Online Monitoring of Financial Functional Time Series

When should a functional representation be refreshed?

Jiajing Sun

joint work with Meiting Zhu, Wolfgang Karl Härdle, Oliver Linton, and Zhuo Lin

Jiajing Sun	University of Chinese Academy of Sciences
Meiting Zhu	Xiamen University
Wolfgang Karl Härdle	Humboldt-Universität zu Berlin; IDA, Bucharest University of Economic Studies
Oliver Linton	University of Cambridge
Zhuo Lin	University of Chinese Academy of Sciences

4th IMA Conference on Mathematics of Finance and Insurance

The Liner Hotel, Liverpool June 2026

Practical question and roadmap

Question

A functional representation is estimated from a stable training sample. When do incoming observations say that it should be refreshed?

- Examples: yield curves, volatility profiles, intraday return curves.
- We need an online monitoring statistic, not only an ex post break test.
- The alarm is a model-maintenance trigger: re-estimate, review specification, or increase risk monitoring.

Construction

1. Fix training-sample FPCA coordinates.
2. Monitor the incoming score stream.
3. Compare HAC, SSMS, and RSMS scaling.

Evidence

1. Simulations: size, size-adjusted power, and delay.
2. Empirical application: S&P 500 COVID return curves.

Why function-valued observations?

- Scalar summaries may hide economically relevant movement.
- Functional time series can change through level, slope, curvature, timing, dispersion, or shape.
- For tractability, we monitor retained FPCA scores estimated from the training sample.

Key modeling choice

Estimate the FPCA basis once on the training sample, then keep it fixed during monitoring.

Curve-valued observations

- yield curves,
- intraday return profiles,
- realized volatility curves,
- term structures of risk measures.

Higher-dimensional objects

- option-implied volatility surfaces,
- limit-order-book shape profiles,
- cross-sectional panels of curves.

A score mean change may represent a shape change, not just a scalar level shift.

What is new, and how this differs

Contribution

1. We formulate online monitoring of a function-valued time series as a fixed-training FPCA score problem.
2. We compare HAC, Shao-type self-normalized, and adjusted-range self-normalized statistics on the same score stream.
3. We study both false-alarm control and post-break timing.
4. We document the method in intraday S&P 500 return curves around the COVID market disruption.

Retrospective change-point testing

- uses a completed sample,
- estimates or tests where a break occurred,
- often focuses on ex post dating.

Online monitoring

- receives observations sequentially,
- compares incoming data to a training sample,
- stops when the current representation is no longer credible.

The stopping time is a decision time, not necessarily the structural break date.

A companion result for general time series

The same adjusted-range self-normalization is also developed for general time series, where the monitored object is a parameter θ defined by an estimating equation

$$\sum_t \psi(X_t, \widehat{\theta}) = 0,$$

covering means, variances, quantiles, and M-, GMM-, or quantile-regression targets.

Same construction	Same conclusion
The same kernel-based HAC, Shao-type, and adjusted-range standardizations, the same early-detection boundary, and the same null-limit, consistency, and detection-delay theory.	The adjusted-range monitor beats Shao-type self-normalization and is competitive with HAC once power is size-adjusted; it is robust to mild training contamination and tuning-free.

- Empirical check there: high-frequency USD/GBP around the 2016 referendum and the 2020 COVID turmoil.

Companion paper: Zhu, Hong, Sun, and Linton (2026), revise & resubmit, *Journal of Business & Economic Statistics*.

From general time series to functional data

In the general theory, FPCA is a bridge

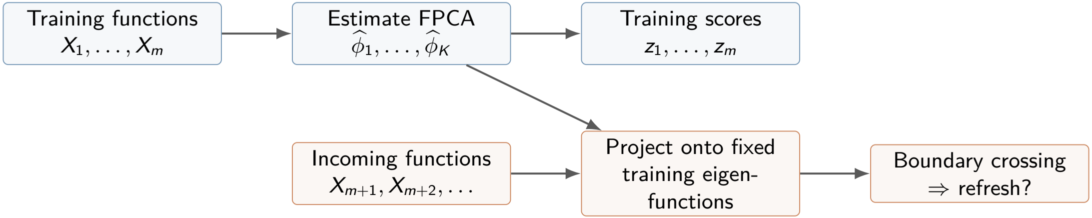
Function-valued data are reduced to FPCA scores, and the score vector is then monitored as a generic parameter.

What this talk makes its object

- Estimate the FPCA basis once on the training sample and freeze it; project incoming curves onto these fixed eigenfunctions, so the alarm reflects new data in fixed coordinates rather than a drifting basis.
- Develop the monitoring theory directly in these fixed functional coordinates, for a fixed number of retained components.
- Read a score-mean change as a shape change of the underlying function.

The rest of the talk makes this functional construction precise.

Training-sample FPCA and online monitoring



- The training sample estimates empirical eigenfunctions $\widehat{\phi}_1, \ldots, \widehat{\phi}_K$.
- During online monitoring, each incoming function is projected onto those same eigenfunctions.
- The question becomes: does the incoming score stream still look like the training-sample score stream?

Score construction and monitoring target

$$a_t = \left(\langle X_t, \widehat{\phi}_1 \rangle, \ldots, \langle X_t, \widehat{\phi}_K \rangle\right)^\top,$$

$$\bar{a}_m = \frac{1}{m}\sum_{t=1}^{m} a_t, \qquad z_t = a_t - \bar{a}_m,$$

$$S_m(k) = \sum_{t=m+1}^{m+k} z_t.$$

Mean-constancy formulation

$$H_0: \; \mathrm{E}X_t = \mu, \quad 1 \le t \le m + \lfloor mT \rfloor,$$

against a post-training mean change at an unknown structural break date.

- The training mean is not treated as negligible.
- It generates the training-centered Brownian limit

$$U_K(s) = B_K(1+s) - (1+s)B_K(1).$$

Monitoring signal and notation

Interpretation

The stopping time is an operational trigger. It says that enough evidence has accumulated to question the current functional representation.

- The structural change may have started before the alarm.
- A practitioner can use the alarm to re-estimate the basis, review specification, or increase risk monitoring.
- This is why detection delay matters in addition to rejection probability.

Object	Meaning
m	training-sample size
T	monitoring horizon in training-sample units
K	number of retained FPCA components
z_t	training-centered retained score vector
$S_m(k)$	monitoring cumulative score after k new observations
γ	boundary exponent for early monitoring times
τ_m	method-specific stopping time

Three normalizers for the same score path

Let $P_t = \sum_{j=1}^{t} z_j$ and $\widetilde{P}_t = \sum_{j=1}^{t} \widehat{W}_m z_j$, where $\widehat{W}_m$ is the training lag-0 transform.

HAC: kernel long-run variance normalizer

$$\widehat{\Gamma}_m = \widehat{\Gamma}_0 + \sum_{\ell=1}^{L_m} K_\ell(\widehat{\Gamma}_\ell + \widehat{\Gamma}_\ell^\top), \quad \widehat{\Gamma}_\ell = \frac{1}{m} \sum_{t=\ell+1}^{m} z_t z_{t-\ell}^\top.$$

Kernel and bandwidth choices enter through K_ℓ and L_m.

SSMS: Shao-type quadratic path self-normalizer

$$D_m = \frac{1}{m^2} \sum_{t=1}^{m} P_t P_t^\top.$$

This uses a quadratic functional of the training cumulative-score path.

RSMS: adjusted-range self-normalizer

$$\widetilde{R}_m = \operatorname{diag}(\widetilde{R}_{m,1}, \ldots, \widetilde{R}_{m,K}), \quad \widetilde{R}_{m,\ell} = \frac{\max_{t\le m} \widetilde{P}_{t,\ell} - \min_{t\le m} \widetilde{P}_{t,\ell}}{\sqrt{m}}.$$

Adjusted range after the lag-0 transform; the statistic uses $\widetilde{R}_m^{-2}$.

Main KS monitoring statistics

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

$$\mathcal{M}_m^{\mathrm{SSMS}}(k) = \frac{S_m(k)^\top D_m^{-1} S_m(k)}{m\, g_\gamma(k/m)^2},$$

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

$$g_\gamma(s) = (1+s)\left(\frac{s}{1+s}\right)^\gamma, \qquad \gamma \in [0, 1/2).$$

- Alarm time: first k such that $\mathcal{M}_m(k)$ exceeds the simulated 5% critical value.
- Weighted-CvM versions integrate the corresponding boundary-scaled path.

The adjusted-range idea

- Transform the retained scores using the training lag-0 covariance.
- For each transformed coordinate, compute the range of its centered cumulative path.
- Put these coordinatewise adjusted ranges on the diagonal of $\widetilde{R}_m$.

RSMS scaling

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

- Shao-type self-normalization uses a quadratic functional of the training cumulative path, which can be large in finite samples.
- The same transformed score path appears in the numerator and in the adjusted-range self-normalizer.
- The goal is stable scaling without kernel and bandwidth choices, not direct LRV estimation.

Theory, critical values, and stopping rules

Finite-horizon null approximation

For fixed $T < \infty$,

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

with analogous limits for SSMS and RSMS after replacing the scaling matrix.

- Fixed alternatives: visible retained-score drift crosses fixed critical values.
- Local alternatives: deterministic local drift is retained in the limit.
- The finite-horizon critical values are simulated from the null limiting distributions.
- Operational point: same score stream, different normalizations, comparable stopping rules.

Assumptions in words

- The retained population score process has a partial-sum approximation.
- The training-sample FPCA eigenfunctions are estimated accurately enough for monitoring.
- HAC requires a consistent training-sample LRV estimator.
- RSMS requires approximate diagonalization after the lag-0 transformation.

Practical check

The paper reports an off-diagonal diagnostic for the RSMS approximate-diagonalization condition.

Simulation design

Data-generating processes

- BB: Brownian-bridge style functional variation.
- fIID: independent functional observations.
- fMA(1): serially dependent functional observations.

Break settings

- Level shift.
- Smooth change.
- Additional localized-change setting.
- Mild contamination in the training sample.

Reported quantities

- empirical size,
- null MAD,
- raw rejection rates,
- SAP,
- ADD.

- 1000 Monte Carlo replications per setting; curves evaluated on a 301-point grid; K selected by 95% FVE.
- The main figures emphasize fMA(1), because serial dependence is where LRV estimation and self-normalization matter most.
- MAD is the mean absolute deviation of null rejection from nominal size; SAP is size-adjusted power; ADD is average detection delay after the break.

Simulation summary

Method	Null MAD	Level shift SAP	Smooth change SAP	Overall SAP
RSMS KS ($\gamma = 0$)	3.7	87.8	54.2	71.0
SSMS KS ($\gamma = 0$)	0.7	82.3	47.1	64.7
HAC KS ($\gamma = 0$)	4.7	91.0	60.3	75.7
RSMS CvM-Late	3.4	86.0	42.4	64.2
SSMS CvM-Late	0.6	81.0	36.9	59.0
HAC CvM-Late	4.0	89.0	47.7	68.3

- HAC has high SAP, but also the largest null-size distortion among the KS statistics.
- Within the self-normalized KS statistics, RSMS improves on SSMS.

Within self-normalized KS

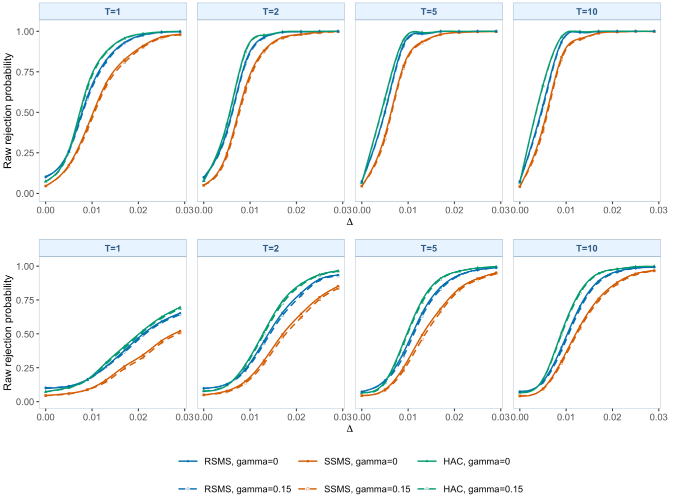
RSMS has higher overall SAP than SSMS: 71.0% vs. 64.7%.

Relative to HAC

HAC has the highest SAP, but also the largest null MAD among the KS statistics.

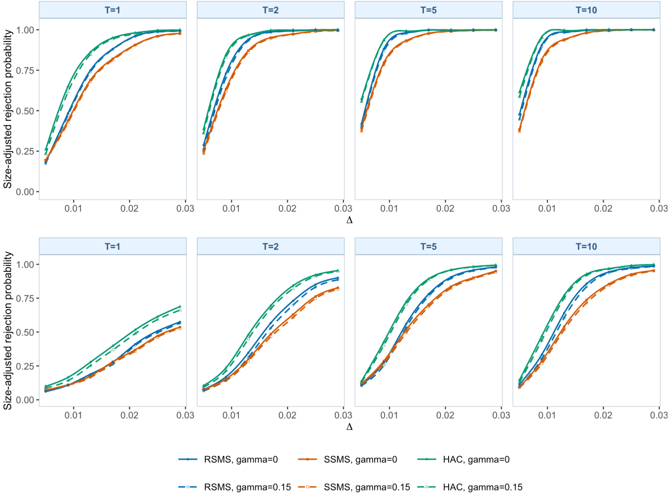
RSMS is the preferred self-normalized KS statistic in the simulation study.

Raw rejection rates under fMA(1) dependence



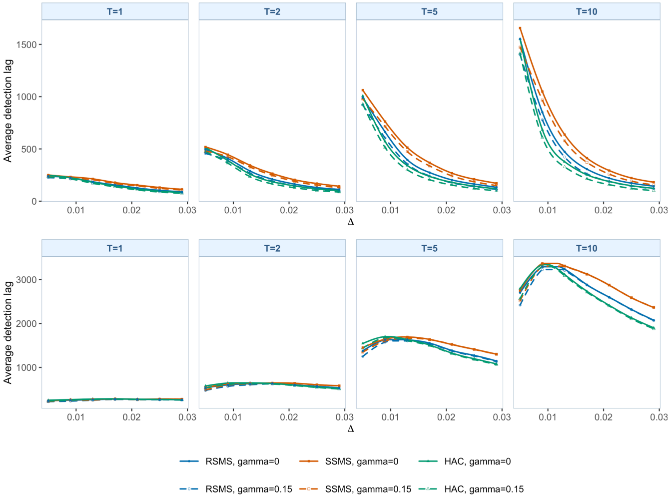
Level-shift and smooth-change settings; columns correspond to $T = 1, 2, 5, 10$. The $\Delta = 0$ point is empirical size, so raw rejection rates mix detection and finite-sample size distortion.

Size-adjusted power under fMA(1) dependence



After empirical size adjustment, HAC receives less benefit from its larger null rejection rate, while RSMS remains above SSMS; the RSMS gain is especially visible under smooth change.

Detection delay matters



Lower ADD means faster post-break intervention. RSMS usually stops earlier than SSMS; HAC can be fast, but that speed comes with larger empirical size distortion. Size, SAP, and ADD have to be read together.

Companion diagnostics and alternative detectors

Weighted-CvM

- KS focuses on the maximum boundary-scaled excursion.
- Weighted-CvM accumulates the standardized path over the monitoring horizon.
- The late-emphasis weight performs best, but results are more sensitive to weight and horizon.

Localized changes

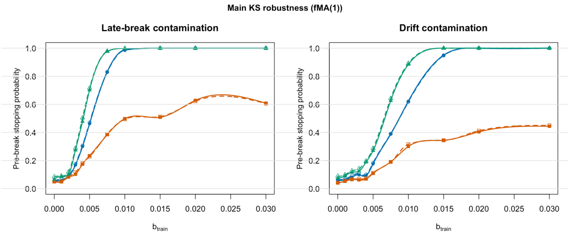
- The appendix also studies disruptions concentrated in part of the function domain.
- HAC can be aggressive; SSMS is conservative; RSMS is favorable within the self-normalized class.

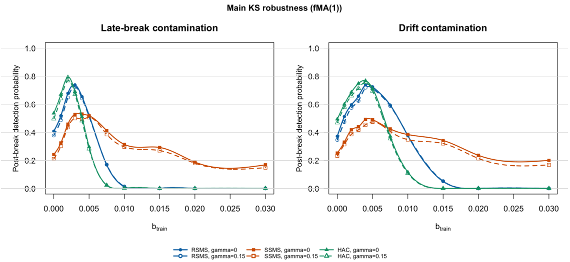
Alternative detector shapes

- Page-CUSUM,
- weighted-CUSUM,
- full-window CUSUM,
- MOSUM and multiscale MOSUM.

For the main recommendation, RSMS KS is cleaner and more transparent.

Mild training-sample contamination

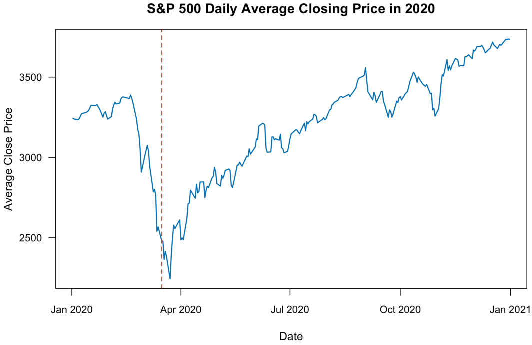


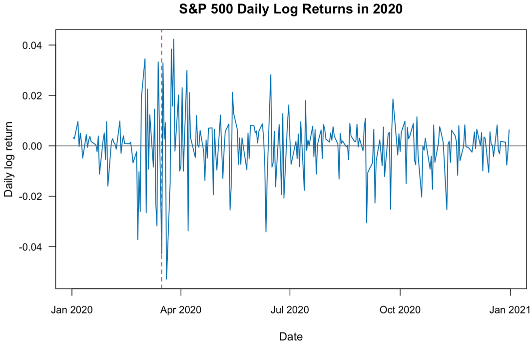


RSMS stays close to HAC in mild post-break detection, but with less pre-break stopping than HAC.

- Empirical monitoring rarely starts from a perfectly clean training sample.
- In the mild contamination range, RSMS keeps post-break detection close to HAC and above SSMS.
- HAC has a stronger tendency to stop before the monitoring break.
- SSMS has fewer pre-break stops, but weaker post-break detection.

Empirical application: S&P 500 in early 2020

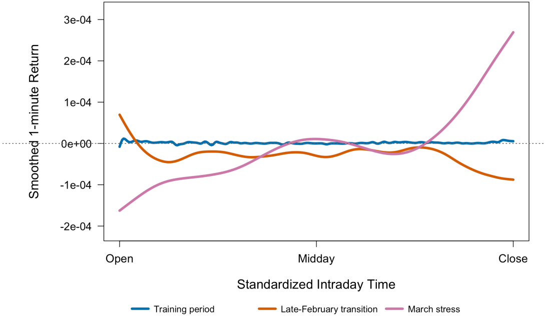




- Licensed 1-minute S&P 500 prices from TickData.
- Training sample: last 50 trading days of 2019.
- Monitoring begins on January 2, 2020.

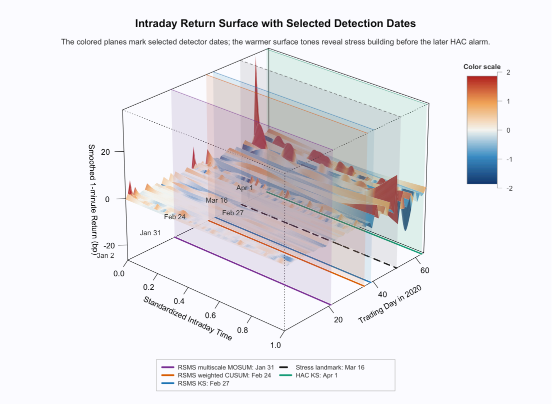
- Adjacent 1-minute log returns.
- Cubic B-spline smoothing; 301-point grid.
- FPCA estimated on 2019 training sample only.
- $T \in \{1, 2, 5\}$; $K = 4$ by 80% FVE.

Why monitor the intraday return curve?



- The opening part becomes sharply more negative.
- The middle of the session becomes more uneven.
- The late-day profile becomes more pronounced.
- The training-sample curve is close to flat.
- The late-February transition already shows clear intraday shape changes.
- The March stress window is not just larger in magnitude; it is different over the day.

Intraday return surface and signal dates



- Figure S5.3 shows the evolving smoothed 1-minute return surface.
- RSMS-based signals appear before the later HAC KS alarm.

Interactive version

Open rotating 3D HTML

`SPX_2020_IntradaySurface_`
`DetectionDates_interactive.html`

Empirical KS monitoring results

Horizon	Standardizer	γ	Stat./CV	First signal
$T = 1$	RSMS	0.00/0.15	1.98 / 1.90	Feb. 26
	SSMS	0.00/0.15	0.05 / 0.05	No signal
	HAC	0.00/0.15	0.65 / 0.63	No signal
$T = 2$	RSMS	0.00/0.15	3.34 / 3.36	Feb. 27
	SSMS	0.00/0.15	0.10 / 0.10	No signal
	HAC	0.00/0.15	1.22 / 1.25	Apr. 2 / Apr. 1
$T = 5$	RSMS	0.00/0.15	2.57 / 2.75	Mar. 13
	SSMS	0.00/0.15	0.08 / 0.09	No signal
	HAC	0.00/0.15	0.96 / 1.05	No / Apr. 2

- RSMS signals at all three horizons.
- SSMS remains below the critical value throughout.
- HAC signals later in this application and requires kernel/bandwidth choices.

- RSMS: Feb. 26 ($T = 1$), Feb. 27 ($T = 2$), Mar. 13 ($T = 5$).
- HAC signals on April 1–2 for the relevant $T = 2$ and $T = 5$ cases.
- SSMS does not cross in the main KS table.

How large are the decision windows?

Window	Days	Avg. $\lvert r\rvert$	Avg. RV	Rel. $\lvert r\rvert$	Rel. RV
Training sample	50	1.13 bp	0.12	1.0×	1.0×
Early 2020 pre-alarm	38	1.91 bp	0.36	1.7×	3.0×
RSMS($T = 2$) to HAC($T = 2$)	25	12.11 bp	11.47	10.7×	94.6×
RSMS($T = 5$) to HAC($T = 5$)	15	13.15 bp	13.32	11.6×	109.8×

- During the RSMS–HAC timing gaps, intraday movements are already very large.
- The earlier RSMS signals are not alarms during a quiet period.
- They leave time for model review before the later HAC signals.

- The RSMS–HAC timing gap is not economically small.
- Average absolute 1-minute returns are about 10–12 times the training-sample level.
- Daily realized variation is about 95–110 times as large.
- In this application, RSMS gives the more useful timing among self-normalized KS statistics and signals earlier than HAC.

Takeaways and model maintenance

What should the audience remember?

1. Online monitoring of financial functional time series can be framed as a fixed-training FPCA score problem.
2. Comparing HAC, SSMS, and RSMS on the same score process separates normalization from representation.
3. RSMS KS is tuning-free relative to HAC LRV estimation and less conservative than SSMS in the simulations.
4. In the S&P 500 application, RSMS gives materially earlier signals during the COVID return-curve transition.

After an alarm

- Check whether the functional representation still fits the incoming data.
- Re-estimate FPCA on a new stable training period if needed.
- Recompute forecasts, risk summaries, and downstream model diagnostics.
- Continue monitoring from the refreshed representation.

The alarm is a model-maintenance trigger, not a final diagnosis.

Limitations and next steps

Limitations

- The main theory is stated for a fixed retained dimension K.
- The monitoring target is mean constancy of the retained score process.
- RSMS requires approximate diagonalization after the lag-0 transformation.
- Critical values are simulated from the limiting distributions.

Next steps

- Treat an alarm as the beginning of model maintenance:
  - choose a new stable training sample,
  - decide whether to re-estimate the FPCA basis,
  - restart monitoring without masking subsequent breaks.
- Monitor economically direct functionals: volatility, covariance, tail risk, dependence, or functional regression stability.
- Localize which part of the domain accounts for the alarm: intraday time, maturity, strike, or location.
- Extend RSMS to panels of functional objects and cross-sectional dependence.

Thank you

Questions and comments are very welcome.

Online Monitoring of Financial Functional Time Series

Resources

Personal webpage

`http://jiajing-sun.com/`

Main paper

`http://jiajing-sun.com/files/when-to-refresh-functional-representation-main-20260615.pdf`

Supplement

`http://jiajing-sun.com/files/when-to-refresh-functional-representation-supplement-20260615.pdf`