m + k^*. \end{cases}$$

(ii) Smooth change:

$$\alpha_t = \alpha_1 + (\alpha_2 - \alpha_1)\Big(1 - e^{-\delta\,(t-(m+k^*))}\Big), \quad t > m + k^*,$$
$$\lambda_t = \lambda_1 + (\lambda_2 - \lambda_1)\Big(1 - e^{-\delta\,(t-(m+k^*))}\Big), \quad t > m + k^*.$$

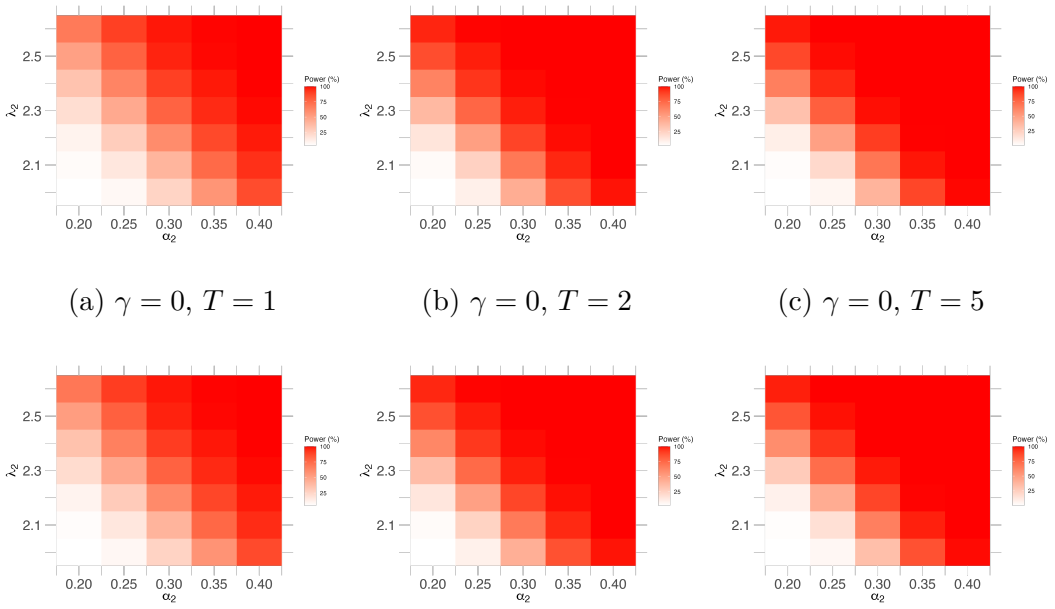
Before the break, set $\alpha_t = \alpha_1 = 0.2$ and $\lambda_t = \lambda_1 = 2$. After $t^* = m + k^*$, vary α_2 and λ_2 on a two–dimensional grid, $\alpha_2 \in \{0.2, 0.25, \ldots, 0.35\}$ and $\lambda_2 \in \{2.0, 2.1, \ldots, 2.5\}$, to systematically explore power across signal strengths. For smooth changes, parameters evolve according to the exponential transition with rate $\delta = 0.05$.

Figures A.1–A.3 report heatmaps of empirical rejection rates for RSMS, SSMS, and CSMS. Within each method, the top row corresponds to $\gamma = 0$ and the bottom to $\gamma = 0.15$; columns index $T = 1, 2, 3$. Color intensity reflects detection power (darker is higher). RSMS and CSMS display rapid, monotone gains in power as the signal strengthens, with upper-right regions approaching 100% rejection. By contrast, SSMS is generally less sensitive, with lighter shades and fewer cells near full rejection.

Third, we examine ARLs for RSMS, SSMS, and CSMS (Figure A.4). We focus on Type (i) (abrupt) breaks, as the qualitative patterns persist under smooth changes. To emphasize early detection, we consider $k^* = 50$ and $T = 1$. The contour plots discretize ARL into five equally spaced bands, with darker shading indicating faster detection (lower ARL).

Each subfigure displays ARL over the post-change grid (α_2, λ_2), with α_2 on the horizontal axis and λ_2 on the vertical axis. Across all procedures and both γ values, ARL declines as α_2 and λ_2 increase, consistent with stronger signals yielding quicker detection.

Comparing the top row ($\gamma = 0$) to the bottom row ($\gamma = 0.15$), introducing a positive γ improves detection for RSMS and CSMS: darker regions (short ARL) expand and lighter regions contract. By contrast, for SSMS, $\gamma = 0.15$ produces a pronounced light band (ARL $\approx$ 400–500), indicating slower detection. Thus, a positive γ benefits early-stage monitoring



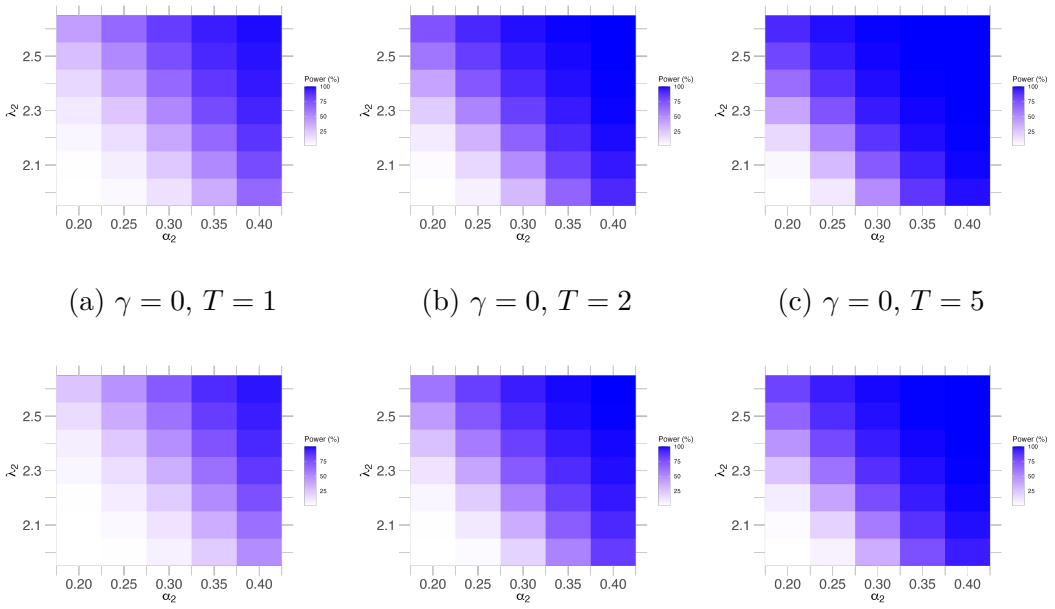
(a) $\gamma = 0$, $T = 1$ (b) $\gamma = 0$, $T = 2$ (c) $\gamma = 0$, $T = 5$

(d) $\gamma = 0.15$, $T = 1$ (e) $\gamma = 0.15$, $T = 2$ (f) $\gamma = 0.15$, $T = 5$

Figure A.1: Empirical rejection rates for RSMS over the post-change grid (α_2, λ_2). Rows: $\gamma = 0$ (top), $\gamma = 0.15$ (bottom). Columns: $T = 1, 2, 5$. Darker shading indicates higher power.

under RSMS and CSMS, but can impair SSMS performance.

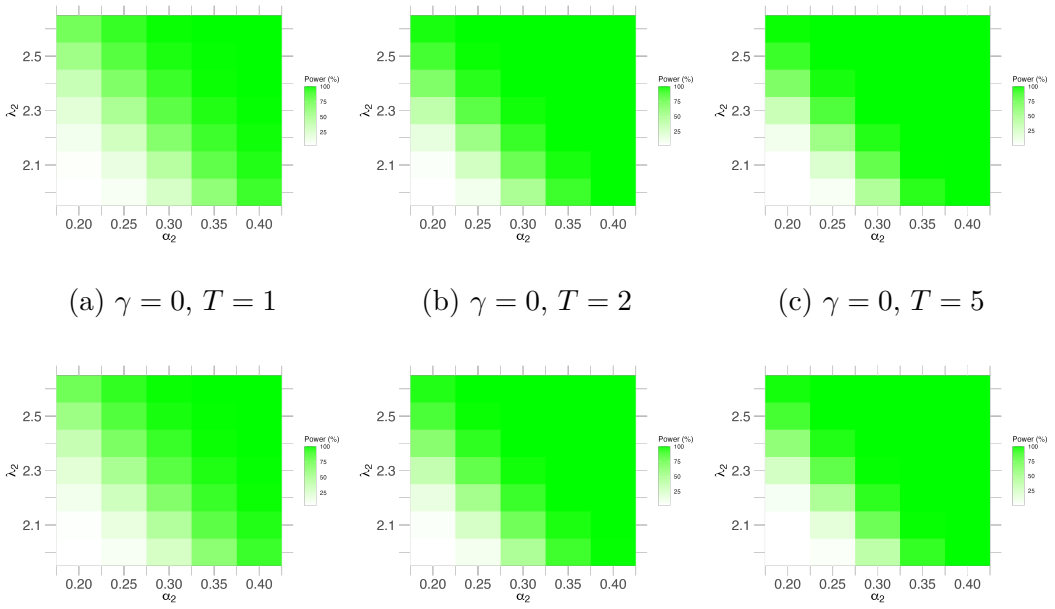
Overall, across a range of designs and evaluation criteria, RSMS achieves a favorable balance of size control and power under both the null and the alternative. It also delivers shorter ARLs, indicating faster detection. Taken together, these features make RSMS the preferred monitoring scheme among the three.



(a) $\gamma = 0$, $T = 1$ (b) $\gamma = 0$, $T = 2$ (c) $\gamma = 0$, $T = 5$

(d) $\gamma = 0.15$, $T = 1$ (e) $\gamma = 0.15$, $T = 2$ (f) $\gamma = 0.15$, $T = 5$

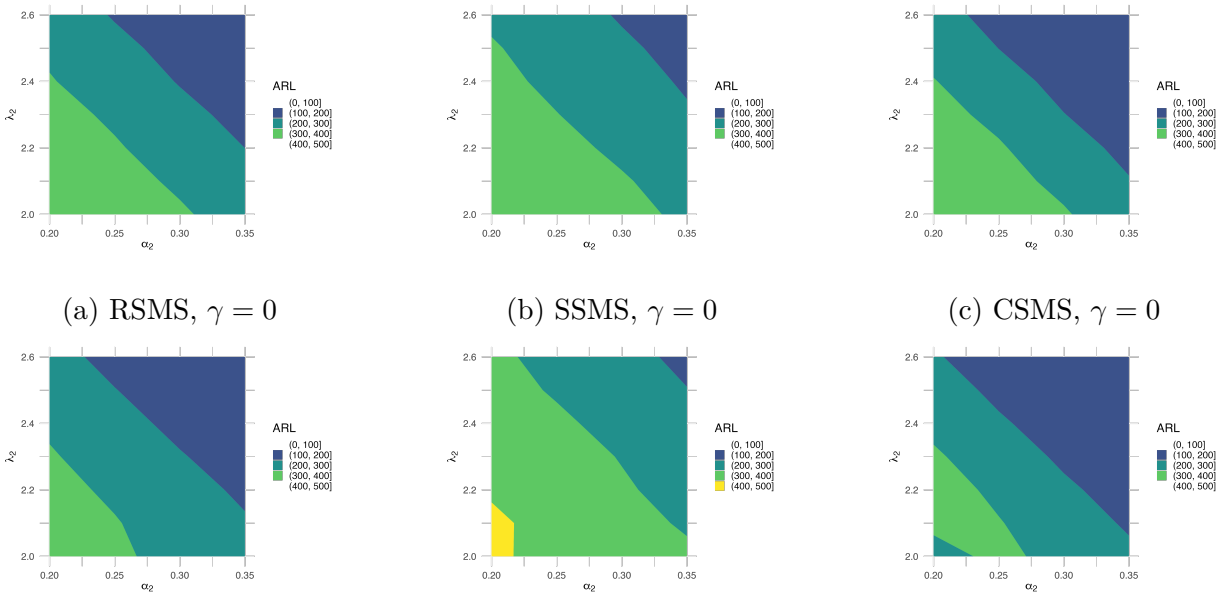
Figure A.2: Empirical rejection rates for SSMS over the post-change grid (α_2, λ_2). Rows: $\gamma = 0$ (top), $\gamma = 0.15$ (bottom). Columns: $T = 1, 2, 5$. Darker shading indicates higher power.



(a) $\gamma = 0$, $T = 1$ (b) $\gamma = 0$, $T = 2$ (c) $\gamma = 0$, $T = 5$

(d) $\gamma = 0.15$, $T = 1$ (e) $\gamma = 0.15$, $T = 2$ (f) $\gamma = 0.15$, $T = 5$

Figure A.3: Empirical rejection rates for CSMS over the post-change grid (α_2, λ_2). Rows: $\gamma = 0$ (top), $\gamma = 0.15$ (bottom). Columns: $T = 1, 2, 5$. Darker shading indicates higher power.



(a) RSMS, $\gamma = 0$ (b) SSMS, $\gamma = 0$ (c) CSMS, $\gamma = 0$

(d) RSMS, $\gamma = 0.15$ (e) SSMS, $\gamma = 0.15$ (f) CSMS, $\gamma = 0.15$

Figure A.4: ARL under Type (i) (abrupt) breaks with $k^* = 50$ and $T = 1$ for RSMS, SSMS, and CSMS. Rows: $\gamma = 0$ (top) and $\gamma = 0.15$ (bottom). Each panel plots ARL over the post-change grid (α_2, λ_2).

A.5 Contaminated Training and Its Effects on Online Change-Point Detection

We follow the contaminated training design used in Figure 1 of Shao & Zhang (2010). Let the training length be m and define

$$X_t = \begin{cases} u_t, & 1 \le t \le m/2, \\ \eta + u_t, & m/2+1 \le t \le m, \end{cases} \qquad u_t = \alpha\, u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0,1) \text{ independently over } t.$$

In our baseline experiments the autoregressive coefficient is $\alpha = 0.5$. Following our simulation notation we write the level shift as $\lambda \equiv \eta$. As shown by Shao & Zhang (2010), the power of the adjusted-range KS statistic can deteriorate to zero as the magnitude of the shift grows; the loss is driven by the concomitant growth of the self-normalizer introduced by Shao (2010). To mitigate this phenomenon, Shao & Zhang (2010) proposed the so-called G test. Hong, Linton, McCabe, Sun & Wang (2024) demonstrate that the adjusted–range–based self-normalization does not suffer the same issue, which makes the adjusted–range–based KS test statistic a robust and power-stable choice for sequential change-point detection under contaminated training.

After training we observe a monitoring stream of length mT, where $T \in \{1, 2, 5\}$ is a fixed horizon multiplier. The (single) true change-point within the monitoring window is set at $k^\star = \lfloor m/4 \rfloor$, and the process continues the AR(1) dynamics but may experience a jump of size

$$\Delta_{\text{test}} = \text{mult} \times \lambda, \qquad \text{mult} \in \{1, 2, 3\}.$$

Formally, for $j = 1, \ldots, mT$,

$$X_{m+j} = \alpha X_{m+j-1} + \varepsilon_{m+j} + \Delta_{\text{test}} \cdot \mathbf{1}\{j > k^\star\}.$$

To remove the location distortion inherited from training, we demean the full stream by the training mean $\bar{X}_m = m^{-1} \sum_{t=1}^{m} X_t$ and apply monitoring to $\tilde{X}_{m+j} = X_{m+j} - \bar{X}_m$.

In the simulation design, we set the training length $m \in \{200, 500\}$, the monitoring horizon $T \in \{1, 2, 5\}$, and the contamination level λ on the grid $\{0, 0.5, 1.0, \ldots, 3.0\}$. The post-change size is indexed by mult $\in \{1, 2, 3\}$ through $\Delta_{\text{test}} = \text{mult} \times \lambda$. The AR parameter is fixed at $\alpha = 0.5$ throughout training and monitoring. For the monitoring rules we adopt a nominal size $\alpha_{\text{sig}} = 0.05$ and set the tuning parameter $\gamma = 0$, under which SSMS attains accurate size.

As emphasized by Shao & Zhang (2010), contamination in the training sample inflates the self-normalizer of naive KS-type procedures, leading to a loss of power. Although Chan et al. (2021) do not explicitly discuss this point, it is clear that self-normalized KS-type monitoring statistics are susceptible to the same power loss. In sequential applications contamination is difficult to avoid because type II errors and rolling re-estimation naturally push recent (possibly shifted) observations into the training window. Consequently, the inflation identified by Shao & Zhang (2010) limits the direct use of the Shao's (2010) self-normalization, motivating robust alternatives.

Table A.4 presents a representative subset of the contaminated-training Monte Carlo, comparing RSMS and SSMS across selected $(\lambda, m, T, \text{mult})$ with $\Delta_{\text{test}} = \text{mult} \times \lambda$ under $\gamma = 0$ (higher detection and lower ARL are preferable).

Across these ten scenarios, RSMS uniformly dominates SSMS. Gains in detection range from roughly 12 to 99 percentage points, and ARL reductions span about 24 to 585 observations. The advantage is especially pronounced when the post–change size is modest (mult = 1): for $(\lambda = 0.5, m = 200, T = 1)$ the detection rates are 44.0% versus 1.7% with ARL 175.49 versus 199.16; for $(\lambda = 1.0, m = 200, T = 1)$ they are 76.8% versus 0.6% with ARL 163.81 versus 199.80. For larger changes (mult = 2), RSMS often attains 100.0% while SSMS lags—e.g., $(\lambda = 1.0, m = 500, T = 2)$ yields 100.0% versus 71.6% with ARL 245.59 versus 802.41. Even when both methods are near certain to detect, RSMS alarms earlier; i.e. in

Table A.4: RSMS vs. SSMS (training sample contaminated): detection rate and ARL across selected contamination levels λ, training length m, monitoring horizon T, and multiplier mult (test change size $\Delta_{\text{test}} = \text{mult} \times \lambda$). Results for $\gamma = 0$.

λ (train)	m	T	mult ($\times\lambda$)	Δ_{test}	RSMS Det.	SSMS Det.	Δ Det. (pp)	RSMS ARL	SSMS ARL	Δ ARL
0.5	200	1	1	0.50	44.0%	1.7%	+42.3	175.49	199.16	-23.67
0.5	500	2	2	1.00	100.0%	58.4%	+41.6	271.52	780.95	-509.43
1.0	200	1	1	1.00	76.8%	0.6%	+76.2	163.81	199.80	-35.99
1.0	500	2	2	2.00	100.0%	71.6%	+28.4	245.59	802.41	-556.82
1.5	200	2	2	3.00	100.0%	69.3%	+30.7	99.30	325.28	-225.98
1.5	500	1	2	3.00	100.0%	9.0%	+91.0	218.72	496.10	-277.38
2.0	200	1	1	2.00	97.4%	0.0%	+97.4	154.11	200.00	-45.89
2.0	500	2	2	4.00	100.0%	88.2%	+11.8	240.64	825.29	-584.65
3.0	200	1	2	6.00	100.0%	4.0%	+96.0	88.03	199.65	-111.62
3.0	500	1	2	6.00	100.0%	0.6%	+99.4	216.93	499.81	-282.88

$(\lambda = 2.0, m = 500, T = 2, \text{mult} = 2)$, ARL equals 240.64 versus 825.29. Overall, RSMS is markedly more robust to contaminated training, delivering higher power and shorter run lengths than SSMS across a broad range of λ, m, and T.