

Research Paper

Clipped self-normalized Max-CUSUM for sparse mean breaks in high-dimensional heavy-tailed locally stationary time series

LEILA MARVIAN MASHHAD*
IMAM REZA INTERNATIONAL UNIVERSITY, MASHHAD, IRAN

Received: June 16, 2026/ Revised: August 05, 2026/ Accepted: September 25, 2026

Abstract: We develop a robust procedure for detecting and locating a single sparse mean break in high-dimensional, locally stationary time series with heavy-tailed innovations and intermittent contamination. The method combines coordinatewise clipping, local self-normalization, and a maximum cumulative sum scan. In contrast with an oracle calibration, the proposed implementation uses a locally centered block-multiplier bootstrap, so it does not require a known covariance structure. We establish negligible clipping bias under finite moments, a Gaussian approximation for the scan statistic, bootstrap validity under the null, a sparse detection condition, and a localization rate. The theory is accompanied by sufficient conditions for time-varying linear and nonlinear filters. A fully specified Monte Carlo study uses 250 replications per main design, reports Wilson confidence intervals and Monte Carlo standard errors, compares oracle and data-driven calibration, and examines sensitivity to clipping, window, ridge, scanning, dimension, and sparsity. The proposed method is competitive with raw max scanning in clean settings and is particularly advantageous for sparse breaks under heavy tails and contamination; dense ℓ_2 aggregation need not be dominated in clean, dense-signal regimes.

Keywords: Change-point detection; Heavy tails; High-dimensional time series; Local stationarity; Robust inference; Self-normalization; Multiplier bootstrap.

Mathematics Subject Classification (2010): 62M10, 62G35, 62H15.

1 Introduction

Abrupt mean changes are difficult to distinguish from smooth time variation when the observed process is high dimensional, serially dependent, and non-Gaussian. This

*Corresponding author: leila.marveian@gmail.com

combination arises naturally in multivariate monitoring, economic panels, sensor arrays, and experimental streams, but the present paper is methodological and does not make any empirical application claim. Classical change-point methods are surveyed by Csorgo and Horvath (1997), while sparse high-dimensional procedures include Cho and Fryzlewicz (2015) and Wang and Samworth (2018). Local stationarity provides a flexible framework for slowly evolving dynamics (Dahlhaus, 1997), and functional dependence measures offer a convenient way to treat nonlinear filters without imposing restrictive mixing assumptions (Wu, 2005).

Recent developments have made high-dimensional change-point analysis substantially more adaptive. For example, data-adaptive multiscale procedures and finite-sample inference for high-dimensional mean vectors clarify how the detection problem changes with signal sparsity and dimensionality (Liu et al., 2020; Yu and Chen, 2020). Complementary work studies sequential monitoring in high dimensions and nonparametric sequential procedures for multivariate time series (Gösmann et al., 2022; Leyli-Abadi et al., 2021).

Robustness is a separate design consideration. Rank- and U -statistic approaches, dependent-wild-bootstrap procedures, and heavy-tailed autoregressive moving average with generalized autoregressive conditional heteroskedasticity (ARMA–GARCH) models provide alternatives to direct clipping (Dehling et al., 2020; Wegner and Wendler, 2024; Bai et al., 2025). Recent work also considers multiple changes under serial dependence, multivariate nonparametric tests, and structured matrix or tensor time series (Diop and Kengne, 2022a; Kojadinovic and Verdier, 2021; Cho and Fryzlewicz, 2023; He et al., 2024; Zhang and Chan, 2025). These strands motivate a method that keeps the sparse max geometry while explicitly stabilizing local observations and local scales.

Broader modern work includes hybrid cumulative sum (CUSUM) algorithms, Bayesian nonparametric modeling for incomplete multivariate series, and recent methodological reviews (Lee et al., 2020; Corradin et al., 2022; Gupta et al., 2024). These approaches address different objectives from the local sparse-break problem considered here, but they reinforce the importance of transparent calibration and reproducible simulation design.

A coordinatewise maximum cumulative sum (Max-CUSUM) scan is well matched to alternatives that affect only a small subset of coordinates. Its weakness is sensitivity to an extreme observation: a single spike can alter both a local mean contrast and its estimated scale. We address this issue by clipping each coordinate after robust centering and scaling, then applying a self-normalized (SN) local max scan. The implementation is completed by a block-multiplier bootstrap that estimates the critical value from the observed series.

The contribution is deliberately targeted. We do not claim uniform superiority over all competitors or all signal geometries. Instead, the method targets sparse mean changes in settings where tail behavior, contamination, and local scale variation can degrade raw max scans. The numerical study makes this distinction explicit by including clean Gaussian and heavy-tailed designs as well as isolated and clustered contamination stress tests.

2 Model and statistic

2.1 Locally stationary sparse-break model

Let $\{X_t\}_{t=1}^n$ be a p -variate series satisfying

$$X_t = \mu(t/n) + \Delta_n \mathbf{1}\{t > k_0\} + G(t/n, \mathcal{F}_t), \quad 1 \leq t \leq n, \quad (1)$$

where k_0 is an unknown change date, $\mu : [0, 1] \rightarrow \mathbb{R}^p$ is a smooth baseline mean, $\Delta_n \in \mathbb{R}^p$ is a jump vector, and $G(u, \mathcal{F}_t)$ is a locally stationary innovation filter driven by $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$. We test $H_0 : \Delta_n = 0$ against sparse alternatives with support $S_n = \{j : \Delta_{n,j} \neq 0\}$ and $s_n = |S_n| \ll p$.

The smooth component in (1) captures gradual evolution in the local mean, volatility, and dependence structure. It is not intended to model abrupt changes; the latter are represented by $\Delta_n \mathbf{1}\{t > k_0\}$. The local window used below limits the contribution of smooth drift to a contrast near a candidate break. This distinction is also relevant for periodic or seasonally evolving series, where a smooth baseline should not be confused with a structural shift (Guo and Modarres, 2020).

2.2 Robust standardization and clipped scan

In practical data analysis, a clipping threshold must be scale free. Let $m_j = \text{median}_{1 \leq t \leq n} X_{t,j}$ and $a_j = 1.4826 \text{ median}_t |X_{t,j} - m_j|$, with a standard-deviation fallback if a_j is numerically zero. This robust standardization follows the general principle that local estimation should not be driven by a small number of extreme observations (Huber, 1964; Catoni, 2012; Diop and Kengne, 2022b). For a dimensionless threshold $\tau > 0$, define

$$Z_{t,j} = m_j + a_j \psi_\tau\{(X_{t,j} - m_j)/a_j\}, \quad \psi_\tau(x) = \text{sgn}(x) \min\{|x|, \tau\}. \quad (2)$$

The theoretical analysis below applies directly after deterministic coordinate scaling; robust scaling in (2) is the implementation counterpart.

For a window h_n and candidate $k \in \mathcal{K}_n = \{h_n, h_n + d, \dots, n - h_n\}$, write $\bar{Z}_{j,L}(k)$ and $\bar{Z}_{j,R}(k)$ for the two adjacent window means, and let $\hat{v}_{j,L}(k)$ and $\hat{v}_{j,R}(k)$ be their sample variances. Define

$$D_j(k) = \frac{\sqrt{h_n}\{\bar{Z}_{j,R}(k) - \bar{Z}_{j,L}(k)\}}{\{\hat{v}_{j,L}(k) + \hat{v}_{j,R}(k) + \lambda_n\}^{1/2}}, \quad T_n = \max_{k \in \mathcal{K}_n} \max_{1 \leq j \leq p} |D_j(k)|. \quad (3)$$

The estimated break date is $\hat{k} = \arg \max_{k \in \mathcal{K}_n} \max_j |D_j(k)|$. The ridge λ_n prevents a division by a nearly zero local variance; it is not a regularizer for the mean contrast.

3 Data-driven calibration and implementation

3.1 Locally centered block-multiplier bootstrap

The Gaussian proxy in the asymptotic theory is useful for analysis but its covariance is unknown. We therefore use a practical block-multiplier bootstrap. Let $\tilde{Z}_t = Z_t - \hat{\mu}_t$,

where $\hat{\mu}_t$ is the coordinatewise running median with span $2h_n + 1$. Partition $\{1, \dots, n\}$ into consecutive blocks of length ℓ_n , draw independent $\xi_b \sim N(0, 1)$ for the blocks, and set $Z_t^* = \xi_{b(t)} \tilde{Z}_t$. Applying (3) to $\{Z_t^*\}$ while retaining the observed local variance estimates gives T_n^* . Conditional on the data, let $q_{1-\alpha}^*$ be an empirical upper quantile of T_n^* . Related resampling ideas for dependent time-series statistics and modern robust functional procedures support this choice of a dependence-aware calibration device (Ma and Zhang, 2024; Wegner and Wendler, 2024).

Our finite-sample implementation uses 99 bootstrap draws in the main experiment and the upper empirical 0.99 quantile as a conservative estimate of the nominal 0.95 critical quantile. This pre-specified upper-quantile rule is used uniformly across methods and scenarios to stabilize size with overlapping windows. The calibration comparison separately reports an oracle Gaussian benchmark and the block-multiplier rule. In larger applications, 499 or more bootstrap draws are recommended.

Implementation steps. Robustly center and scale each coordinate, form $Z_{t,j}$ using (2), and compute the cumulative-sum implementation of T_n and $\hat{k}$. Then obtain locally centered residuals with a running median. For $b = 1, \dots, B$, generate block multipliers, calculate $T_{n,b}^*$, and reject when T_n exceeds the pre-specified upper bootstrap critical value $q_{1-\alpha}^{*,U}$.

The scan uses cumulative sums of $Z_{t,j}$ and $Z_{t,j}^2$. Hence its time complexity is $O(np + np|K_n|/h_n)$ when window means are updated recursively, and $O(np)$ for a fixed grid step. With B bootstrap draws, the calibration cost is $O(Bnp)$ up to the number of scan locations. Memory usage is $O(np)$ if bootstrap draws are processed sequentially.

3.2 Practical tuning

A useful default is $\tau \in [2, 3]$ after robust scaling, h_n between $\max\{40, 4\log(np)\}$ and $n/8$, a scan step of 5 for long series, and $\lambda_n = 10^{-6}$ times a typical squared robust scale. The window must be long enough to average dependence and short enough that the smooth baseline is approximately constant. The requirement $\log(np) = o(h_n)$ motivates the first part of this rule. A smaller h_n is appropriate for closely spaced changes, whereas a larger h_n improves weak-signal detection when the baseline varies slowly.

4 Assumptions and theoretical properties

We state conditions for a single-break problem. They are high level but are accompanied by checkable examples.

A1: moments and functional dependence. For some $q > 4$, $\sup_{u,j} \|G_j(u, \mathcal{F}_0)\|_q < \infty$. If $\mathcal{F}_0^{(m)}$ replaces ε_{-m} by an independent copy, the physical dependence coefficients satisfy $\delta_q(m) = \sup_u \left\| G(u, \mathcal{F}_0) - G(u, \mathcal{F}_0^{(m)}) \right\|_q \leq C\rho^m$ for $0 < \rho < 1$.

A2: local regularity and nondegeneracy. Both $\mu(\cdot)$ and $G(\cdot, \mathcal{F}_0)$ are Lipschitz in rescaled time in the required norms, and each local coordinate variance is bounded

above and away from zero.

A3: growth and tuning. $h_n \rightarrow \infty$, $h_n/n \rightarrow 0$, $\log(np) = o(h_n)$, $\sqrt{h_n}\tau_n^{1-q} = o(1)$, and the bounded-array Gaussian approximation remainder vanishes. A sufficient explicit version is $\tau_n^2 \log^7(np) = o(h_n)$ together with an m_n -dependent approximation satisfying $\log(np)^{1/2} \rho^{m_n} = o(1)$.

A4: break separation. $k_0 \in [h_n, n - h_n]$ and the minimum standardized jump $\delta_n = \min_{j \in S_n} |\Delta_{n,j}|/\sigma_j$ is positive.

Functional dependence is used instead of a mixing coefficient because it directly quantifies the error of replacing a nonlinear filter by an m -dependent approximation. This makes it particularly convenient for locally stationary filters and gives a transparent route to high-dimensional Gaussian approximation (Wu, 2005; Zhang and Wu, 2017). Assumptions A1–A2 hold, for example, for a stable time-varying vector autoregression $G_t = A(t/n)G_{t-1} + e_t$ with uniformly bounded Lipschitz $A(\cdot)$ and spectral radius below one. They also hold for causal nonlinear recursions that are contractive in the innovations.

Lemma 4.1 (Clipping bias). *Under A1 with $q > 2$, uniformly in u and j ,*

$$\begin{aligned} |\mathbb{E}\psi_{\tau_n}(G_j(u, \mathcal{F}_0)) - \mathbb{E}G_j(u, \mathcal{F}_0)| &\leq C\tau_n^{1-q}, \\ |\mathbb{E}\psi_{\tau_n}^2(G_j(u, \mathcal{F}_0)) - \mathbb{E}G_j^2(u, \mathcal{F}_0)| &\leq C\tau_n^{2-q}. \end{aligned}$$

The first bound uses $|Y|\mathbf{1}\{|Y| > \tau_n\} \leq |Y|^q \tau_n^{1-q}$ and therefore requires $q > 1$, which is implied by A1.

Theorem 4.2 (Gaussian and bootstrap approximation). *Under A1–A3 and H_0 , there exists a centered Gaussian array $\{\Gamma_{j,k}\}$ with the covariance of the linearized clipped contrasts such that, for $G_n = \max_{j,k} |\Gamma_{j,k}|$,*

$$\sup_x |\mathbb{P}(T_n \leq x) - \mathbb{P}(G_n \leq x)| \rightarrow 0.$$

If $\ell_n \rightarrow \infty$, $\ell_n/h_n \rightarrow 0$, and the locally centered block-multiplier construction is used, then

$$\sup_x |\mathbb{P}^*(T_n^* \leq x) - \mathbb{P}(G_n \leq x)| \rightarrow_{\mathbb{P}} 0.$$

The proof combines clipping bias control, an m -dependent approximation, concentration of local variances, and a Gaussian approximation for maxima of weakly dependent bounded arrays; see Chernozhukov et al. (2013); Zhang and Wu (2017); Gösmann et al. (2022); Ma and Zhang (2024). The covariance of the proxy is explicitly the covariance of the linearized local contrast vector $\sqrt{h_n}(\tilde{Z}_{j,R}(k) - \tilde{Z}_{j,L}(k))/\{v_{j,L}(k) + v_{j,R}(k)\}^{1/2}$, including overlap across k and cross-sectional dependence across j .

Theorem 4.3 (Detection and localization). *Under A1–A4, if*

$$\sqrt{h_n}\delta_n \geq C_1 \sqrt{\log(np)} + C_2 \sqrt{h_n}\tau_n^{1-q},$$

then the bootstrap-calibrated test has power tending to one. If additionally $\delta_n \sqrt{h_n}/\sqrt{\log(np)} \rightarrow \infty$, then

$$|\hat{k} - k_0| = O_{\mathbb{P}} \left(\frac{\sqrt{h_n \log(np)}}{\delta_n} \right).$$

The detection scale $\{\log(np)/h_n\}^{1/2}$ is the canonical order for coordinatewise sparse scans. The result does not claim a new minimax lower bound for the full locally stationary and dependent model class. Rather, it shows that the robustification preserves the usual sparse-scan order up to constants and the clipping-bias term. The localization condition has a simple interpretation: the standardized jump must dominate the simultaneous stochastic envelope $\sqrt{\log(np)/h_n}$.

5 Simulation design

5.1 Data-generating mechanisms

All simulations are fully specified here. We generate $n = 800$, $p = 40$, and $k_0 = 400$ unless a sensitivity design changes these values. The innovation filter is

$$G_t = AG_{t-1} + e_t, \quad A = 0.35I_p + 0.06 \text{diag}_1 - 0.04 \text{diag}_{-1},$$

where $\text{Cov}(e_{t,i}, e_{t,j}) = 0.45^{|i-j|}$. Innovations are Gaussian or standardized t_5 . The baseline mean is

$$\mu_j(u) = 0.30 \sin(2\pi u + 0.15j) + 0.10 \frac{j-1}{p-1} \cos(4\pi u).$$

The five affected coordinates have unequal jumps $\Delta_{n,j} = 1.1(1.20, 1.10, 1.00, 0.90, 0.80)$ for $j = 1, \dots, 5$. Isolated contamination occurs at a Bernoulli 0.04 fraction of time points; each selected time receives a signed additive spike of magnitude 12 on three randomly selected coordinates. Clustered contamination uses short consecutive episodes with the same coordinate support. The random seeds are deterministic and the numerical outputs are included with this project.

Table 1: Main simulation settings and implementation details.

Feature	Specification
Main dimensions	$n = 800$, $p = 40$, $k_0 = 400$
Dependence	Stable tridiagonal VAR(1), diagonal 0.35, off-diagonals 0.06 and -0.04
Cross-sectional correlation	Toeplitz, $0.45^{ i-j }$
Smooth baseline	$0.30 \sin(2\pi u + 0.15j) + 0.10(j-1)(p-1)^{-1} \cos(4\pi u)$
Innovations	Gaussian or unit-variance Student t_5
Sparse jump	Five unequal affected coordinates; magnitude profile $(1.20, 1.10, 1.00, 0.90, 0.80) \times 1.1$
Contamination	Isolated or clustered signed spikes of magnitude 12; 3 coordinates per affected time
Default tuning	$\tau = 2.2$, $h_n = 80$, scan step $d = 5$, $\lambda = 10^{-6}$
Calibration	Locally centered block multiplier; block length 10; 99 draws; upper 0.99 empirical quantile
Monte Carlo	250 replications per main design; 80 per sensitivity setting
Software	Python 3.13.13, NumPy 2.3.5, SciPy 1.17.1; Linux x86_64; two CPU cores

We compare clipped SN-Max with Raw SN-Max, a self-normalized ℓ_2 scan (SN-L2), and a spatial-sign self-normalized scan (SS-SN). SS-SN is included as a robust

benchmark based on spatial-sign transformation; it is not tailored to coordinate-sparse alternatives. We additionally report a rank-based SN-Max extension in a dedicated robust-comparison experiment. It first maps each coordinate to normal scores of its within-coordinate ranks and then uses the same local max scan and bootstrap rule. The comparison is intentionally complementary to modern nonparametric multivariate and high-dimensional monitoring approaches rather than a claim that all robustification strategies have the same operating characteristics (Dehling et al., 2020; Diop and Kengne, 2022a; Kojadinovic and Verdier, 2021; Leyli-Abadi et al., 2021). Each procedure uses the same scanning grid and the same block-bootstrap calibration principle.

6 Numerical results

6.1 Calibration and uncertainty

Table 2 reports rejection probabilities with 95% Wilson intervals, addressing Monte Carlo uncertainty directly. The conservative practical bootstrap rule keeps the proposed procedure near or below the nominal 5% level in the null designs. The reported intervals are important: with 250 replications, the uncertainty around a 5% rejection probability is non-negligible and should not be hidden by three-decimal reporting. In this Table, all methods use the same bootstrap calibration rule.

Table 2: Empirical rejection probabilities, with 95% Wilson intervals in brackets.

Scenario	Clipped SN-Max	Raw SN-Max	SN-L2	SS-SN
Gaussian Null	0.024 [0.011, 0.051]	0.024 [0.011, 0.051]	0.000 [0.000, 0.015]	0.000 [0.000, 0.015]
Gaussian Break	0.492 [0.431, 0.554]	0.484 [0.423, 0.546]	0.032 [0.016, 0.062]	0.000 [0.000, 0.015]
Heavy-Tail Null	0.052 [0.031, 0.087]	0.036 [0.019, 0.067]	0.000 [0.000, 0.015]	0.000 [0.000, 0.015]
Heavy-Tail Break	0.612 [0.550, 0.670]	0.572 [0.510, 0.632]	0.032 [0.016, 0.062]	0.012 [0.004, 0.035]
Contaminated Null	0.024 [0.011, 0.051]	0.024 [0.011, 0.051]	0.000 [0.000, 0.015]	0.000 [0.000, 0.015]
Contaminated Break	0.548 [0.486, 0.609]	0.384 [0.326, 0.446]	0.040 [0.022, 0.072]	0.004 [0.001, 0.022]

In the Gaussian break design, Clipped SN-Max and Raw SN-Max have similar power, whereas the dense SN-L2 and spatial-sign scan are conservative under the common finite-sample calibration. Under heavy-tailed and contaminated breaks, clipping improves power over Raw SN-Max while retaining null rejection probabilities close to the target range. Figure 1 displays the estimate and interval for every method–scenario combination, making the finite-sample uncertainty visible rather than relying only on a grouped bar chart. In this figure, points are empirical rejection probabilities and horizontal segments are 95% Wilson intervals.

6.2 Oracle versus data-driven calibration

The oracle calculation uses 1,000 clean Gaussian null replicates from the known data-generating mechanism and is therefore unavailable in genuine applications. The data-

driven bootstrap uses only the observed series. Table 3 shows that the practical calibration has similar null behavior and a modest power cost in the contaminated break design, which is the expected price of covariance-free finite-sample protection.

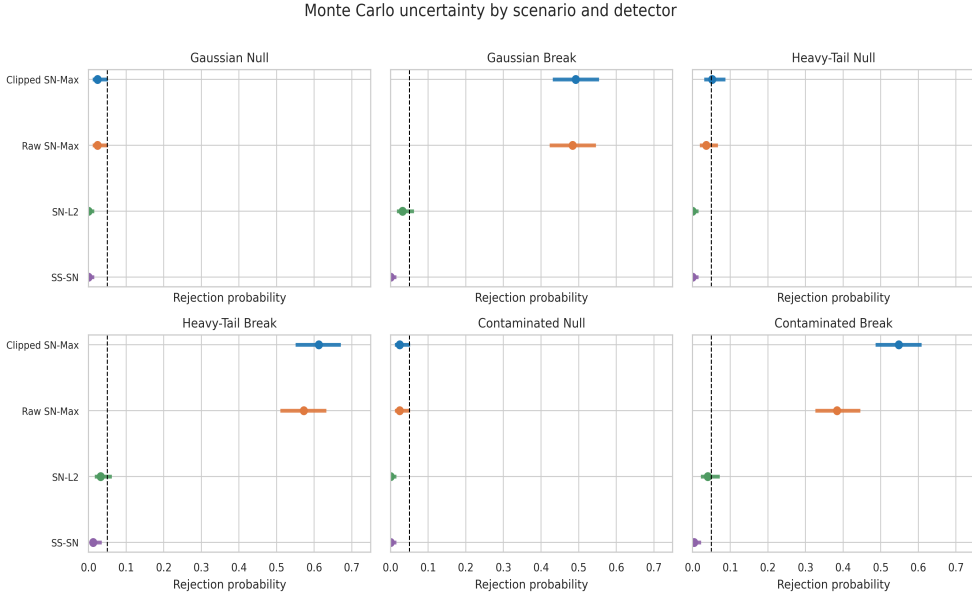


Figure 1: Monte Carlo uncertainty by scenario and detector.

Table 3: Clipped SN-Max: rejection probabilities under oracle Gaussian and data-driven bootstrap calibration.

Scenario	Oracle Gaussian	Block-multiplier bootstrap
Gaussian Null	0.008	0.008
Heavy-Tail Null	0.042	0.033
Contaminated Null	0.058	0.058
Contaminated Break	0.608	0.508

6.3 Additional rank-based robust comparison

To distinguish clipping from an alternative robustification mechanism, we add Rank-SN-Max. This procedure replaces each coordinate by its normal-score rank transform before applying the same SN max scan. The comparison uses 100 replications in the heavy-tailed and contaminated designs, with the same block-multiplier calibration and all other tuning parameters fixed. Table 4 shows that both robust methods improve on Raw SN-Max under heavy tails and contamination. Clipped SN-Max has the largest rejection probability in the two break designs, while the rank method remains a competitive robust alternative.

Table 4: Additional robust comparison. Entries are empirical rejection probabilities from 100 replications.

Scenario	Clipped SN-Max	Raw SN-Max	Rank-SN-Max
Heavy-tail Null	0.050	0.030	0.060
Heavy-tail Break	0.570	0.500	0.540
Contaminated Null	0.050	0.040	0.060
Contaminated Break	0.600	0.520	0.590

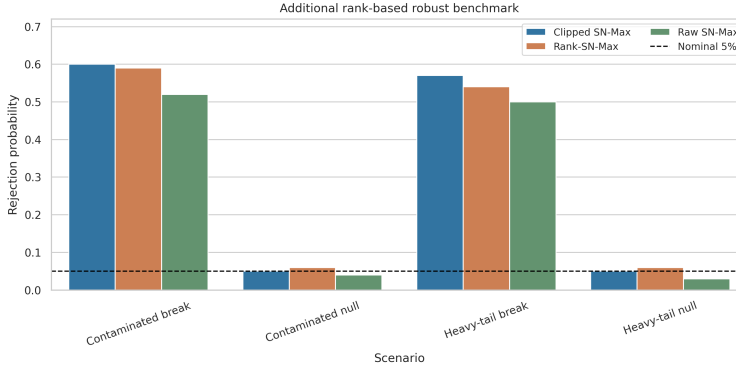


Figure 2: Additional rank-based robust benchmark. The dashed line marks the nominal 5% level.

6.4 Break localization

Table 5 reports mean absolute error with Monte Carlo standard errors in parentheses. In clean Gaussian settings, clipped and raw max scans have essentially equal localization accuracy. Under heavy tails, SN-L2 can be competitive, underscoring that no uniform dominance is claimed. Under contamination, clipping produces the smallest mean absolute error among the four procedures.

Table 5: Mean absolute break-location error; Monte Carlo standard errors in parentheses.

Scenario	Clipped SN-Max	Raw SN-Max	SN-L2	SS-SN
Gaussian Break	28.9 (4.3)	28.8 (4.3)	31.4 (4.1)	40.6 (4.6)
Heavy-Tail Break	25.7 (4.1)	28.6 (4.4)	22.0 (3.5)	26.5 (3.8)
Contaminated Break	21.7 (3.8)	32.7 (4.5)	40.3 (4.6)	30.5 (4.0)

6.5 Robustness frontier, sensitivity, and broader stress tests

Figure 4 summarizes the central robustness trade-off directly. The horizontal coordinate is contaminated-null rejection probability, the vertical coordinate is contaminated-break power, and bubble area is inversely proportional to localization error. The proposed method lies on the favorable part of this finite-sample frontier because it combines a low false-alarm rate with the largest contaminated-break power and the smallest localization error. Better operating points lie toward the upper left and larger bubbles indicate smaller mean absolute localization error.

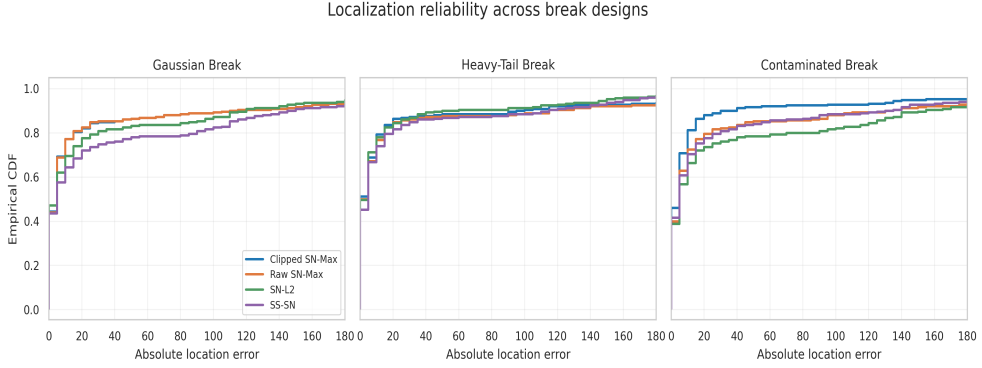


Figure 3: Empirical CDFs of absolute localization error. A curve that rises more quickly corresponds to more reliable localization.

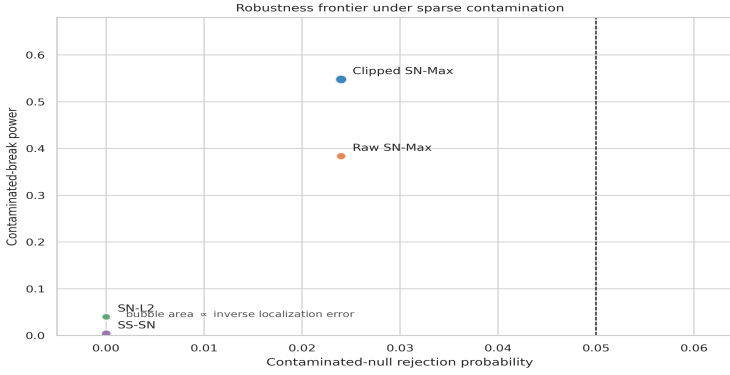


Figure 4: Robustness frontier in the contaminated design.

Figure 5 varies the clipping threshold, window length, scan resolution, ridge, sparsity, and problem scale under contaminated t_5 innovations. Performance is most sensitive to the window length, as expected from the bias–variance tradeoff. In the tested range, $\tau \in [1.8, 2.8]$ yields similar size and power, the scan step has limited impact on detection, and the small ridge has little effect. Power increases with sparsity and with a larger effective window, while the smallest $n = 400, p = 20$ configuration is naturally the most difficult. Each point uses 80 replications.

The primary theoretical target is a single break. We nevertheless conduct an explicit two-break recovery experiment under clustered contamination. Two jumps occur either at $(240, 560)$ or at the more challenging nearby locations $(340, 460)$. Each design uses 100 replications, and a non-maximum-suppression screening rule selects the two strongest local peaks at least 50 observations apart. A recovery is recorded when both selected locations lie within 40 observations of their corresponding true breaks. The experiment is descriptive rather than a replacement for a formal multiple-break theory, but it demonstrates that the clipped local score remains usable inside a segmentation wrapper (Cho and Fryzlewicz, 2023; Tickle et al., 2021).

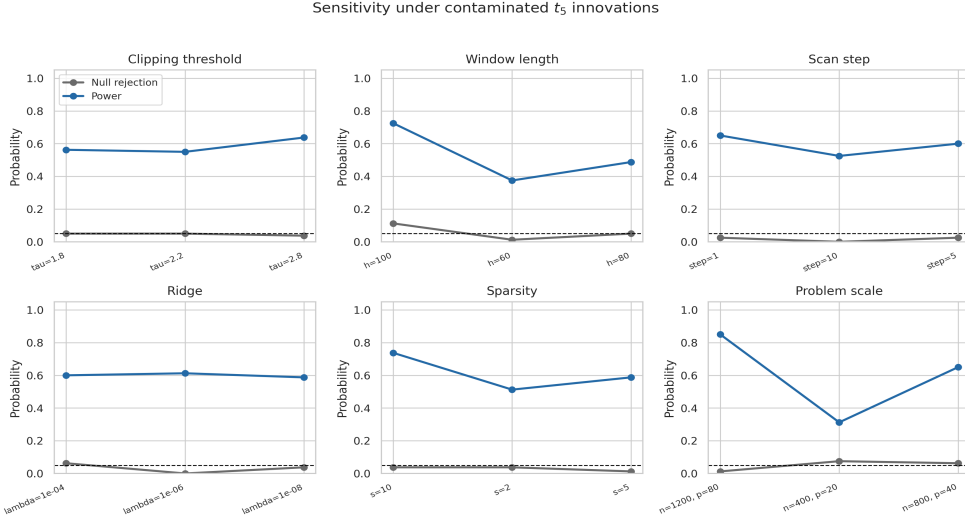


Figure 5: Sensitivity of Clipped SN-Max under contaminated t_5 innovations.

Table 6: Two-break recovery rate under clustered contamination.

Configuration	Clipped SN-Max	Raw SN-Max	Rank-SN-Max
Separated breaks (240, 560)	0.690	0.640	0.680
Nearby breaks (340, 460)	0.700	0.670	0.680

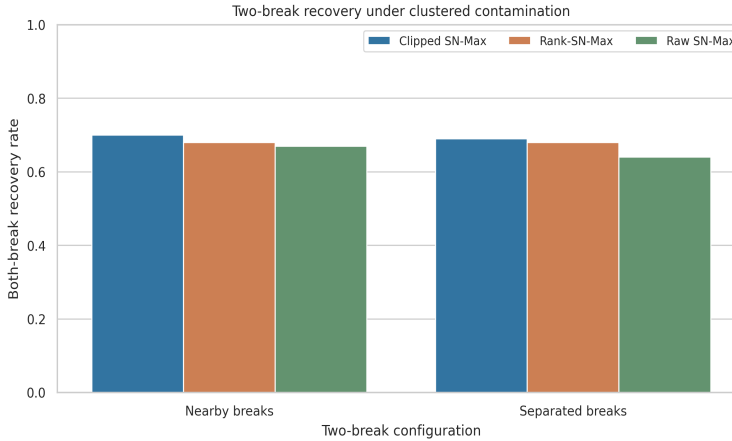


Figure 6: Formal two-break recovery experiment under clustered contamination.

6.6 Computational benchmark

We timed the cumulative-sum scan implementation after 20 warm-up scans and over 100 timed repetitions at each dimension. The benchmark measures the scan stage only,

excluding bootstrap draws because users may choose a different number of bootstrap replications. The clipped method closely tracks Raw SN-Max, whereas rank transformation adds sorting cost. These measurements were obtained on the computing environment reported in Table 1.

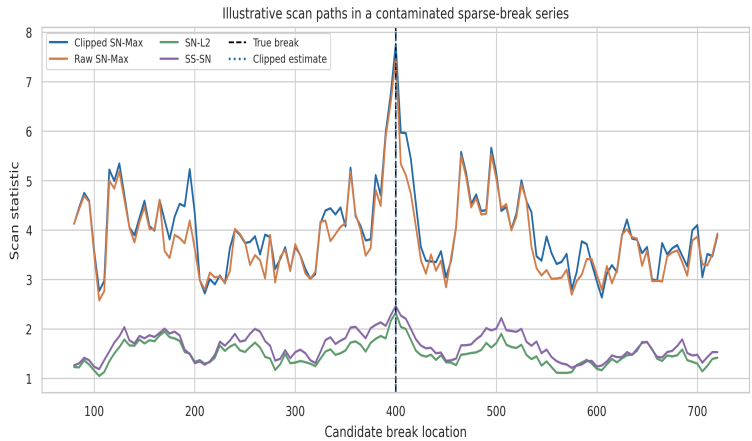


Figure 7: Illustrative scan paths on one contaminated sparse-break realization.

Table 7: Mean scan time in milliseconds after 20 warm-up runs

p	Clipped SN-Max	Raw SN-Max	Rank-SN-Max
20	1.01	0.98	7.14
40	1.92	1.85	12.66
80	3.66	3.45	26.23
120	5.95	4.81	38.91

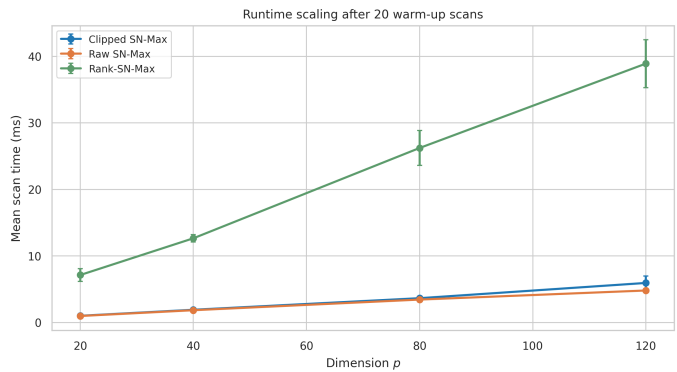


Figure 8: Runtime scaling after 20 warm-up scans. Error bars show one standard deviation across 100 timed scans.

7 Discussion and conclusions

Clipping and SN provide complementary safeguards for sparse high-dimensional scans. Clipping caps pointwise leverage, while SN adapts to local scale. The method is most useful when a sparse jump must be separated from heavy tails or intermittent contamination. In clean Gaussian settings, raw max scanning can perform similarly and dense aggregation can be preferable for dense alternatives; the numerical evidence is interpreted accordingly. Several extensions are natural. A multiple-break procedure can use the clipped local score inside a segmentation scheme, although this requires separate theory. Other directions include data-adaptive block-length selection, covariance and precision-matrix breaks, and structured changes in matrix-valued, multichannel, and tensor time series (He et al., 2024; Zhang and Chan, 2025; Cao et al., 2024). A real-data study should be added only when the application supports a scientifically interpretable sparse structural change and a defensible calibration design.

Acknowledgments

The author thank the reviewers for their constructive comments and valuable suggestions. During manuscript preparation, ChatGPT (OpenAI) was used to assist with grammar and sentence structure.

References

- Bai, Q., Ling, S. and Zhu, K. (2025). Testing and estimation of change point in ARMA model With heavy-tailed G-GARCH noises. *Journal of Time Series Analysis*. doi: 10.1111/jtsa.70039.
- Cao, Z., Seeuws, N., De Vos, M. and Bertrand, A. (2024). Change point detection in multi-channel time series via a time-invariant representation. *IEEE Transactions on Knowledge and Data Engineering*, **36**(12):7743–7756.
- Catoni, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, **48**(4):1148–1185.
- Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, **41**(6):2786–2819.
- Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B*, **77**(2):475–507.
- Cho, H. and Fryzlewicz, P. (2023). Multiple change point detection under serial dependence: Wild contrast maximisation and gappy Schwarz algorithm. *Journal of Time Series Analysis*, **45**(3):479–494.

- Corradin, R., Danese, L. and Ongaro, A. (2022). Bayesian nonparametric change point detection for multivariate time series with missing observations. *International Journal of Approximate Reasoning*, **143**:26–43.
- Csorgo, M. and Horvath, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, **25**(1):1–37.
- Dehling, H., Fried, R. and Wendler, M. (2020). A robust method for shift detection in time series. *Biometrika*, **107**(3):647–660.
- Diop, M.L. and Kengne, W. (2022a). A general procedure for change-point detection in multivariate time series. *TEST*, **32**(1):1–33.
- Diop, M.L. and Kengne, W. (2022b). Epidemic change-point detection in general causal time series. *Statistics & Probability Letters*, **184**:109416.
- Gösmann, J., Stoehr, C., Heiny, J. and Dette, H. (2022). Sequential change point detection in high dimensional time series. *Electronic Journal of Statistics*, **16**(1):3608–3671.
- Guo, L. and Modarres, R. (2020). Nonparametric change point detection for periodic time series. *Canadian Journal of Statistics*, **48**(3):518–534.
- Gupta, M., Wadhvani, R. and Rasool, A. (2024). Comprehensive analysis of change-point dynamics detection in time series data: A review. *Expert Systems with Applications*, **248**:123342.
- He, Y., Kong, X., Trapani, L. and Yu, L. (2024). Online change-point detection for matrix-valued time series with latent two-way factor structure. *The Annals of Statistics*, **52**(4):1646–1670.
- Huber, P.J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, **35**(1):73–101.
- Kojadinovic, I. and Verdier, G. (2021). Nonparametric sequential change-point detection for multivariate time series based on empirical distribution functions. *Electronic Journal of Statistics*, **15**(1):773–829..
- Lee, S., Lee, S. and Moon, M. (2020). Hybrid change point detection for time series via support vector regression and CUSUM method. *Applied Soft Computing*, **89**:106101.
- Leyli-Abadi, M., Samé, A., Oukhellou, L., Cheifetz, N., Mandel, P., Féliers, C. and Heim, V. (2021). Online common change-point detection in a set of nonstationary categorical time series. *Neurocomputing*, **439**:176–196.
- Liu, B., Zhou, C., Zhang, X. and Liu, Y. (2020). A unified data-adaptive framework for high dimensional change point detection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **82**(4):933–963.

- Ma, R. and Zhang, S. (2024). Multiplier subsample bootstrap for statistics of time series. *Journal of Statistical Planning and Inference*, **233**:106183.
- Tickle, S.O., Eckley, I.A. and Fearnhead, P. (2021). A computationally efficient, high-dimensional multiple changepoint procedure with application to global terrorism incidence. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **184**(4):1303–1325.
- Wang, T. and Samworth, R.J. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B*, **80**(1):57–83.
- Wegner, L. and Wendler, M. (2024). Robust change-point detection for functional time series based on U-statistics and dependent wild bootstrap. *Statistical Papers*, **65**(7):4767–4810.
- Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, **102**(40):14150–14154.
- Yu, M. and Chen, X. (2020). Finite sample change point inference and identification for high-dimensional mean vectors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **83**(2):247–270.
- Zhang, D. and Wu, W.B. (2017). Gaussian approximation for high dimensional time series. *The Annals of Statistics*, **45**(5):1895–1919.
- Zhang, X. and Chan, K.S. (2025). Spectral change point estimation for high-dimensional time series by sparse tensor decomposition. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **88**(2):677–696.

Appendix: Proof details and sufficient conditions

7.1 Proof of Lemma 4.1

For $Y = G_j(u, \mathcal{F}_0)$, $|\psi_{\tau_n}(Y) - Y| \leq |Y| \mathbf{1}\{|Y| > \tau_n\}$. Since $|Y| \mathbf{1}\{|Y| > \tau_n\} \leq |Y|^q \tau_n^{1-q}$, the first claim follows. Likewise, $|\psi_{\tau_n}^2(Y) - Y^2| \leq Y^2 \mathbf{1}\{|Y| > \tau_n\} \leq |Y|^q \tau_n^{2-q}$.

Outline for Theorem 4.2

First approximate G by an m_n -dependent filter; A1 gives an exponentially small error. Clipping is one-Lipschitz, so the same approximation applies to the clipped process. Next, the clipped summands are bounded and their high-dimensional maximum admits a Gaussian approximation under A3. The local variance estimators are uniformly consistent because their targets are bounded away from zero. Finally, locally centered block multipliers reproduce the covariance of the m_n -dependent linearization when $\ell_n \rightarrow \infty$ and $\ell_n/h_n \rightarrow 0$. Standard anti-concentration arguments transfer the result from the linearized process to the feasible SN maximum.