Online Monitoring of Distributional Granger Causality:

Self-Normalized Score Monitoring and Anytime-Valid Quantile E-Processes

Jiajing Sun* Abderrahim Taamouti† Yongmiao Hong‡

March 14, 2026

Abstract

We develop a unified Phase-I/Phase-II framework for online monitoring of distributional Granger causality in time series. The object of interest is whether a candidate predictor block Z_{t-1} adds incremental predictive content for a conditional quantile, expectile, or more general elicitable functional of a response Y_t, relative to a baseline block X_{t-1}. Our general backbone is a restricted-model score architecture: estimate the baseline X-only model on a stable training sample, form excluded-block LM/score contributions for the omitted block, and monitor them sequentially. The main branch of the paper is self-normalized monitoring. It applies to quantiles, expectiles, and other elicitable score streams under weak dependence and conditional heteroskedasticity, yields pivotal Kolmogorov–Smirnov and Cramér–von Mises stopping rules, and avoids long-run variance estimation. For quantile and Value-at-Risk monitoring we add a more causality-specific refinement based on conditional calibration of exceedance hits. By betting on the hit process using predictable features extracted from Z_{t-1}, we obtain e-process detectors that remain valid under optional stopping, admit mixture, adaptive, and restart extensions, and deliver a direct tail-risk interpretation of online causality. We provide asymptotic null and local-power results for the self-normalized detectors, finite-sample anytime-valid guarantees for the e-process detectors, detailed appendicial proofs, guidance on critical-value calibration and weight selection, simulation evidence showing how the self-normalized and e-process branches complement each other in abrupt, tail-localized, gradual, and contaminated-break designs, and an hourly BTC/Deribit empirical illustration in which silent e-process benchmarks and more reactive CvM alarms imply economically different alarm-to-trade performance profiles.

Keywords: distributional Granger causality; sequential monitoring; self-normalization; quantile regression; e-process.
JEL: C12, C22, C32.

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

†Management School, University of Liverpool, L69 7ZH, UK. Email: Abderrahim.Taamouti@liverpool.ac.uk.

‡Academy of Mathematics and Systems Science, Chinese Academy of Sciences, No. 55 Zhongguancun East Road, Haidian District, Beijing 100190, China. Email: yh20@cornell.edu.

1 Introduction

Granger causality is fundamentally a statement about the predictive content of an information set for a conditional distribution (Wiener, 1956; Granger, 1969, 1980). In empirical work, however, causality is still often reduced to conditional means or low-dimensional linear dynamics. That reduction is too narrow in many macroeconomic and financial applications, where the predictive role of an omitted block of variables is strongest in the tails rather than at the center of the distribution. Offline distributional-causality methods make this point clearly. Jeong et al. (2012) propose a consistent nonparametric test for Granger non-causality in quantiles, Troster (2018) develops a quantile-regression-based test, Song and Taamouti (2021) studies location-specific quantile measures of causal strength, and Bouezmarni et al. (2024) extends the perspective to expectiles. The present paper asks how to do the same job online, after the forecaster has already been trained and deployed.

The online question is simple to state and nontrivial to answer. After a stable training period, when does a candidate predictor block Z_{t-1} begin to add incremental predictive content for a chosen quantile, expectile, or other elicitable distributional functional of Y_t, relative to a baseline model built only on X_{t-1}? And if the predictive role of Z_{t-1} later weakens or disappears, how should the researcher detect that in real time? This is the monitoring problem faced in Growth-at-Risk surveillance, Value-at-Risk backtesting, tail-spillover detection, and early-warning systems.

Three literatures are especially relevant. First, the distributional-causality literature studies offline predictability in quantiles, expectiles, and related functionals, ranging from nonparametric quantile-causality tests to regression-based tests and quantile-specific causal measures (Jeong et al., 2012; Troster, 2018; Song and Taamouti, 2021; Bouezmarni et al., 2024). Second, the sequential-monitoring literature supplies the online-detection machinery: Chu et al. (1996) gives the classic structural-change monitors, Hoga (2017) and Dette and Gösmann (2020) develop multivariate and parameter-general sequential detectors, and Chan et al. (2021), Zhu et al. (2025), Sun et al. (2026) and Sun and Hong (2026a) show how self-normalization yields pivotal KS/CvM-type monitoring rules without long-run variance tuning. Recent work such as Ghezzi et al. (2025) also emphasizes fast online detection through time-weighted boundaries. Third, the forecast-calibration and e-process literatures show that quantile forecasts can be monitored through exceedance hits, conditional calibration, and nonnegative supermartingales that remain valid under optional stopping. In that branch, Christoffersen (1998) provides the classic interval-forecast coverage perspective, Giacomini and Komunjer (2005) and Gaglianone et al. (2011) connect conditional quantile forecasts to evaluation and VaR backtesting, Gneiting and Resin (2023) sharpens the conditional-calibration diagnostics, and Henzi and Ziegel (2022), Arnold et al. (2023), Casgrain et al. (2024), Grünwald et al. (2024), and Choe and Ramdas (2024) develop sequentially valid testing and e-process tools.

Our objective is to bring these strands together in one coherent econometric framework. The paper is built around a common Phase-I/Phase-II score architecture and two complementary monitoring branches. The common architecture is classical and general: estimate the restricted X-only model on the training sample, freeze the estimator, construct the omitted-block score contribution $Z_{t-1}\psi_\tau(\cdot)$, and monitor cumulative sums of that score online. This is the paper's main

engine.

The first monitoring branch is the general one. We treat the excluded-block score stream as a weakly dependent multivariate process and monitor it by self-normalized Kolmogorov–Smirnov and Cramér–von Mises detectors. This branch is the natural backbone of the paper because it covers quantiles, expectiles, and other elicitable functionals, avoids bandwidth choices, and inherits pivotal Brownian-function limits. The second branch is more specialized and more directly tied to quantile causality. Once the baseline X-only quantile forecast has been issued, "no incremental quantile causality" can be re-expressed as conditional calibration of the hit process after observing Z_{t-1}. That insight leads to quantile e-process monitors, which are exact under optional stopping and particularly attractive for open-end monitoring.

The paper makes four contributions. First, it formulates online distributional predictive causality in a predictive-regression setting that nests VAR-style specifications but is not restricted to them, and it distinguishes clearly between the structural null of no incremental predictive content and the operational null induced by a frozen Phase-I forecaster. Second, it frames self-normalized monitoring as the general branch of the methodology and positions the e-process branch as a quantile/VaR-specific refinement rather than as an unrelated add-on. Third, it develops a unified theory: pivotal null limits and local-power results for the self-normalized statistics, and finite-sample anytime-validity for the quantile e-process branch together with mixture, adaptive, and restart extensions. Fourth, it gives implementation guidance that is explicit enough for empirical work, including how to build the omitted-block score stream, how to calibrate the KS/CvM boundaries, how to choose time weights in the CvM statistics, how to select tail weights across quantile/expectile levels, how to handle feature engineering and calendar gaps in a high-frequency crypto application, and how to translate alarms into four economically interpretable flat/short overlay rules whose performance can be compared detector by detector.

A final point is conceptual. The self-normalized statistics introduced below are not ad hoc. They sit squarely in the sequential self-normalization literature (Shao, 2010; Shao and Zhang, 2010; Shao, 2015; Chan et al., 2021; Hong et al., 2024). In particular, the Brownian functionals that determine our SSMS, RSMS, and CvM critical values are the same ones that appear in Sun and Hong (2026b), so that their critical-value simulation strategy transfers directly to the present paper after replacing factor-model scores by distributional-causality scores. At the same time, the present paper is not a mechanical extension of the factor-model setting. Our contribution is to recast that machinery for online monitoring of distributional Granger causality, where the underlying score process is built from omitted-block predictive content rather than latent-factor estimation errors, and where the null hypotheses, monitoring objectives, and economically relevant alternatives are all different. In this new setting, we show that the same class of Brownian functionals continues to govern the SSMS, RSMS, and CvM limits. This connection is useful because it allows the critical-value simulation strategy of Sun and Hong (2026b) to be carried over, but the mapping itself is nontrivial and depends on the score construction, the structural versus operational null distinction, and the quantile-, expectile-, and macro-finance specifications developed here. Framed this way, the current

paper both builds on the existing self-normalization literature and makes a distinct contribution by extending it to sequential distributional-causality monitoring.

The roadmap is deliberately cumulative. Section 2 introduces the predictive-regression setup, distinguishes the structural and operational nulls, derives the common omitted-block score process, and records the quantile, expectile, and macro-finance special cases. Section 3 then builds the monitoring rules on that score process: SSMS/RSMS/HAC KS and CvM detectors, their critical-value calibration, the admissible time weights, and the quantile e-process mixture, adaptive, and restart specifications. Section 4 states the corresponding guarantees—Brownian-function null limits, local-power behavior, and finite-sample anytime-validity—and explains how each result should be read in practice. Section 5 translates the theory into the N1–N2 size designs, the A1–A5 alternative designs, and the contaminated-training design C1, emphasizing where KS, CvM, and e-process procedures differ. Section 6 then implements the empirical pipeline on hourly BTC/Deribit data in operational detail: source construction, complete-case filtering, rolling-window testing, alarm formation, and performance evaluation across four alarm-to-trade classes. Detailed proofs, calibration tables, supplementary simulation summaries, and the full empirical strategy tables are collected in the appendix.

2 Distributional Granger causality, score construction, and special cases

This section translates online distributional Granger-causality monitoring into a score-monitoring problem. We first set up the predictive-regression environment and distinguish the structural null from the operational online null that is induced by the frozen Phase-I forecaster. We then show how both the self-normalized and e-process branches are driven by the same omitted-block score construction. The final subsection records the quantile, expectile, and macro-finance examples that will be used throughout the paper.

2.1 Predictive regression setup and null hypotheses

Let $\{(Y_t, X_{t-1}, Z_{t-1})\}_{t\in\mathbb{Z}}$ be a strictly stationary time series on $(\Omega, \mathcal{F}, \mathbb{P})$, where $Y_t \in \mathbb{R}$ is the response, $X_{t-1} \in \mathbb{R}^{p_x}$ is a baseline predictor block, and $Z_{t-1} \in \mathbb{R}^{p_z}$ is the candidate block whose incremental predictive role is monitored. We write

$$\mathcal{I}_{t-1}^{X} = \sigma(X_{t-1}, X_{t-2}, \ldots), \qquad \mathcal{I}_{t-1}^{XZ} = \sigma(X_{t-1}, Z_{t-1}, X_{t-2}, Z_{t-2}, \ldots).$$

A standard VAR-type setup is recovered by taking $X_{t-1} = (1, Y_{t-1}, \ldots, Y_{t-p})$ and Z_{t-1} to be lags of an additional process. For notational simplicity we focus on one-step-ahead targets; h-step-ahead versions follow by replacing Y_t with Y_{t+h}.

Let $\mathcal{T} \subset (0, 1)$ index a family of elicitable distributional functionals $\{T_\tau\}_{\tau\in\mathcal{T}}$, such as conditional quantiles or conditional expectiles. Elicitability means that there exists a convex loss ρ_τ and an

identification function $\psi_\tau \in \partial\rho_\tau$ such that the conditional risk $b \mapsto \mathbb{E}[\rho_\tau(Y_t - b) \mid \mathcal{I}_{t-1}]$ is minimized at $b = T_\tau(Y_t \mid \mathcal{I}_{t-1})$ and the conditional first-order condition

$$\mathbb{E}\Big[\psi_\tau\big(Y_t - T_\tau(Y_t \mid \mathcal{I}_{t-1})\big) \,\Big|\, \mathcal{I}_{t-1}\Big] = 0 \tag{1}$$

holds in the usual subgradient sense when necessary (Gneiting and Raftery, 2007; Ehm et al., 2016; Fissler and Ziegel, 2016).

We study the predictive regression

$$T_\tau(Y_t \mid \mathcal{I}_{t-1}^{XZ}) = W_{t-1}^\top \theta_0(\tau), \qquad W_{t-1} = (X_{t-1}^\top, Z_{t-1}^\top)^\top, \qquad \theta_0(\tau) = (\alpha_0(\tau)^\top, \beta_0(\tau)^\top)^\top. \tag{2}$$

The structural null of no incremental functional causality is

$$H_0^{\mathrm{str}} : \qquad \beta_0(\tau) = 0 \quad \text{for all } \tau \in \mathcal{T}. \tag{3}$$

When T_τ ranges over all quantiles or all expectiles, this is a sufficient route to distributional noncausality because those functionals characterize the conditional law.

For online monitoring it is useful to distinguish the structural null from an operational null defined relative to the issued Phase-I forecast sequence. Let $\widehat{\alpha}_m(\tau)$ be the restricted X-only estimator computed on the training sample, define

$$\widehat{m}_{t-1,\tau}^X = X_{t-1}^\top \widehat{\alpha}_m(\tau), \qquad \mathcal{H}_{t-1} = \sigma\big((Y_s, X_{s-1}, Z_{s-1}),\ s \le t-1,\ \widehat{\alpha}_m(\tau), \tau \in \mathcal{T}_m\big),$$

and write the operational null as

$$H_{0,\tau}^{\mathrm{op}} : \qquad \mathbb{E}\big[\psi_\tau\big(Y_t - \widehat{m}_{t-1,\tau}^X\big) \,\big|\, \mathcal{H}_{t-1}\big] = 0 \qquad \text{a.s. for } t \ge m+1. \tag{4}$$

In the quantile case this becomes conditional calibration of the hit process,

$$\mathbb{P}\big(Y_t \le \widehat{q}_{t-1}^X(\tau) \mid \mathcal{H}_{t-1}\big) = \tau, \qquad \widehat{q}_{t-1}^X(\tau) = X_{t-1}^\top \widehat{\alpha}_m(\tau), \tag{5}$$

which is the natural null for the e-process branch.

2.2 Phase-I/Phase-II score construction

Fix a training size m. Phase I consists of the stable training sample $t = 1, \ldots, m$, and Phase II consists of the monitoring period $t = m+1, m+2, \ldots$. We allow both closed-end monitoring up to mT observations for fixed $T > 0$ and open-end monitoring when the horizon is indefinite. Throughout the theory section we treat the open-end design formally as the case $T = \infty$, rather than merely as an informal extension.

Under the structural null, the restricted target satisfies

$$T_\tau(Y_t \mid \mathcal{I}_{t-1}^X) = X_{t-1}^\top \alpha_0(\tau). \tag{6}$$

Estimate $\alpha_0(\tau)$ once on the Phase-I sample by

$$\widehat{\alpha}_m(\tau) \in \operatorname{argmin}_{\alpha \in \mathbb{R}^{p_x}} \sum_{t=1}^{m} \rho_\tau\big(Y_t - X_{t-1}^\top \alpha\big). \tag{7}$$

The Phase-I estimator is then frozen throughout monitoring. For any $t \geq 1$ define the frozen fitted value and residual-direction score

$$\widehat{m}_{t-1,\tau} = X_{t-1}^\top \widehat{\alpha}_m(\tau), \qquad \widehat{u}_{t,\tau} = Y_t - \widehat{m}_{t-1,\tau}, \qquad \widehat{e}_{t,\tau} = \psi_\tau(\widehat{u}_{t,\tau}).$$

The excluded-block LM/score contribution is

$$g_t(\tau) = H_{t-1}\, \widehat{e}_{t,\tau} = H_{t-1}\, \psi_\tau(\widehat{u}_{t,\tau}) \in \mathbb{R}^{d_H}, \tag{8}$$

where $H_{t-1} = Z_{t-1}$ by default and, more generally, $H_{t-1} = h(Z_{t-1})$ can be a predictable feature transformation. The multiplicative role of H_{t-1} is the omitted-block score logic from LM testing: under no incremental causality, the restricted residual-direction score is orthogonal to any predictable feature built from the omitted block.

To monitor several distributional levels jointly, choose a grid $\mathcal{T}_m = \{\tau_1, \ldots, \tau_{d_\tau}\} \subset \mathcal{T}$ and stack the score blocks as

$$\Psi_t = \big(g_t(\tau_1)^\top, \ldots, g_t(\tau_{d_\tau})^\top\big)^\top \in \mathbb{R}^q, \qquad q = d_H d_\tau. \tag{9}$$

Optional tail weighting across τ is introduced by replacing $g_t(\tau_j)$ with $\omega_j g_t(\tau_j)$, where $\omega_j = \omega(\tau_j) \geq 0$ and the choice of ω is discussed in Section 3. The resulting problem is to detect a mean shift in the weakly dependent multivariate score stream $\{\Psi_t\}$.

Remark 1 (What is actually being monitored)**.** The object entering the self-normalized detectors is not the coefficient path $\beta_0(\tau)$ itself but the omitted-block score stream implied by $\beta_0(\tau) = 0$. This distinction matters. Online re-estimation of the unrestricted model at every time point and every τ would be computationally expensive and would obscure the link with classical score testing. Monitoring the excluded-block score keeps the update cost low and makes the general self-normalized branch work uniformly for quantiles, expectiles, and other elicitable functionals.

2.3 Special cases

The two flagship special cases are quantiles/VaR and expectiles. They are written below as examples because the formulas will be referred to repeatedly in the monitoring and theory sections.

Example 1 (Quantile and Value-at-Risk causality)**.** For $\tau \in (0,1)$, the conditional τ-quantile is

$$Q_\tau(Y_t \mid \mathcal{I}_{t-1}) = \inf\{y : \mathbb{P}(Y_t \le y \mid \mathcal{I}_{t-1}) \ge \tau\}.$$

Quantile regression is based on the check loss (Koenker and Bassett, 1978; Koenker, 2005)

$$\rho_\tau^Q(u) = u\big(\tau - \mathbf{1}\{u \le 0\}\big), \qquad \psi_\tau^Q(u) = \tau - \mathbf{1}\{u \le 0\}. \tag{10}$$

Therefore the excluded-block score is

$$g_t(\tau) = H_{t-1}\Big(\tau - \mathbf{1}\{Y_t \le X_{t-1}^\top \widehat{\alpha}_m(\tau)\}\Big). \tag{11}$$

If the response is a loss L_{t+1} and $\tau = 1-\alpha$, then (11) is exactly the online score for Value-at-Risk monitoring at tail probability α. The crucial simplification is that the primitive random object becomes the hit indicator $I_t(\tau) = \mathbf{1}\{Y_t \le \widehat{q}_{t-1}^X(\tau)\}$, which is why the e-process branch can be built from Bernoulli likelihood ratios.

Example 2 (Expectile causality)**.** For $\tau \in (0,1)$, the conditional τ-expectile is (Newey and Powell, 1987)

$$\mu_\tau(Y_t \mid \mathcal{I}_{t-1}) = \operatorname{argmin}_{b\in\mathbb{R}} \mathbb{E}\big[\big|\tau - \mathbf{1}\{Y_t \le b\}\big|(Y_t - b)^2 \mid \mathcal{I}_{t-1}\big]. \tag{12}$$

The asymmetric-squares loss and derivative are

$$\rho_\tau^E(u) = \big|\tau - \mathbf{1}\{u \le 0\}\big|u^2, \qquad \psi_\tau^E(u) = 2u\big|\tau - \mathbf{1}\{u \le 0\}\big|. \tag{13}$$

Hence the excluded-block score becomes

$$g_t(\tau) = 2H_{t-1}\,\widehat{u}_{t,\tau}\,\big|\tau - \mathbf{1}\{\widehat{u}_{t,\tau} \le 0\}\big|. \tag{14}$$

Unlike the quantile case, the expectile score depends on the magnitude of the residual as well as its sign. This smoothness is one reason expectile regressions are often easier to compute and analyze offline (Bouezmarni et al., 2024). Online, however, the same feature means that a finite-sample exact e-process is harder to obtain; the general score-based e-process requires an additional conditional moment generating function bound.

Example 3 (A macro-finance application template)**.** Suppose Y_{t+h} is h-step-ahead GDP growth, an asset loss, or a systemic-stress index. Let X_t collect standard controls such as lags, volatility measures, or common factors, and let Z_t collect the candidate driver of interest, for example policy surprises, funding stress, or external tail-risk measures. Then a distributional predictive regression takes the form

$$T_\tau(Y_{t+h} \mid \mathcal{I}_t) = X_t^\top \alpha(\tau) + Z_t^\top \beta(\tau),$$

and online noncausality monitoring amounts to testing whether $\beta(\tau)$ switches on after the training sample. For quantile Growth-at-Risk one would emphasize low quantiles, say $\mathcal{T}_m = \{0.05, 0.10, 0.20\}$,

whereas for systemic-stress monitoring one would emphasize high quantiles or high expectiles, for example $\mathcal{T}_m = \{0.90, 0.95, 0.99\}$. The same self-normalized and e-process machinery applies once the score stream is formed.

3 Online monitoring procedures

The procedures in this section all start from the same stacked score stream $\{\Psi_t\}$ in (9). The self-normalized branch is the paper's general branch and is the default recommendation for quantiles, expectiles, and other elicitable functionals. The quantile e-process branch is more specialized: it uses the calibration structure of the hit process and is particularly attractive for open-end monitoring. Because several empirical applications run without a pre-specified terminal date, we now write the monitoring rules in a way that covers both finite horizons $T < \infty$ and the formal open-end benchmark $T = \infty$.

3.1 Self-normalized KS/CvM detectors

Let

$$\bar{\Psi}_m = \frac{1}{m}\sum_{t=1}^{m}\Psi_t, \qquad \phi_t = \Psi_t - \bar{\Psi}_m, \qquad S_m(k) = \sum_{t=m+1}^{m+k}\phi_t, \qquad k \geq 1. \tag{15}$$

The detector uses ϕ_t because the training sample supplies the natural estimate of the pre-change mean of the score stream. This centering also makes the framework suitable for both switch-on and switch-off alternatives.

3.1.1 SSMS

Quadratic self-normalization uses the training partial sums of ϕ_t to form

$$D_m = \frac{1}{m^2}\sum_{t=1}^{m}\left(\sum_{j=1}^{t}\phi_j\right)\left(\sum_{j=1}^{t}\phi_j\right)^{\top}. \tag{16}$$

The corresponding KS-type statistic is

$$\mathcal{M}_m^{\mathrm{S}}(k) = \frac{S_m(k)^{\top} D_m^{-1} S_m(k)}{m(1+k/m)^2\left(\frac{k}{k+m}\right)^{2\gamma}}, \qquad \gamma \in [0, 1/2), \tag{17}$$

with stopping time

$$\mathcal{T}_m^{\mathrm{S}} = \inf\left\{1 \leq k \leq mT : \ \mathcal{M}_m^{\mathrm{S}}(k) > c_\alpha^{\mathrm{S}}(T, q, \gamma)\right\}. \tag{18}$$

3.1.2 RSMS

Adjusted-range self-normalization is designed for KS-type monitoring and often improves power in practice, but it is coordinatewise and therefore works best when the long-run covariance of the

working score is diagonal or has first been diagonalized. Using the training window, compute a lag-0 covariance estimate,

$$\widehat{\Sigma}_{\phi,0} = \frac{1}{m}\sum_{t=1}^{m}(\phi_t - \bar{\phi}_m)(\phi_t - \bar{\phi}_m)^\top, \qquad \bar{\phi}_m = \frac{1}{m}\sum_{t=1}^{m}\phi_t,$$

let $\widehat{\Sigma}_{\phi,0} = \widehat{Q}\widehat{D}\widehat{Q}^\top$ be its eigen-decomposition, and define the whitening matrix

$$\widehat{H}_m = (\widehat{D} + \eta I_q)^{-1/2}\widehat{Q}^\top, \tag{19}$$

with a small ridge $\eta > 0$. Set $\widetilde{\phi}_t = \widehat{H}_m\phi_t$ and $\widetilde{S}_m(k) = \sum_{t=m+1}^{m+k}\widetilde{\phi}_t$. For each coordinate ℓ, define the training bridge

$$\widetilde{B}_{m,\ell}(t) = \sum_{j=1}^{t}\widetilde{\phi}_{j,\ell} - \frac{t}{m}\sum_{j=1}^{m}\widetilde{\phi}_{j,\ell}, \qquad t = 1,\dots,m,$$

and the adjusted range

$$\widetilde{R}_{m,\ell} = m^{-1/2}\Big(\max_{1\le t\le m}\widetilde{B}_{m,\ell}(t) - \min_{1\le t\le m}\widetilde{B}_{m,\ell}(t)\Big), \qquad \widetilde{R}_m = \mathrm{diag}(\widetilde{R}_{m,1},\dots,\widetilde{R}_{m,q}). \tag{20}$$

The RSMS statistic is then

$$\widetilde{\mathcal{M}}^{\mathrm{R}}_m(k) = \frac{\widetilde{S}_m(k)^\top\widetilde{R}_m^{-2}\widetilde{S}_m(k)}{m(1+k/m)^2\left(\frac{k}{k+m}\right)^{2\gamma}}, \tag{21}$$

with stopping time

$$\mathcal{T}^{\mathrm{R}}_m = \inf\Big\{1 \le k \le mT :\ \widetilde{\mathcal{M}}^{\mathrm{R}}_m(k) > c^{\mathrm{R}}_\alpha(T,q,\gamma)\Big\}.$$

3.1.3 CvM

To accumulate evidence over monitoring time, define for $k = 1,\dots,mT$

$$\mathcal{I}^{\mathrm{S}}_m(k) = \frac{1}{m}\sum_{j=1}^{k} w\left(\frac{j}{m}\right)\mathcal{M}^{\mathrm{S}}_m(j), \tag{22}$$

$$\mathcal{I}^{\mathrm{R}}_m(k) = \frac{1}{m}\sum_{j=1}^{k} w\left(\frac{j}{m}\right)\widetilde{\mathcal{M}}^{\mathrm{R}}_m(j), \tag{23}$$

with corresponding stopping times

$$\mathcal{T}^{\mathrm{S}}_{m,\mathrm{CvM}} = \inf\Big\{1 \le k \le mT :\ \mathcal{I}^{\mathrm{S}}_m(k) > c^{\mathrm{S}}_{\alpha,\mathrm{CvM}}(T,q;w)\Big\},$$

$$\mathcal{T}^{\mathrm{R}}_{m,\mathrm{CvM}} = \inf\Big\{1 \le k \le mT :\ \mathcal{I}^{\mathrm{R}}_m(k) > c^{\mathrm{R}}_{\alpha,\mathrm{CvM}}(T,q;w)\Big\}.$$

Throughout the CvM theory we set $\gamma = 0$, which is both the cleanest case and the practically recommended one.

For open-end monitoring we use the same score constructions but remove the terminal horizon. The KS-type stopping rules become

$$\mathcal{T}^{\mathrm{S}}_{m,\infty} = \inf\Big\{k \geq 1 :\ \mathcal{M}^{\mathrm{S}}_m(k) > c^{\mathrm{S}}_\alpha(\infty, q, \gamma)\Big\}, \qquad \mathcal{T}^{\mathrm{R}}_{m,\infty} = \inf\Big\{k \geq 1 :\ \widetilde{\mathcal{M}}^{\mathrm{R}}_m(k) > c^{\mathrm{R}}_\alpha(\infty, q, \gamma)\Big\},$$

and

$$\mathcal{T}^{\mathrm{H}}_{m,\infty} = \inf\Big\{k \geq 1 :\ \mathcal{M}^{\mathrm{H}}_m(k) > c^{\mathrm{H}}_\alpha(\infty, q, \gamma)\Big\}.$$

For the open-end CvM rules, let $w_\infty : [0,\infty) \to \mathbb{R}_+$ be an integrable monitoring weight and define

$$\mathcal{I}^{\bullet}_{m,\infty}(k) = \frac{1}{m}\sum_{j=1}^{k} w_\infty\left(\frac{j}{m}\right)\mathcal{M}^{\bullet}_m(j), \qquad \bullet \in \{\mathrm{S}, \mathrm{R}, \mathrm{H}\}.$$

The corresponding stopping times are

$$\mathcal{T}^{\bullet}_{m,\infty,\mathrm{CvM}} = \inf\Big\{k \geq 1 :\ \mathcal{I}^{\bullet}_{m,\infty}(k) > c^{\bullet}_{\alpha,\mathrm{CvM}}(\infty, q; w_\infty)\Big\}.$$

The open-end critical values are again pivotal. The only extra design restriction is that w_∞ must be integrable, because otherwise the CvM detector accumulates nonvanishing null evidence forever and eventually produces a false alarm with probability one.

Remark 2 (Choosing the CvM time weight)**.** It is useful to separate the mathematically admissible weight class from the practically recommended weights. For closed-end monitoring on $[0, T]$ we recommend normalizing all weights so that $\int_0^T w(s)\,ds = T$; this keeps the overall scale of the CvM detector comparable across choices. Writing $r = s/T \in [0, 1]$, the four benchmark weights used in the appendix are

$$w_U(s) \equiv 1, \qquad w_{\mathrm{Early}}(s) = 2(1-r), \qquad w_{\mathrm{Late}}(s) = 2r, \qquad w_{\mathrm{Mid}}(s) = 6r(1-r).$$

The uniform weight w_U is the default omnibus choice when the break date is genuinely unknown. The weight w_{Early} is ideal when one expects predictive content to switch on soon after the training sample, for example after a policy regime change or a crisis onset. The weight w_{Late} is more suitable when the main concern is a delayed break over a fixed horizon. The hump-shaped weight w_{Mid} is useful when one wants power against sustained changes without giving too much influence to the very beginning or the very end of the monitoring period. For open-end monitoring it is useful to preserve the same four labels after reparameterizing time by $x = s/(1+s) \in [0, 1)$. The resulting admissible s-space weights are

$$w_{\infty,U}(s) = \frac{1}{(1+s)^2}, \qquad w_{\infty,\mathrm{Early}}(s) = \frac{2}{(1+s)^3},$$
$$w_{\infty,\mathrm{Late}}(s) = \frac{2s}{(1+s)^3}, \qquad w_{\infty,\mathrm{Mid}}(s) = \frac{6s}{(1+s)^4}.$$

After the change of variables $x = s/(1+s)$ these become exactly the same templates 1, $2(1-x)$,

$2x$, and $6x(1-x)$ on $[0,1)$, which is why they are convenient for both theory and simulation. In quantile work the best weight often depends on where the tail signal is expected to emerge; in expectile work, where the score is smoother and more diffuse, the uniform or middle weight is often the safer benchmark.

Remark 3 (Choosing the cross-τ weight)**.** The cross-τ weight $\omega(\tau)$ determines where in the conditional distribution evidence is emphasized, so it should be chosen with the target application in mind rather than mechanically. For quantiles, a natural benchmark is the variance-stabilizing envelope

$$\omega_Q(\tau) \propto \{\tau(1-\tau)\}^{-1/2}, \qquad \tau \in [\underline{\tau}, \overline{\tau}],$$

because the primitive hit score has variance $\tau(1-\tau)$. This is a good default when one wants an omnibus monitor over a trimmed interval. If the application is downside-risk monitoring, a more targeted choice is

$$\omega_Q^L(\tau) \propto \tau^{-a}\mathbf{1}\{\underline{\tau} \le \tau \le \tau_0\}, \qquad a \in [0, 1/2],$$

which deliberately overweights the lower tail while keeping the weight locally integrable after trimming. For upper-tail surveillance one uses the symmetric counterpart near one. For expectiles, aggressive tail weighting is less attractive because the score depends on residual magnitude and can become unstable in the extremes. A better default is a trimmed bounded envelope or a Phase-I variance standardization such as

$$\omega_E(\tau) \propto \big(\widehat{\mathrm{Var}}_m(g_t(\tau)) \vee c_0\big)^{-1/2},$$

with a small floor $c_0 > 0$ to avoid exploding weights. In short, quantile monitoring often benefits from tail-directed $\omega(\tau)$, whereas expectile monitoring is usually better served by a conservative trimmed grid and mild variance stabilization.

3.2 HAC benchmark, critical values, and practical implementation

Self-normalization is the paper's preferred general solution because it removes long-run variance tuning from the main detector. Still, readers in econometrics often want a familiar studentized benchmark. Let $\Omega = \sum_{\ell=-\infty}^{\infty} \Gamma_\ell$ denote the long-run covariance of $\{\phi_t\}$, where $\Gamma_\ell = \mathbb{E}(\phi_t \phi_{t-\ell}^\top)$. A Phase-I HAC estimator is

$$\widehat{\Gamma}_\ell = \frac{1}{m} \sum_{t=\ell+1}^{m} \phi_t \phi_{t-\ell}^\top, \qquad \ell = 0, 1, \ldots, L_m, \tag{24}$$

$$\widehat{\Omega}_m = \widehat{\Gamma}_0 + \sum_{\ell=1}^{L_m} K\!\left(\frac{\ell}{L_m}\right) \big(\widehat{\Gamma}_\ell + \widehat{\Gamma}_\ell^\top\big),$$

where $K(\cdot)$ is a kernel and L_m is a bandwidth. The HAC KS-type statistic is

$$\mathcal{M}_m^{\mathrm{H}}(k) = \frac{S_m(k)^\top \widehat{\Omega}_m^{-1} S_m(k)}{m(1+k/m)^2 \left(\frac{k}{k+m}\right)^{2\gamma}}, \tag{25}$$

and the HAC CvM statistic is

$$\mathcal{I}_m^{\mathrm{H}}(k) = \frac{1}{m} \sum_{j=1}^{k} w\left(\frac{j}{m}\right) \mathcal{M}_m^{\mathrm{H}}(j). \tag{26}$$

The corresponding stopping rules use critical values $c_\alpha^{\mathrm{H}}(T, q, \gamma)$ and $c_{\alpha,\mathrm{CvM}}^{\mathrm{H}}(T, q; w)$.

Remark 4 (Critical values and the Sun–Hong calibration)**.** The self-normalized null limits in Section 4 are Brownian functionals whose arguments are only (q, T, γ) for KS-type rules and (q, T, w) for CvM-type rules. Hence their critical values are universal once the score dimension, horizon, boundary exponent, and weight function are fixed. The same Brownian-function calibration problem appears in Sun and Hong (2026b). Our contribution is not to invent a new critical-value algorithm, but to show that the very same calibration carries over after replacing factor-model scores by distributional-causality scores. Appendix D reproduces the tabulated values for $q \leq 10$ that are most useful in practice. When an empirical application requires a different q or a non-tabulated horizon T, one simply re-simulates the corresponding Brownian functional.

The practical default is therefore: report SSMS and RSMS as the main detectors, keep HAC as a benchmark, use $w \equiv 1$ unless there is a clear scientific reason to favor early or late changes, and report sensitivity to a small set of γ values such as $\gamma \in \{0, 0.15\}$ for the KS rules.

3.3 Anytime-valid quantile e-process detectors

The self-normalized statistics treat the excluded-block score stream as a generic mean-shift process. For quantiles and VaR there is an alternative viewpoint that is more tightly tied to causality. Once the baseline X-only quantile forecast has been issued, incremental quantile causality of Z_{t-1} is equivalent to asking whether the hit of that issued forecast remains conditionally calibrated after observing Z_{t-1}.

Fix $\tau \in \mathcal{T}_m$ and define the issued forecast and hit process by

$$\widehat{q}_{t-1}^{X}(\tau) = X_{t-1}^\top \widehat{\alpha}_m(\tau), \qquad I_t(\tau) = \mathbf{1}\{Y_t \leq \widehat{q}_{t-1}^{X}(\tau)\}, \qquad t \geq m+1.$$

Under the operational quantile null (5),

$$\mathbb{E}\left[I_t(\tau) - \tau \mid \mathcal{H}_{t-1}\right] = 0. \tag{27}$$

Because the quantile score is $\psi_\tau^Q(u) = \tau - \mathbf{1}\{u \leq 0\}$, the excluded-block score is simply $g_t(\tau) = H_{t-1}(\tau - I_t(\tau))$. The self-normalized and e-process branches therefore start from the same primitive

Algorithm 1 General online monitoring algorithm: SN-SSMS, SN-RSMS, or HAC

Require: Training size m; monitoring horizon mT or open-end horizon; grid $\mathcal{T}_m = \{\tau_1, \ldots, \tau_{d_\tau}\}$; instrument $H_{t-1} = h(Z_{t-1})$; KS exponent γ; CvM time weight w; studentization choice (SSMS, RSMS, or HAC).

Phase I: For each τ_j, compute $\widehat{\alpha}_m(\tau_j)$ from (7).
for $t = 1, \ldots, m$ **do**
    Compute $g_t(\tau_j)$ from (8) and stack Ψ_t as in (9).
end for
Compute $\bar{\Psi}_m$, $\phi_t = \Psi_t - \bar{\Psi}_m$, and the Phase-I studentizer: D_m for SSMS, $(\widehat{H}_m, \widetilde{R}_m)$ for RSMS, or $\widehat{\Omega}_m$ for HAC.
Initialize $S_m(0) = 0$ and, for CvM rules, $\mathcal{I}_m(0) = 0$.
for $k = 1, 2, \ldots$ until horizon or alarm **do**
    Observe $(Y_{m+k}, X_{m+k-1}, Z_{m+k-1})$.
    Compute the new score block Ψ_{m+k} using the frozen Phase-I estimates and set $\phi_{m+k} = \Psi_{m+k} - \bar{\Psi}_m$.
    Update $S_m(k) = S_m(k-1) + \phi_{m+k}$.
    Evaluate the chosen KS-type statistic: SSMS via (17), RSMS via (21), or HAC via (25).
    If a CvM rule is used, update $\mathcal{I}_m(k) = \mathcal{I}_m(k-1) + m^{-1} w(k/m) \mathcal{M}(k)$.
    if the relevant statistic crosses its critical value **then**
        Stop and report alarm time $m + k$.
    end if
end for

surprise $I_t(\tau) - \tau$.

Choose a predictable feature vector $s_{t-1} = h(Z_{t-1}) \in \mathbb{R}^d$. A convenient alternative model is the logistic tilt

$$\operatorname{logit}\big(p_t(\theta, \tau)\big) = \operatorname{logit}(\tau) + \theta^\top s_{t-1}, \qquad \theta \in \mathbb{R}^d, \tag{28}$$

that is,

$$p_t(\theta, \tau) = \frac{\tau e^{\theta^\top s_{t-1}}}{1 - \tau + \tau e^{\theta^\top s_{t-1}}}. \tag{29}$$

For fixed predictable θ, the one-step likelihood ratio is

$$L_t(\theta, \tau) = \left(\frac{p_t(\theta, \tau)}{\tau}\right)^{I_t(\tau)} \left(\frac{1 - p_t(\theta, \tau)}{1 - \tau}\right)^{1 - I_t(\tau)}. \tag{30}$$

Its cumulative product

$$\mathcal{E}_k(\theta, \tau) = \prod_{t=m+1}^{m+k} L_t(\theta, \tau), \qquad \mathcal{E}_0(\theta, \tau) = 1, \tag{31}$$

forms the basic e-process detector.

To make the detector practically useful, we allow four extensions. First, for a finite dictionary

Algorithm 2 Online quantile/VaR monitoring via e-processes

Require: Training size m; false-alarm level α; quantile grid $\{\tau_1, \ldots, \tau_J\}$; feature map $s_{t-1} = h(Z_{t-1})$; either a finite dictionary Θ with mixing weights or a predictable update rule; optional restart set $\mathcal{R}$; optional tail weights ω_j.

Phase I: Estimate the restricted quantile regressions and freeze the forecast rule $\widehat{q}^X_{t-1}(\tau_j) = X_{t-1}^\top \widehat{\alpha}_m(\tau_j)$.
Initialize all component e-processes at one.
for $t = m+1, m+2, \ldots$ **do**
  for $j = 1, \ldots, J$ **do**
    Observe the hit $I_t(\tau_j) = \mathbf{1}\{Y_t \le \widehat{q}^X_{t-1}(\tau_j)\}$.
    Compute the one-step factor from (30) for each active direction or adaptive parameter.
    Update the corresponding product, mixture, restart-bank, and tail-mixture e-processes.
  end for
  if the chosen aggregated e-process exceeds α^{-1} **then**
    Stop and report alarm time t.
  end if
end for

$\Theta = \{\theta^{(1)}, \ldots, \theta^{(L)}\}$ and weights $\pi_\ell \ge 0$ with $\sum_\ell \pi_\ell = 1$, define the mixture

$$\mathcal{E}_k^{\mathrm{mix}}(\tau) = \sum_{\ell=1}^{L} \pi_\ell \, \mathcal{E}_k\big(\theta^{(\ell)}, \tau\big). \tag{32}$$

Second, predictable adaptive betting is permitted: if θ_{t-1} is $\mathcal{H}_{t-1}$-measurable, then

$$\mathcal{E}_k^{\mathrm{ad}}(\tau) = \prod_{t=m+1}^{m+k} L_t(\theta_{t-1}, \tau) \tag{33}$$

remains valid. A simple update is the projected online-logistic step

$$\theta_t = \mathrm{Proj}_\Theta \Big(\theta_{t-1} + \eta_t\big(I_t(\tau) - p_t(\theta_{t-1}, \tau)\big)s_{t-1}\Big). \tag{34}$$

Third, late-onset changes can be handled by a restart bank. For restart times $r \in \mathcal{R}$ with weights $\pi_r \ge 0$, define

$$\mathcal{E}_k^{\mathrm{bank}}(\tau) = \sum_{r \in \mathcal{R}} \pi_r \, \mathcal{E}_{r:k}(\tau), \tag{35}$$

where $\mathcal{E}_{r:k}(\tau)$ starts multiplying one-step factors only from time $m + r$. Fourth, for a quantile grid $\{\tau_1, \ldots, \tau_J\}$ and tail weights $\omega_j \ge 0$ with $\sum_j \omega_j = 1$, define

$$\mathcal{E}_k^{\mathrm{tail}} = \sum_{j=1}^{J} \omega_j \, \mathcal{E}_k(\tau_j). \tag{36}$$

The e-process branch is quantile-specific in its cleanest exact form. For expectiles and more

general elicitable functionals, a score-based exponential supermartingale is still possible when the score projection satisfies a conditional moment generating function bound. The formal statement is postponed to Section 4.

4 Theoretical properties

This section states assumptions, the null limits for the self-normalized and HAC statistics, the power results, and the anytime-validity statements for the e-process branch. It now treats both closed-end monitoring with $T < \infty$ and the formal open-end benchmark $T = \infty$. Because the methodology has two branches, it is useful to separate what is asymptotic from what is finite-sample: the self-normalized and HAC procedures rely on weak convergence, whereas the quantile e-process branch is valid in finite samples under the operational calibration null.

Assumption 1 (Stable training window)**.** The Phase-I window is change-free: there exists a time-invariant coefficient function $\alpha_0(\tau)$ such that

$$T_\tau(Y_t \mid \mathcal{I}_{t-1}^X) = X_{t-1}^\top \alpha_0(\tau), \qquad t = 1, \dots, m, \qquad \tau \in \mathcal{T},$$

and the omitted-block score is centered on the training segment,

$$\mathbb{E}\Big[H_{t-1}\, \psi_\tau\big(Y_t - X_{t-1}^\top \alpha_0(\tau)\big)\Big] = 0, \qquad t = 1, \dots, m, \qquad \tau \in \mathcal{T}.$$

Assumption 1 does not assert that the no-causality null must hold forever. It only says that the window used to estimate and freeze the restricted model is stable. This is indispensable: the self-normalizer and the training mean should represent the pre-change benchmark, not a mixture of regimes.

Assumption 2 (Sequential invariance principle under H_0)**.** Under H_0, there exists a positive definite matrix $\Omega \in \mathbb{R}^{q\times q}$ and a standard q-dimensional Brownian motion $\{B_q(r) : 0 \le r \le 1+T\}$ such that

$$\left\{ \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor mr \rfloor} (\Psi_t - \mu) \right\}_{0\le r\le 1+T} \Rightarrow \Big\{\Omega^{1/2} B_q(r)\Big\}_{0\le r\le 1+T} \quad \text{in } D([0, 1+T], \mathbb{R}^q),$$

where $\mu = \mathbb{E}(\Psi_t)$, typically equal to zero under exact noncausality.

This is the high-level weak-dependence condition behind all self-normalized limits. It summarizes stationarity, finite second moments, and a functional central limit theorem for the stacked score process. We state it directly on Ψ_t because doing so covers quantile, expectile, and other elicitable scores in one line. In concrete models it can be verified under standard mixing or physical-dependence conditions. For quantiles, conditional densities around the target level are additionally needed; for expectiles, finite second moments and differentiability of the criterion are typically enough. When

the monitoring horizon is open-ended, this assumption is interpreted on every compact interval $[0, 1+L]$ with fixed $L < \infty$. That is the standard route for passing from fixed-horizon functional central limit theorems to the $T = \infty$ benchmark.

Assumption 3 (Feasible-oracle equivalence at the partial-sum level)**.** Let $\Psi_t^{(0)}$ denote the infeasible oracle stacked score built with the population restricted coefficient $\alpha_0(\tau)$ in place of $\widehat{\alpha}_m(\tau)$. Then

$$\sup_{0 \le r \le 1+T} \left\| \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor mr \rfloor} \left(\Psi_t - \Psi_t^{(0)} \right) \right\| = o_p(1).$$

Moreover, for each studentizer used in the paper, the feasible and oracle versions are asymptotically equivalent:

$$\left\| D_m - D_m^{(0)} \right\| = o_p(1), \qquad \left\| \widetilde{R}_m - \widetilde{R}_m^{(0)} \right\| = o_p(1), \qquad \left\| \widehat{\Omega}_m - \widehat{\Omega}_m^{(0)} \right\| = o_p(1).$$

Assumption 3 is stronger than a pointwise score approximation and that is exactly what the monitoring problem needs. The statistics depend on partial sums of the score stream and on self-normalizers constructed from the entire training path, so the feasible-oracle gap must be negligible at that level. For quantile regression this type of result comes from Bahadur expansions plus smoothness of the conditional density around the target quantiles; for expectile regression it follows from standard smooth M-estimation expansions. Economically, the assumption says that once the Phase-I estimator is frozen, the online fluctuations are dominated by new incoming observations rather than by first-stage estimation noise. For the $T = \infty$ benchmark the same negligibility is required uniformly on every compact interval $[0, 1+L]$, which is the natural companion to the compact-interval invariance principle just noted.

Assumption 4 (Diagonalization for RSMS)**.** Either (i) the long-run covariance Ω in Assumption 2 is diagonal, or (ii) after the Phase-I whitening step in (19), the transformed score stream has diagonal long-run covariance and the whitening error is asymptotically negligible at the $\sqrt{m}$ scale.

RSMS normalizes each coordinate separately by its adjusted range. That is why RSMS is powerful and robust, but also why some form of diagonalization is needed. The assumption is mild in practice because the whitening step is estimated once from the stable training sample.

Assumption 5 (HAC consistency)**.** Let $\widehat{\Omega}_m$ be the HAC estimator in (24). Then $\widehat{\Omega}_m \xrightarrow{p} \Omega$ in probability.

The HAC benchmark is conceptually straightforward but pays the usual price of kernel and bandwidth choice. Assumption 5 is standard and is imposed only for the benchmark branch, not for the self-normalized branch.

Assumption 6 (CvM weights)**.** For fixed closed-end horizon $T < \infty$, the weight function $w : [0, T] \to \mathbb{R}_+$ is bounded and Riemann integrable. For open-end monitoring, the weight $w_\infty : [0, \infty) \to \mathbb{R}_+$ is nonnegative, bounded on compact sets, and integrable over $[0, \infty)$.

This assumption is not restrictive; it simply formalizes the idea that the CvM detector should accumulate finite weighted evidence. The interesting econometric question is not whether w exists but how it should be chosen. Remark 2 explains that choice.

4.1 Pivotal null limits and critical values

Define the Brownian bridge and monitoring-increment processes

$$B_q^0(r) = B_q(r) - rB_q(1), \qquad r \in [0,1],$$

$$U_q(s) = B_q(1+s) - (1+s)B_q(1), \qquad s \in [0,T],$$

and the Brownian-bridge quadratic form

$$V = \int_0^1 B_q^0(r) B_q^0(r)^\top \, dr. \tag{37}$$

For RSMS, define

$$R = \operatorname{diag}(R_1, \ldots, R_q), \qquad R_\ell = \sup_{0\le r\le 1} B_\ell^0(r) - \inf_{0\le r\le 1} B_\ell^0(r). \tag{38}$$

The matrix V is almost surely positive definite and each R_ℓ is almost surely positive.

Theorem 1 (HAC KS-type null limit)**.** *Under Assumptions 1–5, for $T < \infty$ and $\gamma \in [0, 1/2)$,*

$$\sup_{1\le k\le mT} \mathcal{M}_m^{\mathrm{H}}(k) \xrightarrow{d} \sup_{0<s\le T} \frac{\|U_q(s)\|_2^2}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}}.$$

Theorem 1 is the benchmark result: once the long-run covariance is consistently estimated, the studentized monitoring statistic converges to a nuisance-free Brownian functional. The cost is that the benchmark relies on a tuning parameter in $\widehat{\Omega}_m$.

Theorem 2 (SSMS KS-type null limit)**.** *Under Assumptions 1–3, for $T < \infty$ and $\gamma \in [0, 1/2)$,*

$$\sup_{1\le k\le mT} \mathcal{M}_m^{\mathrm{S}}(k) \xrightarrow{d} \sup_{0<s\le T} \frac{U_q(s)^\top V^{-1} U_q(s)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}}.$$

Theorem 2 is the main self-normalized result. The long-run covariance Ω disappears from the limit because the training bridge normalizer D_m absorbs it. That is the formal reason SSMS is attractive in online work: pivotality is obtained without long-run variance estimation.

Theorem 3 (RSMS KS-type null limit)**.** *Under Assumptions 1–4, for $T < \infty$ and $\gamma \in [0, 1/2)$,*

$$\sup_{1\le k\le mT} \widetilde{\mathcal{M}}_m^{\mathrm{R}}(k) \xrightarrow{d} \sup_{0<s\le T} \frac{U_q(s)^\top R^{-2} U_q(s)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}}.$$

Theorem 3 shows that RSMS is just as pivotal as SSMS after diagonalization, but the self-normalizer is now built from coordinatewise ranges rather than a quadratic form. This is precisely the channel through which RSMS often gains power in KS-type monitoring.

Theorem 4 (HAC CvM-type null limit)**.** *Under the conditions of Theorem 1 and Assumption 6, with* $\gamma = 0$ *and* $T < \infty$*,*

$$\mathcal{I}_m^{\mathrm{H}}(\lfloor ms \rfloor) \xrightarrow{d} \int_0^s w(u) \frac{\|U_q(u)\|_2^2}{(1+u)^2}\, du, \qquad s \in [0, T].$$

The HAC CvM detector accumulates the same studentized evidence over time. Relative to the KS detector it can be more sensitive to gradual or late changes, but it inherits the HAC tuning issue.

Theorem 5 (CvM-SSMS null limit)**.** *Under Assumptions 1–6, with* $\gamma = 0$ *and* $T < \infty$*,*

$$\mathcal{I}_m^{\mathrm{S}}(\lfloor ms \rfloor) \xrightarrow{d} \int_0^s w(u) \frac{U_q(u)^\top V^{-1} U_q(u)}{(1+u)^2}\, du, \qquad s \in [0, T].$$

Theorem 5 shows that the self-normalized CvM detector preserves pivotality while allowing the researcher to emphasize early, middle, or late monitoring times through w.

Theorem 6 (CvM-RSMS null limit)**.** *Under the conditions of Theorem 3 and Assumption 6, with* $\gamma = 0$ *and* $T < \infty$*,*

$$\mathcal{I}_m^{\mathrm{R}}(\lfloor ms \rfloor) \xrightarrow{d} \int_0^s w(u) \frac{U_q(u)^\top R^{-2} U_q(u)}{(1+u)^2}\, du, \qquad s \in [0, T].$$

Theorem 6 is the integrated companion to Theorem 3. It is useful when the researcher expects diffuse but persistent distributional changes rather than a sharp jump that is easiest to catch with a supremum.

Remark 5 (Critical values and weight choice)**.** Theorems 1–6 and Theorems 7–8 imply that all limiting critical values are nuisance-free. What remains is a design choice rather than a nuisance-parameter problem: one chooses (T, γ, w, ω) to match the detection problem of interest. The appendix makes this concrete by reporting Brownian critical values for $q \leq 10$. The finite-horizon tables report $T \in \{1, 2, 5, 10\}$; for KS the open-end values are appended as $T = \infty$, whereas for CvM the open-end values are reported separately because the admissible open-end weights differ from the finite-horizon weights. The first four horizon columns are reproduced from Sun and Hong (2026b), and the open-end entries are generated from the same Brownian architecture after the reparameterization in Proposition 1. In applied quantile work a good starting point is $(w, \omega) = (w_U, \omega_Q)$ on a trimmed lower-tail grid when the break date is unknown, then checking sensitivity to w_{Early} and w_{Late} if the timing of the break is substantively important. In applied expectile work a more conservative default is $(w, \omega) = (w_U, \omega_E)$ or $(w_{\mathrm{Mid}}, \omega_E)$ on a trimmed grid, because expectile scores react more strongly to magnitude outliers in the extreme tails.

4.2 Open-end monitoring and the case $T = \infty$

The finite-horizon results above are the operational default, but in many monitoring applications it is also useful to understand the formal open-end benchmark in which the monitoring system is allowed to run indefinitely. The open-end case is not obtained by simply replacing T with an arbitrarily large number. For KS rules one must control the Brownian tail behavior as $s \to \infty$, and for CvM rules one must choose weights that keep the accumulated null evidence finite. The next proposition and two theorems make those points explicit.

Proposition 1 (Open-end reparameterization)**.** *Let*

$$x = \frac{s}{1+s} \in [0,1), \qquad s = \frac{x}{1-x},$$

and define

$$W_q(x) = \frac{U_q\left(\frac{x}{1-x}\right)}{1+\frac{x}{1-x}} = (1-x)\, U_q\left(\frac{x}{1-x}\right), \qquad x \in [0,1).$$

Then W_q is a standard q-dimensional Brownian motion on $[0,1)$ and is independent of the training bridge $\{B_q^0(r) : 0 \le r \le 1\}$. Consequently, for every symmetric nonnegative definite matrix A and every $\gamma \in [0, 1/2)$,

$$\sup_{0<s<\infty} \frac{U_q(s)^\top A U_q(s)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}} = \sup_{0<x<1} \frac{W_q(x)^\top A W_q(x)}{x^{2\gamma}} \qquad a.s.$$

Proposition 1 is the key simplification behind the $T = \infty$ theory. Instead of simulating Brownian paths on an arbitrarily long interval in s, one can simulate an ordinary Brownian motion on the fixed unit interval in x. This is exactly the device used in the bundled critical-value code.

Theorem 7 (Infinite-horizon KS limits)**.** *Under Assumptions 1–5, interpreted on every compact interval as explained above, and for $\gamma \in [0, 1/2)$,*

$$\begin{aligned}
\sup_{k\ge 1} \mathcal{M}_m^{\mathrm{H}}(k) &\xrightarrow{d} \sup_{0<x<1} \frac{\|W_q(x)\|_2^2}{x^{2\gamma}}, \\
\sup_{k\ge 1} \mathcal{M}_m^{\mathrm{S}}(k) &\xrightarrow{d} \sup_{0<x<1} \frac{W_q(x)^\top V^{-1} W_q(x)}{x^{2\gamma}}, \\
\sup_{k\ge 1} \widetilde{\mathcal{M}}_m^{\mathrm{R}}(k) &\xrightarrow{d} \sup_{0<x<1} \frac{W_q(x)^\top R^{-2} W_q(x)}{x^{2\gamma}}.
\end{aligned}$$

The corresponding open-end critical values are denoted by $c_\alpha^{\mathrm{H}}(\infty, q, \gamma)$, $c_\alpha^{\mathrm{S}}(\infty, q, \gamma)$, and $c_\alpha^{\mathrm{R}}(\infty, q, \gamma)$.

Theorem 7 shows that the open-end KS problem is still pivotal. The only difference from the finite-horizon case is that the monitoring index is now compressed from $s \in [0, \infty)$ to $x \in [0, 1)$. The limit is finite because Brownian motion is continuous at $x = 1$ and because $\gamma < 1/2$ controls the behavior near $x = 0$.

Theorem 8 (Infinite-horizon CvM limits)**.** *Let $w_\infty : [0,\infty) \to \mathbb{R}_+$ satisfy Assumption 6 and define its unit-interval transform by*

$$\widetilde{w}(x) = \frac{w_\infty\left(\frac{x}{1-x}\right)}{(1-x)^2}, \qquad x \in [0,1).$$

Under the conditions of Theorems 4–6, interpreted on every compact interval, we have

$$\begin{aligned}
\sup_{k\geq 1} \mathcal{I}^{\mathrm{H}}_{m,\infty}(k) &\xrightarrow{d} \int_0^1 \widetilde{w}(x)\, \|W_q(x)\|_2^2 \, dx, \\
\sup_{k\geq 1} \mathcal{I}^{\mathrm{S}}_{m,\infty}(k) &\xrightarrow{d} \int_0^1 \widetilde{w}(x) W_q(x)^\top V^{-1} W_q(x)\, dx, \\
\sup_{k\geq 1} \mathcal{I}^{\mathrm{R}}_{m,\infty}(k) &\xrightarrow{d} \int_0^1 \widetilde{w}(x) W_q(x)^\top R^{-2} W_q(x)\, dx.
\end{aligned}$$

In particular, the open-end CvM critical values are pivotal once $(q, \widetilde{w})$ is fixed.

Theorem 8 clarifies why open-end CvM monitoring needs more care than open-end KS monitoring. An unweighted accumulation in the original s-variable would diverge under H_0; the integrability of w_∞ is what prevents that. The transformed unit-interval weight $\widetilde{w}$ is therefore the natural object for both theory and simulation.

Remark 6 (Recommended open-end CvM templates)**.** The four open-end templates introduced in Remark 2,

$$\begin{aligned}
w_{\infty,U}(s) &= \frac{1}{(1+s)^2}, & w_{\infty,\mathrm{Early}}(s) &= \frac{2}{(1+s)^3}, \\
w_{\infty,\mathrm{Late}}(s) &= \frac{2s}{(1+s)^3}, & w_{\infty,\mathrm{Mid}}(s) &= \frac{6s}{(1+s)^4}.
\end{aligned}$$

map into the simple unit-interval weights

$$\begin{aligned}
\widetilde{w}_U(x) &= 1, & \widetilde{w}_{\mathrm{Early}}(x) &= 2(1-x), \\
\widetilde{w}_{\mathrm{Late}}(x) &= 2x, & \widetilde{w}_{\mathrm{Mid}}(x) &= 6x(1-x).
\end{aligned}$$

This is why the bundled code uses the same weight labels for finite and infinite horizons. For quantile monitoring the Late template is attractive when tail spillovers are expected to emerge only after a long calm period, whereas Early or Mid are usually safer for expectile monitoring because expectile scores react more broadly to magnitude shifts.

4.3 Consistency and local power

The main alternative mechanism is a post-change shift in the mean of the excluded-block score stream. Let $s^\star \in (0,T)$ denote the break location in monitoring time.

Assumption 7 (Fixed mean-shift alternative)**.** There exist $s^\star \in (0,T)$ and $\Delta \neq 0$ such that, for $k = \lfloor ms \rfloor$,

$$\mathbb{E}(\Psi_{m+k}) = \begin{cases} \mu + o(1), & 0 \le s \le s^\star, \\ \mu + \Delta + o(1), & s^\star < s \le T, \end{cases}$$

and $\Psi_t - \mathbb{E}\,\Psi_t$ satisfies the same weak-dependence conditions as under H_0.

This assumption is the online translation of "the omitted block starts mattering after the break." It does not require the underlying regression coefficient itself to jump sharply, only that the monitoring score acquire a nonzero mean after the change. That formulation is broad enough to cover both switch-on and switch-off alternatives once the centering is handled appropriately.

Theorem 9 (Consistency)**.** *Under Assumptions 1, 2, 3, and 7,*

$$\mathbb{P}(\mathcal{T}_m^{\mathrm{S}} \le mT \mid H_1) \to 1, \qquad \mathbb{P}(\mathcal{T}_m^{\mathrm{R}} \le mT \mid H_1) \to 1.$$

For the CvM detectors, if in addition there exist an interval $J \subset (s^\star, T]$ with positive Lebesgue measure and a constant $c_w > 0$ such that $w(u) \ge c_w$ for all $u \in J$, then

$$\mathbb{P}(\mathcal{T}_{m,\mathrm{CvM}}^{\mathrm{S}} \le mT \mid H_1) \to 1, \qquad \mathbb{P}(\mathcal{T}_{m,\mathrm{CvM}}^{\mathrm{R}} \le mT \mid H_1) \to 1.$$

Theorem 9 formalizes the basic signal-to-noise logic of online monitoring. After the break, the omitted-block score acquires a mean shift, so the monitoring partial sums pick up a deterministic component of order m. By contrast, every studentizer is built from the stable training window and stays of stochastic order one. The extra positivity condition on w for CvM rules simply ensures that the integrated detector actually places non-negligible mass on some post-break part of the path.

Definition 1 (Local alternatives)**.** A local alternative has post-change drift of order $m^{-1/2}$, that is,

$$\mathbb{E}(\Psi_{m+k}) = \mu + \Delta_m, \qquad \Delta_m = \frac{\delta}{\sqrt{m}}, \qquad k \ge k^\star,$$

for some fixed $\delta \in \mathbb{R}^q$.

Theorem 10 (Local power: drifted Brownian limits)**.** *Under Definition 1 and Assumptions 1–3,*

$$\sup_{1 \le k \le mT} \mathcal{M}_m^{\mathrm{S}}(k) \xrightarrow{d} \sup_{0 < s \le T} \frac{\big(U_q(s) + (s - s^\star)_+ \delta\big)^\top V^{-1} \big(U_q(s) + (s - s^\star)_+ \delta\big)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}}.$$

Analogous drifted limits hold for RSMS and the CvM statistics by replacing $U_q(s)$ with $U_q(s) + (s - s^\star)_+ \delta$ in the corresponding null limits.

Theorem 10 explains how the alternative enters: it tilts the Brownian monitoring increment by a deterministic drift. This form is useful because it links power and delay directly to the direction

and size of the local mean shift. It also explains why CvM time weights matter: a weight that emphasizes early times is more powerful against early drifts, whereas an increasing weight is more useful when the drift turns on late.

4.4 Anytime-validity of the e-process branch

Theorem 11 (Anytime-valid quantile e-process for a fixed direction)**.** *Fix $\tau \in \mathcal{T}_m$ and a predictable direction θ. Under the operational quantile null* (5)*, the process $\{\mathcal{E}_k(\theta,\tau)\}_{k\geq 0}$ in* (31) *is a nonnegative martingale with unit mean. Consequently, for every $\alpha \in (0,1)$,*

$$\mathbb{P}\Big(\sup_{k\geq 1} \mathcal{E}_k(\theta,\tau) \geq \alpha^{-1}\Big) \leq \alpha.$$

The same inequality holds on every finite monitoring horizon.

Theorem 11 is finite-sample and does not rely on asymptotics. This is the main attraction of the e-process branch: false-alarm control is automatic under optional stopping, so the practitioner does not need to pre-specify a monitoring horizon.

Proposition 2 (Mixtures, adaptive betting, restart banks, and tail combinations)**.** *Under* (5)*, each of the processes*

$$\{\mathcal{E}_k^{\mathrm{mix}}(\tau)\}_{k\geq 0}, \qquad \{\mathcal{E}_k^{\mathrm{ad}}(\tau)\}_{k\geq 0}, \qquad \{\mathcal{E}_k^{\mathrm{bank}}(\tau)\}_{k\geq 0}, \qquad \{\mathcal{E}_k^{\mathrm{tail}}\}_{k\geq 0}$$

is an e-process. Hence every threshold rule of the form $\inf\{k : \mathcal{E}_k \geq \alpha^{-1}\}$ controls false alarms at level α.

The proposition explains why the specialized quantile branch can still be flexible. One may mix directions, learn the direction online, hedge against unknown break times by restarts, and combine evidence across several quantile levels, all without losing anytime-validity.

Proposition 3 (General score-based e-process under a conditional mgf bound)**.** *Fix $\tau \in \mathcal{T}_m$ and let $x_t(\tau) = b_{t-1}^\top s_{t-1}\psi_\tau\big(Y_t - \widehat{m}_{t-1,\tau}^X\big)$ be a predictable score projection. Suppose that under the operational null* (4) *there exists a predictable function $\psi_t(\lambda,\tau)$ such that, for every admissible predictable λ_{t-1},*

$$\mathbb{E}\Big[\exp\{\lambda_{t-1}x_t(\tau)\}\,\Big|\,\mathcal{H}_{t-1}\Big] \leq \exp\{\psi_t(\lambda_{t-1},\tau)\}.$$

Then

$$\mathcal{E}_k^E(\tau) := \prod_{t=m+1}^{m+k} \exp\Big\{\lambda_{t-1}x_t(\tau) - \psi_t(\lambda_{t-1},\tau)\Big\} \tag{39}$$

is an e-process. In the conditional sub-Gaussian case, one may take $\psi_t(\lambda,\tau) = \lambda^2\nu_t^2(\tau)/2$.

Proposition 3 is the general expectile extension. It is not as exact and clean as the Bernoulli-hit construction because the score is unbounded, but it shows that the e-process idea is not limited to quantiles. What changes is the price of validity: one needs an exponential-moment bound.

Remark 7 (Detection delay and information growth)**.** If the true post-change hit probability is $p_t^*(\tau) = \mathbb{P}(I_t(\tau) = 1 \mid \mathcal{H}_{t-1})$, then the expected log-growth of the one-step factor equals

$$\mathbb{E}\left[\log L_t(\theta, \tau) \mid \mathcal{H}_{t-1}\right] = \mathrm{KL}\left(\mathrm{Bern}(p_t^*(\tau)) \,\|\, \mathrm{Bern}(\tau)\right) - \mathrm{KL}\left(\mathrm{Bern}(p_t^*(\tau)) \,\|\, \mathrm{Bern}(p_t(\theta, \tau))\right).$$

Thus the e-process grows quickly when the chosen betting model approximates the post-change hit probability well. In the simple constant-alternative case $p_t^*(\tau) = p^* \neq \tau$, the detection-delay heuristic is

$$\mathbb{E}_1[\mathcal{T}_\alpha] \approx \frac{\log(1/\alpha)}{\mathrm{KL}(\mathrm{Bern}(p^*) \,\|\, \mathrm{Bern}(\tau))}.$$

This information-theoretic view of delay is distinct from the Brownian-drift view of the self-normalized branch and helps explain why the two branches are complementary.

5 Simulation studies

This section reports finite-sample evidence for the monitoring procedures developed in Section 3. The experiments are designed to mirror the theoretical setup as closely as possible: the restricted X-only model is estimated once on the Phase-I sample by (7), the estimator is frozen, the omitted-block score vector Ψ_t is constructed as in (8)–(9), and the monitoring rules are then run sequentially over Phase II. Throughout, the nominal level is $\alpha = 0.05$, and every reported number is based on 1,000 Monte Carlo replications. The objective is not just to compare raw power, but also to understand the size–power–delay trade-off across the general self-normalized/HAC branch and the quantile-specific e-process branch.

5.1 Design of the Monte Carlo experiments

The Monte Carlo design uses a simple predictive-regression backbone that is rich enough to generate mean, scale, and tail-only forms of distributional predictive causality. We take

$$X_{t-1} = (1, Y_{t-1})^\top, \qquad n_{\mathrm{mon}} = mT, \qquad m \in \{200, 500\}, \qquad T \in \{1, 2, 5\},$$

with a burn-in of 200 observations. The candidate predictor evolves as

$$Z_t = \phi_z Z_{t-1} + \varepsilon_{z,t}, \qquad \phi_z = 0.5. \tag{40}$$

In the homoskedastic designs, $\varepsilon_{z,t} \sim N(0, 1)$. In the GARCH-type designs N2 and A2, $\varepsilon_{z,t} = \sigma_{z,t} e_{z,t}$ with

$$\sigma_{z,t}^2 = 0.10 + 0.85\sigma_{z,t-1}^2 + 0.05\varepsilon_{z,t-1}^2, \qquad e_{z,t} \sim N(0, 1). \tag{41}$$

The response follows an AR(1) backbone with $\phi_y = 0.5$,

$$Y_t = \phi_y Y_{t-1} + \xi_t, \qquad \phi_y = 0.5, \tag{42}$$

where the innovation component ξ_t is design-specific. For post-break designs, the break fraction $\lambda \in \{0.25, 0.50, 0.75\}$ determines

$$k^\star = \lfloor \lambda m T \rfloor, \qquad D_t = \mathbf{1}\{t > m + k^\star\}.$$

Signal strength is indexed by $c \in \{0.10, 0.25, 0.50\}$.

The null and alternative designs are as follows.

The null designs are N1 and N2 in (43)–(44). These two designs are used to assess finite-sample size. Design N1 provides the baseline homoskedastic benchmark. Design N2 retains the same conditional mean specification but adds GARCH-type time-varying volatility, so it serves as a size experiment under conditional heteroskedasticity.

$$\text{N1:} \qquad Y_t = 0.5 Y_{t-1} + u_t, \qquad u_t \sim N(0,1), \tag{43}$$

$$\text{N2:} \qquad Y_t = 0.5 Y_{t-1} + \sigma_{y,t} u_t, \qquad \sigma^2_{y,t} = 0.10 + 0.85 \sigma^2_{y,t-1} + 0.05 \varepsilon^2_{y,t-1}, \tag{44}$$

The alternative designs are A1–A5 and C1 in (45)–(50). Designs A1 and A2 in (45)–(46) study abrupt location shifts, with A2 augmenting A1 by allowing for time-varying volatility. Design A3 in (47) is a symmetric scale alternative, so that Z_{t-1} affects the conditional distribution only through dispersion. Design A4 in (48) is a downside-tail alternative, in which predictive content is activated only when Z_{t-1} is negative. Design A5 in (49) replaces the abrupt break by a gradual ramp-up over the first 20% of the monitoring window. Finally, design C1 in (50) lets the break start before Phase II begins and therefore contaminates the end of the training sample.

$$\text{A1:} \qquad Y_t = 0.5 Y_{t-1} + c D_t Z_{t-1} + u_t, \tag{45}$$

$$\text{A2:} \qquad Y_t = 0.5 Y_{t-1} + c D_t Z_{t-1} + \sigma_{y,t} u_t, \tag{46}$$

$$\text{A3:} \qquad Y_t = 0.5 Y_{t-1} + \sigma_t u_t, \qquad \sigma^2_t = 1 + c D_t Z^2_{t-1}, \tag{47}$$

$$\text{A4:} \qquad Y_t = 0.5 Y_{t-1} + \sigma_t u_t, \qquad \sigma^2_t = 1 + c D_t (Z^-_{t-1})^2, \quad Z^-_{t-1} = \min(Z_{t-1}, 0), \tag{48}$$

$$\text{A5:} \qquad Y_t = 0.5 Y_{t-1} + \beta_t Z_{t-1} + u_t, \qquad \beta_t = c \min\left\{ \frac{(t - m - k^\star)_+}{L}, 1 \right\}, \quad L = \lfloor 0.2 m T \rfloor, \tag{49}$$

$$\text{C1:} \qquad Y_t = 0.5 Y_{t-1} + c\, \mathbf{1}\{t > \lfloor 0.8 m \rfloor\} Z_{t-1} + u_t. \tag{50}$$

Taken together, these alternatives allow us to compare the monitoring procedures under abrupt coefficient shifts, heteroskedastic location shifts, symmetric scale changes, downside-tail effects, gradual switch-on, and training-sample contamination.

For quantile monitoring we use the grid

$$\mathcal{T}_Q = \{0.05, 0.10, 0.50, 0.90, 0.95\},$$

and for expectile monitoring

$$\mathcal{T}_E = \{0.10, 0.25, 0.50, 0.75, 0.90\}.$$

The default instrument is $H_{t-1} = Z_{t-1}$, except in the downside-tail design A4, where we use the asymmetric instrument

$$H_{t-1} = (Z_{t-1}, Z_{t-1}^{-})^\top$$

to match the structural asymmetry in (48). The uploaded summary archive contains the full KS boundary-exponent grid $\gamma \in \{0, 0.15\}$, the full CvM weight grid, and the quantile e-process variants. The pooled tables below therefore report both KS calibrations, keep the late-weight CvM rules as the benchmark integrated-statistic competitors in the location and tail designs, and report the full CvM weight grid in A5. The accompanying figures display the full method collection.

Under N1–N2 we report rejection frequencies. Under the post-break designs A1–A5 we report the detection rate

$$\mathbb{P}(k^\star < \widehat{k} \le mT)$$

and the conditional detection delay

$$\mathbb{E}(\widehat{k} - k^\star \mid k^\star < \widehat{k} \le mT).$$

For the contaminated-training design C1 the break has already occurred when monitoring starts, so the reported delay is measured from the beginning of Phase II, that is, $\mathbb{E}(\widehat{k} \mid \widehat{k} \le mT)$.

The main-text tables below pool over T and m to keep the discussion readable. Appendix F reports the same SSMS/RSMS procedures separately by horizon T, with distinct columns for $m = 200$ and $m = 500$. Those disaggregated tables show a clear pattern: larger m usually lowers null rejection and strengthens detection in the abrupt, gradual, and contaminated designs, while the effect is weaker or mixed in the pure scale and lower-tail designs A3–A4. Because the break date and monitoring span both scale with m, the raw delays in monitoring periods typically increase with m even when the larger training sample improves detection.

5.2 Size under the null

We begin with the two null designs N1–N2. This is the right place to calibrate the later power comparisons, because the Monte Carlo evidence reveals a clear size–power trade-off across the procedures. Table 1 and Figure 1 summarize the results.

Three facts stand out. First, the quantile e-process branch remains essentially exact, with pooled rejection frequencies of 4.6% and 4.7% for the mixture and restart-bank rules. Second, within the KS branch the larger boundary exponent is uniformly more liberal under the null: SSMS-KS rises from 6.8% to 7.3% in the quantile designs and from 7.2% to 7.9% in the expectile designs when one moves from $\gamma = 0$ to $\gamma = 0.15$; the same monotone increase appears for RSMS-KS and HAC-KS. Third, among the asymptotic procedures SSMS remains the best-calibrated branch, while RSMS,

Table 1: Null rejection frequencies in the pooled size designs N1–N2.

Method	Quantile (%)	Expectile (%)
SSMS-KS ($\gamma = 0$)	6.8	7.2
SSMS-KS ($\gamma = 0.15$)	7.3	7.9
RSMS-KS ($\gamma = 0$)	13.2	12.3
RSMS-KS ($\gamma = 0.15$)	14.6	13.5
HAC-KS ($\gamma = 0$)	12.3	13.7
HAC-KS ($\gamma = 0.15$)	13.8	15.5
SSMS-CvM-Late	6.1	6.5
RSMS-CvM-Late	11.5	10.6
HAC-CvM-Late	10.8	12.0
E-Proc Mix	4.6	–
E-Proc Bank	4.7	–

Notes: The table averages rejection frequencies across the two null designs N1 and N2, training sizes $m \in \{200, 500\}$, and horizons $T \in \{1, 2, 5\}$. The nominal level is 5%. Quantile results use the grid $\{0.05, 0.10, 0.50, 0.90, 0.95\}$ and expectile results use the grid $\{0.10, 0.25, 0.50, 0.75, 0.90\}$. The e-process rows are quantile-only because the exact betting construction is implemented for quantile hits.

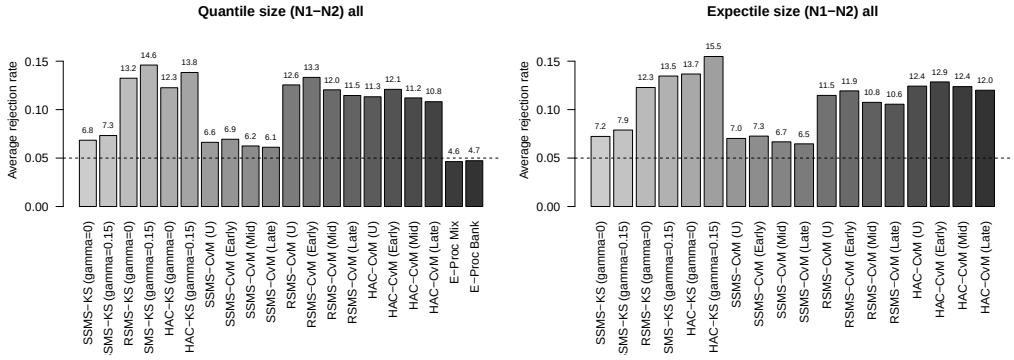


Figure 1: Average null rejection frequencies in the pooled N1–N2 designs across the full method grid. The dashed line marks the nominal 5% level.

HAC, and the late-weight CvM rules are more aggressive.

These size results discipline the later power comparisons. Moving from $\gamma = 0$ to $\gamma = 0.15$ does not improve calibration; it simply tilts the KS branch toward a more liberal stopping rule. For that reason the discussion below treats $\gamma = 0$ as the conservative baseline and $\gamma = 0.15$ as a speed-oriented sensitivity check.

5.3 Abrupt location alternatives

We next turn to the abrupt-switch designs A1 and A2 in (45)–(46). These are the cleanest alternatives for generic online monitoring, because the omitted block starts mattering through a standard predictive-regression coefficient after the break. Table 2 and Figure 2 summarize the results.

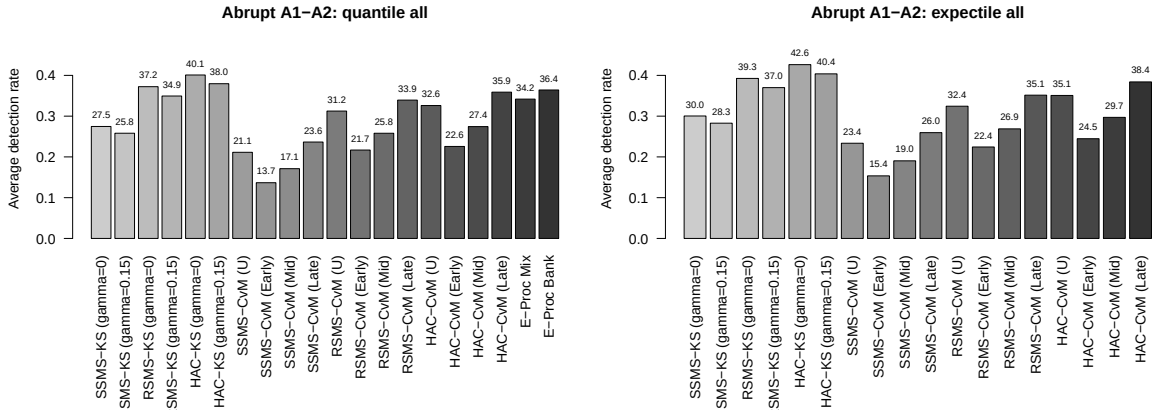
Table 2: Abrupt location alternatives A1–A2: pooled detection rates and conditional delays.

	Quantile		Expectile	
Method	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	27.5	237.9	30.0	236.1
SSMS-KS ($\gamma = 0.15$)	25.8	236.5	28.3	234.4
RSMS-KS ($\gamma = 0$)	37.2	223.8	39.3	220.6
RSMS-KS ($\gamma = 0.15$)	34.9	219.5	37.0	217.2
HAC-KS ($\gamma = 0$)	40.1	222.2	42.6	215.9
HAC-KS ($\gamma = 0.15$)	38.0	218.7	40.4	212.1
SSMS-CvM-Late	23.6	356.6	26.0	349.8
RSMS-CvM-Late	33.9	351.5	35.1	346.6
HAC-CvM-Late	35.9	354.1	38.4	347.8
E-Proc Mix	34.2	236.4	–	–
E-Proc Bank	36.4	230.1	–	–

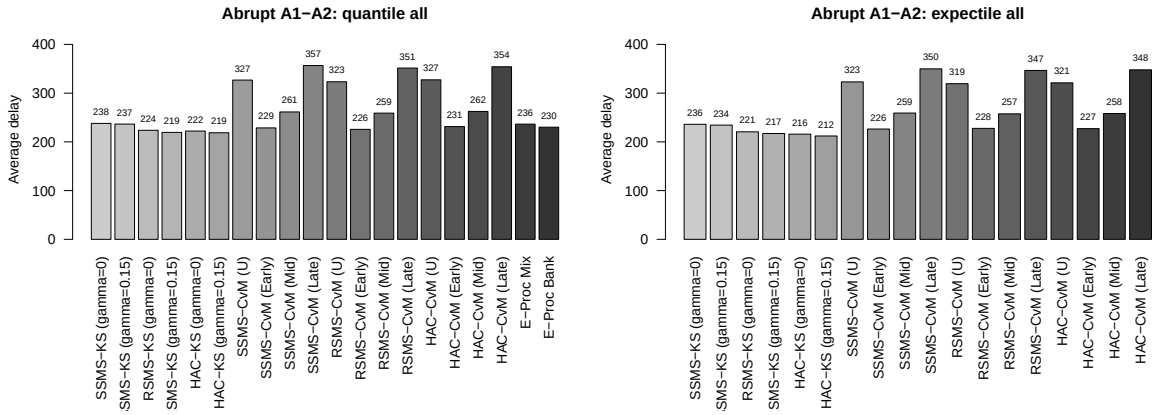
Notes: The table pools over the abrupt designs A1 and A2, signal strengths $c \in \{0.10, 0.25, 0.50\}$, break fractions $\lambda \in \{0.25, 0.50, 0.75\}$ with $k^\star = \lfloor \lambda m T \rfloor$, training sizes $m \in \{200, 500\}$, and horizons $T \in \{1, 2, 5\}$. "Delay is the conditional mean stopping delay $\mathbb{E}(\widehat{k} - k^\star \mid k^\star < \widehat{k} \leq mT)$.

The abrupt-design evidence again points toward the KS branch as the default detector when the omitted predictor enters through a standard coefficient shift. HAC-KS with $\gamma = 0$ is the strongest asymptotic rule, with pooled detection rates of 40.1% in the quantile experiments and 42.6% in the expectile experiments, together with average conditional delays of about 222 and 216 monitoring periods. RSMS-KS with $\gamma = 0$ is close behind at 37.2% and 39.3%, while SSMS-KS is materially more conservative at 27.5% and 30.0%.

Raising the boundary exponent to $\gamma = 0.15$ does not improve abrupt-break detection. Detection rates fall for all three KS branches, although delays shorten slightly. For example, quantile RSMS-KS moves from 37.2% detection at $\gamma = 0$ to 34.9% at $\gamma = 0.15$, while expectile HAC-KS moves from 42.6% to 40.4%. The late-weight CvM rules remain informative but are not leading in A1–A2, whereas the quantile e-process restart bank reaches a pooled detection rate of 36.4% with an average delay of about 230 periods. For generic abrupt alternatives, the preferred benchmark therefore



(a) Detection rates.



(b) Conditional delays.

Figure 2: Pooled performance across the full method grid in the abrupt location designs A1–A2.

remains the $\gamma = 0$ KS branch, especially HAC-KS and then RSMS-KS.

5.4 Tail and scale alternatives

The next set of experiments isolates distributional predictive content that is not well described as a mean shift. Design A3 in (47) is a symmetric scale alternative, while A4 in (48) is a lower-tail alternative that activates only when Z_{t-1} is negative. These are the designs in which distributional monitoring should be most informative relative to mean-based thinking. Table 3 and Figure 3 report the pooled results.

Table 3: Tail and scale alternatives: pooled detection rates and conditional delays.

Panel A: A3 pure scale / symmetric-tail design

	Quantile		Expectile	
Method	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	6.7	207.1	13.3	201.6
SSMS-KS ($\gamma = 0.15$)	6.0	195.6	12.3	190.7
RSMS-KS ($\gamma = 0$)	12.5	207.5	20.7	197.6
RSMS-KS ($\gamma = 0.15$)	11.2	197.4	19.5	188.4
HAC-KS ($\gamma = 0$)	12.8	209.6	23.2	199.6
HAC-KS ($\gamma = 0.15$)	11.7	196.6	22.2	187.7
SSMS-CvM-Late	7.2	370.7	10.6	366.1
RSMS-CvM-Late	13.3	373.3	17.3	368.7
HAC-CvM-Late	13.2	383.0	19.2	375.0
E-Proc Mix	7.8	209.1	–	–
E-Proc Bank	8.3	218.9	–	–

Panel B: A4 downside-tail design with asymmetric instrument

	Quantile		Expectile	
Method	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	22.4	223.5	29.6	218.2
SSMS-KS ($\gamma = 0.15$)	20.1	220.4	27.6	214.1
RSMS-KS ($\gamma = 0$)	30.6	206.5	38.6	205.7
RSMS-KS ($\gamma = 0.15$)	26.8	204.8	35.6	201.0
HAC-KS ($\gamma = 0$)	32.9	204.3	41.5	193.6
HAC-KS ($\gamma = 0.15$)	28.8	201.9	37.9	189.3
SSMS-CvM-Late	27.0	342.7	28.8	348.0
RSMS-CvM-Late	41.2	331.1	41.8	341.9
HAC-CvM-Late	44.7	332.7	47.7	337.1
E-Proc Mix	12.7	263.1	–	–
E-Proc Bank	13.6	261.8	–	–

Notes: Panel A corresponds to the symmetric scale design $\sigma_t^2 = 1 + cD_t Z_{t-1}^2$. Panel B corresponds to the downside-tail design $\sigma_t^2 = 1 + cD_t(Z_{t-1}^-)^2$ with $Z_{t-1}^- = \min(Z_{t-1}, 0)$ and instrument $H_{t-1} = (Z_{t-1}, Z_{t-1}^-)^\top$. Both panels pool over $c \in \{0.10, 0.25, 0.50\}$, $\lambda \in \{0.25, 0.50, 0.75\}$, $m \in \{200, 500\}$, and $T \in \{1, 2, 5\}$.

The A3 scale design remains the hardest alternative in the Monte Carlo study. Even after restoring the SSMS rows and the $\gamma = 0.15$ KS calibrations, pooled quantile detection rates stay

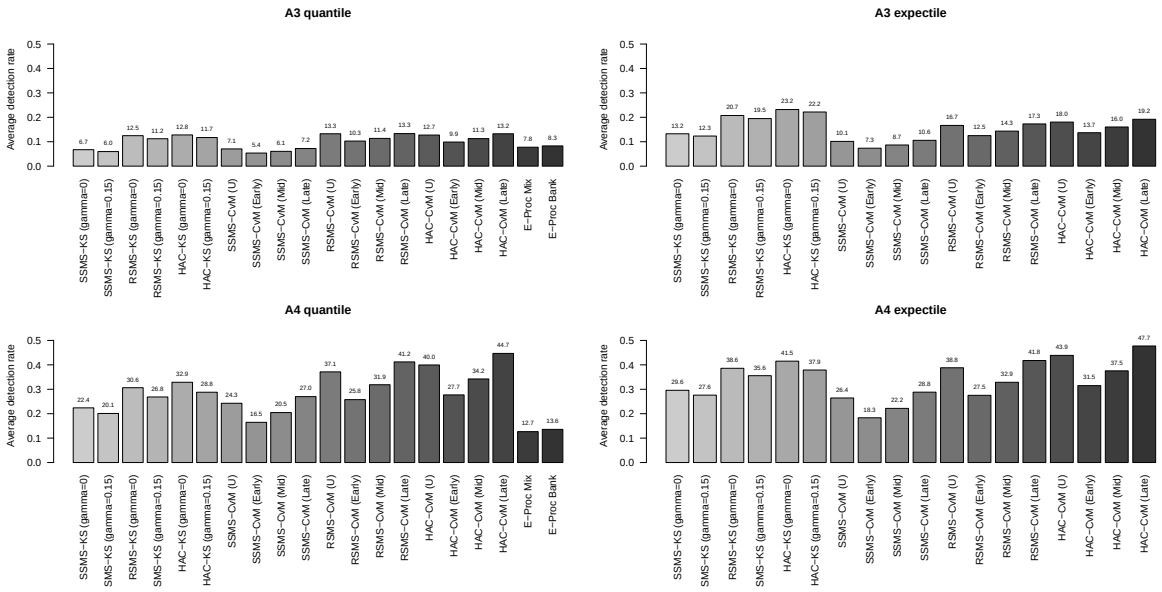


Figure 3: Pooled detection rates across the full method grid in the tail and scale designs A3–A4.

modest. The strongest quantile performance in A3 comes from the late-weight CvM rules, at 13.3% for RSMS-CvM-Late and 13.2% for HAC-CvM-Late, while the KS branch ranges from 6.7% to 12.8% under $\gamma = 0$ and is uniformly a little weaker under $\gamma = 0.15$. The expectile branch is more informative in this setting, with HAC-KS reaching a pooled detection rate of 23.2% at $\gamma = 0$.

Design A4 is much more decisive. Once predictive content is concentrated in the lower tail and the instrument is matched to that asymmetry through (Z_{t-1}, Z_{t-1}^{-}), the late-weight CvM rules become first-order. HAC-CvM-Late is the strongest detector in both model classes, with pooled detection rates of 44.7% for quantiles and 47.7% for expectiles, compared with 32.9% and 41.5% for HAC-KS at $\gamma = 0$. The $\gamma = 0.15$ KS calibrations again trade a small delay reduction for lower detection. The practical implication is therefore sharper than before: when omitted predictive content is believed to act through an asymmetric tail channel rather than a center-of-distribution coefficient shift, the late-weight CvM branch should be treated as a primary specification rather than as a robustness check.

5.5 Gradual switch-on and the role of the CvM time weight

Design A5 in (49) replaces the abrupt break by a ramp-up over the first 20% of the monitoring period. This is the design in which the choice of the CvM time weight w is most substantively important. Table 4 and Figure 4 summarize the SSMS, RSMS, and HAC CvM results across the four benchmark weights.

Two regularities stand out. First, detection rates increase monotonically as the weight shifts from Early to Mid to Uniform to Late for all three asymptotic branches. In the quantile case, SSMS-CvM rises from 9.2% under the Early weight to 17.5% under the Late weight, RSMS-CvM rises from 16.1% to 27.0%, and HAC-CvM rises from 16.7% to 28.4%. The same monotone pattern appears in

Table 4: Gradual alternative A5: CvM weight trade-off for SSMS, RSMS, and HAC monitors.

Panel A: Quantile

	SSMS-CvM		RSMS-CvM		HAC-CvM	
Weight	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
Early	9.2	234.3	16.1	231.3	16.7	240.1
Mid	11.9	271.3	19.7	273.5	20.6	276.5
Uniform	15.5	341.4	24.6	337.5	25.3	343.6
Late	17.5	368.9	27.0	365.6	28.4	371.5

Panel B: Expectile

	SSMS-CvM		RSMS-CvM		HAC-CvM	
Weight	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
Early	10.1	235.3	16.4	238.0	18.1	238.6
Mid	13.0	271.1	19.9	275.0	22.0	275.9
Uniform	16.8	340.3	25.3	341.4	27.3	343.7
Late	18.9	367.8	27.7	367.3	30.5	369.6

Notes: The table averages over the gradual design A5, where the post-break coefficient rises linearly over a ramp window of length $L = \lfloor 0.2mT \rfloor$. All entries pool over $c \in \{0.10, 0.25, 0.50\}$, $\lambda \in \{0.25, 0.50, 0.75\}$, $m \in \{200, 500\}$, and $T \in \{1, 2, 5\}$.

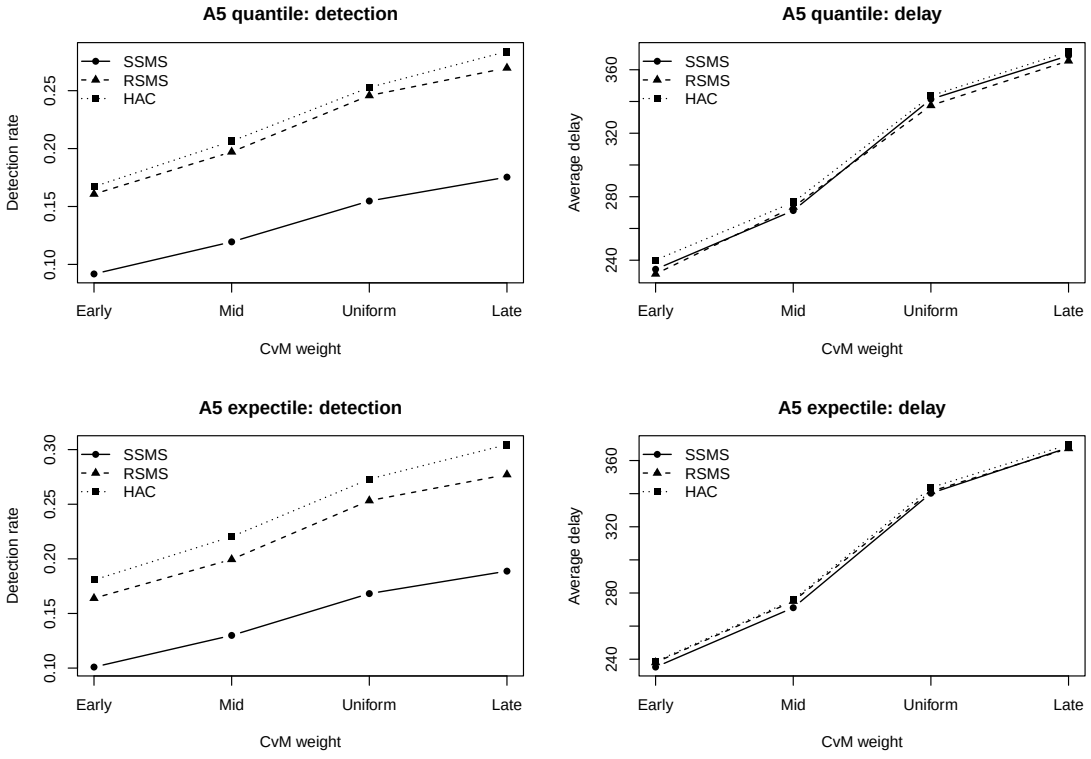


Figure 4: CvM weight trade-off across the full self-normalized/HAC grid in the gradual design A5.

the expectile experiments, where HAC-CvM rises from 18.1% to 30.5%. Second, the conditional delay also rises as the weight moves toward late monitoring times. This is not a contradiction. A late-weight rule is willing to wait longer for persistent evidence and therefore captures more difficult gradual breaks, but among the breaks that it detects, it stops later on average.

This weight trade-off is exactly the behavior suggested by the theory in Section 4. Early weights are better aligned with a concern for fast response near the beginning of the monitoring window; Late weights are better aligned with the objective of maximizing eventual detection against alternatives whose drift accumulates only slowly. The inclusion of the SSMS branch shows that this monotone ranking is not a HAC artifact: it is common to all three asymptotic branches. HAC-KS with $\gamma = 0$ still has the strongest pooled average detection rate among the broad detectors in A5, but the point of Table 4 is different. It shows that within the CvM branch the choice of w has economically meaningful consequences and should be treated as part of the model-design problem rather than as a nuisance detail.

5.6 Training contamination

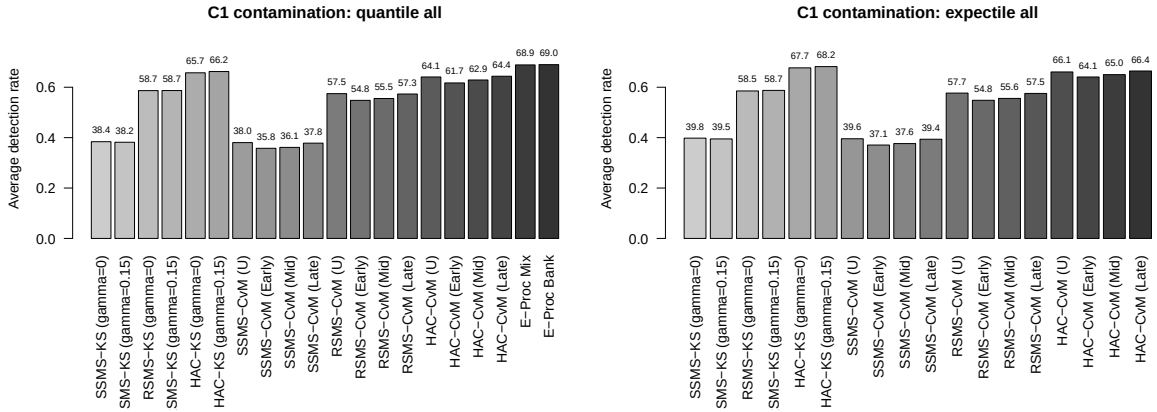
The last design, C1 in (50), lets the predictive relationship start inside the last 20% of the training sample. This deliberately violates the idealized assumption that Phase I is perfectly stable and asks how much the detectors deteriorate when the frozen restricted fit is already mildly contaminated before monitoring begins. Table 5 and Figure 5 report the results.

Table 5: Training-contamination design C1: pooled detection rates and conditional delays.

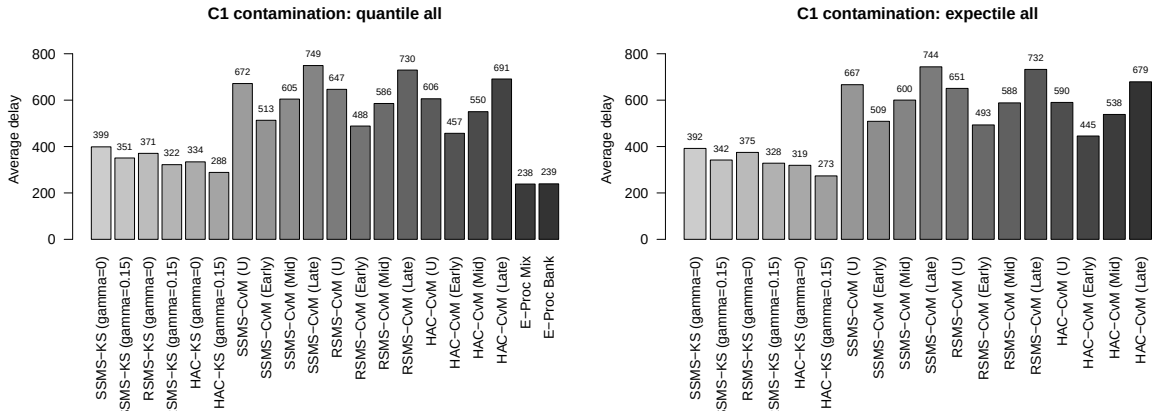
	Quantile		Expectile	
Method	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	38.4	398.7	39.8	391.9
SSMS-KS ($\gamma = 0.15$)	38.2	350.6	39.5	341.9
RSMS-KS ($\gamma = 0$)	58.7	370.6	58.5	374.7
RSMS-KS ($\gamma = 0.15$)	58.7	321.9	58.7	328.2
HAC-KS ($\gamma = 0$)	65.7	334.0	67.7	319.1
HAC-KS ($\gamma = 0.15$)	66.2	288.5	68.2	273.4
SSMS-CvM-Late	37.8	749.1	39.4	743.6
RSMS-CvM-Late	57.3	729.7	57.5	732.5
HAC-CvM-Late	64.4	691.1	66.4	679.0
E-Proc Mix	68.9	238.3	–	–
E-Proc Bank	69.0	239.2	–	–

Notes: In design C1 the break starts inside the last 20% of the training period, so the monitoring sample begins after the predictive relationship has already switched on. The reported delay is therefore measured from the beginning of monitoring, i.e. $\mathbb{E}(\widehat{k} \mid \widehat{k} \leq mT)$. All entries pool over $c \in \{0.10, 0.25, 0.50\}$, $m \in \{200, 500\}$, and $T \in \{1, 2, 5\}$.

This is the design in which the quantile e-process branch looks especially attractive. In the quantile experiments, the mixture and restart-bank e-processes have pooled detection rates of 68.9% and 69.0%, respectively, with average delays around 239 monitoring periods. Both outperform the



(a) Detection rates.



(b) Conditional delays.

Figure 5: Performance across the full method grid in the contaminated-training design C1.

asymptotic KS and CvM rules in delay. Among the asymptotic procedures, HAC-KS is strongest, with pooled detection rates of 65.7% at $\gamma = 0$ and 66.2% at $\gamma = 0.15$; the increase in the boundary exponent leaves detection almost unchanged but shortens the pooled delay from 334.0 to 288.5. RSMS-KS and SSMS-KS show the same pattern, with little change in detection and materially shorter delays under $\gamma = 0.15$.

The economic interpretation is clear. When monitoring begins from a mildly contaminated baseline, a more liberal boundary can be worthwhile because the break is already active at time zero. Even so, the e-process branch remains the most attractive quantile-based device in C1 because it combines the highest pooled detection rates with by far the shortest delays. In the expectile experiments, where the exact hit-based e-process is unavailable, HAC-KS remains the strongest of the asymptotic detectors, and the $\gamma = 0.15$ calibration again shortens delay markedly, from 319.1 to 273.4, with a small increase in pooled detection.

The simulation evidence points to a clear but now more nuanced division of labor between the two branches of the paper. First, the newly integrated $\gamma = 0.15$ runs do not overturn the baseline conclusions: relative to $\gamma = 0$, they are uniformly more liberal under the null, usually a little faster conditional on detection, and generally a little weaker in the abrupt and tail alternatives. The contaminated-training design C1 is the main exception, because once the break is already active at the start of monitoring the more liberal boundary can shorten delay without materially sacrificing detection.

Second, HAC-KS and RSMS-KS with $\gamma = 0$ remain the strongest broad-purpose detectors for generic monitoring problems. Third, the CvM branch is valuable when the economic question is explicitly about *when* the omitted predictive content emerges. The A4 and A5 designs show that a late-weight CvM detector can be materially better than KS rules when the break is tail-localized or accumulates gradually. Fourth, the quantile e-process branch is not merely a theoretical add-on. It achieves near-nominal size under the null and performs especially well in the contaminated-training setting, while remaining competitive in abrupt alternatives.

These findings motivate the practical recommendations used in the rest of the paper. For generic monitoring problems, HAC-KS or RSMS-KS with $\gamma = 0$ are reasonable starting points, while the $\gamma = 0.15$ versions are best interpreted as sensitivity checks when the researcher is willing to trade some size control for speed. When scientific prior information suggests that predictive content is concentrated in the lower tail or is likely to switch on only gradually, a late-weight CvM rule deserves to be reported alongside KS. When exact finite-sample calibration, open-end validity, or robustness to training contamination is central, the quantile e-process restart bank becomes particularly attractive.

6 Empirical illustration: BTC downside-risk monitoring with Deribit option signals

We study whether hourly BTC Deribit option-market variables add incremental predictive content for the lower tail of BTC spot returns once a spot-only baseline forecaster has already been trained. The response variable is the next-hour BTC spot return Y_t. The baseline block X_{t-1} contains lagged and realized spot-market controls from Binance BTCUSDT hourly candles, while the omitted block Z_{t-1} contains five option-market signals extracted from BTC Deribit trades. After merging the spot and option panels and applying complete-case filtering, the final estimation sample contains 3,289 hourly observations spanning 2025-05-01 01:00:00 to 2025-09-15 08:00:00 UTC, covering approximately 4.5 months. This duration is chosen to include periods of both heightened and moderate volatility in the BTC market, ensuring that the sample captures meaningful variation for tail-risk monitoring while providing sufficient observations for robust rolling-window estimation and predictive causality analysis.

More specifically, the baseline block is

$$X_{t-1} = \left(\mathrm{ret}_{t-1}, \mathrm{mom24}_{t-1}, RV^{24}_{t-1}, Range_{t-1}, \log V^{\mathrm{spot}}_{t-1}\right)^\top,$$

where ret_{t-1} is the one-hour log return of BTCUSDT in the previous hour; $\mathrm{mom24}_{t-1}$ is the 24-hour momentum, computed as the cumulative log return over the previous 24 hours; RV^{24}_{t-1} is the 24-hour realized variance, calculated as the sum of squared intraday hourly returns within the previous 24 hours; $Range_{t-1}$ is the previous-hour high-low range, capturing short-term volatility and microstructure effects; and $\log V^{\mathrm{spot}}_{t-1}$ is the logarithm of hourly spot trading volume, which proxies for market liquidity and activity.

The candidate block is

$$Z_{t-1} = \left(IV - RV\ \mathrm{spread}_{t-1}, Skew_{t-1}, TermSlope_{t-1}, PCImb_{t-1}, \log Activity_{t-1}\right)^\top,$$

where $IV - RV\ \mathrm{spread}_{t-1}$ is the difference between the implied volatility extracted from BTC options and the realized variance of BTC spot returns over the past hour, capturing the market-implied perception of near-term tail risk; $Skew_{t-1}$ is the option-implied skewness of the BTC return distribution, constructed from the slope of out-of-the-money put and call implied volatilities, which reflects asymmetric downside risk; $TermSlope_{t-1}$ is the slope of the option term structure, i.e., the difference between short-term and long-term ATM implied volatilities, which signals expected changes in volatility over the horizon; $PCImb_{t-1}$ is the put-call imbalance, calculated as the volume-weighted difference between put and call trades, capturing directional sentiment in the option market; and $\log Activity_{t-1}$ is the log of total option transaction volume or count within the previous hour, representing option-market activity and liquidity.

The spot leg is constructed from hourly BTCUSDT candles downloaded through the Binance Spot API. The option leg is constructed from BTC option transactions obtained through the

Blockchain Research Center accessible-data service under "BTC Deribit Transactions / BTC Deribit Option Data".[1] The raw option records arrive as segmented CSV deliveries and are parsed into hourly UTC features before merging with the spot panel.

6.1 Data construction

The decision time is the end of each completed UTC hour. On the spot side, we align Binance candles to completed one-hour intervals and define the response as the next-hour BTC return. The baseline block uses only information available at the decision time: the lagged one-hour return, a 24-hour momentum measure, 24-hour annualized realized volatility, the contemporaneous high-low range, and log quote volume. On the option side, each Deribit trade is time-stamped in UTC and aggregated to completed clock hours. The preprocessing retains observations with strictly positive time-to-maturity, implied volatility, traded amount, and underlying index price. Contracts are classified as near-dated when time-to-maturity lies in $[7, 45]$ days and mid-dated when it lies in $[46, 90]$ days; an option is treated as approximately at the money when $|\log(K/S)| \leq 0.05$, where K is strike and S is the contemporaneous Deribit index price. Hourly implied-volatility levels are amount-weighted averages computed separately for the near-dated and mid-dated groups.

From those hourly sufficiency statistics we construct the five omitted-block variables used in the monitoring tests. The IV-RV spread is near-dated ATM implied volatility minus 24-hour realized volatility. The skew proxy is near OTM put IV minus near OTM call IV. The term slope is near-dated ATM IV minus mid-dated ATM IV. The put-call imbalance is $((P - C)/(P + C))$ using near-dated traded amounts. Log option activity is $\log(1 + \text{total option amount})$. To avoid dropping isolated hours mechanically, implied-volatility levels are filled only over short local outages, while pure flow variables are left at zero in hours with no qualifying trade. After merging the spot and option panels, we keep only complete cases for the response, the baseline controls, and the omitted-block features required by the monitoring tests. That final filter is what reduces the usable estimation sample to 3,289 observed merged hours.

6.2 Testing protocol

The testing protocol mirrors the theory sections closely. Monitoring is carried out on the lower-tail quantile grid $\{0.05, 0.10\}$. Each rolling experiment estimates the restricted X-only quantile model on a 90-day training window ($m = 2160$ hours), then monitors for 14 days ($mT = 336$ hours), and refits every 24 hours. Given the available sample, only 34 rolling windows fit once the 90-day training window and 14-day monitoring horizon are imposed. In every rolling window, the X and Z regressors are standardized using training-sample moments only, the restricted quantile regressions are estimated at $\tau \in \{0.05, 0.10\}$, and the excluded-block score vectors $Z_{t-1}\psi_\tau(Y_t - X_{t-1}^\top \widehat{\alpha}_m(\tau))$ are formed hour by hour during monitoring. Because the omitted block has five variables and two quantiles are monitored jointly, the stacked score dimension is $q = 10$.

[1] `https://blockchain-research-center.com/`

For the self-normalized branch we compute SSMS, RSMS, and HAC studentizers from the training scores, form the closed-end KS and CvM monitoring paths over the 14-day horizon, and compare them with simulated Brownian critical values calibrated at $\alpha = 0.05$ for $(q, T) = (10, 336/2160)$. In parallel we run the three quantile e-process detectors (mixture, adaptive, and restart-bank). An alarm is recorded at the first boundary crossing in a given rolling window, and any resulting trading action starts at the next tradable merged-sample hour. The empirical application therefore evaluates 21 detectors and 84 detector-strategy combinations: flat and short reactions, each with 6-hour and 24-hour holding periods.

6.3 Empirical results

Table 6: Descriptive statistics for the BTC spot–Deribit option estimation panel

Variable	N	Mean	SD	P5	Median	P95
Panel A: response and spot-only baseline controls						
Next-hour BTC return Y_t	3289	0.0001	0.0035	-0.0052	0.0000	0.0053
Lagged BTC return	3289	0.0001	0.0035	-0.0052	0.0000	0.0053
24-hour return momentum	3289	0.0004	0.0082	-0.0132	0.0003	0.0138
24-hour realized volatility (ann.)	3289	0.3086	0.1225	0.1439	0.2919	0.5505
Hourly high-low range	3289	0.0048	0.0032	0.0016	0.0040	0.0106
Log spot quote volume	3289	17.8463	0.7098	16.7327	17.8100	19.0959
Panel B: Deribit signals entering the omitted block Z_{t-1}						
Put-call imbalance	3289	-0.1187	0.4890	-0.8700	-0.1354	0.7396
Log option activity	3289	7.3354	0.8195	6.1224	7.3392	8.5978
Option skew proxy	3289	0.0205	0.0327	-0.0264	0.0179	0.0779
Term-structure slope	3289	-0.0360	0.0198	-0.0688	-0.0356	-0.0033
IV-RV spread	3289	0.0668	0.1170	-0.1632	0.0836	0.2333
Panel C: auxiliary implied-volatility levels used to construct the signals						
Near-dated ATM IV	3289	0.3754	0.0401	0.3169	0.3705	0.4523
Mid-dated ATM IV	3289	0.4114	0.0388	0.3555	0.4057	0.4801

Notes: The final complete-case estimation panel contains 3,289 hourly observations from 2025-05-01 01:00:00 to 2025-09-15 08:00:00 UTC. The spot leg uses hourly BTCUSDT candles from the Binance Spot API. The option leg uses BTC option transactions from the Blockchain Research Center accessible-data service ("BTC Deribit Transactions / BTC Deribit Option Data"; `https://blockchain-research-center.com/`), which were parsed into hourly option features from segmented raw CSV deliveries. Panel B lists the five variables included in the omitted block Z_{t-1}; Panel C lists the underlying ATM implied-volatility levels used to form the IV-RV spread and term-slope signals. Light-gray gaps in Figure 6 correspond to real coverage outages inherited from the raw option files rather than zero-trading periods.

Table 6 shows that the spot-only baseline controls remain economically plausible over this sample: the hourly return distribution is centered near zero but visibly dispersed, realized volatility is elevated, and quote volume remains high. The option block also displays substantial variation. The put-call imbalance is negative on average, the IV-RV spread is positive on average, and both skew and term structure move materially over time. These are not smooth background regressors. They move episodically, which is exactly the sort of omitted-block variation that online monitoring procedures are designed to detect.

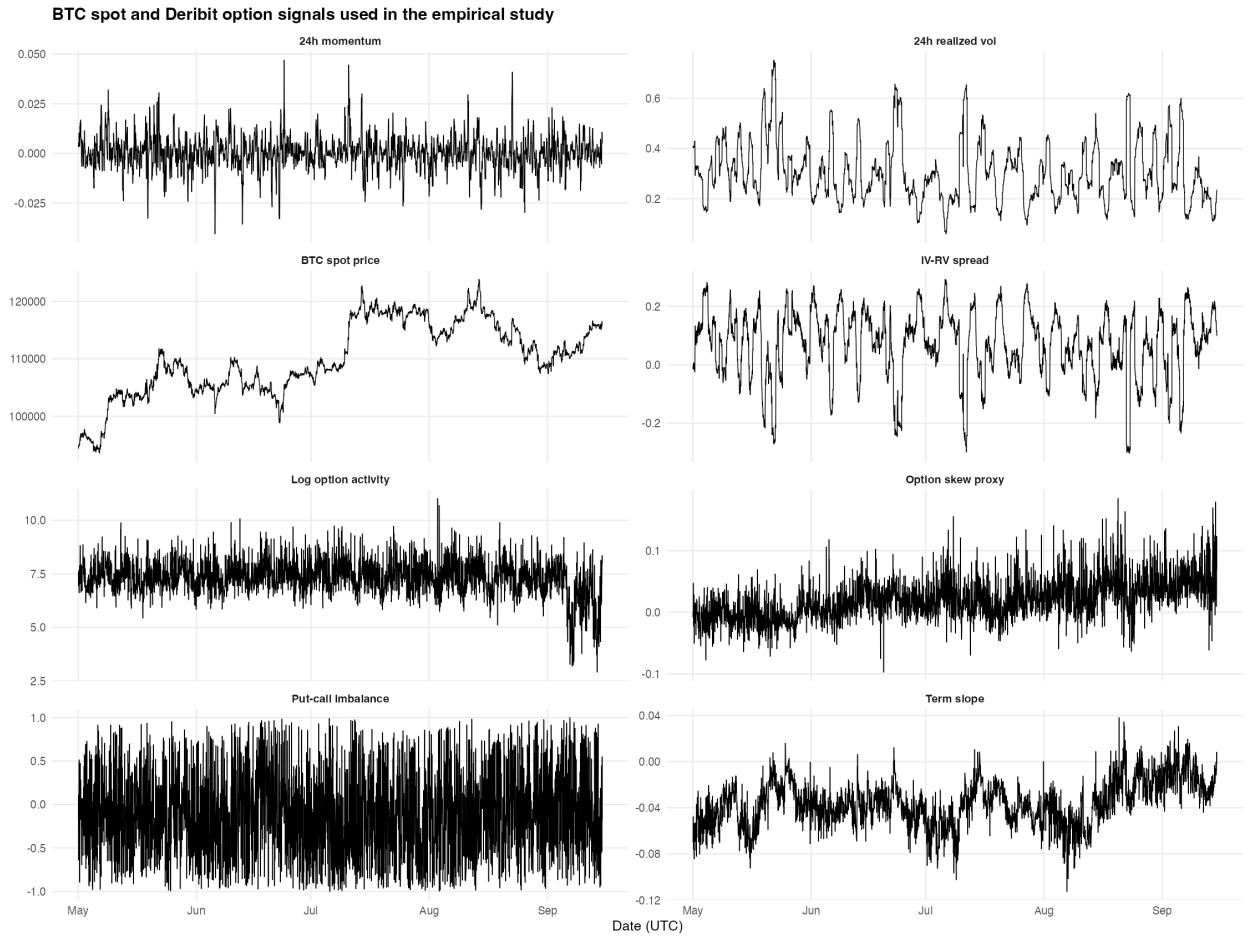


Figure 6: BTC spot and Deribit option signals on the estimation sample. The eight panels report, from top-left to bottom-right, 24-hour momentum, 24-hour realized volatility, BTC spot price, the IV-RV spread, log option activity, the skew proxy, put-call imbalance, and the term slope. Light shaded bands mark larger merged-data gaps inherited from the raw option files. A line segment that traverses a shaded band simply joins the last observed value before the gap to the first observed value after the gap; it is not an imputed within-gap path.

Figure 6 makes the data environment transparent. Spot controls move persistently, while the option-based omitted-block signals are more episodic and exhibit visibly clustered bursts. That is the empirical setting in which the monitoring statistics should matter most: if Deribit information adds incremental predictive content, it is likely to do so in intermittent episodes rather than as a permanent shift in every hour.

Table 7: Alarm-to-trade rules used in the empirical backtest

Rule class	Exposure after an alarm	Hold length	Interpretation
Flat 6h	Move BTC exposure from +1 to 0	6 observed merged hours	Tactical de-risking rule for brief downside-warning episodes. It asks whether the alarm is useful as a short-lived risk filter rather than as a directional short signal.
Flat 24h	Move BTC exposure from +1 to 0	24 observed merged hours	Persistent de-risking rule. It is designed for alarms that identify downside conditions lasting beyond a few hours.
Short 6h	Move BTC exposure from +1 to −1	6 observed merged hours	Aggressive short-horizon overlay. It tests whether an alarm should be interpreted as an immediate bearish signal rather than merely as a cue to step aside.
Short 24h	Move BTC exposure from +1 to −1	24 observed merged hours	Persistent bearish overlay. It is the most demanding rule because it requires alarm information to carry over into a longer short position.

Notes: Trading begins at the next tradable merged-sample hour after the first threshold crossing within a rolling monitoring window. If multiple alarms overlap, the alarm state is extended rather than stacked. All returns and drawdown measures are computed on observed merged hours only, exactly matching the available spot–option overlap in the final estimation panel.

Table 7 reports the four alarm-to-trade rules used in the empirical backtest. The flat rules move exposure from +1 to 0 to test whether an alarm signals de-risking opportunities, with 6-hour and 24-hour holds targeting short-lived versus more persistent episodes. The short rules move exposure from +1 to -1, translating alarms into aggressive bearish overlays over the same horizons. Overlapping alarms extend the exposure rather than stacking, ensuring economically meaningful positions.

Table 8 is the main empirical comparison because it reports, detector by detector, how much each alarm-to-trade rule earned on the realized sample. Three results stand out. First, all three e-process detectors remain silent on this sample, so they coincide with the no-alarm buy-and-hold benchmark: a cumulative return of 21.5%, an annualized return of 57.4%, a Sortino ratio of 2.859, and a maximum drawdown of 13.3%.[2] In other words, the e-process branch functions here as a conservative benchmark rather than as an active alarm generator.

Second, the strongest raw-return overlays come from the CvM branch, especially RSMS-CvM (Late). Its cumulative returns are 27.8% for the flat 6-hour rule, 29.3% for the flat 24-hour rule,

[2]The Sortino ratio measures mean hourly strategy return relative to downside semideviation, so a higher value means more return per unit of harmful downside volatility.

Table 8: Detector-by-detector performance across the four alarm-to-trade rules

	Flat 6h		Flat 24h		Short 6h		Short 24h	
Method	Total (%)	Sortino	Total (%)	Sortino	Total (%)	Sortino	Total (%)	Sortino
EP-Adap.	21.5	2.859	21.5	2.859	21.5	2.859	21.5	2.859
EP-Bank	21.5	2.859	21.5	2.859	21.5	2.859	21.5	2.859
EP-Mix	21.5	2.859	21.5	2.859	21.5	2.859	21.5	2.859
SSMS-KS (0)	22.0	2.911	22.5	2.962	22.5	2.963	23.4	3.065
SSMS-KS (0.15)	21.1	2.817	21.9	2.900	20.7	2.773	22.2	2.934
RSMS-KS (0)	18.8	2.557	19.9	2.694	16.1	2.246	18.2	2.497
RSMS-KS (0.15)	20.6	2.758	23.2	3.059	19.7	2.654	24.9	3.240
HAC-KS (0)	19.7	2.657	20.0	2.693	17.9	2.450	18.4	2.513
HAC-KS (0.15)	19.3	2.613	23.0	3.032	17.1	2.360	24.4	3.179
SSMS-CvM (U)	21.7	2.882	19.1	2.597	21.9	2.898	16.8	2.330
SSMS-CvM (E)	22.4	2.984	25.7	3.364	23.3	3.079	29.9	3.809
SSMS-CvM (M)	23.5	3.106	23.2	3.091	25.5	3.324	24.8	3.251
SSMS-CvM (L)	22.6	2.982	21.6	2.875	23.8	3.104	21.7	2.883
RSMS-CvM (U)	22.2	2.935	19.8	2.684	22.8	3.005	17.8	2.400
RSMS-CvM (E)	22.5	2.970	25.6	3.297	23.5	3.076	29.7	3.727
RSMS-CvM (M)	22.7	2.989	19.9	2.696	23.9	3.119	18.0	2.430
RSMS-CvM (L)	**27.8**	3.568	**29.3**	3.766	**34.4**	4.241	**37.3**	4.583
HAC-CvM (U)	22.1	2.928	24.5	3.220	22.7	2.993	27.5	3.537
HAC-CvM (E)	17.1	2.363	19.3	2.607	12.7	1.787	16.8	2.252
HAC-CvM (M)	17.9	2.451	18.7	2.544	14.2	1.989	15.8	2.161
HAC-CvM (L)	25.8	3.370	24.6	3.265	30.2	3.832	27.7	3.586

Notes: Total return is the cumulative sample return earned by the corresponding alarm-to-trade rule over the observed merged hours from 2025-05-01 to 2025-09-15; it therefore answers directly how much each rule earned on the realized sample. Sortino is computed from the hourly strategy return series. Boldface marks the highest total return within a rule class. The three e-process rules never alarm on this sample, so their entries coincide with the no-alarm buy-and-hold benchmark. Abbreviations: U = Uniform, E = Early, M = Mid, L = Late, and the number in parentheses for KS rules is the boundary exponent γ.

34.4% for the short 6-hour rule, and 37.3% for the short 24-hour rule, with the corresponding Sortino ratios increasing from 3.568 to 4.583. HAC-CvM (Late) is the next-strongest return generator and is particularly competitive when the evaluation also rewards precision and moderate alarm counts. The broad pattern is therefore that the most profitable overlays are produced by detectors that allow cumulative score evidence to build, rather than by the most conservative hit-based monitors.

Third, raw return and implementation balance are not identical objectives. Appendix Table G.1 shows that HAC-CvM (Late) is the best practical-score configuration in the flat classes once precision, capture, and drawdown are ranked jointly, whereas SSMS-CvM (Mid) is the most balanced short-side rule. This difference matters because the best-return rule is not automatically the most attractive deployable rule. On this sample, longer 24-hour holds also dominate the shorter 6-hour holds more often than not, especially on the short side, which suggests that the downside episodes identified by the Deribit block tend to persist beyond a few hours once they begin.

The appendix reports the complete 84-rule leaderboard, family-level summaries, and auxiliary empirical figures, and it also clarifies why some detector-specific equity panels appear to show only one line. Appendix Tables G.1 and G.2 summarize the same economic trade-off from two complementary angles. The Sortino ratio measures average hourly strategy return relative to downside semideviation, so it rewards return earned per unit of harmful downside risk rather than total volatility. The practical score is a lower-is-better average of within-class ranks for Sortino, maximum drawdown, precision, and event capture, so it is designed to summarize deployability rather than raw return alone. Economically, it is meant to summarize whether a detector is not just profitable on one sample path but also disciplined enough to be plausible as a deployable risk-management overlay.

Appendix Figures G.1 and G.2 show that alarm activity is sparse and clustered in only a subset of rolling windows. That timing pattern is economically important. The e-process rules never alarm, several KS and SSMS-CvM rules alarm only once or twice, and the more active RSMS/HAC CvM rules account for most of the economically meaningful deviation from the benchmark. This sparsity also explains the detector-by-detector equity panels in Appendix Figure G.4: when the black benchmark and colored strategy path coincide exactly, the rendered figure only shows the colored line because it sits on top. Economically, those overlapping stretches are hours with no active alarm-driven exposure change, so the strategy is simply matching buy-and-hold. The appendix therefore does not just add extra pictures. It documents why different detectors translate into different trading overlays and why, in this sample, most of the economic dispersion comes from a relatively small set of CvM alarms.

7 Conclusion

This concluding section gathers the paper's main messages and clarifies how the two monitoring branches should be read together.

This paper proposes a unified framework for online monitoring of distributional Granger causality.

The general branch of the paper is self-normalized score monitoring: estimate the restricted baseline model on a stable training window, construct the omitted-block score stream, and monitor it sequentially through SSMS and RSMS KS/CvM detectors. That branch is deliberately broad and is the main methodological contribution because it covers quantiles, expectiles, and more general elicitable functionals under weak dependence without long-run variance tuning. The quantile e-process branch is more specialized. It exploits the fact that quantile causality can be reformulated as conditional calibration of forecast hits and therefore brings finite-sample anytime-validity, restart schemes, and a direct tail-risk interpretation.

The paper also clarifies how the two branches should be read together. Self-normalization is the general econometric machinery. The e-process branch is not a replacement for that machinery; it is a quantile/VaR-specific refinement that uses extra structure when that structure is available. The simulation evidence sharpens this division of labor but also reveals a finite-sample size–power trade-off: HAC and RSMS KS rules with $\gamma = 0$ emerge as the strongest general-purpose detectors, the $\gamma = 0.15$ KS rules are modestly faster but more liberal and are most attractive when monitoring begins after contamination may already have started, Late-weight CvM rules are particularly useful when predictive content accumulates in the tails or turns on gradually, and the quantile e-process branch is especially attractive for exact size control, open-end validity, and contaminated-training scenarios.

The BTC/Deribit illustration points to the same complementarity in practice. On the hourly merged panel of 3,289 observations from 2025-05-01 to 2025-09-15, only 34 rolling monitoring windows fit once the training and monitoring horizons are imposed. In that short option-coverage sample, all three e-process detectors remain silent and therefore reproduce the no-alarm buy-and-hold benchmark, which earns 21.5% cumulatively (57.4% annualized) with a Sortino ratio of 2.859. The strongest raw-return overlays come instead from the self-normalized CvM branch, especially RSMS-CvM (Late), whose cumulative return reaches 37.3% under the short 24-hour rule. At the same time, the practical-score comparison shows that HAC-CvM (Late) is the most balanced flat overlay and SSMS-CvM (Mid) is the most balanced short overlay once precision, capture, and drawdown are taken seriously. The appendix makes clear why these results look the way they do economically: alarm activity is sparse, many methods shadow the long-only benchmark for most of the sample, and only a subset of CvM rules generate enough clustered alarms to produce visible return dispersion. The empirical message is therefore aligned with the theoretical one: the self-normalized and e-process branches are complementary because they react to different temporal patterns of predictive deterioration and can serve different operational roles, from conservative benchmarking to more aggressive alarm-driven overlays. Natural extensions include fully data-adaptive tail-weighting schemes, richer multi-asset empirical panels, and continuum-of-τ monitoring.

Acknowledgements

The research is supported by National Natural Science Foundation of China (NSFC) grants No. 71988101 and No. 72173120.

References

Arnold, S., A. Henzi, and J. F. Ziegel (2023). Sequentially valid tests for forecast calibration. *The Annals of Applied Statistics 17*(3), 1909–1935.

Bouezmarni, T., M. Doukali, and A. Taamouti (2024). Testing Granger non-causality in expectiles. *Econometric Reviews 43*(1), 30–51.

Casgrain, P., M. Larsson, and J. F. Ziegel (2024). Sequential testing for elicitable functionals via supermartingales. *Bernoulli 30*(2), 1347–1374.

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

Choe, Y. J. and A. Ramdas (2024). Comparing sequential forecasters. *Operations Research 72*(4), 1368–1387.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*(4), 841–862.

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

Dette, H. and J. Gösmann (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.

Ehm, W., T. Gneiting, A. Jordan, and F. Krüger (2016). Of quantiles and expectiles: Consistent scoring functions, choquet representations and forecast rankings. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(3), 505–562.

Fissler, T. and J. F. Ziegel (2016). Higher order elicitability and osband's principle. *The Annals of Statistics 44*(4), 1680–1707.

Gaglianone, W. P., L. R. Lima, O. Linton, and D. R. Smith (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics 29*(1), 150–160.

Ghezzi, F., E. Rossi, and L. Trapani (2025). Fast on-line changepoint detection using heavily-weighted CUSUM and veto-based decision rules. *Journal of Econometrics 251*, 106071.

Giacomini, R. and I. Komunjer (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economic Statistics 23*(4), 416–431.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Gneiting, T. and J. Resin (2023). Regression diagnostics meets forecast evaluation: conditional calibration, reliability diagrams, and coefficient of determination. *Electronic Journal of Statistics 17*(2), 3226–3286.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*(3), 424–438.

Granger, C. W. J. (1980). Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control 2*, 329–352.

Grünwald, P., R. de Heide, and W. M. Koolen (2024). Safe testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 86*(5), 1091–1128.

Henzi, A. and J. F. Ziegel (2022). Valid sequential inference on probability forecast performance. *Biometrika 109*(3), 647–663.

Hoga, Y. (2017). Monitoring multivariate time series. *Journal of Multivariate Analysis 155*, 105–121.

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

Jeong, K., W. K. Härdle, and S. Song (2012). A consistent nonparametric test for causality in quantile. *Econometric Theory 28*(4), 861–887.

Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Newey, W. K. and J. L. Powell (1987). Asymmetric least squares estimation and testing. *Econometrica 55*(4), 819–847.

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. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association 110*(512), 1797–1817.

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

Song, X. and A. Taamouti (2021). Measuring Granger causality in quantiles. *Journal of Business & Economic Statistics 39*(4), 937–952.

Sun, J. and Y. Hong (2026a). Online monitoring of loading-space instability in large factor models via self-normalized DFT spectral scores. Working paper.

Sun, J. and Y. Hong (2026b). Online monitoring of loading-space instability in large factor models via self-normalized DFT spectral scores. Working paper, submitted to the *Journal of Econometrics*; available as SSRN 6359276.

Sun, J., M. Zhu, W. K. Härdle, and Z. Lin (2026). Online monitoring via streaming curves. Submitted to *Journal of Econometrics*.

Troster, V. (2018). Testing for Granger-causality in quantiles. *Econometric Reviews 37*(8), 850–866.

Wiener, N. (1956). The theory of prediction. In E. F. Beckenbach (Ed.), *Modern Mathematics for Engineers*, pp. 165–190. New York: McGraw–Hill.

Zhu, M., Y. Hong, J. Sun, and O. Linton (2025). Sequential change-point detection in time series: An adjusted-range-based self-normalization approach. Manuscript, revise and resubmit at Journal of Business & Economic Statistics.

A Proofs for the self-normalized and HAC results

Throughout the appendix, weak convergence statements are understood in the Skorokhod space $D(\cdot)$ with the usual J_1 topology. Because all Gaussian limits below have continuous sample paths, the continuous mapping theorem can be applied through the sup norm on compact sets.

A.1 Auxiliary weak limits

Recall that $\phi_t = \Psi_t - \bar{\Psi}_m$ and $S_m(k) = \sum_{t=m+1}^{m+k} \phi_t$. Define

$$G_m(r) = \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor mr \rfloor} (\Psi_t - \mu), \qquad r \in [0, 1+T],$$

so Assumption 2 states $G_m(\cdot) \Rightarrow \Omega^{1/2} B_q(\cdot)$.

Lemma A.1 (Training bridge process and D_m)**.** *Under Assumption 2:*

(i) the training bridge process

$$H_m(r) = \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor mr \rfloor} \phi_t, \qquad r \in [0, 1],$$

satisfies $H_m(\cdot) \Rightarrow \Omega^{1/2} B_q^0(\cdot)$ in $D([0,1], \mathbb{R}^q)$;

(ii) the quadratic self-normalizer obeys

$$D_m = \frac{1}{m^2} \sum_{t=1}^{m} \left(\sum_{j=1}^{t} \phi_j \right) \left(\sum_{j=1}^{t} \phi_j \right)^\top \xrightarrow{d} \Omega^{1/2} V \Omega^{1/2};$$

(iii) consequently, $\mathbb{P}(\det D_m > 0) \to 1$ and

$$D_m^{-1} \xrightarrow{d} \Omega^{-1/2} V^{-1} \Omega^{-1/2}.$$

Proof. Write $\phi_t = (\Psi_t - \mu) - (\bar{\Psi}_m - \mu)$ and note that

$$\sqrt{m}(\bar{\Psi}_m - \mu) = G_m(1) + o_p(1).$$

Hence, for $r \in [0, 1]$,

$$\begin{aligned} H_m(r) &= \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor mr \rfloor} (\Psi_t - \mu) - \frac{\lfloor mr \rfloor}{m} \cdot \frac{1}{\sqrt{m}} \sum_{t=1}^{m} (\Psi_t - \mu) \\ &= G_m(r) - \frac{\lfloor mr \rfloor}{m} G_m(1). \end{aligned} \tag{A.1}$$

Because $\sup_{r\in[0,1]} |\lfloor mr\rfloor/m - r| \le 1/m$ and $\sup_{r\in[0,1]} \|G_m(r)\| = O_p(1)$ under the FCLT,

$$\sup_{r\in[0,1]} \big\|H_m(r) - \big(G_m(r) - rG_m(1)\big)\big\| = o_p(1).$$

The map $x(\cdot) \mapsto x(\cdot) - (\cdot)x(1)$ is continuous at continuous limits, so Assumption 2 yields

$$H_m(\cdot) \Rightarrow \Omega^{1/2} B_q^0(\cdot), \qquad B_q^0(r) = B_q(r) - rB_q(1),$$

which proves (i).

Let $S_t^{\mathrm{tr}} = \sum_{j=1}^t \phi_j$. Then

$$\begin{aligned} D_m &= \frac{1}{m^2}\sum_{t=1}^m S_t^{\mathrm{tr}}(S_t^{\mathrm{tr}})^\top = \frac{1}{m}\sum_{t=1}^m \Big(\frac{S_t^{\mathrm{tr}}}{\sqrt{m}}\Big)\Big(\frac{S_t^{\mathrm{tr}}}{\sqrt{m}}\Big)^\top \\ &= \int_0^1 \Big(\frac{S_{\lceil mr\rceil}^{\mathrm{tr}}}{\sqrt{m}}\Big)\Big(\frac{S_{\lceil mr\rceil}^{\mathrm{tr}}}{\sqrt{m}}\Big)^\top dr. \end{aligned} \tag{A.2}$$

Moreover,

$$\sup_{r\in[0,1]} \left\| \frac{S_{\lceil mr\rceil}^{\mathrm{tr}}}{\sqrt{m}} - H_m(r) \right\| \le \frac{1}{\sqrt{m}} \max_{1\le t\le m} \|\phi_t\| = o_p(1),$$

under the moment condition implicit in Assumption 2. Hence the integrand in (A.2) can be replaced by $H_m(r)H_m(r)^\top$ at $o_p(1)$ cost. The map $x(\cdot) \mapsto \int_0^1 x(r)x(r)^\top dr$ is continuous on $C([0,1],\mathbb{R}^q)$, so part (i) implies

$$D_m \xrightarrow{d} \int_0^1 \Omega^{1/2} B_q^0(r) B_q^0(r)^\top \Omega^{1/2} dr = \Omega^{1/2} V \Omega^{1/2},$$

which proves (ii).

By Lemma A.3, V is positive definite almost surely. Therefore $\Omega^{1/2}V\Omega^{1/2}$ is positive definite almost surely, and matrix inversion is continuous on the set of invertible matrices. This yields (iii). □

Lemma A.2 (Monitoring partial sums)**.** *Under Assumption 2, in* $D([0,T],\mathbb{R}^q)$*,*

$$\left\{\frac{1}{\sqrt{m}} S_m(\lfloor ms\rfloor)\right\}_{0\le s\le T} \Rightarrow \left\{\Omega^{1/2} U_q(s)\right\}_{0\le s\le T}, \qquad U_q(s) = B_q(1+s) - (1+s)B_q(1).$$

Proof. For $k = \lfloor ms\rfloor$,

$$\begin{aligned} \frac{1}{\sqrt{m}} S_m(k) &= \frac{1}{\sqrt{m}} \sum_{t=m+1}^{m+k} (\Psi_t - \mu) - \frac{k}{m}\cdot\frac{1}{\sqrt{m}}\sum_{t=1}^m (\Psi_t - \mu) \\ &= \Big(G_m(1+s) - G_m(1)\Big) - sG_m(1) + o_p(1) \\ &= G_m(1+s) - (1+s)G_m(1) + o_p(1), \end{aligned}$$

where the $o_p(1)$ term is uniform in $s \in [0,T]$ and comes only from the floor operation. The map $x(\cdot) \mapsto x(1+\cdot) - (1+\cdot)x(1)$ is continuous at continuous limits, so Assumption 2 implies the stated convergence. □

Lemma A.3 (Invertibility of V)**.** *The matrix $V = \int_0^1 B_q^0(r) B_q^0(r)^\top dr$ is almost surely positive definite.*

Proof. Fix $a \in \mathbb{R}^q$ with $a \neq 0$ and consider the scalar Gaussian bridge $Z_a(r) = a^\top B_q^0(r)$. For any $r \in (0,1)$,

$$\mathrm{Var}\big(Z_a(r)\big) = r(1-r)a^\top a > 0,$$

so $Z_a(r)$ is nondegenerate. Since $Z_a(\cdot)$ has continuous sample paths and is not almost surely identically zero, there exists a neighborhood of some $r^\star \in (0,1)$ on which $Z_a(r)^2 > 0$ with probability one. Therefore

$$a^\top V a = \int_0^1 (a^\top B_q^0(r))^2 dr = \int_0^1 Z_a(r)^2 dr > 0$$

almost surely. Because this holds for every $a \neq 0$, the matrix V is positive definite almost surely. □

Lemma A.4 (Weighted Riemann-sum approximation)**.** *Let $f_m : [0,T] \to \mathbb{R}$ be random functions and let $f : [0,T] \to \mathbb{R}$ be continuous. Suppose*

$$\sup_{0 \le s \le T} |f_m(s) - f(s)| \to_p 0.$$

If $w : [0,T] \to \mathbb{R}$ is bounded and Riemann integrable, then

$$R_m(s) := \frac{1}{m} \sum_{j=1}^{\lfloor ms \rfloor} w(j/m) f_m(j/m)$$

satisfies

$$\sup_{0 \le s \le T} \left| R_m(s) - \int_0^s w(u) f(u)\, du \right| \to_p 0.$$

Proof. Fix $s \in [0,T]$ and decompose

$$\begin{aligned} R_m(s) - \int_0^s w(u) f(u)\, du = \frac{1}{m} \sum_{j=1}^{\lfloor ms \rfloor} w(j/m)\big(f_m(j/m) - f(j/m)\big) \\ + \left[\frac{1}{m} \sum_{j=1}^{\lfloor ms \rfloor} w(j/m) f(j/m) - \int_0^s w(u) f(u)\, du \right]. \end{aligned} \tag{A.3}$$

For the first term, let $\|w\|_\infty = \sup_{0 \le u \le T} |w(u)| < \infty$. Then, uniformly in s,

$$\left| \frac{1}{m} \sum_{j=1}^{\lfloor ms \rfloor} w(j/m)\big(f_m(j/m) - f(j/m)\big) \right| \le \|w\|_\infty \frac{\lfloor mT \rfloor}{m} \sup_{0 \le u \le T} |f_m(u) - f(u)|.$$

Because $\lfloor mT \rfloor / m \le T$, the right-hand side is $o_p(1)$.

For the second term in (A.3), the integrand $u \mapsto w(u)f(u)$ is Riemann integrable because w is Riemann integrable and f is continuous. Hence the deterministic Riemann sums converge uniformly:

$$\sup_{0 \le s \le T} \left| \frac{1}{m} \sum_{j=1}^{\lfloor ms \rfloor} w(j/m) f(j/m) - \int_0^s w(u) f(u)\, du \right| \to 0.$$

Combining the two bounds proves the lemma. □

A.2 Proof of Theorem 1

Proof. We divide the argument into four steps.

Lemma A.2 yields

$$\left\{ m^{-1/2} S_m(\lfloor ms \rfloor) \right\}_{0 \le s \le T} \Rightarrow \left\{ \Omega^{1/2} U_q(s) \right\}_{0 \le s \le T} \quad \text{in } D([0,T], \mathbb{R}^q).$$

Because the limit has continuous paths, we may work with the uniform topology on $[0,T]$.

Assumption 5 gives $\widehat{\Omega}_m \xrightarrow{p} \Omega$, hence

$$\widehat{\Omega}_m^{-1} \xrightarrow{p} \Omega^{-1}$$

by continuity of matrix inversion on the cone of positive definite matrices. Therefore the pair

$$\left(\left\{ m^{-1/2} S_m(\lfloor ms \rfloor) \right\}_{0 \le s \le T},\ \widehat{\Omega}_m^{-1} \right)$$

converges jointly to

$$\left(\left\{ \Omega^{1/2} U_q(s) \right\}_{0 \le s \le T},\ \Omega^{-1} \right).$$

For $s \in [0,T]$, let $s_m = \lfloor ms \rfloor / m$ and define

$$\Phi_m(s) = \frac{\left(m^{-1/2} S_m(\lfloor ms \rfloor)\right)^\top \widehat{\Omega}_m^{-1} \left(m^{-1/2} S_m(\lfloor ms \rfloor)\right)}{(1+s_m)^2 \left(\frac{s_m}{1+s_m}\right)^{2\gamma}}.$$

By the joint convergence above,

$$\Phi_m(s) \Rightarrow \Phi(s) := \frac{U_q(s)^\top U_q(s)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}} \qquad \text{for each } s > 0.$$

To upgrade this to a process statement, note that the map

$$(x, A) \mapsto \left\{ \frac{x(s)^\top A x(s)}{(1+s)^2 (s/(1+s))^{2\gamma}} \right\}_{0 < s \le T}$$

is continuous whenever $x(\cdot)$ is continuous and A is positive definite.

It remains to check continuity at $s = 0$. Since $U_q(0) = 0$ and Brownian increments satisfy $\|U_q(s)\|_2^2 = O(s)$ almost surely as $s \downarrow 0$, the numerator is of order s, whereas the denominator behaves like $s^{2\gamma}$ up to a bounded multiplicative factor. Because $\gamma < 1/2$, we have $s^{1-2\gamma} \to 0$, so $\Phi(s) \to 0$ as $s \downarrow 0$. Hence Φ extends continuously to $[0, T]$ by setting $\Phi(0) = 0$. The supremum functional is continuous on $C([0, T])$, so

$$\sup_{1 \le k \le mT} \mathcal{M}_m^{\mathrm{H}}(k) = \sup_{0 < s \le T} \Phi_m(s) \xrightarrow{d} \sup_{0 < s \le T} \Phi(s),$$

which is exactly the asserted limit. □

A.3 Proof of Theorem 2

Proof. The proof is slightly more delicate than the HAC proof because the studentizer is random and constructed from the same training path.

Let $S_m^{(0)}(k)$ and $D_m^{(0)}$ denote the monitoring partial sum and training self-normalizer computed from the oracle scores $\Psi_t^{(0)}$. Assumption 3 implies

$$\sup_{0 \le s \le T} \left\| m^{-1/2} S_m(\lfloor ms \rfloor) - m^{-1/2} S_m^{(0)}(\lfloor ms \rfloor) \right\| = o_p(1)$$

and

$$\left\| D_m - D_m^{(0)} \right\| = o_p(1).$$

Hence it is enough to prove the theorem for the oracle version, after which the feasible statistic follows by Slutsky's theorem. To lighten notation, we now suppress the oracle superscript.

Lemma A.2 yields

$$\left\{ m^{-1/2} S_m(\lfloor ms \rfloor) \right\}_{0 \le s \le T} \Rightarrow \left\{ \Omega^{1/2} U_q(s) \right\}_{0 \le s \le T}.$$

Lemma A.1 gives

$$D_m^{-1} \xrightarrow{d} \Omega^{-1/2} V^{-1} \Omega^{-1/2}.$$

Because both objects are measurable functionals of the same score array, joint convergence holds:

$$\left(\left\{ m^{-1/2} S_m(\lfloor ms \rfloor) \right\}_{0 \le s \le T}, D_m^{-1} \right) \Rightarrow \left(\left\{ \Omega^{1/2} U_q(s) \right\}_{0 \le s \le T}, \Omega^{-1/2} V^{-1} \Omega^{-1/2} \right).$$

For $s > 0$ with $s_m = \lfloor ms \rfloor / m$,

$$\mathcal{M}_m^{\mathrm{S}}(\lfloor ms \rfloor) = \frac{\left(m^{-1/2} S_m(\lfloor ms \rfloor) \right)^\top D_m^{-1} \left(m^{-1/2} S_m(\lfloor ms \rfloor) \right)}{(1 + s_m)^2 \left(\frac{s_m}{1 + s_m} \right)^{2\gamma}}.$$

Applying the continuous mapping theorem to the jointly convergent pair above yields the pointwise

limit

$$\mathcal{M}_m^{\mathrm{S}}(\lfloor ms \rfloor) \Rightarrow \frac{\left(\Omega^{1/2} U_q(s)\right)^\top \Omega^{-1/2} V^{-1} \Omega^{-1/2} \left(\Omega^{1/2} U_q(s)\right)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}}$$
$$= \frac{U_q(s)^\top V^{-1} U_q(s)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}}.$$

This is the pivotal nuisance cancellation that makes SSMS attractive.

Define

$$\Phi^{\mathrm{S}}(s) = \frac{U_q(s)^\top V^{-1} U_q(s)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}}, \qquad s > 0.$$

As in the HAC proof, $\|U_q(s)\|_2^2 = O(s)$ almost surely near zero, while the denominator behaves like $s^{2\gamma}$ up to a bounded factor. Since $\gamma < 1/2$, $\Phi^{\mathrm{S}}(s) \to 0$ as $s \downarrow 0$. Thus the limit process extends continuously to $[0, T]$ by setting $\Phi^{\mathrm{S}}(0) = 0$. The supremum map is continuous on $C([0, T])$, so

$$\sup_{1 \le k \le mT} \mathcal{M}_m^{\mathrm{S}}(k) = \sup_{0 < s \le T} \mathcal{M}_m^{\mathrm{S}}(\lfloor ms \rfloor) \xrightarrow{d} \sup_{0 < s \le T} \Phi^{\mathrm{S}}(s),$$

which proves the theorem. □

A.4 Proof of Theorem 3

Proof. Under Assumption 4, after the one-time Phase-I whitening step we may analyze a transformed score sequence whose long-run covariance is diagonal. We write the transformed quantities again as $\widetilde{\phi}_t$ and $\widetilde{S}_m(k)$ for notational simplicity, and denote the diagonal long-run covariance by

$$\Omega = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_q^2).$$

Lemma A.2, applied to the whitened score process, yields

$$\left\{ m^{-1/2} \widetilde{S}_m(\lfloor ms \rfloor) \right\}_{0 \le s \le T} \Rightarrow \left\{ \Omega^{1/2} U_q(s) \right\}_{0 \le s \le T}.$$

For each coordinate ℓ, define the scaled training bridge

$$\widetilde{H}_{m,\ell}(r) = \frac{1}{\sqrt{m}} \widetilde{B}_{m,\ell}(\lfloor mr \rfloor), \qquad r \in [0, 1].$$

By the same bridge argument as in Lemma A.1,

$$\widetilde{H}_{m,\ell}(\cdot) \Rightarrow \sigma_\ell B_\ell^0(\cdot) \quad \text{in } D([0,1], \mathbb{R}).$$

The map $x(\cdot) \mapsto \sup_{0\le r\le 1} x(r) - \inf_{0\le r\le 1} x(r)$ is continuous on $C([0,1])$, so

$$\widetilde{R}_{m,\ell} = \sup_{0\le r\le 1} \widetilde{H}_{m,\ell}(r) - \inf_{0\le r\le 1} \widetilde{H}_{m,\ell}(r) \xrightarrow{d} \sigma_\ell R_\ell.$$

Therefore

$$\widetilde{R}_m^{-2} \xrightarrow{d} \Omega^{-1/2} R^{-2} \Omega^{-1/2}.$$

Joint convergence of the monitoring partial sums and adjusted ranges gives, for $s > 0$,

$$\begin{aligned}
\widetilde{\mathcal{M}}_m^{\mathrm{R}}(\lfloor ms \rfloor) &= \frac{\big(m^{-1/2}\widetilde{S}_m(\lfloor ms \rfloor)\big)^\top \widetilde{R}_m^{-2} \big(m^{-1/2}\widetilde{S}_m(\lfloor ms \rfloor)\big)}{(1+s_m)^2 \left(\frac{s_m}{1+s_m}\right)^{2\gamma}} \\
&\Rightarrow \frac{\big(\Omega^{1/2}U_q(s)\big)^\top \Omega^{-1/2} R^{-2} \Omega^{-1/2} \big(\Omega^{1/2}U_q(s)\big)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}} \\
&= \frac{U_q(s)^\top R^{-2} U_q(s)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}},
\end{aligned}$$

where $s_m = \lfloor ms \rfloor / m$.

Because each $R_\ell > 0$ almost surely, the limit is well defined. The same small-s argument used in the previous proofs shows that the limit process extends continuously to zero when $\gamma < 1/2$. Hence the supremum functional is continuous and

$$\sup_{1\le k\le mT} \widetilde{\mathcal{M}}_m^{\mathrm{R}}(k) \xrightarrow{d} \sup_{0<s\le T} \frac{U_q(s)^\top R^{-2} U_q(s)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}},$$

which proves the theorem. □

A.5 Proof of Theorem 4

Proof. Set

$$\mathcal{M}^{\mathrm{H}}(s) = \frac{\|U_q(s)\|_2^2}{(1+s)^2}, \qquad s \in [0,T].$$

By Theorem 1 with $\gamma = 0$,

$$\sup_{0\le s\le T} \left|\mathcal{M}_m^{\mathrm{H}}(\lfloor ms \rfloor) - \mathcal{M}^{\mathrm{H}}(s)\right| \to_p 0.$$

Define $f_m(s) = \mathcal{M}_m^{\mathrm{H}}(\lfloor ms \rfloor)$ and $f(s) = \mathcal{M}^{\mathrm{H}}(s)$. Then Lemma A.4 applies directly because w is bounded and Riemann integrable by Assumption 6. Therefore,

$$\sup_{0\le s\le T} \left| \frac{1}{m} \sum_{j=1}^{\lfloor ms \rfloor} w(j/m) \mathcal{M}_m^{\mathrm{H}}(j) - \int_0^s w(u) \mathcal{M}^{\mathrm{H}}(u)\, du \right| \to_p 0.$$

The left-hand side is exactly

$$\sup_{0\le s\le T}\left|\mathcal{I}_m^{\mathrm{H}}(\lfloor ms\rfloor)-\int_0^s w(u)\frac{\|U_q(u)\|_2^2}{(1+u)^2}\,du\right|,$$

which proves the claimed functional convergence. □

A.6 Proof of Theorem 5

Proof. Let

$$\mathcal{M}^{\mathrm{S}}(s)=\frac{U_q(s)^\top V^{-1}U_q(s)}{(1+s)^2},\qquad s\in[0,T].$$

By Theorem 2 with $\gamma=0$,

$$\sup_{0\le s\le T}\left|\mathcal{M}_m^{\mathrm{S}}(\lfloor ms\rfloor)-\mathcal{M}^{\mathrm{S}}(s)\right|\to_p 0.$$

Apply Lemma A.4 with $f_m(s)=\mathcal{M}_m^{\mathrm{S}}(\lfloor ms\rfloor)$ and $f(s)=\mathcal{M}^{\mathrm{S}}(s)$. Since w satisfies Assumption 6, we obtain

$$\sup_{0\le s\le T}\left|\mathcal{I}_m^{\mathrm{S}}(\lfloor ms\rfloor)-\int_0^s w(u)\mathcal{M}^{\mathrm{S}}(u)\,du\right|\to_p 0.$$

Substituting the definition of $\mathcal{M}^{\mathrm{S}}$ yields

$$\mathcal{I}_m^{\mathrm{S}}(\lfloor ms\rfloor)\xrightarrow{d}\int_0^s w(u)\frac{U_q(u)^\top V^{-1}U_q(u)}{(1+u)^2}\,du,\qquad s\in[0,T],$$

as asserted. □

A.7 Proof of Theorem 6

Proof. Define

$$\mathcal{M}^{\mathrm{R}}(s)=\frac{U_q(s)^\top R^{-2}U_q(s)}{(1+s)^2},\qquad s\in[0,T].$$

Theorem 3 with $\gamma=0$ implies

$$\sup_{0\le s\le T}\left|\widetilde{\mathcal{M}}_m^{\mathrm{R}}(\lfloor ms\rfloor)-\mathcal{M}^{\mathrm{R}}(s)\right|\to_p 0.$$

Applying Lemma A.4 with $f_m(s)=\widetilde{\mathcal{M}}_m^{\mathrm{R}}(\lfloor ms\rfloor)$ and $f(s)=\mathcal{M}^{\mathrm{R}}(s)$ gives

$$\sup_{0\le s\le T}\left|\mathcal{I}_m^{\mathrm{R}}(\lfloor ms\rfloor)-\int_0^s w(u)\mathcal{M}^{\mathrm{R}}(u)\,du\right|\to_p 0.$$

Hence

$$\mathcal{I}_m^{\mathrm{R}}(\lfloor ms\rfloor)\xrightarrow{d}\int_0^s w(u)\frac{U_q(u)^\top R^{-2}U_q(u)}{(1+u)^2}\,du,\qquad s\in[0,T],$$

which proves the theorem. □

A.8 Proof of Theorem 9

Proof. We give the argument for SSMS; the RSMS proof is identical after replacing D_m by the adjusted-range normalizer. Let $k^\star = \lfloor ms^\star \rfloor$.

Write

$$S_m(k) = \sum_{j=1}^{k} \big(\phi_{m+j} - \mathbb{E}\,\phi_{m+j}\big) + \sum_{j=1}^{k} \mathbb{E}\,\phi_{m+j} =: A_m(k) + B_m(k).$$

Because the training window is stable, $\bar{\Psi}_m \xrightarrow{p} \mu$ and therefore $\mathbb{E}\,\phi_{m+j} = \mathbb{E}\,\Psi_{m+j} - \mu + o(1)$. Under Assumption 7, for $j > k^\star$,

$$\mathbb{E}\,\phi_{m+j} = \Delta + o(1),$$

whereas for $j \le k^\star$ the mean is $o(1)$. Hence, for any fixed $s_1 \in (s^\star, T]$ and $k_1 = \lfloor ms_1 \rfloor$,

$$B_m(k_1) = \sum_{j=k^\star+1}^{k_1} \{\Delta + o(1)\} = (k_1 - k^\star)\Delta + o(m).$$

Since $k_1 - k^\star \asymp m$, we obtain $\|B_m(k_1)\| \asymp m$.

The centered component $A_m(k)$ satisfies the same weak-dependence conditions as under the null, so

$$\sup_{0 \le s \le T} \Big\| m^{-1/2} A_m(\lfloor ms \rfloor) \Big\| = O_p(1).$$

Therefore,

$$m^{-1/2} S_m(k_1) = m^{-1/2} A_m(k_1) + m^{-1/2} B_m(k_1) = O_p(1) + \sqrt{m}\,(s_1 - s^\star)\Delta + o(\sqrt{m}),$$

which implies

$$\|m^{-1/2} S_m(k_1)\| \to_p \infty.$$

The matrix D_m is computed from the stable training sample only. By Lemma A.1,

$$D_m \xrightarrow{d} \Omega^{1/2} V \Omega^{1/2}, \qquad D_m^{-1} = O_p(1).$$

Hence

$$\mathcal{M}_m^{\mathrm{S}}(k_1) = \frac{\big(m^{-1/2} S_m(k_1)\big)^\top D_m^{-1} \big(m^{-1/2} S_m(k_1)\big)}{(1 + k_1/m)^2 \left(\frac{k_1}{k_1+m}\right)^{2\gamma}} \to_p \infty,$$

because the denominator is a fixed positive constant at $s_1 > 0$.

Since the critical value is fixed,

$$\mathbb{P}\left(\sup_{1 \le k \le mT} \mathcal{M}_m^{\mathrm{S}}(k) > c_\alpha^{\mathrm{S}}(T, q, \gamma) \right) \to 1.$$

This proves consistency of the SSMS KS detector. For the CvM rule, use the additional assumption

that there exists an interval $J \subset (s^\star, T]$ of positive length on which $w(u) \geq c_w > 0$. By the same argument as above, for every fixed $s \in J$,

$$\mathcal{M}_m^{\mathrm{S}}(\lfloor ms \rfloor) \to_p \infty.$$

Because the statistic is nonnegative,

$$\mathcal{I}_m^{\mathrm{S}}(\lfloor m \sup J \rfloor) \geq \frac{1}{m} \sum_{j:\, j/m \in J} w(j/m) \mathcal{M}_m^{\mathrm{S}}(j) \geq c_w \frac{1}{m} \sum_{j:\, j/m \in J} \mathcal{M}_m^{\mathrm{S}}(j).$$

The set $\{j : j/m \in J\}$ contains order-m indices, and on each fixed compact subinterval of J the summands diverge in probability. Consequently the lower bound diverges in probability, so the CvM detector is also consistent. The RSMS argument is unchanged except for replacing the quadratic self-normalizer by the adjusted-range normalizer. □

A.9 Proof of Theorem 10

Proof. Again we present the argument for SSMS; the RSMS and CvM cases follow by the same substitutions as under the null.

Let $k^\star = \lfloor ms^\star \rfloor$. For $k = \lfloor ms \rfloor$,

$$S_m(k) = \sum_{t=m+1}^{m+k} \big(\Psi_t - \mathbb{E}\,\Psi_t\big) - k(\bar{\Psi}_m - \mu) + \sum_{t=m+1}^{m+k} \big(\mathbb{E}\,\Psi_t - \mu\big).$$

The first two terms are exactly the centered terms that generated the null limit in Lemma A.2. Hence, after scaling by $\sqrt{m}$,

$$\left\{ \frac{1}{\sqrt{m}} \left[\sum_{t=m+1}^{m+\lfloor ms \rfloor} \big(\Psi_t - \mathbb{E}\,\Psi_t\big) - \lfloor ms \rfloor (\bar{\Psi}_m - \mu) \right] \right\}_{0 \leq s \leq T} \Rightarrow \{\Omega^{1/2} U_q(s)\}_{0 \leq s \leq T}.$$

Under Definition 1,

$$\mathbb{E}\,\Psi_t - \mu = \begin{cases} 0, & t \leq m + k^\star, \\ \delta/\sqrt{m}, & t > m + k^\star. \end{cases}$$

Therefore, for $k = \lfloor ms \rfloor$,

$$\begin{aligned} \frac{1}{\sqrt{m}} \sum_{t=m+1}^{m+k} \big(\mathbb{E}\,\Psi_t - \mu\big) &= \frac{(\lfloor ms \rfloor - k^\star)_+}{\sqrt{m}} \cdot \frac{\delta}{\sqrt{m}} \\ &\to (s - s^\star)_+ \delta, \end{aligned}$$

uniformly on compact subsets of $[0, T]$.

Adding the centered term and the deterministic drift gives

$$\left\{\frac{1}{\sqrt{m}}S_m(\lfloor ms\rfloor)\right\}_{0\le s\le T} \Rightarrow \left\{\Omega^{1/2}\big(U_q(s)+(s-s^\star)_+\delta\big)\right\}_{0\le s\le T}.$$

By Assumption 3, the same limit holds for the feasible statistic. The training-window self-normalizer is unaffected by the monitoring-period local drift, so Lemma A.1 still gives

$$D_m^{-1} \xrightarrow{d} \Omega^{-1/2}V^{-1}\Omega^{-1/2}.$$

Proceeding exactly as in the proof of Theorem 2, we obtain

$$\begin{aligned}\mathcal{M}_m^{\mathrm{S}}(\lfloor ms\rfloor) &\Rightarrow \frac{\big(\Omega^{1/2}(U_q(s)+(s-s^\star)_+\delta)\big)^\top\Omega^{-1/2}V^{-1}\Omega^{-1/2}\big(\Omega^{1/2}(U_q(s)+(s-s^\star)_+\delta)\big)}{(1+s)^2\left(\frac{s}{1+s}\right)^{2\gamma}}\\ &= \frac{\big(U_q(s)+(s-s^\star)_+\delta\big)^\top V^{-1}\big(U_q(s)+(s-s^\star)_+\delta\big)}{(1+s)^2\left(\frac{s}{1+s}\right)^{2\gamma}}.\end{aligned}$$

Taking the supremum over $s\in(0,T]$ gives the asserted drifted Brownian limit. The RSMS and CvM results follow by replacing V^{-1} with R^{-2} or integrating the corresponding path against w. □

B Proofs for the e-process results

This appendix proves the finite-sample validity statements for the e-process branch. Unlike the self-normalized results, the arguments here are martingale arguments rather than weak-convergence arguments, so the proofs proceed by conditioning on the past and verifying the supermartingale recursion one step at a time.

B.1 Proof of Theorem 11

Proof. Fix $\tau\in\mathcal{T}_m$ and a predictable direction θ.

Under the operational quantile null,

$$\mathbb{P}\big(I_t(\tau)=1\mid\mathcal{H}_{t-1}\big)=\tau,\qquad \mathbb{P}\big(I_t(\tau)=0\mid\mathcal{H}_{t-1}\big)=1-\tau.$$

Conditioning on $\mathcal{H}_{t-1}$ and using the definition of $L_t(\theta,\tau)$,

$$\begin{aligned}\mathbb{E}\left[L_t(\theta,\tau)\mid\mathcal{H}_{t-1}\right] &= \tau\left(\frac{p_t(\theta,\tau)}{\tau}\right)+(1-\tau)\left(\frac{1-p_t(\theta,\tau)}{1-\tau}\right)\\ &= p_t(\theta,\tau)+1-p_t(\theta,\tau)=1.\end{aligned}$$

Because θ is predictable, $L_t(\theta,\tau)$ is $\mathcal{H}_t$-measurable and nonnegative. Writing

$$\mathcal{E}_k(\theta,\tau) = \mathcal{E}_{k-1}(\theta,\tau)\, L_{m+k}(\theta,\tau),$$

we obtain

$$\begin{aligned}\mathbb{E}\left[\mathcal{E}_k(\theta,\tau) \mid \mathcal{H}_{m+k-1}\right] &= \mathcal{E}_{k-1}(\theta,\tau)\,\mathbb{E}\left[L_{m+k}(\theta,\tau) \mid \mathcal{H}_{m+k-1}\right] \\ &= \mathcal{E}_{k-1}(\theta,\tau).\end{aligned}$$

Thus $\{\mathcal{E}_k(\theta,\tau)\}_{k\ge 0}$ is a nonnegative martingale with $\mathbb{E}\,\mathcal{E}_k(\theta,\tau) = 1$ for every k.

Ville's inequality for nonnegative supermartingales now gives

$$\mathbb{P}\Big(\sup_{k\ge 1} \mathcal{E}_k(\theta,\tau) \ge \alpha^{-1}\Big) \le \alpha.$$

Since the event $\{\sup_{1\le j\le K} \mathcal{E}_j(\theta,\tau) \ge \alpha^{-1}\}$ is contained in $\{\sup_{k\ge 1} \mathcal{E}_k(\theta,\tau) \ge \alpha^{-1}\}$, the same bound holds on every finite monitoring horizon. This proves the theorem. □

B.2 Proof of Proposition 2

Proof. We verify each construction separately.

Mixtures. For every fixed direction $\theta^{(\ell)}$, Theorem 11 shows that $\{\mathcal{E}_k(\theta^{(\ell)},\tau)\}_{k\ge 0}$ is a nonnegative martingale with unit mean. Therefore, with deterministic weights $\pi_\ell \ge 0$ satisfying $\sum_\ell \pi_\ell = 1$,

$$\mathcal{E}_k^{\mathrm{mix}}(\tau) = \sum_{\ell=1}^{L} \pi_\ell \mathcal{E}_k(\theta^{(\ell)},\tau)$$

is nonnegative and

$$\mathbb{E}\left[\mathcal{E}_k^{\mathrm{mix}}(\tau) \mid \mathcal{H}_{m+k-1}\right] = \sum_{\ell=1}^{L} \pi_\ell\, \mathbb{E}\left[\mathcal{E}_k(\theta^{(\ell)},\tau) \mid \mathcal{H}_{m+k-1}\right] = \sum_{\ell=1}^{L} \pi_\ell \mathcal{E}_{k-1}(\theta^{(\ell)},\tau) = \mathcal{E}_{k-1}^{\mathrm{mix}}(\tau).$$

Hence the mixture is itself an e-process.

Predictable adaptation. Suppose now that θ_{t-1} is $\mathcal{H}_{t-1}$-measurable. Conditional on $\mathcal{H}_{t-1}$, the value of θ_{t-1} is fixed, so the calculation in the proof of Theorem 11 gives

$$\mathbb{E}\left[L_t(\theta_{t-1},\tau) \mid \mathcal{H}_{t-1}\right] = 1.$$

Therefore

$$\mathbb{E}\left[\mathcal{E}_k^{\mathrm{ad}}(\tau) \mid \mathcal{H}_{m+k-1}\right] = \mathcal{E}_{k-1}^{\mathrm{ad}}(\tau)\,\mathbb{E}\left[L_{m+k}(\theta_{m+k-1},\tau) \mid \mathcal{H}_{m+k-1}\right] = \mathcal{E}_{k-1}^{\mathrm{ad}}(\tau),$$

so the adaptive product is again an e-process.

Restart banks. Fix a restart time $r \in \mathcal{R}$. By construction, $\mathcal{E}_{r:k}(\tau) = 1$ for $k < r$, and for $k \geq r$ it is the product of valid one-step factors from time $m + r$ onward. Thus, from time r onward, it is just another e-process started at one. A convex combination of such restarted e-processes,

$$\mathcal{E}_k^{\text{bank}}(\tau) = \sum_{r \in \mathcal{R}} \pi_r \mathcal{E}_{r:k}(\tau),$$

inherits the e-process property by the same linearity argument as for mixtures.

Tail combinations. For each grid point τ_j, the process $\{\mathcal{E}_k(\tau_j)\}_{k \geq 0}$ is an e-process. If $\omega_j \geq 0$ and $\sum_j \omega_j = 1$, then

$$\mathcal{E}_k^{\text{tail}} = \sum_{j=1}^{J} \omega_j \mathcal{E}_k(\tau_j)$$

is nonnegative and satisfies

$$\mathbb{E}\left[\mathcal{E}_k^{\text{tail}} \mid \mathcal{H}_{m+k-1}\right] = \sum_{j=1}^{J} \omega_j \, \mathbb{E}\left[\mathcal{E}_k(\tau_j) \mid \mathcal{H}_{m+k-1}\right] = \sum_{j=1}^{J} \omega_j \mathcal{E}_{k-1}(\tau_j) = \mathcal{E}_{k-1}^{\text{tail}}.$$

Hence it is also an e-process.

Finally, once each process has been shown to be an e-process, Ville's inequality implies that any stopping rule of the form $\inf\{k : \mathcal{E}_k \geq \alpha^{-1}\}$ controls false alarms at level α. □

B.3 Proof of Proposition 3

Proof. Define the one-step multiplicative factor

$$M_t(\tau) = \exp\Big\{\lambda_{t-1} x_t(\tau) - \psi_t(\lambda_{t-1}, \tau)\Big\}.$$

Because λ_{t-1} is predictable, both λ_{t-1} and $\psi_t(\lambda_{t-1}, \tau)$ are $\mathcal{H}_{t-1}$-measurable. Therefore,

$$\begin{aligned}
\mathbb{E}\left[M_t(\tau) \mid \mathcal{H}_{t-1}\right] &= \exp\big\{-\psi_t(\lambda_{t-1}, \tau)\big\}\, \mathbb{E}\Big[\exp\{\lambda_{t-1} x_t(\tau)\} \mid \mathcal{H}_{t-1}\Big] \\
&\leq \exp\big\{-\psi_t(\lambda_{t-1}, \tau)\big\} \exp\big\{\psi_t(\lambda_{t-1}, \tau)\big\} \\
&= 1,
\end{aligned}$$

where the inequality is exactly the conditional mgf bound assumed in the proposition.

Now write

$$\mathcal{E}_k^E(\tau) = \mathcal{E}_{k-1}^E(\tau)\, M_{m+k}(\tau).$$

Conditioning on $\mathcal{H}_{m+k-1}$ gives

$$\mathbb{E}\left[\mathcal{E}_k^E(\tau) \mid \mathcal{H}_{m+k-1}\right] = \mathcal{E}_{k-1}^E(\tau)\,\mathbb{E}\left[M_{m+k}(\tau) \mid \mathcal{H}_{m+k-1}\right] \le \mathcal{E}_{k-1}^E(\tau).$$

Thus $\{\mathcal{E}_k^E(\tau)\}_{k\ge 0}$ is a nonnegative supermartingale and therefore an e-process.

In the conditional sub-Gaussian case the standard bound

$$\mathbb{E}\left[\exp\{\lambda x_t(\tau)\} \mid \mathcal{H}_{t-1}\right] \le \exp\{\lambda^2 \nu_t^2(\tau)/2\}$$

shows that one may set $\psi_t(\lambda, \tau) = \lambda^2 \nu_t^2(\tau)/2$. This is just the familiar exponential-supermartingale construction specialized to the score projection $x_t(\tau)$. □

C Open-end monitoring and the limit $T = \infty$

This appendix gives a more detailed derivation of the open-end results stated in Section 4. The main idea is simple but important: after reparameterizing the monitoring time by $x = s/(1+s)$, the infinite horizon $s \in [0, \infty)$ is compressed to the unit interval $x \in [0, 1)$. That turns the open-end Brownian functionals into ordinary functionals of Brownian motion on a compact interval, which is why the critical-value simulation problem remains manageable.

C.1 Proof of Proposition 1

Fix $x \in [0, 1)$ and write $s = s(x) = x/(1-x)$. By definition,

$$W_q(x) = \frac{U_q(s)}{1+s}, \qquad U_q(s) = B_q(1+s) - (1+s)B_q(1).$$

We verify three claims: the covariance structure of W_q, its independence from the training bridge B_q^0, and the change-of-variables identity for the quadratic-form functionals.

We first compute the covariance of U_q. Consider one coordinate and let $0 \le s_1 \le s_2$. Using $\mathrm{Cov}(B(a), B(b)) = \min(a, b)$,

$$\begin{aligned}
\mathrm{Cov}\big(U(s_1), U(s_2)\big) &= \mathrm{Cov}\big(B(1+s_1) - (1+s_1)B(1),\, B(1+s_2) - (1+s_2)B(1)\big) \\
&= \mathrm{Cov}(B(1+s_1), B(1+s_2)) - (1+s_2)\,\mathrm{Cov}(B(1+s_1), B(1)) \\
&\qquad - (1+s_1)\,\mathrm{Cov}(B(1), B(1+s_2)) + (1+s_1)(1+s_2)\,\mathrm{Var}(B(1)) \\
&= (1+s_1) - (1+s_2) - (1+s_1) + (1+s_1)(1+s_2) \\
&= s_1(1+s_2).
\end{aligned}$$

Dividing by $(1+s_1)(1+s_2)$ gives

$$\mathrm{Cov}\big(W(x_1), W(x_2)\big) = \frac{s_1(1+s_2)}{(1+s_1)(1+s_2)} = \frac{s_1}{1+s_1} = x_1 = \min(x_1, x_2),$$

where $x_i = s_i/(1+s_i)$. Thus each coordinate of W_q is a standard Brownian motion on $[0,1)$, and the q-dimensional process is a standard q-variate Brownian motion because the coordinates inherit the independence of the underlying Brownian motion B_q.

We next show independence from the training bridge. For $r \in [0,1]$ and one coordinate,

$$B^0(r) = B(r) - rB(1).$$

Hence

$$\operatorname{Cov}\big(B^0(r), W(x)\big) = \frac{\operatorname{Cov}(B(r) - rB(1), U(s))}{1+s}.$$

The numerator is

$$\begin{aligned}\operatorname{Cov}(B(r), U(s)) &= \operatorname{Cov}(B(r), B(1+s)) - (1+s)\operatorname{Cov}(B(r), B(1)) \\ &= r - (1+s)r = -rs,\end{aligned}$$

while

$$\begin{aligned}\operatorname{Cov}(rB(1), U(s)) &= r\operatorname{Cov}(B(1), B(1+s)) - (1+s)r\operatorname{Var}(B(1)) \\ &= r - (1+s)r = -rs.\end{aligned}$$

Subtracting yields $\operatorname{Cov}(B^0(r), W(x)) = 0$ for all (r,x). Because (B_q^0, W_q) is jointly Gaussian, zero cross-covariance implies independence.

Finally, for any symmetric nonnegative definite matrix A,

$$\frac{U_q(s)^\top A U_q(s)}{(1+s)^2\left(\frac{s}{1+s}\right)^{2\gamma}} = \frac{W_q(x)^\top A W_q(x)}{x^{2\gamma}}, \qquad x = \frac{s}{1+s}.$$

Taking the supremum over $s \in (0,\infty)$ is the same as taking the supremum over $x \in (0,1)$ because $x \mapsto s(x)$ is a bijection between those intervals. This proves the proposition.

C.2 Proof of Theorem 7

We write the proof for SSMS; the RSMS and HAC cases follow by the same argument after replacing V^{-1} with R^{-2} or I_q.

For $x \in (0,1)$ define the transformed monitoring index

$$k_m(x) = \left\lfloor \frac{mx}{1-x} \right\rfloor.$$

This is the discrete analogue of the mapping $s = x/(1-x)$. Also define the transformed feasible statistic

$$\mathbb{M}_m^{\mathrm{S}}(x) = \mathcal{M}_m^{\mathrm{S}}\big(k_m(x)\big), \qquad x \in (0,1).$$

Fix any $\varepsilon \in (0,1)$ and set $T_\varepsilon = (1-\varepsilon)/\varepsilon$. Then $x \le 1-\varepsilon$ is equivalent to $s \le T_\varepsilon$, so the finite-horizon SSMS theorem applied with $T = T_\varepsilon$ yields

$$\sup_{0<x\le 1-\varepsilon} \mathbb{M}_m^{\mathrm{S}}(x) = \sup_{1\le k\le mT_\varepsilon} \mathcal{M}_m^{\mathrm{S}}(k) \xrightarrow{d} \sup_{0<s\le T_\varepsilon} \frac{U_q(s)^\top V^{-1} U_q(s)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}}.$$

By Proposition 1, the right-hand side equals

$$\sup_{0<x\le 1-\varepsilon} \frac{W_q(x)^\top V^{-1} W_q(x)}{x^{2\gamma}}.$$

We now let $\varepsilon \downarrow 0$. Because W_q has continuous sample paths on $[0,1]$ and V is almost surely positive definite, the map

$$x \mapsto \frac{W_q(x)^\top V^{-1} W_q(x)}{x^{2\gamma}}$$

is almost surely continuous on $(0,1]$. Near $x = 1$ there is no difficulty because $W_q(1)$ is finite almost surely. Near $x = 0$, Brownian motion satisfies $W_q(x) = O_{\text{a.s.}}(x^{1/2-\eta})$ for every $\eta > 0$ along sufficiently fine neighborhoods of zero; since $\gamma < 1/2$, the factor $x^{-2\gamma}$ does not destroy finiteness. Therefore the displayed path is almost surely bounded on $(0,1]$, and

$$\sup_{0<x\le 1-\varepsilon} \frac{W_q(x)^\top V^{-1} W_q(x)}{x^{2\gamma}} \uparrow \sup_{0<x<1} \frac{W_q(x)^\top V^{-1} W_q(x)}{x^{2\gamma}} \qquad \text{a.s. as } \varepsilon \downarrow 0.$$

To transfer this limit back to the feasible statistic, use a standard ε–η truncation argument. Given $\eta > 0$, choose ε so small that the probability mass contributed by the Brownian tail $x \in (1-\varepsilon, 1)$ is at most η in the limit. On the compact interval $(0, 1-\varepsilon]$ we already have weak convergence from the finite-horizon theorem. Since the open-end statistic is the supremum over all such compact pieces and the limit process is continuous and bounded on $(0,1)$, letting $\varepsilon \downarrow 0$ yields

$$\sup_{k\ge 1} \mathcal{M}_m^{\mathrm{S}}(k) \xrightarrow{d} \sup_{0<x<1} \frac{W_q(x)^\top V^{-1} W_q(x)}{x^{2\gamma}}.$$

The RSMS case is identical after replacing the quadratic self-normalizer by the diagonal range normalizer. The HAC case is even simpler because the normalizing matrix is deterministic in the limit and the numerator becomes $\|W_q(x)\|_2^2$. This proves Theorem 7.

C.3 Proof of Theorem 8

Again we present the argument for SSMS; the RSMS and HAC cases are the same after the obvious substitutions.

Start from the open-end CvM statistic

$$\mathcal{I}^{\mathrm{S}}_{m,\infty}(k) = \frac{1}{m}\sum_{j=1}^{k} w_\infty\left(\frac{j}{m}\right)\mathcal{M}^{\mathrm{S}}_m(j).$$

Because every summand is nonnegative, the path $k \mapsto \mathcal{I}^{\mathrm{S}}_{m,\infty}(k)$ is nondecreasing. Hence

$$\sup_{k\geq 1}\mathcal{I}^{\mathrm{S}}_{m,\infty}(k) = \lim_{K\to\infty}\mathcal{I}^{\mathrm{S}}_{m,\infty}(K).$$

Fix $\varepsilon \in (0,1)$ and take $K = \lfloor mT_\varepsilon \rfloor$ with $T_\varepsilon = (1-\varepsilon)/\varepsilon$. Then the finite-horizon CvM theorem gives

$$\mathcal{I}^{\mathrm{S}}_{m,\infty}(\lfloor mT_\varepsilon \rfloor) \xrightarrow{d} \int_0^{T_\varepsilon} w_\infty(s)\frac{U_q(s)^\top V^{-1}U_q(s)}{(1+s)^2}\,ds.$$

Now change variables from s to $x = s/(1+s)$. Since $s = x/(1-x)$ and $ds = dx/(1-x)^2$,

$$\begin{aligned}\int_0^{T_\varepsilon} w_\infty(s)\frac{U_q(s)^\top V^{-1}U_q(s)}{(1+s)^2}\,ds &= \int_0^{1-\varepsilon}\frac{w_\infty\left(\frac{x}{1-x}\right)}{(1-x)^2}W_q(x)^\top V^{-1}W_q(x)\,dx \\ &= \int_0^{1-\varepsilon}\widetilde{w}(x)\,W_q(x)^\top V^{-1}W_q(x)\,dx.\end{aligned}$$

Because $\widetilde{w}$ is nonnegative and integrable on $[0,1)$ by assumption, while $x \mapsto W_q(x)^\top V^{-1}W_q(x)$ is almost surely continuous and bounded on $[0,1]$, the product is integrable. Therefore

$$\int_0^{1-\varepsilon}\widetilde{w}(x)\,W_q(x)^\top V^{-1}W_q(x)\,dx \uparrow \int_0^{1}\widetilde{w}(x)\,W_q(x)^\top V^{-1}W_q(x)\,dx \qquad \text{a.s.}$$

as $\varepsilon \downarrow 0$, by monotone convergence.

The final step is the same truncation argument as in the KS proof. On each compact interval $(0, 1-\varepsilon]$ we have convergence from the finite-horizon theorem. The nonnegative accumulation property lets us pass from compact horizons to the open-end limit by letting $\varepsilon \downarrow 0$. Thus

$$\sup_{k\geq 1}\mathcal{I}^{\mathrm{S}}_{m,\infty}(k) \xrightarrow{d} \int_0^{1}\widetilde{w}(x)\,W_q(x)^\top V^{-1}W_q(x)\,dx.$$

The RSMS and HAC cases follow by replacing V^{-1} with R^{-2} or I_q.

A final remark explains why the integrability restriction is essential. If one were to use the unweighted original-horizon choice $w_\infty(s) \equiv 1$, then $\widetilde{w}(x) = (1-x)^{-2}$, which is not integrable near $x = 1$. Since $W_q(1)$ is nondegenerate, the corresponding open-end CvM integral diverges with positive probability and in fact leads to eventual false alarms under H_0. This is why open-end CvM monitoring requires horizon-decaying weights.

D Tabulated asymptotic critical values

This appendix records the asymptotic critical values used for the monitoring statistics when the stacked score dimension satisfies $q \leq 10$. The finite-horizon tables report $T \in \{1, 2, 5, 10\}$. For KS we also append the open-end $T = \infty$ values, while for CvM the open-end $T = \infty$ values are reported separately because they use admissible horizon-decaying weights rather than the finite-horizon weights. For $T = \infty$, the results are taken from the critical value tabulations in Sun and Hong (2026b), because the self-normalized online monitoring statistics are identical. The open-end ($T = \infty$) entries are simulated directly from the reparameterized unit-interval Brownian representation in Proposition 1. We report the 5% and 10% significance thresholds. For the finite-horizon CvM statistics we use the four normalized time weights

$$w_U(s) \equiv 1, \qquad w_{\text{Early}}(s) = 2(1-r), \qquad w_{\text{Mid}}(s) = 6r(1-r), \qquad w_{\text{Late}}(s) = 2r, \qquad r = s/T.$$

For the open-end CvM statistics we use their admissible counterparts

$$w_{\infty,U}(s) = \frac{1}{(1+s)^2}, \qquad w_{\infty,\text{Early}}(s) = \frac{2}{(1+s)^3},$$
$$w_{\infty,\text{Late}}(s) = \frac{2s}{(1+s)^3}, \qquad w_{\infty,\text{Mid}}(s) = \frac{6s}{(1+s)^4}.$$

Two practical remarks are useful. First, following Sun–Hong (2026), the SSMS and RSMS KS critical values are reported separately for $\gamma = 0$ and $\gamma = 0.15$. For SSMS, the finite-horizon values for $T \in \{1, 2, 5, 10\}$ are taken from the Sun–Hong tabulations, and the open-end $T = \infty$ values are reported in the same split-by-γ format for $q \leq 10$. Second, our finite-horizon CvM theory uses $\gamma = 0$, so only the $\gamma = 0$ finite-horizon CvM tables are reported below even when the archive contains additional entries. The separate open-end CvM tables were simulated from the bundled code using 5,000 Monte Carlo draws with training-grid size 1,500 and open-end monitoring grid size 2,000.

Table D.1: SSMS KS critical values for $\gamma = 0$. Finite-horizon values for $T \in \{1, 2, 5, 10\}$ are from Sun–Hong (2026); the $T = \infty$ column reports the corresponding open-end calibration.

	$T=1$		$T=2$		$T=5$		$T=10$		$T=\infty$	
q	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	32.7258	22.1591	45.8954	30.7701	53.7975	37.2349	62.0154	42.4631	65.9610	44.7320
2	66.6268	48.5072	92.9504	67.4811	110.5803	82.7167	123.7205	91.4123	146.7400	102.8720
3	108.9289	83.4714	147.4078	110.9598	186.7649	142.1685	206.8708	154.6299	227.8760	173.4590
4	159.2089	124.6814	206.3431	163.4394	270.1868	211.2319	306.5439	232.1413	323.6320	247.0100
5	218.5677	170.3210	280.5398	222.7883	354.8600	285.8250	406.7011	321.1939	444.4140	352.7790
6	279.1300	225.4792	362.5287	294.1692	464.3820	372.5477	515.4844	414.5120	565.3490	456.1710
7	348.2288	284.5871	447.7015	369.7363	570.9342	463.5198	636.8051	522.3679	698.9010	568.3740
8	420.9176	347.0096	538.7612	450.8265	701.0175	574.2443	777.0256	641.8620	838.7720	695.2910
9	496.6725	416.5989	645.2948	540.3669	829.8143	689.1502	924.5440	766.9828	1003.2170	833.9390
10	583.7221	485.5654	766.5619	631.8476	969.2089	810.8890	1085.0200	907.4008	1187.9140	995.5880

Table D.2: SSMS KS critical values for $\gamma = 0.15$. Finite-horizon values for $T \in \{1, 2, 5, 10\}$ are from Sun–Hong (2026); the $T = \infty$ column reports the corresponding open-end calibration.

	$T = 1$		$T = 2$		$T = 5$		$T = 10$		$T = \infty$	
q	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	43.5841	30.2456	55.5930	39.1039	61.6436	43.3043	68.9495	47.9834	72.9220	49.2450
2	88.7728	65.1229	112.1324	81.9121	125.0469	95.0829	135.8995	102.4043	153.5570	111.9650
3	141.6571	110.4683	176.2494	134.8827	209.2979	160.2459	224.5483	171.1760	239.4010	183.3990
4	209.5982	164.0076	246.8474	196.8125	303.3772	237.6468	330.9420	255.2317	343.4570	263.4310
5	281.7272	223.8936	336.5669	265.5484	396.8837	321.9688	441.2219	349.2220	462.5860	375.1900
6	359.9652	293.9266	427.7711	350.1683	511.7752	415.0403	553.7712	449.1153	589.7830	480.0480
7	446.5727	369.6907	531.3681	441.0598	629.1840	514.5814	686.1611	564.0993	722.8380	593.6210
8	541.1048	448.1467	635.3852	533.0248	761.9476	639.3377	836.4094	689.3037	870.2260	724.6980
9	642.8357	533.7764	755.5265	636.4059	912.9737	762.0676	984.8651	825.3698	1034.8360	868.3870
10	746.6774	624.8431	889.4109	744.4037	1057.0473	892.2286	1155.6050	973.0264	1230.0280	1030.7010

Table D.3: RSMS KS critical values for $\gamma = 0$.

	$T = 1$		$T = 2$		$T = 5$		$T = 10$		$T = \infty$	
q	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	1.993	1.478	2.780	2.036	3.401	2.514	3.965	2.902	4.254	3.123
2	2.913	2.251	3.986	3.059	5.071	3.906	5.718	4.459	6.271	4.806
3	3.750	2.976	4.998	3.966	6.509	5.106	7.297	5.827	7.909	6.376
4	4.439	3.594	5.916	4.817	7.712	6.233	8.754	7.054	9.303	7.627
5	5.079	4.139	6.758	5.598	8.773	7.230	9.981	8.223	10.758	8.814
6	5.753	4.772	7.560	6.351	9.811	8.206	11.181	9.308	12.224	9.990
7	6.376	5.355	8.455	7.061	10.910	9.173	12.407	10.362	13.239	11.130
8	6.845	5.893	9.254	7.793	11.991	10.180	13.486	11.424	14.579	12.285
9	7.420	6.451	10.016	8.558	12.886	11.063	14.398	12.413	15.644	13.382
10	7.980	6.918	10.719	9.227	13.844	11.974	15.574	13.429	16.879	14.545

Table D.4: RSMS KS critical values for $\gamma = 0.15$.

	$T = 1$		$T = 2$		$T = 5$		$T = 10$		$T = \infty$	
q	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	2.659	1.983	3.349	2.490	3.833	2.893	4.381	3.278	4.631	3.395
2	3.838	2.961	4.746	3.708	5.688	4.397	6.205	4.914	6.610	5.117
3	4.852	3.893	5.912	4.739	7.185	5.733	7.931	6.336	8.331	6.665
4	5.712	4.652	7.000	5.719	8.497	6.935	9.434	7.591	9.721	7.957
5	6.502	5.347	7.988	6.608	9.577	8.025	10.695	8.867	11.160	9.300
6	7.360	6.134	8.877	7.431	10.701	9.033	11.950	9.984	12.608	10.406
7	8.091	6.819	9.892	8.303	11.933	10.079	13.175	11.083	13.665	11.557
8	8.778	7.495	10.816	9.098	13.110	11.137	14.244	12.184	14.959	12.710
9	9.431	8.223	11.631	9.983	14.049	12.040	15.301	13.161	16.054	13.873
10	10.124	8.783	12.427	10.799	15.023	13.008	16.493	14.197	17.363	14.984

Table D.5: HAC KS critical values for $\gamma = 0$.

	$T=1$		$T=2$		$T=5$		$T=10$		$T=\infty$	
q	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	2.471	1.906	3.351	2.578	4.128	3.178	4.755	3.607	5.071	3.856
2	3.557	2.836	4.779	3.855	6.008	4.884	6.692	5.405	7.303	5.858
3	4.482	3.736	5.917	4.927	7.635	6.295	8.387	6.975	9.364	7.577
4	5.324	4.494	7.106	5.947	9.035	7.634	10.020	8.395	10.960	9.192
5	6.198	5.217	8.088	6.923	10.351	8.839	11.450	9.780	12.568	10.610
6	6.929	5.977	9.145	7.811	11.633	10.045	12.858	11.110	14.245	12.126
7	7.756	6.666	10.250	8.793	12.917	11.198	14.354	12.335	15.601	13.390
8	8.473	7.348	11.180	9.693	14.170	12.324	15.674	13.683	16.927	14.886
9	9.168	8.028	12.108	10.683	15.242	13.433	16.990	14.898	18.287	16.219
10	9.883	8.705	13.004	11.560	16.295	14.546	18.245	16.096	19.674	17.661

Table D.6: HAC KS critical values for $\gamma = 0.15$.

	$T=1$		$T=2$		$T=5$		$T=10$		$T=\infty$	
q	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	3.212	2.529	3.997	3.145	4.680	3.613	5.189	3.999	5.368	4.135
2	4.601	3.739	5.679	4.574	6.647	5.466	7.197	5.888	7.593	6.183
3	5.786	4.818	6.968	5.868	8.349	6.943	9.013	7.501	9.672	7.916
4	6.816	5.736	8.289	7.009	9.844	8.369	10.677	8.976	11.247	9.564
5	7.896	6.684	9.412	8.107	11.246	9.676	12.168	10.429	12.970	11.058
6	8.830	7.608	10.688	9.165	12.725	11.005	13.641	11.778	14.575	12.471
7	9.843	8.487	11.909	10.261	14.008	12.203	15.137	13.101	16.022	13.812
8	10.673	9.273	12.936	11.276	15.277	13.381	16.513	14.520	17.383	15.296
9	11.556	10.177	14.030	12.397	16.500	14.591	17.992	15.675	18.742	16.589
10	12.435	11.013	15.007	13.399	17.650	15.757	19.175	16.992	20.284	18.029

Table D.7: SSMS CvM critical values under the U time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	12.483	7.967	39.042	24.903	130.761	86.491	340.484	216.863
2	27.843	19.444	85.697	59.086	303.141	213.612	737.393	517.809
3	47.984	35.354	143.039	106.987	534.842	385.628	1293.803	938.700
4	71.826	55.250	206.982	160.079	794.204	602.697	1949.494	1458.233
5	99.427	77.870	286.612	223.035	1079.308	837.449	2702.193	2062.131
6	132.188	103.957	383.462	301.194	1408.869	1120.060	3473.633	2751.090
7	166.746	134.522	476.707	390.094	1756.622	1412.217	4391.147	3499.741
8	204.034	166.542	582.953	478.144	2188.439	1776.620	5339.763	4376.824
9	244.005	202.436	708.213	585.900	2633.803	2157.359	6461.590	5264.469
10	286.584	241.174	837.923	694.421	3119.958	2569.750	7645.609	6335.862

Table D.8: SSMS CvM critical values under the Early time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	9.095	5.919	29.018	18.350	113.675	72.833	282.573	180.373
2	21.471	14.915	65.270	46.537	257.409	183.230	650.860	450.259
3	36.388	26.440	109.067	82.472	437.568	324.006	1094.516	797.562
4	53.031	41.050	163.586	126.420	653.351	486.059	1619.269	1236.508
5	72.560	57.323	224.436	177.056	894.236	687.770	2244.281	1764.355
6	95.386	76.324	296.696	235.271	1155.688	932.093	2903.559	2333.992
7	118.053	96.272	374.785	301.552	1474.926	1185.315	3644.617	2935.543
8	146.165	119.743	462.233	378.476	1828.196	1476.534	4465.723	3629.699
9	174.413	144.823	559.215	457.401	2217.746	1786.795	5434.110	4465.270
10	206.235	173.596	658.242	546.508	2587.728	2145.562	6503.477	5349.945

Table D.9: SSMS CvM critical values under the Mid time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	13.390	8.368	38.963	24.769	142.889	92.446	348.924	216.495
2	30.446	21.359	93.376	64.098	331.030	229.952	804.268	565.117
3	51.531	37.633	155.556	114.536	560.134	416.139	1346.948	993.097
4	74.757	57.890	227.702	172.180	857.669	643.026	2064.897	1526.924
5	104.975	80.308	314.359	240.273	1176.354	905.223	2736.415	2147.468
6	135.764	108.800	409.305	322.474	1504.148	1193.020	3564.315	2833.736
7	171.619	138.804	516.239	415.278	1892.283	1525.102	4507.004	3591.053
8	211.302	172.425	630.966	516.320	2352.093	1905.872	5556.841	4479.052
9	252.432	209.945	777.066	628.124	2833.143	2313.330	6726.250	5442.135
10	300.251	250.803	921.475	755.026	3297.312	2722.918	7941.618	6492.952

Table D.10: SSMS CvM critical values under the Late time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	16.762	10.526	47.160	29.224	162.606	102.972	379.998	236.890
2	37.718	25.900	111.034	76.965	378.923	262.692	863.124	589.770
3	61.891	45.954	182.100	134.622	627.723	460.823	1444.747	1053.595
4	93.105	71.429	273.511	203.375	941.682	705.142	2124.516	1614.477
5	129.796	101.346	377.888	287.836	1280.277	998.006	2958.968	2259.654
6	170.440	133.057	500.548	390.694	1686.899	1319.448	3907.671	3066.846
7	214.220	171.765	633.499	500.289	2141.948	1698.678	4924.550	3893.173
8	261.794	213.616	770.665	620.538	2599.323	2099.527	5979.125	4828.629
9	315.203	259.247	909.615	762.886	3147.959	2541.806	7131.018	5932.784
10	372.872	309.458	1086.613	894.263	3690.121	3034.550	8472.348	7062.362

Table D.11: RSMS CvM critical values under the U time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	0.808	0.556	2.530	1.708	8.955	6.097	23.207	15.766
2	1.311	0.976	3.889	2.922	14.638	10.840	36.302	27.171
3	1.737	1.363	5.105	4.064	19.868	15.166	48.600	37.329
4	2.113	1.710	6.236	5.057	23.823	19.265	59.891	47.034
5	2.498	2.026	7.341	6.015	27.857	22.836	69.867	56.883
6	2.853	2.375	8.477	6.921	31.483	26.268	79.376	64.915
7	3.183	2.697	9.509	7.889	35.687	29.930	88.720	73.305
8	3.507	2.997	10.539	8.826	39.293	33.385	98.133	82.512
9	3.858	3.313	11.523	9.836	42.743	36.695	105.911	90.160
10	4.154	3.600	12.326	10.716	46.261	39.969	114.774	98.268

Table D.12: RSMS CvM critical values under the Early time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	0.583	0.397	1.884	1.301	7.541	5.028	19.606	13.272
2	0.950	0.713	3.025	2.287	12.402	9.106	31.220	23.303
3	1.263	0.982	4.019	3.144	16.253	12.666	40.911	32.048
4	1.554	1.235	4.915	3.983	19.542	15.930	50.315	39.845
5	1.817	1.509	5.719	4.770	22.738	18.943	59.217	47.949
6	2.071	1.741	6.546	5.465	25.915	21.734	67.004	55.793
7	2.319	1.975	7.358	6.179	28.879	24.349	74.947	62.896
8	2.575	2.205	8.092	6.888	32.276	27.331	82.644	70.093
9	2.806	2.430	8.821	7.548	34.945	30.266	90.747	77.671
10	3.022	2.628	9.548	8.218	38.468	33.033	97.795	84.395

Table D.13: RSMS CvM critical values under the Mid time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	0.855	0.591	2.554	1.749	9.726	6.542	23.913	16.115
2	1.424	1.030	4.343	3.169	15.676	11.777	38.900	28.881
3	1.832	1.452	5.637	4.367	20.771	16.149	51.552	39.824
4	2.276	1.814	6.814	5.497	25.419	20.215	62.231	49.761
5	2.642	2.170	7.956	6.527	30.221	24.130	72.754	58.941
6	3.017	2.510	9.019	7.557	34.311	27.965	81.813	68.009
7	3.362	2.832	10.100	8.514	38.234	31.567	92.342	76.118
8	3.716	3.167	11.080	9.483	42.230	35.367	101.412	85.565
9	4.039	3.496	12.230	10.472	46.183	38.862	110.527	93.929
10	4.357	3.788	13.231	11.485	49.495	42.257	119.999	103.166

Table D.14: RSMS CvM critical values under the Late time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	1.087	0.740	3.157	2.138	11.067	7.463	25.900	16.994
2	1.756	1.285	5.053	3.687	17.622	13.072	42.270	30.365
3	2.316	1.788	6.730	5.131	23.166	18.017	54.609	42.156
4	2.782	2.226	8.148	6.509	28.626	22.580	67.263	52.424
5	3.246	2.668	9.508	7.733	33.369	27.002	78.419	63.652
6	3.689	3.046	10.818	8.950	38.284	31.426	88.823	72.965
7	4.141	3.473	12.156	10.061	43.130	35.316	99.348	83.034
8	4.576	3.834	13.431	11.241	46.992	38.967	108.880	91.962
9	4.938	4.232	14.643	12.343	50.959	43.123	119.413	100.909
10	5.380	4.613	15.799	13.476	55.154	46.607	128.888	109.968

Table D.15: HAC CvM critical values under the U time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	1.054	0.741	3.110	2.248	11.215	8.158	29.125	20.565
2	1.611	1.269	4.861	3.796	17.983	14.097	44.837	34.199
3	2.160	1.745	6.308	5.189	23.798	19.177	57.425	47.063
4	2.630	2.177	7.728	6.458	28.889	24.023	70.274	58.065
5	3.116	2.594	8.974	7.636	33.529	28.533	82.069	69.133
6	3.538	3.053	10.344	8.827	38.378	32.860	93.285	79.109
7	3.983	3.428	11.772	10.066	43.211	37.061	105.164	89.992
8	4.380	3.841	12.900	11.253	47.685	41.234	116.442	100.545
9	4.786	4.209	14.202	12.464	51.778	45.646	126.768	110.314
10	5.226	4.610	15.270	13.678	55.886	49.416	136.757	120.351

Table D.16: HAC CvM critical values under the Early time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	0.739	0.539	2.336	1.699	9.339	6.668	24.955	17.882
2	1.156	0.931	3.660	2.879	14.780	11.624	38.830	29.858
3	1.562	1.261	4.811	3.980	19.541	15.819	50.302	40.942
4	1.892	1.588	5.894	4.957	23.659	19.775	61.034	50.825
5	2.254	1.888	6.943	5.903	27.749	23.619	71.812	60.156
6	2.558	2.212	7.867	6.798	31.728	27.240	81.736	69.384
7	2.882	2.507	8.933	7.745	35.609	30.795	91.259	78.555
8	3.181	2.785	9.922	8.659	39.348	34.240	101.073	88.407
9	3.478	3.073	10.923	9.607	42.825	37.728	110.922	96.398
10	3.789	3.347	11.786	10.505	46.327	41.283	118.980	105.248

Table D.17: HAC CvM critical values under the Mid time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	1.097	0.775	3.307	2.367	12.117	8.703	30.949	21.996
2	1.675	1.338	5.120	4.025	19.190	15.042	47.791	36.350
3	2.251	1.817	6.658	5.463	25.318	20.357	61.362	49.714
4	2.736	2.272	8.183	6.802	30.679	25.507	74.738	61.755
5	3.265	2.711	9.541	8.032	35.736	30.397	87.093	73.225
6	3.698	3.163	10.912	9.305	40.990	34.823	99.346	83.866
7	4.157	3.588	12.324	10.638	46.025	39.339	111.219	95.192
8	4.599	3.986	13.695	11.885	50.626	43.866	123.175	106.639
9	5.008	4.388	15.009	13.152	54.998	48.475	134.480	116.885
10	5.439	4.796	16.179	14.392	59.592	52.604	145.235	127.247

Table D.18: HAC CvM critical values under the Late time weight.

	$T=1$		$T=2$		$T=5$		$T=10$	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	1.362	0.964	3.947	2.847	13.427	9.652	33.704	23.550
2	2.097	1.647	6.206	4.727	21.463	16.633	51.297	38.957
3	2.810	2.252	7.892	6.472	28.234	22.759	65.868	53.168
4	3.388	2.830	9.685	7.969	34.360	28.291	79.785	66.073
5	4.022	3.363	11.245	9.514	39.912	33.717	92.500	78.391
6	4.577	3.906	12.978	10.970	45.671	38.617	106.071	89.582
7	5.170	4.420	14.624	12.472	51.238	43.780	119.654	101.430
8	5.674	4.934	16.165	13.913	56.416	48.741	131.957	113.510
9	6.200	5.405	17.691	15.496	61.144	53.619	143.529	125.185
10	6.715	5.920	19.076	16.951	65.989	58.195	155.433	136.107

D.1 Open-end CvM critical values

Because the open-end CvM detector uses the admissible horizon-decaying weights

$$w_{\infty,U}(s) = (1+s)^{-2}, \qquad w_{\infty,\mathrm{Early}}(s) = 2(1+s)^{-3},$$
$$w_{\infty,\mathrm{Mid}}(s) = 6s(1+s)^{-4}, \qquad w_{\infty,\mathrm{Late}}(s) = 2s(1+s)^{-3}.$$

its $T=\infty$ critical values are reported separately below rather than as additional finite-horizon columns.

Table D.19: Open-end SSMS CvM critical values ($T=\infty$) under the admissible time weights.

	U		Early		Mid		Late	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	20.465	13.057	13.175	8.603	20.488	12.919	28.227	17.434
2	47.967	32.468	30.894	21.528	47.564	32.387	66.861	44.393
3	78.242	58.469	50.193	37.731	77.286	58.572	108.909	80.033
4	115.193	88.067	72.441	57.414	114.214	88.897	161.351	121.122
5	161.092	127.214	102.305	80.961	159.864	125.468	223.399	175.643
6	210.942	166.187	134.615	109.724	211.175	168.748	292.804	228.952
7	264.344	215.600	169.724	138.480	265.143	213.548	363.357	296.067
8	327.032	266.196	209.491	171.499	329.009	267.078	450.771	367.413
9	400.105	326.834	258.212	212.902	400.981	326.275	547.072	448.393
10	472.607	398.409	305.257	257.858	476.005	398.340	654.227	542.649

Table D.20: Open-end RSMS CvM critical values ($T = \infty$) under the admissible time weights.

	U		Early		Mid		Late	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	1.355	0.930	0.892	0.604	1.364	0.933	1.885	1.287
2	2.199	1.634	1.405	1.065	2.180	1.631	3.082	2.244
3	2.880	2.266	1.836	1.461	2.840	2.254	3.955	3.141
4	3.585	2.839	2.289	1.829	3.557	2.826	4.975	3.898
5	4.197	3.423	2.714	2.231	4.246	3.427	5.788	4.700
6	4.784	4.051	3.094	2.591	4.778	4.034	6.570	5.455
7	5.348	4.533	3.464	2.929	5.375	4.522	7.351	6.099
8	5.862	5.004	3.789	3.255	5.900	5.015	8.121	6.768
9	6.449	5.543	4.222	3.605	6.506	5.568	8.782	7.511
10	7.026	6.086	4.564	3.968	7.084	6.123	9.593	8.330

Table D.21: Open-end HAC CvM critical values ($T = \infty$) under the admissible time weights.

	U		Early		Mid		Late	
q	5%	10%	5%	10%	5%	10%	5%	10%
1	1.651	1.213	1.067	0.780	1.666	1.217	2.329	1.672
2	2.613	2.045	1.696	1.321	2.655	2.083	3.669	2.822
3	3.441	2.807	2.204	1.799	3.452	2.798	4.813	3.899
4	4.234	3.529	2.726	2.263	4.228	3.525	5.825	4.821
5	4.960	4.274	3.242	2.770	5.022	4.242	6.776	5.842
6	5.756	4.880	3.698	3.192	5.758	4.902	7.846	6.663
7	6.375	5.507	4.153	3.612	6.416	5.518	8.718	7.565
8	7.032	6.165	4.550	4.015	7.054	6.166	9.608	8.414
9	7.757	6.808	5.071	4.464	7.798	6.850	10.572	9.283
10	8.528	7.473	5.502	4.877	8.454	7.486	11.541	10.225

E Continuum-of-τ monitoring

This appendix sketches how the monitoring framework can be formulated for a continuum of distributional levels $\tau \in \mathcal{T} \subset (0, 1)$. The main text focuses on a finite grid because that is the directly implementable object. The continuum formulation is conceptually useful because it clarifies how dense grids approximate ideal sup- and integral-type detectors over τ.

For $\tau \in \mathcal{T}$, define the oracle restricted residual

$$u^0_{t,\tau} = Y_t - X^\top_{t-1}\alpha_0(\tau)$$

and the oracle score

$$g^0_t(\tau) = H_{t-1}\,\psi_\tau\big(u^0_{t,\tau}\big) \in \mathbb{R}^{d_H}. \tag{E.1}$$

Let $\phi^0_t(\tau) = g^0_t(\tau) - \mathbb{E}\,g^0_t(\tau)$ and view $\tau \mapsto g^0_t(\tau)$ as a random element of $\ell^\infty(\mathcal{T})^{d_H}$. A continuum version of Assumption 3 is the uniform negligibility condition

$$\sup_{\tau\in\mathcal{T}} \max_{1\le t\le m(1+T)} \big\|g_t(\tau) - g^0_t(\tau)\big\| = o_p(1). \tag{E.2}$$

Consider the partial-sum process indexed by (r, τ),

$$\mathbb{U}_m(r,\tau) = \frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor mr \rfloor} \phi_t^0(\tau), \qquad (r,\tau) \in [0, 1+T] \times \mathcal{T}.$$

A continuum theory is obtained by strengthening Assumption 2 to the functional invariance principle

$$\{\mathbb{U}_m(r,\tau)\}_{(r,\tau)\in[0,1+T]\times\mathcal{T}} \Rightarrow \{\mathbb{U}(r,\tau)\}_{(r,\tau)\in[0,1+T]\times\mathcal{T}} \quad \text{in } \ell^\infty([0,1+T]\times\mathcal{T})^{d_H}, \tag{E.3}$$

where $\mathbb{U}$ is a mean-zero Gaussian process with covariance kernel

$$\mathrm{Cov}\left(\mathbb{U}(r_1,\tau_1), \mathbb{U}(r_2,\tau_2)\right) = \min(r_1, r_2)\, \Xi(\tau_1, \tau_2),$$

for an appropriate long-run covariance kernel Ξ. The key technical ingredient is stochastic equicontinuity in τ.

Define the bridge and monitoring-increment processes

$$\mathbb{U}^0(r,\tau) = \mathbb{U}(r,\tau) - r\mathbb{U}(1,\tau), \qquad \mathbb{U}^\Delta(s,\tau) = \mathbb{U}(1+s,\tau) - (1+s)\mathbb{U}(1,\tau).$$

For each τ, define the self-normalizer matrix

$$V(\tau) = \int_0^1 \mathbb{U}^0(r,\tau)\mathbb{U}^0(r,\tau)^\top dr.$$

The continuum SSMS limit is then

$$\mathcal{M}_\infty^{\mathrm{S}} = \sup_{0<s\le T} \sup_{\tau\in\mathcal{T}} \frac{\mathbb{U}^\Delta(s,\tau)^\top V(\tau)^{-1} \mathbb{U}^\Delta(s,\tau)}{(1+s)^2 \left(\frac{s}{1+s}\right)^{2\gamma}}. \tag{E.4}$$

Similarly, a continuum CvM limit with cross-τ weight ω is

$$\mathcal{I}_\infty^{\mathrm{S}}(s) = \int_0^s w(u) \left[\int_{\mathcal{T}} \omega(\tau) \frac{\mathbb{U}^\Delta(u,\tau)^\top V(\tau)^{-1} \mathbb{U}^\Delta(u,\tau)}{(1+u)^2}\, d\tau \right] du. \tag{E.5}$$

Under (E.2)–(E.3) and stochastic equicontinuity, dense finite grids approximate these continuum monitors. This observation justifies the finite-grid procedures in the main text while clarifying their functional meaning.

F Disaggregated simulation tables by horizon and training size

This appendix unpacks the pooled simulation summaries in Section 5 by keeping the monitoring horizon T and training size m separate. For compactness, the tables focus on the SSMS/RSMS statistics requested in the revision: KS rules for $\gamma \in \{0, 0.15\}$ and CvM rules under the Uniform,

Early, Mid, and Late weights.

The disaggregated tables in this appendix broadly support the patterns discussed in the main text. In particular, larger m often improves detection in the abrupt, gradual, and contaminated designs, while its effect is weaker and sometimes mixed in A3–A4. The role of T is also design-dependent, with longer horizons typically helping more when the distributional change is persistent. We therefore view these tables mainly as a complement to the pooled results, showing how the same qualitative conclusions vary across specific (T, m) combinations.

F.1 Disaggregated null results

Table F.1: Null designs N1–N2: separate results for $T = 1$ and $m \in \{200, 500\}$.

	Quantile rejection (%)		Expectile rejection (%)	
Method	$m = 200$	$m = 500$	$m = 200$	$m = 500$
SSMS-KS ($\gamma = 0$)	7.4	6.2	8.9	6.0
SSMS-KS ($\gamma = 0.15$)	8.3	6.4	9.6	6.6
RSMS-KS ($\gamma = 0$)	17.1	12.2	15.9	10.6
RSMS-KS ($\gamma = 0.15$)	19.1	12.6	17.8	11.2
SSMS-CvM (U)	7.6	6.9	8.6	6.6
SSMS-CvM (Early)	8.3	7.3	9.2	7.1
SSMS-CvM (Mid)	7.8	6.6	8.8	6.6
SSMS-CvM (Late)	7.3	6.6	8.4	6.3
RSMS-CvM (U)	17.2	12.6	16.1	10.5
RSMS-CvM (Early)	18.6	12.7	17.0	10.8
RSMS-CvM (Mid)	16.6	12.2	15.2	10.2
RSMS-CvM (Late)	16.0	11.9	15.0	10.0

Notes: The table reports pooled rejection frequencies (in percent) over the two null designs and the full grid of distributional levels; the only dimensions kept separate are the horizon T and the training size m. *Comment:* Moving from $m = 200$ to $m = 500$ lowers the null rejection frequency for 12/12 quantile methods and 12/12 expectile methods at this horizon, so the larger training sample is slightly more conservative here.

Table F.2: Null designs N1–N2: separate results for $T = 2$ and $m \in \{200, 500\}$.

	Quantile rejection (%)		Expectile rejection (%)	
Method	$m = 200$	$m = 500$	$m = 200$	$m = 500$
SSMS-KS ($\gamma = 0$)	8.0	6.6	8.4	7.0
SSMS-KS ($\gamma = 0.15$)	8.9	6.6	9.9	7.2
RSMS-KS ($\gamma = 0$)	16.1	11.5	16.4	9.9
RSMS-KS ($\gamma = 0.15$)	18.1	12.2	18.0	10.2
SSMS-CvM (U)	8.0	7.0	8.6	7.2
SSMS-CvM (Early)	8.1	7.1	8.9	6.8
SSMS-CvM (Mid)	7.2	6.2	7.8	6.5
SSMS-CvM (Late)	6.2	6.3	7.0	6.2
RSMS-CvM (U)	15.0	11.0	14.8	9.7
RSMS-CvM (Early)	15.6	11.2	14.6	9.6
RSMS-CvM (Mid)	14.4	10.5	13.9	8.8
RSMS-CvM (Late)	12.6	10.0	13.5	8.5

Notes: The table reports pooled rejection frequencies (in percent) over the two null designs and the full grid of distributional levels; the only dimensions kept separate are the horizon T and the training size m. *Comment:* Moving from $m = 200$ to $m = 500$ lowers the null rejection frequency for 11/12 quantile methods and 12/12 expectile methods at this horizon, so the larger training sample is slightly more conservative here.

Table F.3: Null designs N1–N2: separate results for $T = 5$ and $m \in \{200, 500\}$.

	Quantile rejection (%)		Expectile rejection (%)	
Method	$m = 200$	$m = 500$	$m = 200$	$m = 500$
SSMS-KS ($\gamma = 0$)	7.3	5.6	7.1	6.0
SSMS-KS ($\gamma = 0.15$)	8.0	5.7	7.8	6.2
RSMS-KS ($\gamma = 0$)	14.3	8.3	13.2	7.7
RSMS-KS ($\gamma = 0.15$)	16.6	9.1	15.1	8.4
SSMS-CvM (U)	5.5	4.8	5.7	5.5
SSMS-CvM (Early)	6.0	4.8	6.5	5.3
SSMS-CvM (Mid)	5.3	4.5	5.3	5.2
SSMS-CvM (Late)	5.5	4.9	5.5	5.3
RSMS-CvM (U)	11.5	8.0	10.4	7.3
RSMS-CvM (Early)	13.5	8.4	12.0	7.5
RSMS-CvM (Mid)	11.0	7.5	9.9	6.5
RSMS-CvM (Late)	10.8	7.5	9.8	6.7

Notes: The table reports pooled rejection frequencies (in percent) over the two null designs and the full grid of distributional levels; the only dimensions kept separate are the horizon T and the training size m. *Comment:* Moving from $m = 200$ to $m = 500$ lowers the null rejection frequency for 12/12 quantile methods and 12/12 expectile methods at this horizon, so the larger training sample is slightly more conservative here.

F.2 Disaggregated abrupt results

Table F.4: Abrupt location designs A1–A2: separate results for $T = 1$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	19.6	55.7	29.1	144.4	22.0	55.1	31.7	142.9
SSMS-KS ($\gamma = 0.15$)	18.5	53.4	27.7	139.9	20.6	52.9	30.2	138.0
RSMS-KS ($\gamma = 0$)	31.4	52.5	39.3	135.5	33.0	52.7	41.4	135.0
RSMS-KS ($\gamma = 0.15$)	28.9	49.9	37.4	130.6	30.6	50.4	39.5	130.2
SSMS-CvM (U)	16.9	70.7	22.6	182.5	18.8	69.9	24.6	179.3
SSMS-CvM (Early)	11.3	48.3	14.8	128.8	12.7	47.8	16.3	125.3
SSMS-CvM (Mid)	13.8	55.8	18.4	144.6	15.3	55.3	20.1	142.7
SSMS-CvM (Late)	18.9	77.4	25.0	195.8	20.9	76.3	27.4	193.2
RSMS-CvM (U)	29.2	69.4	31.6	177.3	29.6	68.9	32.9	178.9
RSMS-CvM (Early)	20.7	47.4	22.0	126.9	20.8	47.3	22.5	126.0
RSMS-CvM (Mid)	24.3	55.1	26.0	142.9	24.2	55.2	26.9	142.6
RSMS-CvM (Late)	31.5	76.1	34.4	192.7	32.2	75.8	35.9	192.7

Notes: The table reports pooled detection rates (in percent) and conditional delays over designs A1–A2, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* At this horizon, increasing the training size from $m = 200$ to $m = 500$ raises the detection rate for all reported SSMS/RSMS methods in both the quantile and expectile panels. Raw delays also increase in periods because the break date and monitoring window are longer when m is larger.

Table F.5: Abrupt location designs A1–A2: separate results for $T = 2$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	22.3	105.1	34.1	259.9	25.1	105.0	36.5	256.5
SSMS-KS ($\gamma = 0.15$)	20.4	100.9	32.3	256.9	23.1	101.6	34.7	252.5
RSMS-KS ($\gamma = 0$)	32.4	98.7	43.7	245.3	34.7	98.4	45.3	241.8
RSMS-KS ($\gamma = 0.15$)	29.6	96.5	41.4	238.7	31.8	95.3	43.0	234.9
SSMS-CvM (U)	18.4	140.9	25.8	352.7	20.6	138.3	28.5	349.8
SSMS-CvM (Early)	11.4	93.4	16.8	245.9	13.5	94.5	18.6	244.9
SSMS-CvM (Mid)	14.4	111.7	20.7	283.3	16.7	110.9	22.9	281.2
SSMS-CvM (Late)	19.4	154.9	27.8	384.9	21.9	151.6	30.1	379.2
RSMS-CvM (U)	29.8	135.5	34.5	348.3	31.4	136.8	35.8	343.4
RSMS-CvM (Early)	20.7	93.2	23.5	245.0	22.0	95.0	24.7	248.3
RSMS-CvM (Mid)	24.7	108.0	28.2	279.0	26.0	110.3	29.7	278.2
RSMS-CvM (Late)	31.5	151.1	36.9	379.3	32.8	150.3	38.2	374.1

Notes: The table reports pooled detection rates (in percent) and conditional delays over designs A1–A2, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* At this horizon, increasing the training size from $m = 200$ to $m = 500$ raises the detection rate for all reported SSMS/RSMS methods in both the quantile and expectile panels. Raw delays also increase in periods because the break date and monitoring window are longer when m is larger.

Table F.6: Abrupt location designs A1–A2: separate results for $T = 5$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	23.3	246.4	36.3	615.9	26.2	248.9	38.7	608.0
SSMS-KS ($\gamma = 0.15$)	21.5	246.4	34.5	621.6	24.3	245.6	36.8	616.0
RSMS-KS ($\gamma = 0$)	32.4	229.5	44.2	581.1	34.8	231.6	46.3	564.1
RSMS-KS ($\gamma = 0.15$)	30.1	227.6	42.1	573.4	32.4	230.7	44.6	561.4
SSMS-CvM (U)	17.4	348.9	25.7	865.1	19.6	346.9	28.1	854.0
SSMS-CvM (Early)	11.0	241.1	16.8	614.5	12.6	232.4	18.4	613.4
SSMS-CvM (Mid)	14.2	277.4	21.2	695.8	16.1	278.9	23.0	685.7
SSMS-CvM (Late)	20.6	384.3	30.2	942.3	23.1	383.2	32.3	915.7
RSMS-CvM (U)	28.3	342.6	34.0	866.9	29.4	341.8	35.5	846.7
RSMS-CvM (Early)	19.9	233.5	23.3	608.3	20.2	238.7	24.4	611.0
RSMS-CvM (Mid)	23.5	275.4	28.0	693.7	24.4	273.6	30.0	685.0
RSMS-CvM (Late)	31.4	378.5	37.8	931.1	32.5	374.9	39.4	911.7

Notes: The table reports pooled detection rates (in percent) and conditional delays over designs A1–A2, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* At this horizon, increasing the training size from $m = 200$ to $m = 500$ raises the detection rate for all reported SSMS/RSMS methods in both the quantile and expectile panels. Raw delays also increase in periods because the break date and monitoring window are longer when m is larger.

F.3 Disaggregated scale-design results

Table F.7: Scale design A3: separate results for $T = 1$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	9.3	50.9	7.8	135.0	17.4	48.3	16.1	122.5
SSMS-KS ($\gamma = 0.15$)	8.4	47.9	7.4	124.3	16.5	46.0	15.8	115.3
RSMS-KS ($\gamma = 0$)	18.6	49.3	14.4	130.4	27.9	47.5	23.9	119.9
RSMS-KS ($\gamma = 0.15$)	17.2	46.8	13.4	118.9	26.2	44.7	23.2	111.7
SSMS-CvM (U)	9.4	70.3	7.2	180.5	13.9	69.0	11.7	180.6
SSMS-CvM (Early)	7.4	45.6	5.4	121.0	10.3	44.6	8.7	119.1
SSMS-CvM (Mid)	8.3	54.4	6.2	139.7	12.1	53.7	10.4	142.1
SSMS-CvM (Late)	9.6	78.4	7.8	200.2	14.8	76.5	12.8	198.5
RSMS-CvM (U)	19.0	68.8	13.0	182.9	23.9	68.1	17.8	179.1
RSMS-CvM (Early)	14.5	44.5	10.3	120.4	17.9	46.2	13.5	121.3
RSMS-CvM (Mid)	16.5	53.8	11.3	145.1	20.5	53.9	15.8	140.5
RSMS-CvM (Late)	19.6	76.7	13.9	200.7	25.3	75.5	18.9	195.7

Notes: The table reports pooled detection rates (in percent) and conditional delays over design A3, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* For A3, the larger training sample does not help at $T = 1$: only 0/12 quantile methods and 0/12 expectile methods improve in detection when moving from $m = 200$ to $m = 500$, while raw delays rise in periods.

Table F.8: Scale design A3: separate results for $T = 2$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	7.4	92.9	7.0	228.2	15.1	86.1	14.0	228.6
SSMS-KS ($\gamma = 0.15$)	6.5	89.3	6.1	216.2	13.7	83.8	12.9	214.6
RSMS-KS ($\gamma = 0$)	15.5	90.5	11.1	238.8	24.4	86.1	20.9	225.5
RSMS-KS ($\gamma = 0.15$)	13.6	86.0	9.8	219.2	22.9	80.7	19.6	213.1
SSMS-CvM (U)	7.6	138.6	7.2	361.9	11.9	137.1	10.1	358.4
SSMS-CvM (Early)	5.6	90.4	5.3	238.2	8.4	89.6	7.0	233.7
SSMS-CvM (Mid)	6.5	110.1	6.1	288.1	9.8	108.1	8.2	279.6
SSMS-CvM (Late)	7.4	157.0	6.7	399.6	11.3	153.6	9.9	397.4
RSMS-CvM (U)	16.3	137.8	11.3	363.7	20.8	135.0	15.1	357.0
RSMS-CvM (Early)	12.4	90.3	8.4	245.9	15.2	88.3	10.9	241.7
RSMS-CvM (Mid)	13.9	109.0	9.6	289.1	17.9	107.8	12.7	280.3
RSMS-CvM (Late)	15.6	156.8	10.6	403.3	20.8	151.8	15.0	398.0

Notes: The table reports pooled detection rates (in percent) and conditional delays over design A3, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* For A3, the larger training sample does not help at $T = 2$: only 0/12 quantile methods and 0/12 expectile methods improve in detection when moving from $m = 200$ to $m = 500$, while raw delays rise in periods.

Table F.9: Scale design A3: separate results for $T = 5$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	4.4	197.0	4.5	538.5	9.0	189.0	7.9	535.2
SSMS-KS ($\gamma = 0.15$)	3.6	184.3	3.8	511.5	8.1	178.7	6.9	505.9
RSMS-KS ($\gamma = 0$)	8.6	191.3	6.7	545.0	14.9	199.2	12.4	507.1
RSMS-KS ($\gamma = 0.15$)	7.5	180.8	5.8	532.9	13.6	191.0	11.3	489.4
SSMS-CvM (U)	5.5	355.8	5.5	888.9	7.3	350.1	5.9	858.8
SSMS-CvM (Early)	4.3	229.2	4.1	569.3	5.4	219.9	4.3	569.1
SSMS-CvM (Mid)	4.6	278.1	4.6	703.5	6.3	274.7	5.1	692.8
SSMS-CvM (Late)	5.8	398.5	6.0	990.8	7.9	392.4	6.7	978.5
RSMS-CvM (U)	11.4	343.1	8.6	910.5	12.9	350.2	9.5	902.9
RSMS-CvM (Early)	9.1	217.1	7.0	558.9	10.2	222.7	7.3	580.3
RSMS-CvM (Mid)	9.6	268.0	7.2	722.4	11.0	274.8	8.2	710.9
RSMS-CvM (Late)	11.5	391.3	8.8	1010.8	13.4	391.3	10.2	1000.1

Notes: The table reports pooled detection rates (in percent) and conditional delays over design A3, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* For A3, the larger training sample does not help at $T = 5$: only 3/12 quantile methods and 0/12 expectile methods improve in detection when moving from $m = 200$ to $m = 500$, while raw delays rise in periods.

F.4 Disaggregated tail-design results

Table F.10: Lower-tail design A4: separate results for $T = 1$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	23.5	48.2	21.6	137.2	26.8	49.0	30.1	134.0
SSMS-KS ($\gamma = 0.15$)	20.6	46.7	20.2	130.5	24.6	47.2	28.6	127.4
RSMS-KS ($\gamma = 0$)	31.8	45.1	32.7	126.6	37.3	46.6	40.1	124.3
RSMS-KS ($\gamma = 0.15$)	27.2	43.8	29.8	121.1	33.8	44.1	38.1	119.2
SSMS-CvM (U)	27.6	62.4	19.7	178.7	26.4	65.6	24.7	177.2
SSMS-CvM (Early)	19.6	41.9	14.1	118.3	19.3	44.7	17.6	123.2
SSMS-CvM (Mid)	23.7	49.4	16.6	141.8	22.7	51.9	20.8	140.2
SSMS-CvM (Late)	30.2	69.6	21.4	195.7	28.3	72.8	27.0	192.9
RSMS-CvM (U)	41.7	58.7	31.6	173.4	41.3	63.6	35.1	174.8
RSMS-CvM (Early)	29.4	39.6	23.1	115.9	29.8	42.7	25.6	119.3
RSMS-CvM (Mid)	36.0	46.2	26.7	136.5	34.8	50.2	29.6	139.0
RSMS-CvM (Late)	46.3	66.1	34.2	189.7	44.4	70.7	37.7	190.1

Notes: The table reports pooled detection rates (in percent) and conditional delays over design A4, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* For A4, the effect of m is mixed at $T = 1$: 2/12 quantile methods and 4/12 expectile methods improve when moving from $m = 200$ to $m = 500$. The gains are therefore more method-dependent than in the abrupt, gradual, and contaminated designs.

Table F.11: Lower-tail design A4: separate results for $T = 2$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	22.6	91.5	24.0	260.5	27.6	92.2	33.4	243.1
SSMS-KS ($\gamma = 0.15$)	19.7	89.8	22.2	251.9	25.5	89.4	31.9	234.9
RSMS-KS ($\gamma = 0$)	29.3	84.5	34.4	235.5	36.8	84.0	42.7	229.3
RSMS-KS ($\gamma = 0.15$)	24.3	84.9	31.4	229.5	32.9	82.3	40.5	220.8
SSMS-CvM (U)	27.6	121.3	21.5	345.1	27.7	130.7	27.1	347.1
SSMS-CvM (Early)	18.9	83.5	13.8	241.3	19.1	89.6	18.1	240.4
SSMS-CvM (Mid)	22.5	95.7	17.1	277.3	22.6	104.1	21.6	273.8
SSMS-CvM (Late)	29.9	137.1	22.7	380.7	28.8	144.8	28.7	378.6
RSMS-CvM (U)	41.8	113.2	34.0	337.7	42.1	124.3	37.6	341.7
RSMS-CvM (Early)	29.0	75.6	23.3	228.2	30.5	84.1	26.1	241.9
RSMS-CvM (Mid)	36.0	89.8	28.3	267.1	35.7	99.1	31.1	274.2
RSMS-CvM (Late)	46.3	130.6	35.9	374.0	44.4	139.7	39.5	375.5

Notes: The table reports pooled detection rates (in percent) and conditional delays over design A4, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* For A4, the effect of m is mixed at $T = 2$: 4/12 quantile methods and 4/12 expectile methods improve when moving from $m = 200$ to $m = 500$. The gains are therefore more method-dependent than in the abrupt, gradual, and contaminated designs.

Table F.12: Lower-tail design A4: separate results for $T = 5$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	18.9	216.4	23.7	587.0	25.7	219.8	33.9	571.2
SSMS-KS ($\gamma = 0.15$)	16.0	214.6	21.9	588.8	23.2	217.0	31.9	568.4
RSMS-KS ($\gamma = 0$)	23.8	199.9	31.9	547.3	32.6	203.3	42.1	546.6
RSMS-KS ($\gamma = 0.15$)	19.6	202.2	28.8	547.5	29.0	203.0	39.1	536.5
SSMS-CvM (U)	27.3	291.7	22.1	847.0	26.1	319.6	26.7	861.4
SSMS-CvM (Early)	18.0	189.7	14.4	564.1	17.6	216.1	18.3	601.7
SSMS-CvM (Mid)	23.9	227.8	19.0	664.7	22.3	254.0	23.1	685.7
SSMS-CvM (Late)	31.8	333.0	25.9	940.0	29.8	359.2	30.4	939.5
RSMS-CvM (U)	39.9	275.0	33.7	813.1	40.0	307.7	36.7	843.3
RSMS-CvM (Early)	27.0	186.0	22.7	555.2	27.7	208.1	25.5	578.4
RSMS-CvM (Mid)	35.3	215.6	28.9	648.0	34.4	244.3	31.4	668.2
RSMS-CvM (Late)	46.9	315.1	37.5	911.0	44.1	348.9	40.5	926.7

Notes: The table reports pooled detection rates (in percent) and conditional delays over design A4, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* For A4, the effect of m is mixed at $T = 5$: 4/12 quantile methods and 8/12 expectile methods improve when moving from $m = 200$ to $m = 500$. The gains are therefore more method-dependent than in the abrupt, gradual, and contaminated designs.

F.5 Disaggregated gradual-design results

Table F.13: Gradual design A5: separate results for $T = 1$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	14.8	59.1	22.8	156.1	16.6	61.2	25.0	156.5
SSMS-KS ($\gamma = 0.15$)	13.5	57.3	21.6	150.7	15.3	59.6	23.7	151.5
RSMS-KS ($\gamma = 0$)	25.5	57.0	32.6	149.5	26.5	58.5	34.1	151.8
RSMS-KS ($\gamma = 0.15$)	23.2	55.2	30.8	143.9	24.5	57.4	32.3	147.2
SSMS-CvM (U)	12.7	71.7	16.9	187.4	14.2	72.8	18.3	191.3
SSMS-CvM (Early)	8.6	48.9	9.7	132.5	9.0	47.6	11.2	129.4
SSMS-CvM (Mid)	10.4	56.8	13.2	149.0	11.2	56.9	14.2	152.3
SSMS-CvM (Late)	14.1	79.4	19.2	201.4	15.6	79.6	20.5	204.5
RSMS-CvM (U)	23.5	70.5	25.5	185.4	23.9	71.9	25.9	188.2
RSMS-CvM (Early)	16.1	46.5	17.1	130.0	16.1	50.5	16.9	133.0
RSMS-CvM (Mid)	19.2	55.1	20.8	149.1	19.1	57.4	20.3	148.1
RSMS-CvM (Late)	25.3	77.4	28.1	198.5	26.0	78.7	28.3	199.8

Notes: The table reports pooled detection rates (in percent) and conditional delays over design A5, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* At this horizon, increasing the training size from $m = 200$ to $m = 500$ raises the detection rate for all reported SSMS/RSMS methods in both the quantile and expectile panels. Raw delays also increase in periods because the break date and monitoring window are longer when m is larger.

Table F.14: Gradual design A5: separate results for $T = 2$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	16.6	117.8	26.2	291.0	18.7	114.6	28.7	299.8
SSMS-KS ($\gamma = 0.15$)	14.6	114.1	24.5	289.6	16.9	114.4	27.0	294.8
RSMS-KS ($\gamma = 0$)	27.5	112.8	35.4	289.7	28.2	112.2	36.9	287.4
RSMS-KS ($\gamma = 0.15$)	24.7	110.2	33.6	289.5	26.2	108.8	34.7	283.0
SSMS-CvM (U)	13.0	144.0	19.1	366.0	14.6	141.9	20.4	370.3
SSMS-CvM (Early)	7.6	93.9	11.5	256.2	8.9	96.6	11.8	255.9
SSMS-CvM (Mid)	9.8	113.1	14.4	291.7	11.2	115.0	15.5	294.9
SSMS-CvM (Late)	13.8	157.5	20.9	396.8	15.3	158.0	22.1	399.5
RSMS-CvM (U)	24.4	142.3	27.1	368.1	24.8	141.6	28.4	369.0
RSMS-CvM (Early)	15.4	95.0	17.1	264.8	16.0	95.9	17.6	263.0
RSMS-CvM (Mid)	19.1	113.5	21.3	295.5	19.6	113.7	21.8	296.7
RSMS-CvM (Late)	25.5	157.5	29.0	397.5	26.2	156.2	30.1	397.2

Notes: The table reports pooled detection rates (in percent) and conditional delays over design A5, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* At this horizon, increasing the training size from $m = 200$ to $m = 500$ raises the detection rate for all reported SSMS/RSMS methods in both the quantile and expectile panels. Raw delays also increase in periods because the break date and monitoring window are longer when m is larger.

Table F.15: Gradual design A5: separate results for $T = 5$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	17.4	283.6	28.0	726.7	19.2	275.3	30.1	717.3
SSMS-KS ($\gamma = 0.15$)	15.5	278.8	26.4	731.6	17.6	270.6	28.5	727.8
RSMS-KS ($\gamma = 0$)	26.2	278.4	35.2	720.4	27.4	274.2	37.8	692.0
RSMS-KS ($\gamma = 0.15$)	24.0	273.0	33.7	717.0	25.4	273.5	36.0	690.4
SSMS-CvM (U)	12.1	362.4	18.9	917.1	13.5	363.2	20.0	902.0
SSMS-CvM (Early)	6.9	245.3	10.8	629.0	8.0	242.2	11.7	639.9
SSMS-CvM (Mid)	9.3	287.8	14.6	729.3	10.6	285.5	15.3	722.1
SSMS-CvM (Late)	14.6	396.9	22.7	981.3	16.1	395.1	23.6	970.3
RSMS-CvM (U)	21.6	353.0	25.4	905.8	22.0	359.6	26.9	917.9
RSMS-CvM (Early)	14.0	237.2	16.9	614.6	14.4	239.0	17.3	646.9
RSMS-CvM (Mid)	17.4	285.4	20.5	742.2	17.6	288.9	21.3	745.6
RSMS-CvM (Late)	24.7	392.3	29.1	970.4	24.9	392.8	30.9	979.1

Notes: The table reports pooled detection rates (in percent) and conditional delays over design A5, the full grid of distributional levels, signal strengths, and break fractions; only T and m are kept separate. *Comment:* At this horizon, increasing the training size from $m = 200$ to $m = 500$ raises the detection rate for all reported SSMS/RSMS methods in both the quantile and expectile panels. Raw delays also increase in periods because the break date and monitoring window are longer when m is larger.

F.6 Disaggregated contamination results

Table F.16: Contaminated-training design C1: separate results for $T = 1$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	29.1	116.4	34.3	306.9	30.8	114.8	34.8	300.2
SSMS-KS ($\gamma = 0.15$)	29.9	102.3	34.4	275.8	31.4	100.7	34.7	269.6
RSMS-KS ($\gamma = 0$)	50.0	108.4	55.5	287.7	49.3	111.7	54.2	295.3
RSMS-KS ($\gamma = 0.15$)	50.5	92.9	55.5	261.3	50.0	97.4	53.7	263.1
SSMS-CvM (U)	29.5	154.8	34.4	394.1	30.7	152.1	34.9	387.5
SSMS-CvM (Early)	27.7	120.1	31.2	308.9	28.2	118.7	31.8	301.6
SSMS-CvM (Mid)	27.9	137.8	32.7	351.2	29.1	136.6	32.9	343.8
SSMS-CvM (Late)	29.3	167.0	35.0	423.1	31.2	165.6	35.4	418.8
RSMS-CvM (U)	49.0	150.7	54.0	382.7	48.2	151.2	52.6	387.8
RSMS-CvM (Early)	45.9	115.5	50.4	300.4	44.6	116.2	48.2	302.4
RSMS-CvM (Mid)	46.8	134.5	51.9	342.4	45.4	135.1	49.9	345.2
RSMS-CvM (Late)	49.0	164.6	55.0	414.7	47.9	164.2	53.7	418.7

Notes: The table reports pooled detection rates (in percent) and conditional delays over design C1 and the full grid of distributional levels and signal strengths; only T and m are kept separate. *Comment:* At this horizon, increasing the training size from $m = 200$ to $m = 500$ raises the detection rate for all reported SSMS/RSMS methods in both the quantile and expectile panels. Raw delays also increase in periods because the break date and monitoring window are longer when m is larger.

G Auxiliary empirical summaries and figures

This appendix complements Section 6 by unpacking how the alarms are distributed across detectors and rolling windows and by documenting the full strategy leaderboard. Two summary statistics deserve explicit interpretation.

Taken together, Tables G.1 and G.2 echo the main-text comparison. The return leaders are

Table F.17: Contaminated-training design C1: separate results for $T = 2$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	36.2	195.6	41.9	518.5	38.6	190.9	43.4	496.2
SSMS-KS ($\gamma = 0.15$)	35.8	170.1	40.8	459.5	37.9	165.5	42.7	443.1
RSMS-KS ($\gamma = 0$)	57.1	184.7	62.9	480.4	56.9	180.1	63.0	478.4
RSMS-KS ($\gamma = 0.15$)	57.8	159.7	62.2	420.4	57.4	155.4	62.6	426.5
SSMS-CvM (U)	35.1	292.5	41.8	747.6	37.5	290.6	44.1	742.2
SSMS-CvM (Early)	32.7	224.9	38.4	581.5	34.3	219.8	40.9	576.6
SSMS-CvM (Mid)	33.0	262.4	39.3	670.9	35.4	262.5	41.5	666.0
SSMS-CvM (Late)	33.8	326.6	40.5	823.8	36.4	325.1	42.3	815.6
RSMS-CvM (U)	55.5	288.9	62.3	722.6	56.2	285.8	62.9	726.1
RSMS-CvM (Early)	51.7	218.8	59.0	562.9	52.9	217.1	59.3	567.8
RSMS-CvM (Mid)	52.9	260.9	60.0	650.4	53.9	258.0	61.2	656.8
RSMS-CvM (Late)	54.0	321.4	62.0	801.6	55.0	319.6	62.2	803.6

Notes: The table reports pooled detection rates (in percent) and conditional delays over design C1 and the full grid of distributional levels and signal strengths; only T and m are kept separate. *Comment:* At this horizon, increasing the training size from $m = 200$ to $m = 500$ raises the detection rate for all reported SSMS/RSMS methods in both the quantile and expectile panels. Raw delays also increase in periods because the break date and monitoring window are longer when m is larger.

Table F.18: Contaminated-training design C1: separate results for $T = 5$ and $m \in \{200, 500\}$.

	Quantile, $m = 200$		Quantile, $m = 500$		Expectile, $m = 200$		Expectile, $m = 500$	
Method	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay	Det. (%)	Delay
SSMS-KS ($\gamma = 0$)	40.5	364.4	48.3	890.2	42.8	357.0	48.4	892.6
SSMS-KS ($\gamma = 0.15$)	40.8	307.8	47.3	788.3	42.5	298.9	47.8	773.7
RSMS-KS ($\gamma = 0$)	60.0	339.8	66.4	822.6	60.6	336.7	67.2	846.2
RSMS-KS ($\gamma = 0.15$)	60.3	278.5	65.9	718.7	61.5	286.8	67.1	740.2
SSMS-CvM (U)	39.1	714.1	48.3	1726.5	41.8	701.4	48.3	1725.9
SSMS-CvM (Early)	38.2	537.8	46.3	1305.8	40.5	529.7	46.6	1306.4
SSMS-CvM (Mid)	37.2	643.0	46.8	1562.0	40.6	637.0	46.4	1555.6
SSMS-CvM (Late)	39.3	798.9	48.9	1955.5	42.1	791.3	48.7	1945.3
RSMS-CvM (U)	57.8	679.9	66.2	1654.5	59.2	683.1	67.1	1670.1
RSMS-CvM (Early)	56.8	501.0	64.9	1231.6	58.5	508.5	65.5	1246.7
RSMS-CvM (Mid)	56.6	619.9	65.0	1505.4	57.5	620.1	65.4	1513.6
RSMS-CvM (Late)	57.6	778.3	66.3	1897.4	59.2	779.8	67.2	1909.0

Notes: The table reports pooled detection rates (in percent) and conditional delays over design C1 and the full grid of distributional levels and signal strengths; only T and m are kept separate. *Comment:* At this horizon, increasing the training size from $m = 200$ to $m = 500$ raises the detection rate for all reported SSMS/RSMS methods in both the quantile and expectile panels. Raw delays also increase in periods because the break date and monitoring window are longer when m is larger.

Table G.1: Best practical-score configuration within each alarm-to-trade class

Rule class	Best practical-score method	Ann. return (%)	Total (%)	Sortino	Max DD (%)	Precision (%)	Capture (%)	Alarms
Buy-and-hold benchmark	No-alarm / always long	57.4	21.5	2.859	13.3	–	–	0
Flat 6h	HAC-CvM (Late)	66.6	25.8	3.370	11.6	33.3	2.4	6
Flat 24h	HAC-CvM (Late)	64.0	24.6	3.265	11.6	66.7	4.2	6
Short 6h	SSMS-CvM (Mid)	66.0	25.5	3.324	11.6	66.7	1.2	3
Short 24h	SSMS-CvM (Mid)	64.6	24.8	3.251	11.6	100.0	4.2	3

Notes: Practical score is the equally weighted average of within-class ranks for Sortino, maximum drawdown, precision, and event capture. The buy-and-hold benchmark is the no-alarm path; on this sample all three e-process rules coincide with that benchmark because they never cross their boundaries. Total return and annualized return are both reported so that the table can be read either as a sample-period performance summary or as an annualized comparison.

Table G.2: Average performance by detector family and trading rule

Rule class	Family	Ann. return (%)	Total (%)	Sortino	Max DD (%)	Precision (%)	Capture (%)	Avg. alarms
Flat 6h	E-process	57.4	21.5	2.859	13.3	–	–	0.0
	SSMS-KS	57.5	21.6	2.864	13.2	0.0	0.0	1.5
	RSMS-KS	53.3	19.7	2.658	14.6	10.0	0.3	4.0
	HAC-KS	52.9	19.5	2.635	14.7	16.7	0.3	2.5
	SSMS-CvM	59.6	22.6	2.988	12.6	45.8	0.8	2.5
	RSMS-CvM	62.3	23.8	3.116	12.3	12.5	0.5	5.8
	HAC-CvM	55.4	20.7	2.778	14.0	19.6	0.9	4.8
Flat 24h	E-process	57.4	21.5	2.859	13.3	–	–	0.0
	SSMS-KS	58.8	22.2	2.931	12.8	25.0	0.3	1.5
	RSMS-KS	57.4	21.6	2.877	13.2	73.3	1.8	4.0
	HAC-KS	57.2	21.5	2.862	13.3	100.0	1.5	2.5
	SSMS-CvM	59.2	22.4	2.982	13.1	79.2	2.1	2.5
	RSMS-CvM	61.6	23.6	3.111	13.0	34.2	2.1	5.8
	HAC-CvM	57.7	21.8	2.909	13.2	46.7	1.8	4.8
Short 6h	E-process	57.4	21.5	2.859	13.3	–	–	0.0
	SSMS-KS	57.6	21.6	2.868	13.2	0.0	0.0	1.5
	RSMS-KS	49.3	17.9	2.450	15.9	10.0	0.3	4.0
	HAC-KS	48.4	17.5	2.405	16.2	16.7	0.3	2.5
	SSMS-CvM	61.9	23.6	3.101	12.2	45.8	0.8	2.5
	RSMS-CvM	67.2	26.1	3.360	11.8	12.5	0.5	5.8
	HAC-CvM	53.5	20.0	2.650	15.2	19.6	0.9	4.8
Short 24h	E-process	57.4	21.5	2.859	13.3	–	–	0.0
	SSMS-KS	60.2	22.8	3.000	12.3	25.0	0.3	1.5
	RSMS-KS	57.4	21.6	2.869	13.6	73.3	1.8	4.0
	HAC-KS	57.1	21.4	2.846	13.5	100.0	1.5	2.5
	SSMS-CvM	61.0	23.3	3.068	13.5	79.2	2.1	2.5
	RSMS-CvM	65.9	25.7	3.285	13.7	34.2	2.1	5.8
	HAC-CvM	58.1	22.0	2.884	13.9	46.7	1.8	4.8

Notes: Each entry averages over all methods inside the named family and rule class. The e-process family contains the mixture, adaptive, and restart-bank e-processes; the KS and CvM families are split by studentizer because their alarm intensity differs materially in practice. Precision and capture averages omit undefined no-alarm rows.

mostly CvM rules, especially in the RSMS family, but the methods that look most balanced once drawdown and alarm quality are taken seriously are HAC-CvM (Late) for flat overlays and SSMS-CvM (Mid) for short overlays. By contrast, the e-process rows coincide with the buy-and-hold benchmark because there are no alarms at all in this sample. In finance terms, the appendix therefore shows a clean split between conservative detectors that preserve long exposure and more reactive detectors that occasionally transform the signal into economically meaningful de-risking or short overlays.

Figure G.1: BTC spot price and alarm timing across all methods. The top panel plots the BTC spot price on the observed merged sample. The lower raster marks the first alarm time within each rolling monitoring window for detectors that trigger at least once. Detectors with no alarms are therefore absent from the raster and appear only as blank rows in the companion heatmap. Light shaded vertical bands mark longer merged-data gaps. The figure is useful for reading the timing and clustering of alarms, whereas the tables remain the primary source for economic-performance comparisons.

The two alarm-timing figures are best read together. Figure G.1 shows *when* the economically relevant alarms arrive in calendar time, whereas Figure G.2 shows *which* detectors are active across the 34 rolling experiments. The common message is alarm sparsity. Several detectors never fire,

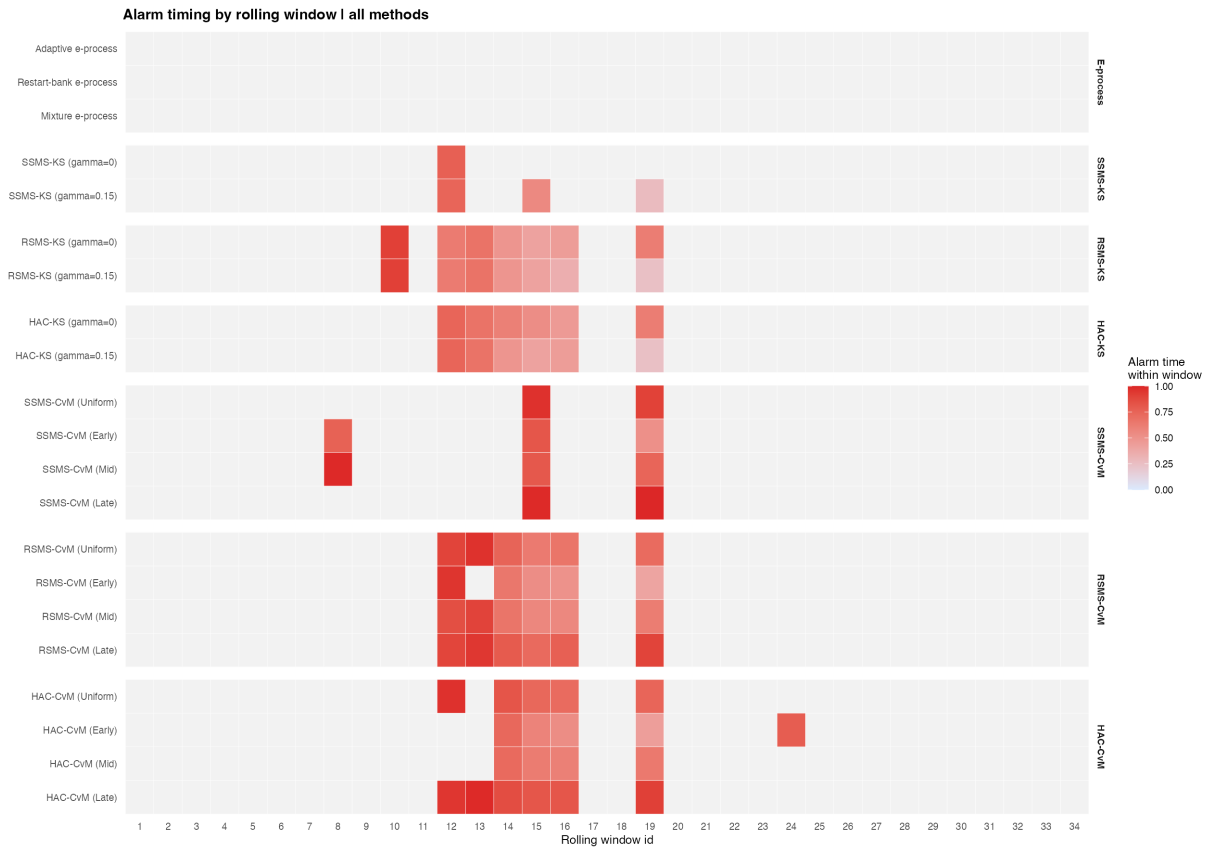


Figure G.2: Alarm-incidence heatmap by detector and rolling window. A colored cell indicates that the detector issues a first alarm in that rolling monitoring window; a blank cell means no alarm in that window. Entirely blank rows correspond to silent detectors, including all three e-process rules. The heatmap makes clear that alarm activity is sparse and clustered in only a subset of windows, which is why many strategy paths remain close to the benchmark for most of the sample.

several others fire only once or twice, and most of the action is concentrated in a small group of CvM methods. This timing structure explains the detector-by-detector equity panels below: where the black buy-and-hold benchmark is not visible, it is typically lying exactly underneath the colored strategy line. Those overlapping segments correspond to hours in which the alarm rule does not create an active exposure change, so the strategy is simply matching buy-and-hold. Visible divergences appear only when a detector generates a sufficient cluster of alarms to move the rule away from the benchmark.

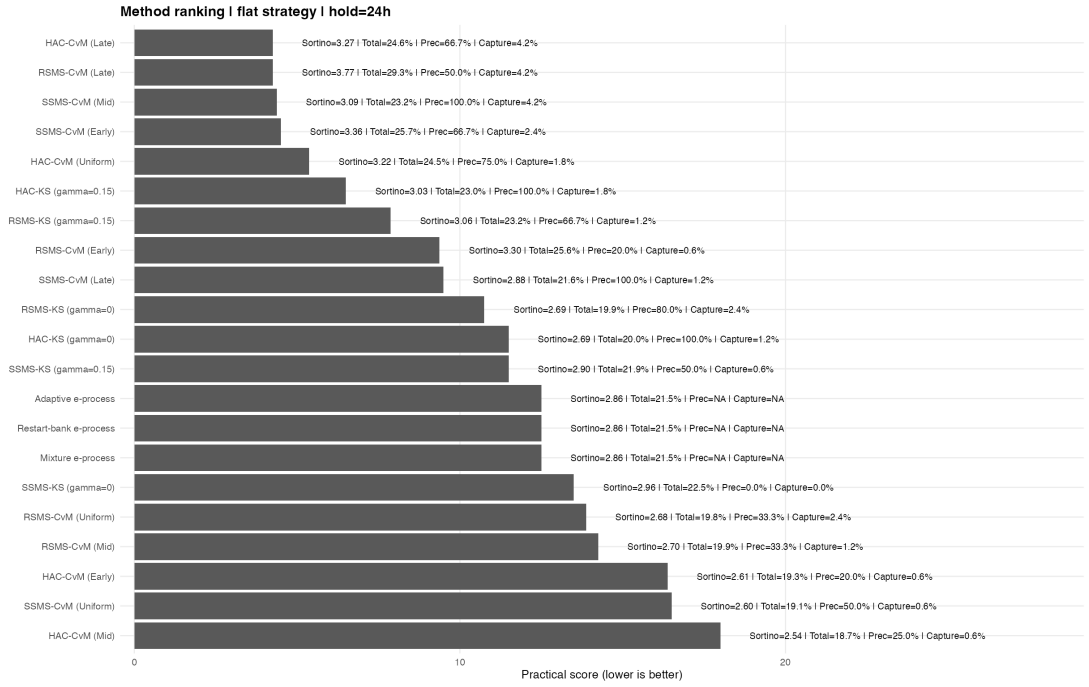


Figure G.3: Practical-score ranking for the flat 24-hour rule. Lower bars are better because the practical score averages within-class ranks for Sortino, drawdown, precision, and event capture. The text annotations report each method's Sortino ratio, total return, precision, and event capture. The figure is useful for visual ranking, but the appendix tables remain the authoritative source because they report the same quantities for every rule class.

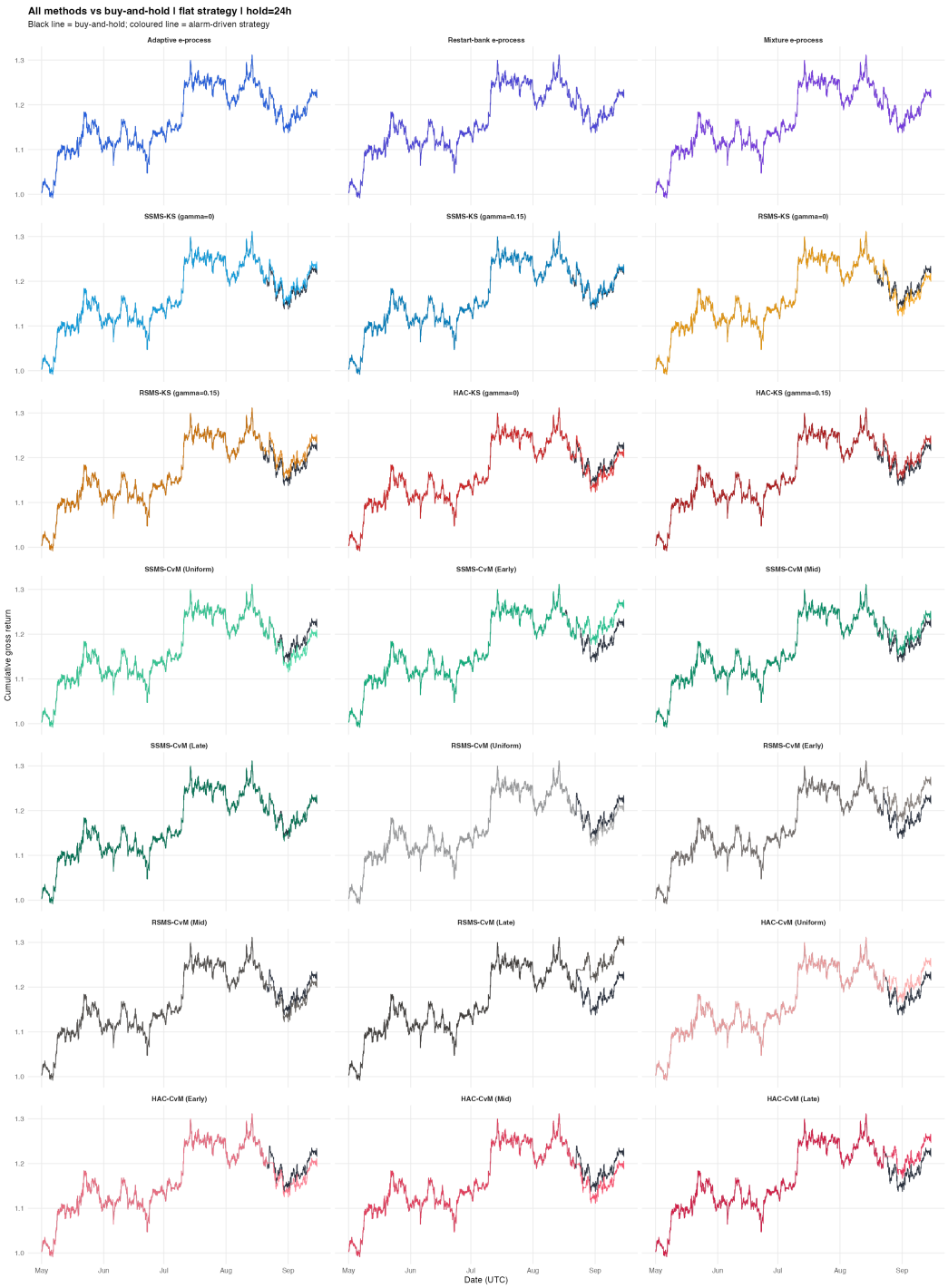


Figure G.4: Detector-by-detector equity curves for the flat 24-hour rule. Each panel compares the buy-and-hold benchmark (black line) with the alarm-driven strategy implied by one detector (colored line), evaluated on observed merged hours only. When the two paths overlap exactly, the rendered plot shows only the colored line because it lies on top of the black benchmark. Such overlap means the rule is behaving just like buy-and-hold at that time, i.e. there is no active alarm-driven departure from the benchmark. The panels therefore visualize alarm timing as much as they visualize return differences.

H Complete empirical strategy tables

For completeness, this appendix reports the full 84-rule empirical leaderboard. The tables are grouped by the alarm-to-trade action taken after a first boundary crossing: move flat for 6 hours, move flat for 24 hours, go short for 6 hours, and go short for 24 hours. A row with zero alarms should be read as benchmark replication, whereas differences across rows reveal how the same monitoring evidence maps into economically different exposure rules.

Table H.1: Full empirical results: flat-on-alarm, 6-hour hold

Method	Ann. return (%)	Total (%)	Ending gross	Sortino	Max DD (%)	Precision (%)	Capture (%)	Alarms	Practical score
HAC-CvM (Late)	66.6	25.8	1.258	3.370	11.6	33.3	2.4	6	2.375
RSMS-CvM (Late)	70.8	27.8	1.278	3.568	11.6	33.3	1.2	6	2.625
SSMS-CvM (Mid)	61.7	23.5	1.235	3.106	11.8	66.7	1.2	3	2.625
SSMS-CvM (Early)	59.3	22.4	1.224	2.984	12.6	66.7	1.2	3	3.875
HAC-CvM (Uniform)	58.7	22.1	1.221	2.928	12.8	25.0	0.6	4	8.125
SSMS-CvM (Uniform)	57.8	21.7	1.217	2.882	13.1	50.0	0.6	2	8.125
RSMS-CvM (Uniform)	58.8	22.2	1.222	2.935	12.8	16.7	0.6	6	8.375
RSMS-CvM (Mid)	60.0	22.7	1.227	2.989	12.4	0.0	0.0	6	9.250
RSMS-CvM (Early)	59.6	22.5	1.225	2.970	12.6	0.0	0.0	5	10.250
SSMS-CvM (Late)	59.8	22.6	1.226	2.982	12.8	0.0	0.0	2	10.500
HAC-KS ($\gamma = 0.15$)	52.4	19.3	1.193	2.613	14.9	33.3	0.6	3	12.125
SSMS-KS ($\gamma = 0$)	58.4	22.0	1.220	2.911	12.9	0.0	0.0	1	12.250
Adaptive e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	13.000
Restart-bank e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	13.000
Mixture e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	13.000
RSMS-KS ($\gamma = 0$)	51.3	18.8	1.188	2.557	15.2	20.0	0.6	5	13.500
HAC-CvM (Early)	47.3	17.1	1.171	2.363	15.7	20.0	0.6	5	14.250
SSMS-KS ($\gamma = 0.15$)	56.5	21.1	1.211	2.817	13.5	0.0	0.0	2	14.750
RSMS-KS ($\gamma = 0.15$)	55.4	20.6	1.206	2.758	13.9	0.0	0.0	3	15.250
HAC-KS ($\gamma = 0$)	53.3	19.7	1.197	2.657	14.6	0.0	0.0	2	15.750
HAC-CvM (Mid)	49.2	17.9	1.179	2.451	15.9	0.0	0.0	4	17.500

Notes: This table reports all 21 detector configurations for the flat-on-alarm rule with a 6-hour holding period. Total return and ending gross are measured over the observed merged hours of the empirical sample. Practical score is the equally weighted average of within-class ranks for Sortino, maximum drawdown, precision, and event capture. A row with zero alarms is mechanically identical to buy-and-hold; on this sample the three e-process rules provide that benchmark.

Table H.2: Full empirical results: flat-on-alarm, 24-hour hold

Method	Ann. return (%)	Total (%)	Ending gross	Sortino	Max DD (%)	Precision (%)	Capture (%)	Alarms	Practical score
RSMS-CvM (Late)	73.7	29.3	1.293	3.766	11.6	50.0	4.2	6	4.250
HAC-CvM (Late)	64.0	24.6	1.246	3.265	11.6	66.7	4.2	6	4.250
SSMS-CvM (Mid)	61.0	23.2	1.232	3.091	12.0	100.0	4.2	3	4.375
SSMS-CvM (Early)	66.3	25.7	1.257	3.364	11.6	66.7	2.4	3	4.500
HAC-CvM (Uniform)	63.8	24.5	1.245	3.220	11.6	75.0	1.8	4	5.375
HAC-KS ($\gamma = 0.15$)	60.5	23.0	1.230	3.032	12.2	100.0	1.8	3	6.500
RSMS-KS ($\gamma = 0.15$)	61.1	23.2	1.232	3.059	12.0	66.7	1.2	3	7.875
RSMS-CvM (Early)	66.1	25.6	1.256	3.297	11.6	20.0	0.6	5	9.375
SSMS-CvM (Late)	57.6	21.6	1.216	2.875	13.7	100.0	1.2	2	9.500
RSMS-KS ($\gamma = 0$)	53.7	19.9	1.199	2.694	14.4	80.0	2.4	5	10.750
SSMS-KS ($\gamma = 0.15$)	58.1	21.9	1.219	2.900	13.0	50.0	0.6	2	11.500
HAC-KS ($\gamma = 0$)	54.0	20.0	1.200	2.693	14.4	100.0	1.2	2	11.500
Adaptive e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	12.500
Restart-bank e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	12.500
Mixture e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	12.500
SSMS-KS ($\gamma = 0$)	59.4	22.5	1.225	2.962	12.6	0.0	0.0	1	13.500
RSMS-CvM (Uniform)	53.3	19.8	1.198	2.684	14.5	33.3	2.4	6	13.875
RSMS-CvM (Mid)	53.5	19.9	1.199	2.696	14.4	33.3	1.2	6	14.250
HAC-CvM (Early)	52.1	19.3	1.193	2.607	14.2	20.0	0.6	5	16.375
SSMS-CvM (Uniform)	52.1	19.1	1.191	2.597	15.0	50.0	0.6	2	16.500
HAC-CvM (Mid)	51.0	18.7	1.187	2.544	15.3	25.0	0.6	4	18.000

Notes: This table reports all 21 detector configurations for the flat-on-alarm rule with a 24-hour holding period. Total return and ending gross are measured over the observed merged hours of the empirical sample. Practical score is the equally weighted average of within-class ranks for Sortino, maximum drawdown, precision, and event capture. A row with zero alarms is mechanically identical to buy-and-hold; on this sample the three e-process rules provide that benchmark.

Table H.3: Full empirical results: short-on-alarm, 6-hour hold

Method	Ann. return (%)	Total (%)	Ending gross	Sortino	Max DD (%)	Precision (%)	Capture (%)	Alarms	Practical score
SSMS-CvM (Mid)	66.0	25.5	1.255	3.324	11.6	66.7	1.2	3	2.375
HAC-CvM (Late)	75.8	30.2	1.302	3.832	11.6	33.3	2.4	6	2.500
RSMS-CvM (Late)	84.2	34.4	1.344	4.241	11.6	33.3	1.2	6	2.750
SSMS-CvM (Early)	61.2	23.3	1.233	3.079	12.0	66.7	1.2	3	4.125
HAC-CvM (Uniform)	60.0	22.7	1.227	2.993	12.4	25.0	0.6	4	8.125
SSMS-CvM (Uniform)	58.1	21.9	1.219	2.898	13.0	50.0	0.6	2	8.125
RSMS-CvM (Uniform)	60.2	22.8	1.228	3.005	12.4	16.7	0.6	6	8.375
RSMS-CvM (Mid)	62.6	23.9	1.239	3.119	11.6	0.0	0.0	6	9.250
SSMS-CvM (Late)	62.3	23.8	1.238	3.104	12.3	0.0	0.0	2	10.250
RSMS-CvM (Early)	61.7	23.5	1.235	3.076	11.8	0.0	0.0	5	10.250
HAC-KS ($\gamma = 0.15$)	47.4	17.1	1.171	2.360	16.4	33.3	0.6	3	12.125
SSMS-KS ($\gamma = 0$)	59.5	22.5	1.225	2.963	12.6	0.0	0.0	1	12.250
Adaptive e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	13.000
Restart-bank e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	13.000
Mixture e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	13.000
RSMS-KS ($\gamma = 0$)	45.2	16.1	1.161	2.246	17.1	20.0	0.6	5	13.500
HAC-CvM (Early)	37.2	12.7	1.127	1.787	18.2	20.0	0.6	5	14.250
SSMS-KS ($\gamma = 0.15$)	55.7	20.7	1.207	2.773	13.8	0.0	0.0	2	14.750
RSMS-KS ($\gamma = 0.15$)	53.3	19.7	1.197	2.654	14.6	0.0	0.0	3	15.250
HAC-KS ($\gamma = 0$)	49.3	17.9	1.179	2.450	15.9	0.0	0.0	2	15.750
HAC-CvM (Mid)	41.0	14.2	1.142	1.989	18.5	0.0	0.0	4	17.500

Notes: This table reports all 21 detector configurations for the short-on-alarm rule with a 6-hour holding period. Total return and ending gross are measured over the observed merged hours of the empirical sample. Practical score is the equally weighted average of within-class ranks for Sortino, maximum drawdown, precision, and event capture. A row with zero alarms is mechanically identical to buy-and-hold; on this sample the three e-process rules provide that benchmark.

Table H.4: Full empirical results: short-on-alarm, 24-hour hold

Method	Ann. return (%)	Total (%)	Ending gross	Sortino	Max DD (%)	Precision (%)	Capture (%)	Alarms	Practical score
SSMS-CvM (Mid)	64.6	24.8	1.248	3.251	11.6	100.0	4.2	3	3.750
RSMS-CvM (Late)	90.0	37.3	1.373	4.583	11.6	50.0	4.2	6	4.625
HAC-CvM (Late)	70.6	27.7	1.277	3.586	11.6	66.7	4.2	6	4.625
SSMS-CvM (Early)	75.1	29.9	1.299	3.809	11.6	66.7	2.4	3	4.875
HAC-KS ($\gamma = 0.15$)	63.7	24.4	1.244	3.179	11.6	100.0	1.8	3	5.625
HAC-CvM (Uniform)	70.3	27.5	1.275	3.537	11.6	75.0	1.8	4	5.750
RSMS-KS ($\gamma = 0.15$)	64.8	24.9	1.249	3.240	11.6	66.7	1.2	3	7.500
SSMS-CvM (Late)	57.7	21.7	1.217	2.883	14.2	100.0	1.2	2	9.500
RSMS-CvM (Early)	74.8	29.7	1.297	3.727	11.6	20.0	0.6	5	9.750
RSMS-KS ($\gamma = 0$)	50.1	18.2	1.182	2.497	15.6	80.0	2.4	5	10.750
HAC-KS ($\gamma = 0$)	50.5	18.4	1.184	2.513	15.5	100.0	1.2	2	11.000
SSMS-KS ($\gamma = 0.15$)	58.9	22.2	1.222	2.934	12.8	50.0	0.6	2	11.500
Adaptive e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	12.500
Restart-bank e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	12.500
Mixture e-process	57.4	21.5	1.215	2.859	13.3	–	–	0	12.500
SSMS-KS ($\gamma = 0$)	61.5	23.4	1.234	3.065	11.9	0.0	0.0	1	13.500
RSMS-CvM (Uniform)	49.2	17.8	1.178	2.400	15.9	33.3	2.4	6	13.875
RSMS-CvM (Mid)	49.7	18.0	1.180	2.430	15.7	33.3	1.2	6	14.750
SSMS-CvM (Uniform)	46.8	16.8	1.168	2.330	16.7	50.0	0.6	2	16.250
HAC-CvM (Early)	46.9	16.8	1.168	2.252	15.2	20.0	0.6	5	16.625
HAC-CvM (Mid)	44.6	15.8	1.158	2.161	17.3	25.0	0.6	4	18.000

Notes: This table reports all 21 detector configurations for the short-on-alarm rule with a 24-hour holding period. Total return and ending gross are measured over the observed merged hours of the empirical sample. Practical score is the equally weighted average of within-class ranks for Sortino, maximum drawdown, precision, and event capture. A row with zero alarms is mechanically identical to buy-and-hold; on this sample the three e-process rules provide that benchmark.