Robustness of Change-Point Detection Methods under Violations of Gaussianity and Independence

Yianna Ioannou

Faculty of Mathematics and Statistics
University of Cyprus, Nicosia

Research project submitted for the requirements of MAS499

Supervisor: Dr. Andreas Anastasiou

December 2025

Abstract

Change-point detection is a well-studied problem in statistics and signal processing, going back several decades. It aims to find the time points in a data sequence where parametric or non-parametric changes occur, such as changes in the mean, changes in the variance, and more. Change-point detection has been applied in many areas, including finance, biology, and environmental science. Many classical methods are developed under the assumptions that the noise is Gaussian and independent. In practice, these assumptions are often not realistic. We evaluate eight modern change-point detection methods to assess their robustness when classical modelling assumptions are violated. In particular, we study their performance under deviations from Gaussianity (while retaining i.i.d. structure) and under deviations from independence (while keeping Gaussian noise). The methods to be examined can detect changes in various signal structures; we focus on the examples of piecewise-constant signals. We compare the methods across several metrics for detection accuracy and computational cost. Our results provide a clear picture of how well these methods perform when the framework of Gaussian i.i.d. noise does not hold.

Contents

1 Introduction

1.1 Problem setting

Offline change-point detection has been of interest to statisticians for many decades. The goal of parametric change-point detection is to identify when and where abrupt changes occur in the statistical properties of a data sequence. Typical applications arise in biology (e.g., copy number variation (A.B. Olshen et al., 2004) or ion channels (T. Hotz et al., 2013)), finance (C.-J Kim et al., 2005), and environmental science (e.g., climate data analysis (J. Reeves et al., 2007) or environmental monitoring systems (M. Londschien et al., 2019)), among many others. Change-point detection plays a central role in the analysis of temporal, spatial, and genomic data, and continues to attract significant attention due to its importance and relevance across scientific disciplines.

A common observation model is

$$X_t = f_t + \varepsilon_t, \qquad t = 1, \dots, T, \tag{1}$$

where X_t are the observed data, f_t is an unknown deterministic signal and ε_t represents random noise. The random noise ε_t is assumed to follow an independent and identically distributed Gaussian distribution with mean zero and variance σ^2. Depending on the application, different classes of signals f_t are of interest. Two important examples are:

Piecewise-constant signals: $f_t = \mu_j \;$ for $t = r_{j-1}+1, \dots, r_j \;$ and $\; f_{r_j} \neq f_{r_{j+1}}, \;\; j = 1, \dots, N+1$, with $r_0 = 0$ and $r_{N+1} = T$. Here, the signal consists of constant mean segments separated by change-points at r_j.

Continuous, piecewise-linear signals: $f_t = \mu_{j,1} + \mu_{j,2}\, t \;$ for $t = r_{j-1}+1$, with the additional constraint of $\mu_{k,1} + \mu_{k,2r_k} = \mu_{k+1,1} + \mu_{k+1,2r_k}$, for $k = 1, 2, \dots N$. The change-points r_k satisfy $f_{r_k-1} + f_{r_k+1} \neq 2f_{r_k}$. In this case, the signal is a sequence of linear trends joined together at unknown knots r_j.

A fundamental quantity used by many multiple change-point detection methods is the CUSUM (Cumulative SUM) statistic. For any interval $[s, e]$ with $1 \leq s < b < e \leq T$, the CUSUM statistic at location b is defined as

$$\widetilde{X}_{s,e}^{b} = \sqrt{\frac{e-b}{n(b-s+1)}} \sum_{t=s}^{b} X_t - \sqrt{\frac{b-s+1}{n(e-b)}} \sum_{t=b+1}^{e} X_t, \tag{2}$$

where $s \leq b < e$ with $n = e - s + 1$. This statistic measures the discrepancy between weighted sample means on the left and right subsegments of $[s, e]$ split at b. A large value of $|\widetilde{X}_{s,e}^{b}|$ indicates evidence for a change in the mean near location b. Due to its sensitivity to mean shifts, the CUSUM statistic serves as the fundamental building block for a wide range of modern change-point procedures. In the statistical problem, the change-points are unknown and must be estimated from the data by suitable detection methods. Many classical approaches to this problem are developed under two standard assumptions on the noise:

1. the noise ε_t is Gaussian;

2. the noise ε_t is independent.

These assumptions are convenient because they allow one to prove theoretical consistency. However, in many real-world data sets, the noise may be heavy-tailed, skewed, or dependent in time. It is therefore important to understand how robust different change-point detection methods are when these assumptions are not satisfied.

1.2 Scope of this Study

In this report, we work in the offline change-point detection setting for one-dimensional signals; see model (1), for f_t being piecewise-constant. Our scope is to evaluate the accuracy of each method in its ideal environment and to assess its robustness when deviating from the standard assumptions of Gaussianity and independence, which we examine through simulated data. We study the performance of the following change-point detection methods: Isolate-Detect (`ID`)(A. Anastasiou and P. Fryzlewicz, 2022), Wild Binary Segmentation (`WBS`)(P. Fryzlewicz, 2014), Wild Binary Segmentation 2 (`WBS2`)(P. Fryzlewicz, 2020), Narrowest-Over-Threshold (`NOT`)(R. Baranowski et al., 2019), Moving Sum(`MOSUM`)(B. Eichinger and C. Kirch, 2018), Data-adaptive structural change-point detection via isolation (`DAIS`) (A. Anastasiou and S. Loizidou, 2025), Linearly Penalized Segmentation (`PELT`)(R. Killick et al., 2012), and Seeded Binary Segmentation (`SeedBS`)(S. Kovács et al., 2020). To assess the methods, we conduct a simulation study in which, for each scenario, we evaluate both the accuracy of the estimated number of change-points and the accuracy of their estimated locations; we also report the computational cost of each procedure. The results of such a study allow us to identify which procedures remain robust under deviations from Gaussianity and independence.

1.3 Importance of assumptions and robustness

In addition to noise properties, several practical factors also influence how well changes can be detected. An important one is the signal-to-noise ratio (SNR), which reflects how large the change in the mean is compared to the variability of the noise. A high SNR, where the jump size is large and the noise variance is small, makes changes easier to detect. Conversely, a low SNR makes detection more difficult. The location of the change-point also matters; when a change occurs very close to the beginning or end of the sequence, many algorithms struggle to detect it. Let $r_0 = 0$ and $r_{N+1} = T$, and define $\delta_T = \min_{j=1,\dots,N+1}(r_j - r_{j-1})$, the minimum spacing between two adjacent change-points. The detectability of the signal is closely linked to the relationship between δ_T and the minimum jump size $\underline{f}_T$, which **must satisfy** $\sqrt{\delta_T}\,\underline{f}_T \geq \underline{C}\sqrt{\log T}$, for a sufficiently large constant C. Thus, larger spacing δ_T or larger jump magnitude $\underline{f}_T$ makes change-points easier to detect. The length of the time series is another factor, as longer data typically provides more information and leads to more reliable results.

1.4 Overview of methods

Several modern change-point detection methods have been developed, each using a different strategy to identify structural breaks in the signal. Below, we give a brief overview of the methods considered in this study, along with their main references.

Wild Binary Segmentation (WBS) (P. Fryzlewicz, 2014) and **WBS2** (P. Fryzlewicz, 2020). WBS aims to detect change-points by examining random sub-intervals of the data. The method draws M random intervals $[s_j, e_j]$, for $j \in 1,\dots,M$, with $1 \leq s_j < e_j \leq T$, and on each such interval it computes the CUSUM statistic (2) for all candidate locations b, where $b \in \{s_j,\dots,e_j-1\}$. For each interval, the maximiser $b_j^\star = \arg\max_{s_j \leq b < e_j} |\widetilde{X}_{s_j,e_j}^{b}|$ is recorded, and among all these $b_j^\star$ we take the one with the largest absolute contrast, denoted $b^\star$. If this maximal contrast exceeds a threshold ζ_T, then $b^\star$ is declared a change-point. The procedure is then repeated recursively on the left

and right sub-intervals around the estimated change-point $b^\star$.
WBS2 refines this idea while keeping the same CUSUM contrast. In the original WBS, a single large collection of M random intervals is drawn on $[1, T]$ and reused throughout the recursion. By contrast, in WBS2, a relatively small number K of random sub-intervals is drawn within the current segment. It then uses the first change-point candidate to split the data into two parts, and again recursively draws the same number K of sub-intervals to the left and to the right of this change-point candidate, and so on. This process continues until no further change-points are found. This adaptive mechanism prevents the random interval set from missing regions that contain closely spaced change-points, since new intervals are created in segments that appear noisy or structurally complex. Also, WBS2 replaces fixed thresholding or BIC-like criteria with the Steepest Drop to Low Levels (SDLL) model selection rule. SDLL orders the CUSUM magnitudes and identifies the location of a sharp drop, separating true change-points from noise, and uses thresholding only as a secondary check. This approach is particularly effective when many change-points are present: WBS2 maintains accuracy and avoids over-segmentation compared to WBS, especially when there are no change-points, and SDLL is easier to calibrate than penalty-based alternatives.

Isolate-Detect (ID) (A. Anastasiou and P. Fryzlewicz, 2022). The ID method proceeds in two stages. First, in the *isolation* stage, ID constructs a structured collection of expanding intervals that are highly likely to contain at most one true change-point. Given a sequence of length T and an expansion parameter $\lambda_T > 0$, two families of intervals are formed: the right-expanding intervals, $R_j = [1, \min\{j\lambda_T, T\}]$ and the left-expanding intervals, $L_j = [\max\{1, T - j\lambda_T + 1\}, T]$ for $j = 1, \ldots, K$, where $K = \lceil T/\lambda_T \rceil$. These intervals are ordered as $S_{R,L} = \{R_1, L_1, R_2, L_2, \ldots, R_K, L_K\}$. Secondly, in the *detection* stage, ID scans the intervals in $S_{R,L}$ and, for each interval $[s, e]$, evaluates the CUSUM function $|\widetilde{X}^b_{s,e}|$ for all candidate change-points $b \in \{s, \ldots, e-1\}$. The candidate change-point is $b^* = \arg\max_{b \in \{s,\ldots,e-1\}} |\widetilde{X}^b_{s,e}|$. If $\beta^* > \zeta_T$, where $\zeta_T = C\sqrt{2\log T}$ is a threshold depending on the noise level and the length of the sequence, then b^* is declared a detected change-point. If not, then the next interval in $S_{R,L}$ is tested. Upon detection, ID makes a new start from the end-point (or start-point) of the right (or left) expanding interval where the detection occurred. Upon the correct choice of ζ_T, ID ensures that we work on intervals with at most one change-point, which was our aim. This iterative boundary update ensures that each subsequent search focuses only on the part of the signal that may still contain undetected changes, avoiding unnecessary re-examination of previously scanned regions.

Narrowest-Over-Threshold (NOT) (R. Baranowski et al., 2019). The NOT procedure detects multiple change-points in two steps. In the first step, rather than analysing the full sample at once, we randomly extract subsamples $(Y_{s+1}, \ldots, Y_e)'$, where the pair (s, e) is drawn uniformly from $\{0, \ldots, T-1\} \times \{1, \ldots, T\}$ with $0 \le s < e \le T$ and $e - s \ge 2(d-1)$, where d is known and typically small. Since the signal we are working with is piecewise constant, each segment is described by a single parameter. (its constant level), so we set $d = 1$. With $d = 1$, the identifiability condition simply requires that the minimum distance between consecutive change-points is at least 1, ensuring that every segment contains at least one observation. Let $\ell(Y_{s+1:e}; \Theta)$ denote the likelihood of a parametric model with parameter Θ fitted to the data $(Y_{s+1}, \ldots, Y_e)$. For each interval $[s, e]$, we compute for all candidate change-points $b \in \{s+d, \ldots, e-d\}$, the generalised log-likelihood ratio (GLR) statistic

$$\mathcal{R}^b_{s,e}(Y) = 2\log\left[\frac{\sup_{\Theta^1,\Theta^2}\left\{\ell(Y_{s+1:b};\Theta^1)\,\ell(Y_{b+1:e};\Theta^2)\right\}}{\sup_{\Theta}\ell(Y_{s+1:e};\Theta)}\right]. \tag{3}$$

This whole procedure is repeated for M randomly drawn intervals $(s_1, e_1), \ldots, (s_M, e_M)$. In the second step, all quantities $\mathcal{R}_{s_m,e_m}(Y) = \max_{b\in\{s_m+d-1,\ldots,e_m-d\}}\mathcal{R}^b_{s_m,e_m}(Y)$, for $m = 1, \ldots, M$, are compared to a given threshold ζ_T. Only those intervals with $\mathcal{R}_{s_m,e_m}(Y) > \zeta_T$ are retained, and among these significant intervals we pick the one with the smallest length, say $[s^*_m, e^*_m]$. The corresponding change-point estimate is then $b^* = \arg\max_b \mathcal{R}^b_{s^*_m,e^*_m}(Y)$. Once a change-point b^* has been identified in $[s^*_m, e^*_m]$, the same NOT procedure is applied recursively to the subintervals to the left and to the right of b^*, and the recursion stops in any segment as soon as no further intervals within that segment yield a GLR exceeding ζ_T. In this way, NOT builds up an estimate of the entire set of change-points by repeatedly focusing on the narrowest intervals that provide statistically significant evidence of a change, whereas WBS searches for change-points by examining many randomly drawn intervals and selecting the one with the maximum contrast function value.

Moving Sum (MOSUM) (B. Eichinger and C. Kirch, 2018). The MOSUM method chooses a window length G such that $2G < n$, where n is the total number of observations, and for each point k computes the MOSUM statistic

$$W_{k,n}(G) = \frac{\sqrt{G}}{\hat{\tau}_k}\left\|\Sigma^{-1/2}\big(\hat{\beta}^+(k) - \hat{\beta}^-(k)\big)\right\|, \qquad G \le k \le n-G,$$

where $\hat{\beta}^+_k$ and $\hat{\beta}^-_k$ are the least squares estimators and Σ is a 2×2 diagonal matrix, based on the distribution of $\hat{\beta}^+_k - \hat{\beta}^-_k$ under $H_0 : J_n = 0$, where J_n is the total number of change-points and k_j their locations. The term $\hat{\tau}_k$ denotes a (possibly) location-dependent estimator of τ. For some $\alpha \in (0, 1)$, H_0 is rejected if $W_n(G) := \max_{G\le k\le n-G} W_{k,n}(G)$ exceeds a critical value $C_n(G, \alpha)$ obtained from the asymptotic null distribution of $W_n(G)$, ensuring that the test controls the family-wise error rate at level α when scanning $k = G, \ldots, n-G$. Flagged points exceeding the threshold are treated as candidate change-points. Once candidate points are found, the method groups candidates that are close to each other (within some distance depending on G) to avoid detecting multiple versions of the same change. The data are then split at the detected change-points, and the procedure is applied recursively to the resulting segments, continuing until no MOSUM statistic exceeds the threshold.

Data-adaptive structural change-point detection via isolation (DAIS) (A. Anastasiou and S. Loizidou, 2025). DAIS initially examines the interval $[1, T]$ and identifies regions where changes are likely to occur. We denote $d_{s,e}$ as the location of the largest difference, in absolute value, in the interval [s,e], where $1 \le s < e \le T$, and is defined as

$$d_{s,e} = \arg\max_{t\in\{s,\,s+1,\,\ldots,\,e-1\}} |X_{t+1} - X_t|$$

A mean change at time r_i creates a jump in the observations, which in turn produces a pronounced spike in the first difference $|X_{t+1} - X_t|$. DAIS uses the index $d_{s,e}$ to capture this spike within the interval $[s, e]$, treating it as the most likely location of a change. These spike locations then guide the construction of the expanding left and right intervals that form the core isolation step of the algorithm. At each expanded interval, DAIS evaluates the CUSUM function 2 at all candidate locations and declares a change-point when the CUSUM statistic exceeds the threshold ζ_T, where $\zeta_t \in \mathbb{R}^+$. Once a change-point has been detected, the interval is updated so that it becomes more tightly centered around the suspected location, allowing the change to be isolated in a smaller region for more precise estimation. The procedure is then repeated. If the interval contains multiple change-points, DAIS adaptively splits it into further sub-intervals, rather than following a predetermined scanning structure. This data-adaptive mechanism is the key novelty of the algorithm and contrasts with existing methods in the literature that follow a fixed workflow and do not adapt to the underlying data structure regarding change-point locations, if any exist.

Linearly Penalized Segmentation (PELT) (R. Killick et al., 2012). PELT formulates change-point detection as the minimization of a cost function consisting of a goodness-of-fit term and a penalty on the number of detected change-points. The penalty can be selected using criteria such as BIC, MBIC, AIC, or chosen manually. A lower penalty tends to identify more change-points, whereas a higher penalty favors segmentations with minimal complexity. PELT achieves exact solutions efficiently using a pruning technique that avoids unnecessary calculations, making it suitable for large data sets.

Seeded Binary Segmentation (SeedBS) (S. Kovács et al., 2020). Seeded Binary Segmentation is a fast and flexible change-point detection method designed for large-scale problems. Similar to Wild Binary Segmentation (WBS) and the Narrowest-Over-Threshold (NOT) method, SeedBS first identifies candidate split points across many intervals and then refines them into a final selection. Its key innovation lies in the deterministic construction of seeded intervals, which may be pre-computed in advance. This yields near-linear computational cost, independent of the true number of change-points, while maintaining strong statistical accuracy. Compared to other methods, SeedBS offers robust performance across a wide range of scenarios, making it effective for both large univariate signals and high-dimensional situations where model fitting is computationally expensive.

All of these methods are well-established in the literature and are commonly used in both theoretical studies and applied settings. By comparing multiple approaches, we aim to gain a better understanding of how different detection strategies respond when the classical Gaussianity and independence assumptions are violated.

2 Gaussian i.i.d. Noise

2.1 Model and simulation function

We consider the model in (1), where f_t is a piecewise-constant signal with structural changes in the mean at unknown locations. The noise sequence $\{\varepsilon_t\}$ is assumed to follow an independent and identically distributed Gaussian distribution with mean zero and variance σ^2. This setting corresponds to the classical baseline model in change-point analysis, under which many detection procedures are theoretically justified and serve as a reference for assessing robustness under model deviations.

To evaluate the performance of the methods, we carried out a simulation study. First, $m = 500$ independent time series are generated under the same pre-specified signal. For each replication, the accuracy of the methods is assessed through four key metrics.

Key Metrics

Let K and $\widehat{K}$ denote the true and estimated numbers of change-points, and let $\mathcal{T}$ and $\widehat{\mathcal{T}}$

be the corresponding sets of true and estimated change-point locations.

The **dnc** metric provides a numerical measure of the difference between the estimated and true number of change-points, serving as a global indicator of detection accuracy. We show the distribution of the estimation error $\widehat{K} - K$ across the discrete categories ($\leq -3, -2, -1, 0, 1, 2, \geq 3$). Negative values correspond to under-detection, positive values correspond to over-detection, and concentration at 0 reflects correct detection. The **mse** metric represents the mean squared error of the estimated signal compared to the true signal, providing a measure of overall signal reconstruction accuracy by averaging the squared differences across all time points. The **dh** metric quantifies the Hausdorff distance between the sets $\widehat{\mathcal{T}}$ and $\mathcal{T}$, providing a measure of location accuracy. Finally, the **time** metric reflects computational complexity and reports the average time required to detect change-points. Lower values indicate faster execution and lower computational cost. For `mse`, `dh`, and `time`, lower values correspond to better performance.

2.2 Signals and experimental design

We evaluated the methods on three representative signals with different structural characteristics.
Before presenting the results, we clarify the notation used for the different method variants. Methods whose names end in `th` use a threshold-based penalty function to determine when to stop adding change-points, whereas methods ending in `ic` or `sic` rely on an information-criterion penalty, specifically the Schwarz Information Criterion (SIC). In addition, several change-point detection methods execute extremely quickly, resulting in recorded run times of `0.000`; these values are not exactly zero but fall below the reporting precision (approximately $< 10^{-3}$), reflecting the very low computational cost of these algorithms.

Signal 1: First, we examine Signal 1, which contains two ($K = 2$) large and well-separated change-points, with true change-point locations: $\mathcal{T} = \{100, 200\}$. Such signals are typically easy to detect and provide a clear baseline for comparing method accuracy.

$$f_t = \begin{cases} 0, & 1 \leq t \leq 100 \\ 5, & 101 \leq t \leq 200 \\ -3, & 201 \leq t \leq 300 \end{cases}$$

Table 1: Signal 1: Results under Gaussian i.i.d. noise.

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	331	86	49	34	0.0237	0.177	0.020
wbssbic	0	0	474	21	5	0	0.0140	0.025	0.020
notic	0	0	478	19	3	0	0.0134	0.022	0.038
IsolateDetect_sic	0	0	468	23	8	1	0.0149	0.030	0.002
IsolateDetect_th	0	0	431	40	22	7	0.0177	0.062	0.001
wbs2	0	0	458	16	14	12	0.0174	0.041	0.000
mosum	0	0	441	58	1	0	0.0133	0.058	0.000
dais	0	0	437	45	13	5	0.0161	0.061	0.000
pelt	0	0	500	0	0	0	0.0113	0.000	0.003
seedBS_th	0	0	374	73	32	21	0.0178	0.125	0.018
seedBS_ic	0	0	483	10	6	1	0.0132	0.024	0.008

Discussion Under Signal 1, most methods show their distribution of the estimation error $\widehat{K} - K$ concentrated at 0, indicating accurate estimation of the number of change-points. `pelt` is exact on all replications (500 in 0; `DH`$= 0$) and achieves the lowest `MSE`. The next most reliable performers are `seedBS_ic`, `wbsbic`, `notic`, `wbs2`, `IsolateDetect_sic`, and `IsolateDetect_th`, all with very small right tails and low location error. `mosum` and `dais` are also stable but show slightly higher `MSE` and `DH` than the best group. Greater sensitivity appears for `wbsth`, and `seedBS_th`, which display larger estimation error: $1, 2, \geq 3$, (over-segmentation) and hence increased `DH`. Computationally, nearly all procedures run extremely fast in this ideal scenario, with `wbs2`, `mosum`, and `dais` being the most efficient, while `notic` is the slowest among the group but still fast in absolute terms.

Signal 2: Next, we examine Signal 2, which contains two ($K = 2$) short and slight changes in the mean, with true change-point locations: $\mathcal{T} = \{496, 504\}$. This scenario is more challenging for most methods, as short segments are harder to detect and may lead to underestimation or instability.

$$f_t = \begin{cases} 0, & 1 \leq t \leq 496 \\ 1, & 497 \leq t \leq 504 \\ 0, & 505 \leq t \leq 1000 \end{cases}$$

Table 2: Signal 2: Results under Gaussian i.i.d. noise.

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	279	76	106	25	10	4	0.0130	0.738	0.029
wbssbic	478	10	12	0	0	0	0.0095	0.988	0.029
notic	481	8	11	0	0	0	0.0094	0.988	0.038
IsolateDetect_sic	475	12	13	0	0	0	0.0095	0.986	0.008
IsolateDetect_th	376	29	80	8	6	1	0.0112	0.825	0.011
wbs2	392	42	36	11	8	11	0.0113	0.878	0.000
mosum	459	38	3	0	0	0	0.0094	0.972	0.001
dais	409	23	62	4	2	0	0.0105	0.881	0.000
pelt	500	0	0	0	0	0	0.0089	1.016	0.008
seedBS_th	317	106	58	13	5	1	0.0107	0.774	0.021
seedBS_ic	483	8	9	0	0	0	0.0094	0.997	0.010

Discussion Under Signal 2, most methods show their distribution of the estimation error $\widehat{K} - K$ concentrated at -2, indicating substantial under-detection across many methods. `pelt`, `wbsbic`, `notic`, `IsolateDetect_sic`, `mosum`, and `seedBS_ic` concentrated heavily at $K - \hat{K} = -2$. In contrast, `wbsth`, `IsolateDetect_th`, `seedBS_th`, `dais`, and `wbs2` exhibit relatively more mass at $K - \hat{K} = 0$ compared to the other methods.However, the overall concentration is still low, indicating that these methods rarely recover the correct number of change-points under this signal. The `MSE` values are broadly similar, with slight gains for `pelt`, `wbssic`, `notic`, `IsolateDetect_sic`, and `mosum`, indicating highly accurate reconstruction of the underlying signal. Computation times are uniformly small, with near-linear methods (`mosum`, `wbs2`, `dais`) being fastest.

Signal 3: Finally, we examine Signal 3, which contains fourteen ($K = 14$) closely spaced change-points, with true change-point locations: $\mathcal{T} = \{10, 20, 30, \ldots, 140\}$. This high-frequency setting is the most difficult, requiring methods to balance sensitivity with the risk of over-segmentation.

Table 3: Signal 3: Results under Gaussian i.i.d. noise.

Method	≤ -3	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	9	417	65	7	2	0.0252	0.213	0.014
wbsbic	0	0	1	302	142	40	15	0.0242	0.229	0.014
notic	0	0	1	455	43	1	0	0.0213	0.179	0.060
IsolateDetect_sic	0	0	1	442	52	5	0	0.0202	0.168	0.003
IsolateDetect_th	0	0	4	479	17	0	0	0.0216	0.178	0.002
wbs2	0	0	10	421	60	6	3	0.0250	0.211	0.000
mosum	0	1	31	313	129	22	4	0.0247	0.265	0.000
dais	0	0	7	473	19	0	1	0.0226	0.183	0.000
pelt	500	0	0	0	0	0	0	2.0189	3.765	0.000
seedBS_th	0	12	56	375	51	5	1	0.0306	0.290	0.012
seedBS_ic	0	0	1	320	135	36	8	0.0250	0.231	0.006

Discussion Under Signal 3, most methods remain stable, and performance is stronger compared to the setting in Table 2. `IsolateDetect_th`, `IsolateDetect_sic`, `notic`, `wbs2`, `wbsth`, and `dais` show their distribution of the estimation error $\widehat{K}-K$ concentrated almost entirely at 0, indicating reliable detection accuracy. In contrast, `mosum`, `wbsbic` `seedBS_th`, and `seedBS_ic` exhibit mass at 0 but also notable proportions at -1, 1, 2, and ≥ 3, meaning these methods have a tendency of over-estimation, with a substantial number of estimates exceeding the true number of change-points. Furthermore, `pelt` has $\widehat{K}-K$ concentrated at "≤ -3" and is the only method that displays extremely large mean squared error (`MSE`) together with the highest `DH` value, indicating heavily inaccurate estimation and very poor localisation accuracy. The remaining methods have relatively small `MSE` and `DH` values. Computation times are uniformly low across all methods.

3 Non-Gaussian Noise

In the previous experiments, the noise term ε_t was assumed to be i.i.d. Gaussian with mean zero and variance one. To investigate the robustness of the detection methodologies under departures from the Gaussianity assumption, we now consider signals contaminated with i.i.d. noise from alternative distributions. Specifically, we generate examples where the noise term follows:

(i) Gamma distribution,

(ii) Chi-square distribution,

(iii) Student-t distribution,

(iv) Uniform distribution.

In addition, we further examine the Student-t distribution in more detail by varying its degrees of freedom, thereby assessing how heavier or lighter tails influence the ability of the detection methods to correctly identify the true change-points.

3.1 Main signal and simulation function

We use the below piecewise-constant signal as the reference signal for all non-Gaussian scenarios. This signal has length $n = 1000$ and contains two ($K = 2$) change-points located at $\mathcal{T} = \{460, 540\}$.
The signal is defined as

$$f_t = \begin{cases} 0, & 1 \leq t \leq 460, \\ 1, & 461 \leq t \leq 540, \\ 0, & 541 \leq t \leq 1000. \end{cases}$$

This signal is chosen because its change-points are large and well separated, making them straightforward to detect under non-Gaussian noise. By keeping the signal fixed and only varying the distribution of the noise, we can isolate and assess the impact of different noise characteristics on the performance of the change-point detection methods. This design ensures that any differences in performance can be attributed to the noise distribution rather than to changes in the underlying signal structure.

Simulation study: We modified the simulation function to accept the noise distribution as an input argument and adjust the generated data accordingly. For each experiment, the observed data are constructed by adding the noise drawn from the chosen distribution to the fixed main signal, where ε_t follows the corresponding non-Gaussian distribution. Each scenario is replicated $m = 500$ times, and the performance of the methods is evaluated using the previously defined key metrics.

3.1.1 Gamma noise (shape = 1, scale = 1)

We now examine the behavior of the change-point detection methods when the noise follows a Gamma distribution with shape parameter $\alpha = 1$ and scale parameter $\lambda = 1$. This choice yields unit variance, ensuring comparability with the other noise settings. The Gamma distribution is positively skewed, which can shift the local mean upward, potentially increasing the occurrence of spurious detections.

Table 4: Main Signal: Results under Gamma noise (with shape =1 and scale = 1).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	0	0	0	500	1.2858	0.942	0.030
wbssbic	0	0	38	2	93	367	1.1104	0.631	0.030
notic	9	0	421	44	19	7	1.0154	0.107	0.039
IsolateDetect_sic	0	0	130	24	100	246	1.0720	0.454	0.007
IsolateDetect_th	0	0	0	0	0	500	1.2260	0.901	0.005
wbs2	0	0	0	0	0	500	1.4298	0.962	0.001
mosum	2	44	403	45	6	0	1.0186	0.078	0.001
dais	0	0	0	1	0	499	1.2306	0.893	0.000
pelt	15	0	276	60	111	38	1.0451	0.231	0.009
seedBS_th	0	0	0	0	0	500	1.3044	0.947	0.021
seedBS_ic	2	0	60	14	95	329	1.1065	0.590	0.010

Discussion Under Gamma noise with shape = 1 and scale = 1, the performance of most methods is unstable due to the positive skewness of the noise. `wbsth`, `wbs2`, `dais`, `IsolateDetect_th` and `seedBS_th` tend to over-segment heavily, as indicated by all of their instances being at $\widehat{K} - K \geq 3$, together with elevated `MSE` values. Methods such as `notic`, and `mosum` have estimation error $K - \widehat{K}$ concentrated primarily at 0, indicating frequent correct detection. The `pelt` method displays more dispersed behavior, with $\widehat{K} - K$ values spread across $\{-2, 0, 1, 2, \geq 3\}$, suggesting reduced robustness under skewed noise. `IsolateDetect_sic` shows a tendency to over-segment, though less severely than `wbsbic` and `seedBS_ic`, which exhibit mild over-segmentation, as reflected by their $\widehat{K} - K$ values distributed over $\{0, 1, 2, \geq 3\}$, with the majority of cases concentrated in ≥ 3. Overall, Gamma noise introduces moderate but noticeable challenges, increasing variability and false detections across methods.

3.1.2 Chi-square noise (df = 0.5)

In this section, we evaluate the performance of the change-point detection methods when the noise follows a Chi-square distribution with $df = 0.5$. This distribution is strongly right-skewed with $\mathbb{E}[X] = 0.5$ and $\mathrm{Var}(X) = 1$. The pronounced skewness and heavy right tail may cause methods to over-segment by interpreting large random fluctuations as structural breaks.

Table 5: Main Signal: Results under Chi-square noise (df= 0.5).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	0	0	0	500	0.7920	0.986	0.030
wbsbic	0	0	0	0	0	500	0.4854	0.924	0.030
notic	75	1	205	112	51	56	0.0300	0.487	0.039
IsolateDetect_sic	0	0	5	0	7	488	0.3211	0.862	0.038
IsolateDetect_th	0	0	0	0	0	500	0.7782	0.985	0.013
wbs2	0	0	0	0	0	500	0.9496	0.993	0.001
mosum	2	69	394	32	3	0	0.0206	0.089	0.001
dais	0	0	0	0	0	500	0.8324	0.986	0.000
pelt	0	0	13	14	62	411	0.2171	0.651	0.008
seedBS_th	0	0	0	0	0	500	0.8709	0.989	0.021
seedBS_ic	0	0	0	0	1	499	0.5159	0.934	0.009

Discussion For Chi-square noise with df= 0.5, the performance of almost all methods worsen notably. The strong skewness of the distribution leads to frequent over-segmentation, particularly for `wbsth`, `wbssic`, `IsolateDetect_sic`, `IsolateDetect_th`, `wbs2`, `dais`,`seedBS_th` and `seedBS_ic`, all of which display large or even all counts in the "$\geq$ 3" column and high MSE values. The `pelt` method performs slightly better than heavily over-segmenting approaches, but still exhibits substantial variability, with a large fraction of outcomes in $\widehat{K} - K \geq 3$, indicating reduced robustness under highly skewed noise. Methods such as `notic` and `mosum` show a mild tendency toward under-segmentation, but more than half of their realizations are concentrated at $\widehat{K} - K = 0$, indicating frequent correct detection of the true change-points. Among these, `mosum` achieves the highest accuracy, with the largest proportion of correctly identified change-points, while also attaining the lowest MSE and DH values.

3.1.3 Student-t Noise (df ={3, 10, 15})

In this section, we investigate the Student-t distribution by varying its degrees of freedom, to understand how tail–heaviness influences change-point detection. We begin with very heavy-tailed noise (df = 3) and gradually increase the degrees of freedom (df = 10 and df = 15), allowing us to observe how the distribution transitions toward Gaussian behaviour as $\nu \to \infty$. This progression is particularly relevant because the t-distribution resembles the normal distribution but has heavier tails; for small ν the presence of extreme values becomes more likely, making detection methods prone to overfitting, whereas for larger ν it approaches $\mathcal{N}(0, 1)$ and behaves more stably. By examining this range of regimes, we assess how the severity of heavy tails affects both the stability and accuracy of the detection methods.

Table 6: Main Signal: Results under Student-t noise (df = 3).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	0	0	0	500	1.2137	0.918	0.025
wbssbic	2	0	10	2	32	454	0.9792	0.785	0.025
notic	412	15	37	27	4	5	0.0776	1.038	0.037
IsolateDetect_sic	15	2	64	11	78	330	0.7475	0.657	0.005
IsolateDetect_th	0	0	0	0	0	500	1.0640	0.876	0.005
wbs2	0	0	0	0	0	500	1.3666	0.938	0.000
mosum	135	208	140	16	1	0	0.0647	0.447	0.000
dais	0	0	0	0	1	499	1.1095	0.881	0.000
pelt	54	0	171	9	155	111	0.7594	0.699	0.001
seedBS_th	0	0	0	0	0	500	1.2264	0.928	0.017
seedBS_ic	9	1	24	0	44	422	0.9946	0.796	0.008

Table 7: Main Signal: Results under Student-t noise (df = 10).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	107	62	110	221	0.0441	0.494	0.024
wbssbic	0	0	416	12	65	7	0.0189	0.098	0.024
notic	3	0	478	14	5	0	0.0125	0.043	0.036
IsolateDetect_sic	0	0	447	25	25	3	0.0165	0.061	0.003
IsolateDetect_th	0	0	221	41	134	104	0.0331	0.310	0.005
wbs2	0	0	191	59	89	161	0.0358	0.379	0.000
mosum	14	75	378	30	3	0	0.0241	0.105	0.000
dais	1	0	261	56	111	71	0.0288	0.261	0.000
pelt	86	0	404	6	4	0	0.0247	0.224	0.001
seedBS_th	1	2	103	94	98	202	0.0357	0.481	0.016
seedBS_ic	8	0	417	12	51	12	0.0203	0.108	0.007

Table 8: Main Signal: Results under Student-t noise (df = 15).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	191	74	130	105	0.0273	0.366	0.025
wbssbic	0	0	464	6	28	2	0.0132	0.048	0.025
notic	1	0	486	9	4	0	0.0112	0.028	0.036
IsolateDetect_sic	0	0	460	26	11	3	0.0147	0.044	0.002
IsolateDetect_th	0	0	325	50	86	39	0.0215	0.196	0.005
wbs2	0	0	270	80	76	74	0.0218	0.264	0.000
mosum	4	56	395	40	4	1	0.0195	0.075	0.000
dais	2	0	320	57	95	26	0.0207	0.202	0.000
pelt	62	0	437	1	0	0	0.0191	0.161	0.001
seedBS_th	0	2	180	116	99	103	0.0222	0.372	0.016
seedBS_ic	6	0	455	8	28	3	0.0143	0.063	0.007

Overall conclusion across degrees of freedom As the degrees of freedom increase from $3 \to 10 \to 15$, the heavy tails of the Student-t distribution become lighter, and its variance decreases, resulting in steady performance improvements across all methods. At df= 3, heavy tails cause over-segmentation ($\hat{K} - K \geq 1$) and localization errors (big values of `DH`), particularly for `wbsth`, `wbsbic`, `IsolateDetect_th`, `wbs2`, `dais`, and `seedBS_th`. `notic` is the only method that tends to under-segment heavily, with $\hat{K} - K = -2$, followed by `mosum` that has $\widehat{K} - K$ values that are distributed over $\{-2, -1, 0, 1\}$. The rest of the methods (`IsolateDetect_sic`, `pelt` and `seedBS_ic`) have $\widehat{K} - K$ values that are distributed over $\{-2, -1, 0, 1, 2\}$. You can observe that as the Student-t distribution becomes closer to the normal distribution ($3 \to 10$), the threshold-based methods (`wbsth`, `IsolateDetect_th`, `seedBS_th`, and `wbs2`) continue to suffer from severe over-segmentation even at lighter tails, while `wbsbic`, `IsolateDetect_sic`, `seedBS_ic`, `notic`, and `pelt` become increasingly stable, achieving low `MSE`, small `DH`, and tightly concentrated $\hat{K} - K$ distributions. Overall, the IC-based methods exhibit the most consistent and accurate performance as the degrees of freedom increase, and the worst methods are `wbsth` and `seedBS_th`, with $\widehat{K} - K$ values that are distributed over $\{\leq -3, -2, -1, 0, 1, 2, \geq 3\}$, which shows that the methods are inconsistent, showing no reliable pattern in estimating the number of change-points.

3.1.4 Uniform Noise (min $= -\sqrt{3}$, max $= \sqrt{3}$)

We now consider a centered uniform distribution with parameters $a = -\sqrt{3}$ and $b = \sqrt{3}$. This choice is particularly important since it matches the first two moments of the standard Normal distribution: zero mean and unit variance, while remaining symmetric. This allows us to assess the performance of the methods in a setting free of mean bias and with a Gaussian-like scale.

Table 9: Main Signal: Results under Uniform noise ($a = -\sqrt{3}$, $b = \sqrt{3}$).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	483	15	2	0	0.0099	0.031	0.025
wbssbic	0	0	494	6	0	0	0.0097	0.019	0.025
notic	0	0	493	7	0	0	0.0098	0.019	0.036
IsolateDetect_sic	0	0	488	12	0	0	0.0108	0.023	0.004
IsolateDetect_th	0	0	483	15	2	0	0.0108	0.026	0.006
wbs2	0	0	496	3	1	0	0.0097	0.017	0.000
mosum	2	33	416	45	4	0	0.0150	0.060	0.000
dais	2	0	488	9	1	0	0.0107	0.025	0.000
pelt	17	0	483	0	0	0	0.0118	0.051	0.001
seedBS_th	0	5	488	6	1	0	0.0106	0.024	0.017
seedBS_ic	0	0	495	5	0	0	0.0099	0.015	0.008

Discussion Under centered Uniform noise with $a = -\sqrt{3}$ and $b = \sqrt{3}$, all methods show highly stable performance across the four evaluation metrics. The distribution of $\hat{K} - K$ is tightly concentrated at 0 for nearly all procedures, indicating that they consistently detect the correct number of change-points. Only `mosum` exhibits a few deviations in $\{-2, -1, 1, 2\}$, but these occurrences remain infrequent. DH values stay uniformly low across methods, reflecting accurate localisation, while MSE values are all close to 10^{-2}, demonstrating strong estimation accuracy in this centered, mean-zero setting.

4 Non-i.i.d. Noise

4.1 Dependence Structures Considered

In the previous section, the noise term ε_t was assumed to be independent and identically distributed (i.i.d.) but non-Gaussian. To further assess the robustness of the change-point detection procedures under violations of the independence assumption, we now consider settings in which the noise exhibits dependence while preserving Gaussian marginal distributions. Specifically, we contaminate the signal with three classical stationary time series models:

(i) Auto-regressive process of order 1 (**AR(1)**),

(ii) Moving-average process of order 1 (**MA(1)**),

(iii) Auto-regressive moving-average process (**ARMA(1,1)**).

These models are fundamental in time series analysis and provide controlled departures from independence.

AR(1) process.

$$X_t = \alpha\, X_{t-1} + \varepsilon_t + \mu_t,$$

where $|\alpha| < 1$ ensures stationarity. This model induces serial correlation between successive noise terms, leading to persistent deviations that may mimic genuine structural breaks.

MA(1) process.

$$X_t = \varepsilon_t + \beta\, \varepsilon_{t-1} + \mu_t,$$

Where the noise depends on current and lagged innovations. Dependence is short-term but can still create locally smooth patterns that affect detection sensitivity.

ARMA(1,1) process.

$$X_t = \alpha\, X_{t-1} + \varepsilon_t + \beta\, \varepsilon_{t-1} + \mu_t.$$

This combines the features of AR and MA, producing both persistent and short-term correlation in the noise.

4.2 Simulation design and results

Throughout this section, we retain the same main signal 3.1 as before, ensuring that any changes in performance can be attributed solely to the noise structure, and follow the same simulation protocol: each setting is repeated 500 independent runs, and the results are evaluated using the same four key criteria. To introduce dependence while preserving Gaussian marginal distributions, we modify the simulation function so that the noise process is generated from classical stationary time-series models. The routine now accepts two parameters, α and β, allowing the noise to follow an AR(1) model when only α is specified, an MA(1) model when only β is set, and an ARMA$(1,1)$ process when both parameters are provided.

4.2.1 AR(1) Noise

We begin our analysis with the autoregressive noise structure, where dependence in the noise is controlled by the parameter α. We vary α over the set

$$\alpha \in \{0.1,\ 0.3,\ 0.5,\ 0.8\}.$$

to examine how dependence affects the performance of different change-point detection methods. The i.i.d. case corresponds to $\alpha = 0$, and for a stationary AR(1) process, stationarity is ensured when $|\alpha| \leq 1$. As $|\alpha|$ increases towards 1, the noise becomes more correlated across time, which generally makes detecting the true change-points more difficult.

Table 10: Main Signal: Results under AR(1) noise ($\alpha = 0.1$).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	142	106	106	146	0.0255	0.459	0.025
wbssbic	1	0	475	11	12	1	0.0127	0.037	0.025
notic	1	0	479	9	9	2	0.0124	0.030	0.037
IsolateDetect_sic	1	0	462	22	13	2	0.0147	0.045	0.005
IsolateDetect_th	0	0	267	77	91	65	0.0228	0.256	0.007
wbs2	0	0	231	79	66	124	0.0260	0.350	0.000
mosum	2	48	344	86	18	2	0.0201	0.126	0.000
dais	0	0	293	87	81	39	0.0196	0.205	0.000
pelt	31	0	469	0	0	0	0.0151	0.085	0.001
seedBS_th	0	2	175	151	102	70	0.0172	0.376	0.016
seedBS_ic	4	0	479	8	9	0	0.0128	0.039	0.007

Table 11: Main Signal: Results under AR(1) noise ($\alpha = 0.3$).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	0	0	0	500	0.1288	0.895	0.024
wbssbic	1	0	358	47	57	37	0.0294	0.181	0.024
notic	1	0	363	40	61	35	0.0293	0.162	0.037
IsolateDetect_sic	2	0	297	56	85	60	0.0367	0.239	0.007
IsolateDetect_th	0	0	2	2	9	487	0.1182	0.823	0.008
wbs2	0	0	0	0	2	498	0.2172	0.907	0.000
mosum	3	32	186	167	68	44	0.0332	0.340	0.000
dais	0	0	6	6	8	480	0.1081	0.805	0.000
pelt	45	1	446	8	0	0	0.0247	0.138	0.001
seedBS_th	0	0	1	2	3	494	0.0885	0.868	0.017
seedBS_ic	13	0	391	40	33	23	0.0279	0.164	0.007

Table 12: Main Signal: Results under AR(1) noise ($\alpha = 0.5$).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	0	0	0	500	0.4313	0.963	0.025
wbssbic	3	3	96	32	57	309	0.1385	0.611	0.025
notic	5	1	110	36	64	284	0.1293	0.573	0.039
IsolateDetect_sic	2	0	45	19	51	383	0.1640	0.687	0.011
IsolateDetect_th	0	0	0	0	0	500	0.4408	0.955	0.010
wbs2	0	0	0	0	0	500	0.6180	0.973	0.000
mosum	2	3	37	68	117	273	0.0834	0.702	0.000
dais	0	0	0	0	0	500	0.4270	0.955	0.000
pelt	78	12	352	41	14	3	0.0549	0.301	0.001
seedBS_th	0	0	0	0	0	500	0.3808	0.961	0.017
seedBS_ic	28	11	157	57	49	198	0.1021	0.531	0.008

Table 13: Main Signal: Results under AR(1) noise ($\alpha = 0.8$).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	0	0	0	500	2.0073	0.982	0.029
wbssbic	0	0	0	0	0	500	1.7368	0.969	0.029
notic	0	0	0	0	0	500	1.5407	0.955	0.043
IsolateDetect_sic	0	0	0	0	0	500	1.8017	0.973	0.032
IsolateDetect_th	0	0	0	0	0	500	2.0783	0.983	0.008
wbs2	0	0	0	0	0	500	2.2139	0.986	0.001
mosum	0	0	0	0	0	500	0.7212	0.911	0.001
dais	0	0	0	0	0	500	2.0681	0.983	0.000
pelt	1	1	9	23	38	428	0.7057	0.803	0.008
seedBS_th	0	0	0	0	0	500	2.0044	0.983	0.021
seedBS_ic	0	0	0	0	0	500	1.6810	0.968	0.010

Discussion This experiment examines the robustness of the competing detection methods when the independence assumption of the noise is violated. For small dependence ($\alpha = 0.1$), all procedures yield satisfactory performance, with small MSE and DH values and almost perfect recovery of the true change-points. The methods, such as `wbssbic` and `seedBS_ic`, `pelt`, `notic`, and `IsolateDeteect_sic`, provide the most stable and precise results, with over 450 concentrating detections at $\hat{K} - K = 0$, indicating very good estimation. In contrast, `wbsth`, and `seedBS_th` show a tendency toward over-segmentation, with $\hat{K} - K$ concentrated primarily at $0, 1, 2$ and "≥ 3". As the autocorrelation increases to $\alpha = 0.3$ and $\alpha = 0.5$, dependence between observations makes the noise persist over time, effectively increasing the estimation error: $\hat{K} - K$. Under these conditions, `wbsth`, `wbs2`, `dais`, `IsolateDetect_th`, `mosum`, and `seedBS_th` worsen more rapidly, producing additional false positives and elevated MSE. On the other hand, `wbssbic` and `seedBS_ic`, `pelt`, `notic`, and `IsolateDeteect_sic` retain moderate robustness, with limited inflation in `MSE` and `DH`. The penalized segmentation `pelt` outperforms all the methods so far, with the most observations concentrated at $\hat{K} - K = 0$, and the lowest `MSE`. When α approaches 0.8, the dependence structure becomes dominant. In this regime, all methods degrade substantially, with MSE values near 2.00 and DH approaching 1, and $\hat{K} - K$ distributed at "≥ 3", except `pelt` that stands out as the only approach that $\hat{K} - K$ does not concentrate entirely at 0, with its errors spread across nearby categories, and also retains meaningful accuracy, exhibiting the lowest `MSE` and `DH` values and producing change-point estimates that remain close to the true locations.

Overall, methods relying on thresholding are the least robust to dependence, as serial correlation causes systematic over-detection and high MSE, whereas `pelt` demonstrates strong resilience and consistent localization performance even under extreme autocorrelation.

4.2.2 ARMA(1,1) Noise

After completing the analysis for the AR(1) dependence structure, we now extend the investigation to the auto-regressive and moving-average dependence. In this setting, the noise follows an ARMA(1,1) process,where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, with $\alpha = 0.1$ and $\beta = 0.4$. This model combines both persistent and short-term correlation, allowing us to assess the joint effect of autoregression and moving-average dependence on change-point detection performance.

Table 14: Results for the main signal contaminated with ARMA(1) noise ($\alpha = 0.5$ and $\beta = 0.4$).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	0	0	0	500	1.1131	0.979	0.040
wbssbic	3	0	7	5	17	468	0.4822	0.862	0.040
notic	4	0	19	11	24	442	0.4135	0.819	0.057
IsolateDetect_sic	0	0	3	4	9	484	0.5316	0.882	0.036
IsolateDetect_th	0	0	0	0	0	500	1.1813	0.979	0.016
wbs2	0	0	0	0	0	500	1.4393	0.987	0.001
mosum	0	1	10	35	40	414	0.1863	0.793	0.001
dais	0	0	0	0	0	500	1.1757	0.980	0.000
pelt	119	30	200	73	52	26	0.1296	0.567	0.015
seedBS_th	0	0	0	0	0	500	1.1259	0.981	0.031
seedBS_ic	21	8	31	29	35	376	0.3683	0.808	0.014

Discussion Introducing ARMA(1,1) dependence with $\alpha = 0.5$ and $\beta = 0.4$ strengthens both short and long-range temporal correlation, making change-point detection considerably more challenging. Methods such as `wbsth`, `IsolateDetect_th`, `wbs2`,`dais`, and `seedBS_th` consistently overestimate the number of change-points, as shown by the occurrences of $\hat{K} - K \geq 3$ and their higher MSE values. `wbssbic`, `IsolateDetect_sic`, `seedBS-ic`, and `notic` exhibit moderate estimation error. Meanwhile, `pelt` continues to stand out under correlated settings, returning the lowest DH and MSE values, and producing $\hat{K} - K$ errors distributed over $\{-2, -1, 0, 1, 2, 3\}$. Overall, ARMA(1,1) dependence makes change-point detection substantially harder, degrading both accuracy and localisation across most methods. Most procedures display high DH values (close to 1), indicating difficulty in identifying the true change-point locations.

4.2.3 MA(1) Noise

In this subsection, we compare the effect of short–run and long–run dependence on change–point detection performance by analysing MA(1) and AR(1) noise processes with $\beta = 0.4$ and $\alpha = 0.4$, respectively. In an MA(1) model, the autocorrelation structure is limited strictly to lag 1, meaning the dependence is short–term and decays immediately. Beyond the first lag, the noise behaves almost as if it were i.i.d., resulting in minimal spread of noise through time. Consequently, change–points remain relatively sharp, with reduced blurring around true jump locations, which typically yields lower MSE and DH values. This behaviour is visible in the results, where most methods demonstrate accurate localisation and stable performance under MA(1) noise.

In contrast, AR(1) models exhibit long–run dependence, as the correlation propagates recursively across time steps. Even moderate AR coefficients (e.g., $\alpha = 0.4$) induce persistent temporal structure, making change–points harder to localise. Thus, for equal parameter magnitudes, AR(1) processes create a more challenging environment for change–point detection than MA(1), highlighting the stronger disruptive effect of persistent dependence.

Table 15: Main Signal: Results under AR(1) noise ($a = 0.4$).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	0	0	0	500	0.2502	0.937	0.030
wbssbic	1	0	225	51	101	122	0.0591	0.375	0.030
notic	2	1	251	49	76	121	0.0585	0.331	0.041
IsolateDetect_sic	1	0	169	34	93	203	0.0777	0.447	0.014
IsolateDetect_th	0	0	0	0	0	500	0.2471	0.917	0.004
wbs2	0	0	0	0	0	500	0.4151	0.961	0.000
mosum	2	22	117	129	108	122	0.0507	0.508	0.001
dais	0	0	0	0	0	500	0.2339	0.917	0.000
pelt	70	6	409	14	1	0	0.0365	0.225	0.008
seedBS_th	0	0	0	0	0	500	0.1986	0.932	0.022
seedBS_ic	20	8	297	62	43	70	0.0470	0.304	0.010

Table 16: Main Signal: Results under MA(1) noise ($b = 0.4$).

Method	-2	-1	0	1	2	≥ 3	**MSE**	**DH**	**time**
wbsth	0	0	0	0	0	500	0.1435	0.899	0.041
wbssbic	3	1	394	29	45	28	0.0271	0.137	0.041
notic	2	1	396	36	47	18	0.0266	0.121	0.053
IsolateDetect_sic	4	1	345	58	55	37	0.0326	0.168	0.013
IsolateDetect_th	0	0	0	1	8	491	0.1337	0.842	0.009
wbs2	0	0	0	0	3	497	0.2678	0.923	0.001
mosum	2	36	249	134	56	23	0.0317	0.288	0.001
dais	0	0	1	3	13	483	0.1221	0.821	0.000
pelt	59	0	439	2	0	0	0.0254	0.157	0.016
seedBS_th	0	0	0	1	1	498	0.1065	0.889	0.032
seedBS_ic	11	5	407	40	27	10	0.0260	0.134	0.015

Discussion Under MA(1) noise with $\beta = 0.4$, dependence is limited to the first lag, after which the noise behaves almost independently. This short–range dependence preserves the sharpness of true change-points and therefore supports accurate estimation. This behaviour is reflected in the tables, where most methods, specifically the isolation-based methods, seem to almost correctly finding the true change-points($\texttt{dnc} = \{0, 1\}$), accompanied by low MSE values and small Hausdorff distances. In particular, `wbssbic`,`notic`, `IsolateDetect_sic`, and `seedBS_ic` achieve highly concentrated mass at $dnc = 0$, alongside very low reconstruction error and localization error, confirming strong performance under mild dependence. `pelt` also performs exceptionally well, showing very good recovery in both count accuracy and location. In contrast, `wbsth`,`IsolateDetect_th`, `wbs2`, `dais`, and `seedBS_th` exhibit heavy overestimation, with most mass assigned to the ≥ 3, indicating limited robustness even in the presence of only short-term autocorrelation.

Comparing this to the previously studied AR(1) case, it is evident that MA(1) noise produces far more favourable conditions for change-point detection. AR(1) dependence accumulates over time, leading to increased MSE, higher Hausdorff distances, and greater spread in the count of change-points. Importantly, `seedBS_ic` and `notic` remain among the most reliable methods in both regimes, with `pelt` excelling particularly under MA(1) and AR(1). Overall, the results confirm that short–term dependence has a limited impact on modern estimators, whereas persistent autocorrelation, as in the AR(1) setting, significantly increases detection difficulty and degrades both count accuracy and localization precision.

5 Real Data Example

We examine the performance of five change-point detection methods that proved to be the most robust under deviations from Gaussianity and Dependence, `pelt`, `wbssic`, `IsolateDetect_sic`, `seedBS_ic`, and `notic`. We assess how well they perform on daily EUR/GBP exchange rate data covering the period from 29 February to 01 August 2016. The data were obtained from the European Central Bank's official reference exchange rate dataset and represent daily spot rates of the British Pound (GBP) against the Euro (EUR). This period is of particular interest due to the significant economic and political events surrounding the United Kingdom's decision to leave the European Union (Brexit), which led to pronounced fluctuations in the exchange rate. The objective of this analysis is to evaluate how effectively, accurately, and efficiently the considered change-point detection methods identify changes in the mean of the series, with special focus on detecting the major shift in March–June 2016 when the referendum outcome caused a sharp depreciation of the British Pound. This real data example complements the simulation results by providing empirical evidence of the methods' robustness in a highly volatile financial setting.

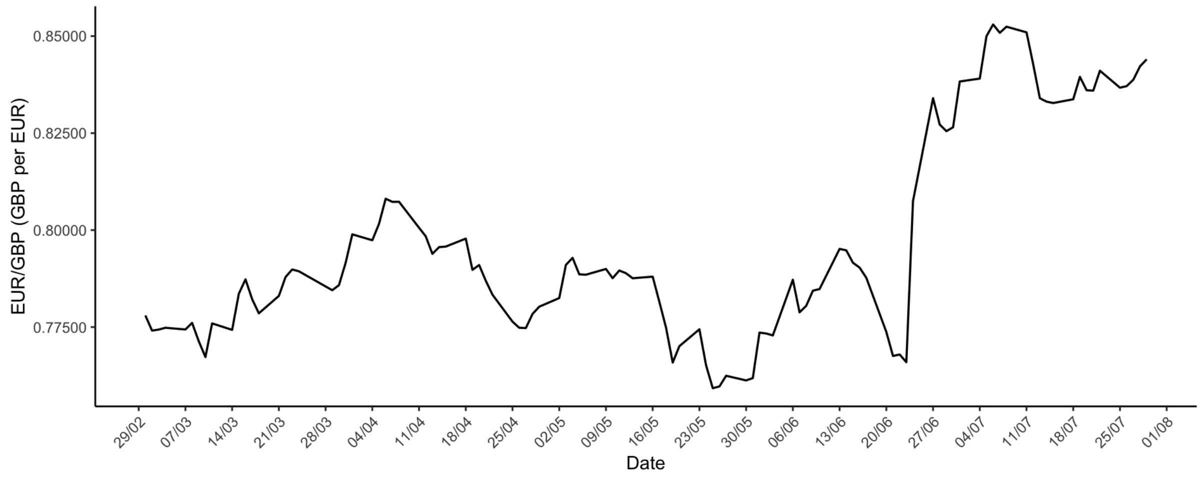

Figure 1: EUR/GBP daily spot rate from 29 February to 01 August 2016, highlighting the sharp appreciation of the euro against the British pound.

WBS (Information Criterion)

The `wbsbic` method identifies seventeen change-points in the mean of the EUR/GBP exchange rate during the March–July 2016 period. These detections include several fluctuations throughout April, May, and early June, as well as the major structural movements associated with the Brexit referendum. Importantly, the method successfully captures the key breakpoint on **23/06/2016** at **0.76595**, marking the sharp depreciation of the Pound on the referendum day. It also flags **24/06/2016** at **0.80750**, which reflects the immediate continuation of the same shock rather than a distinct structural change. The subsequent important change-point on **30/06/2016** at **0.82650** corresponds to the point at which the exchange rate reached its post-referendum peak and stabilised at a higher level. Overall, these results indicate that `wbsbic` can detect the key economic shifts in this period, while also introducing some additional short-term changes driven by the volatility surrounding the Brexit event.

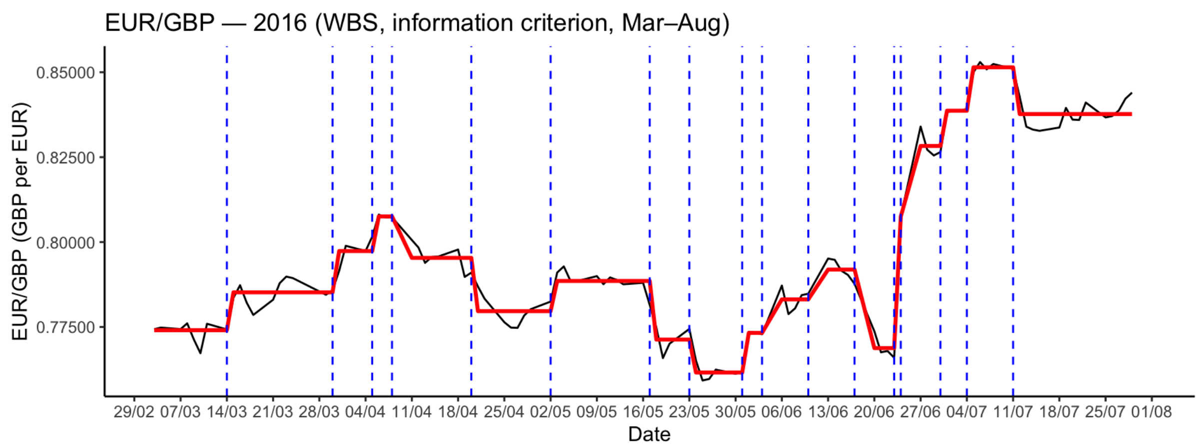

Figure 2: EUR/GBP daily spot rate from March to July 2016, with change-points estimated using the Wild Binary Segmentation method.

Narrowest Over Threshold (Information Criterion)

The `notic` method identifies twelve change-points in the EUR/GBP series during March–July 2016, showing less over-segmentation than `wbsbic`. The two important structural changes detected by `notic` occur on **23/06/2016** at **0.76595** and on **30/06/2016** at **0.82650**. Both of these dates correspond to the major Brexit-related shifts in the exchange rate, and importantly, they coincide with the key change-points also identified by `wbsbic`. The first reflects the sharp depreciation of the Pound on the referendum day, while the second captures the post-referendum peak in early summer. Overall, `notic` successfully recovers the main structural breaks while producing fewer additional detections than `wbsbic`, resulting in a clearer and more concise segmentation of the series.

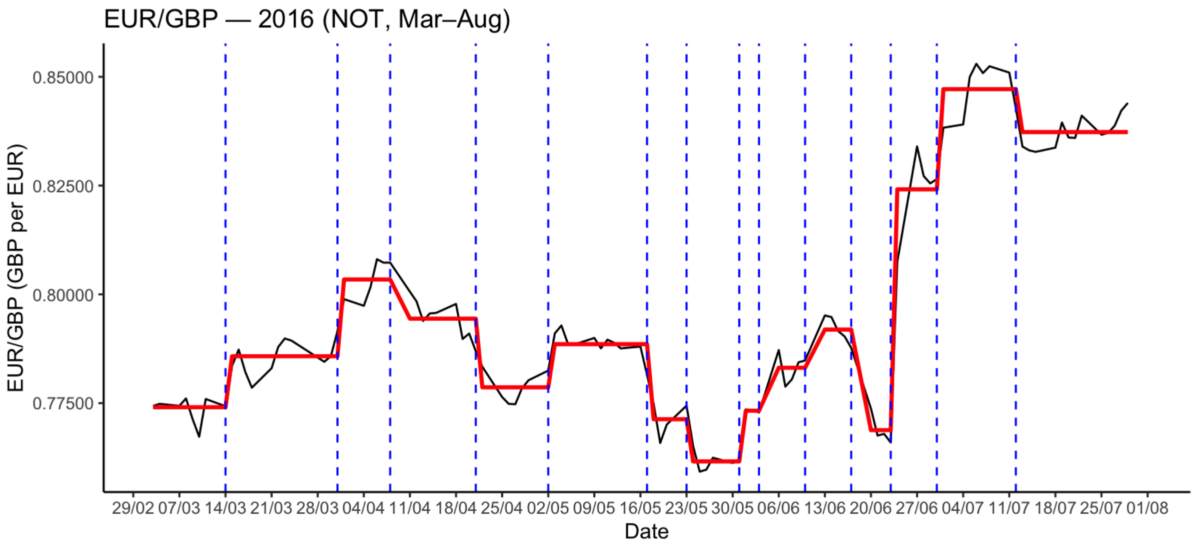

Figure 3: EUR/GBP daily spot rate from March to July 2016, with change-points estimated using the Narrowest Over Threshold method.

Linearly Penalized Segmentation

The `pelt` method identifies only a single change-point in the EUR/GBP exchange rate over the March–July 2016 period. This change occurs on **24/06/2016** with EUR/GBP value **0.80750**, corresponding to the immediate and dramatic market reaction following the Brexit referendum result. Unlike other procedures that detect several additional adjustments before and after this event, `pelt` isolates only the dominant structural break, effectively capturing the major shift in mean level but overlooking the smaller fluctuations and subsequent stabilisation movements. This behaviour reflects `pelt` 's more conservative segmentation under this noise structure.

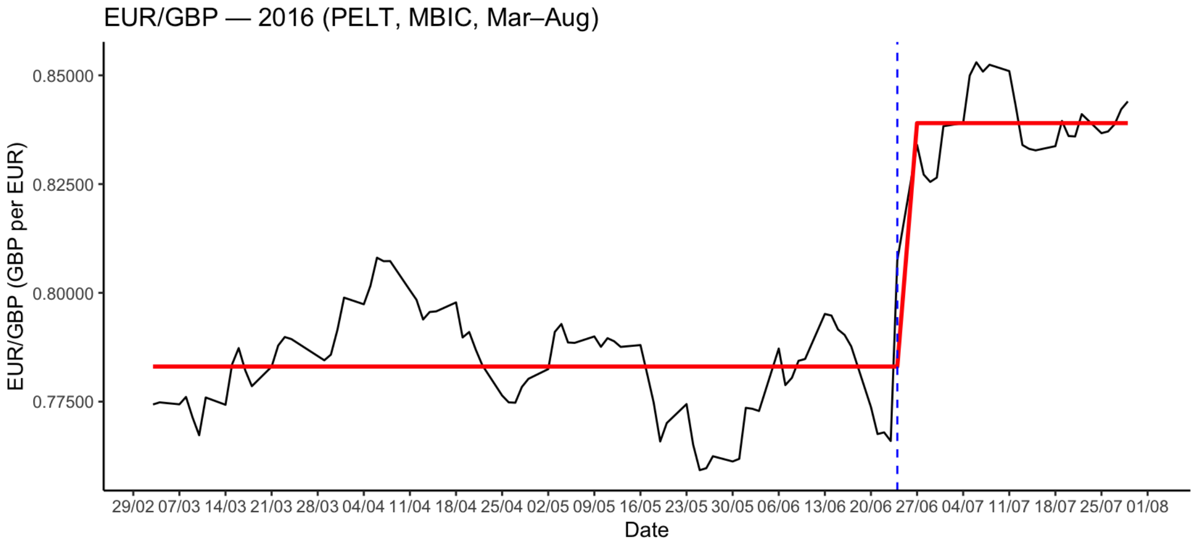

Figure 4: EUR/GBP daily spot rate from March to July 2016, with change-points estimated using the Linearly Penalized Segmentation method.

Seeded BS (Information Criterion)

The `seedBS_ic` method identifies 22 change-points, showing strong over-segmentation of the EUR/GBP series. Even so, it does detect the two key structural breaks of interest: **23/06/2016** at **0.76595**, corresponding to the Brexit referendum shock, and **30/06/2016** at **0.82650**, marking the post-referendum peak. However, although it correctly locates the first major change, `seedBS_ic` also detects **24/06/2016** as an additional change-point, a behaviour consistent with `wbsbic`, despite this not being a true structural break but simply part of the same immediate market reaction. Overall, while the method captures the important movements, its tendency to detect many extra and unnecessary change-points reduces the clarity of its segmentation.

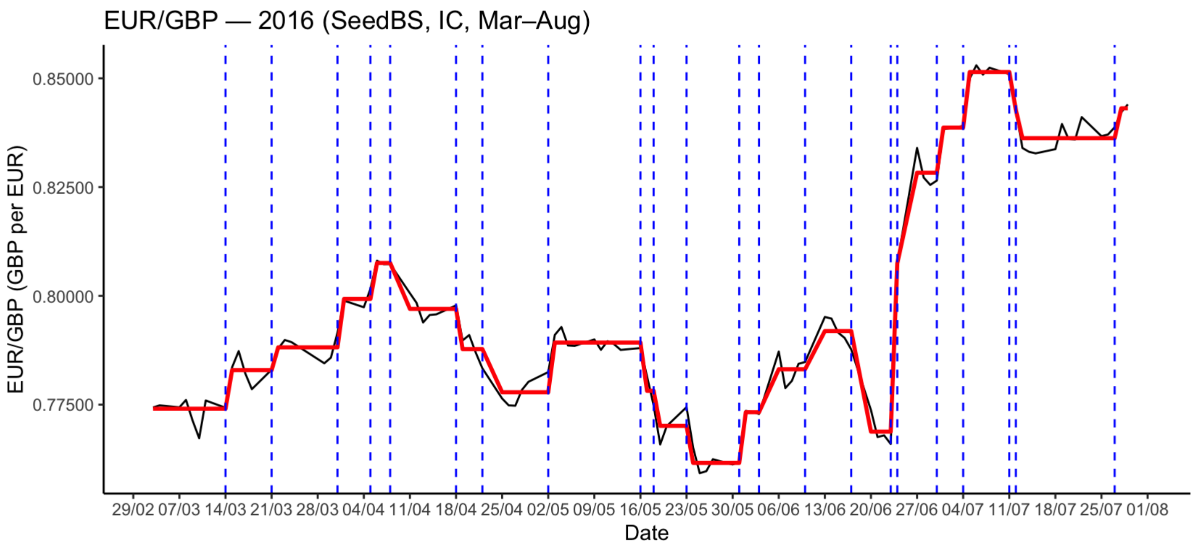

Figure 5: EUR/GBP daily spot rate from March to July 2016, with change-points estimated using the Seeded Binary Segmentation method, with the information criterion.

IsolateDetect (Information Criterion)

The `IsolateDetect_sic` method identifies 14 change-points in the EUR/GBP series over the March–July 2016 period, producing a moderate level of segmentation compared to the more aggressive procedures such as `SeedBS_ic`. Among these detections, the two most important structural changes occur on **23/06/2016** at **0.76595** and **30/06/2016** at **0.82650**, matching the key historical events previously discussed. These detections show that `IsolateDetect_sic` accurately recovers the principal movements in the exchange rate while avoiding excessive over-segmentation. Overall, the method provides a balanced and interpretable segmentation, capturing the essential structural breaks without introducing an unnecessary number of change-points.

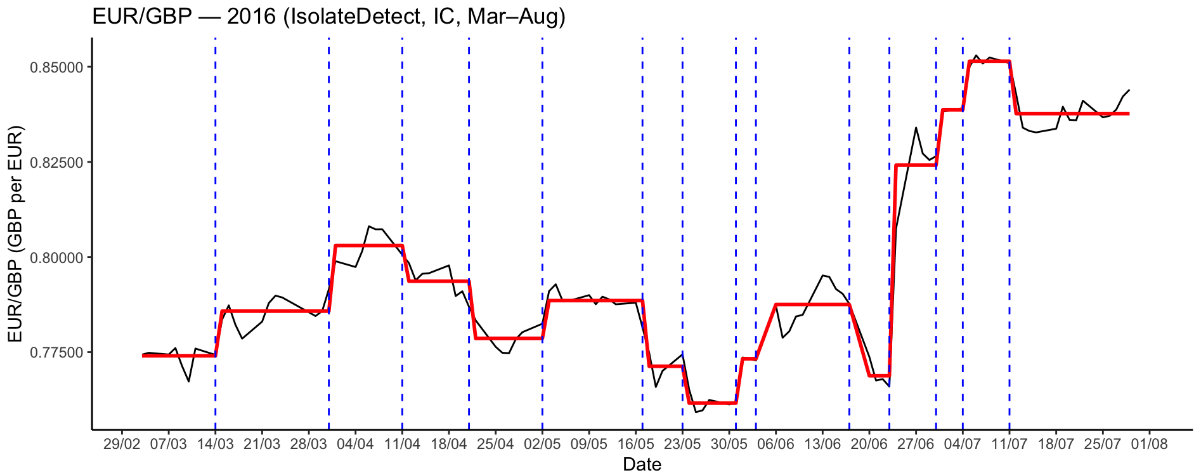

Figure 6: EUR/GBP daily spot rate from March to July 2016, with change-points estimated using the Isolate Detect method.

6 Discussion and Conclusions

Under Gaussian i.i.d. noise, all three signals reveal a clear ranking in method robustness, with `pelt`, `wbsbic`, `notic`, `IsolateDetect_sic`, `IsolateDetect_th`, `wbs2`, and `seedBS_ic` consistently forming the most reliable group. For large and well-separated change-points (Signal 1), most procedures concentrate their estimation error $\widehat{K} - K$ at 0, with `pelt` achieving perfect detection and the lowest `MSE`, and the remaining top-performing methods showing only negligible right tails and small localisation error. When the signal becomes weaker with short and low-amplitude changes (Signal 2), almost all procedures substantially under-estimate the number of change-points, with distributions heavily concentrated at -2. Only a small subset (`wbsth`, `IsolateDetect_th`, `seedBS_th`, `dais`, `wbs2`) displays a slight shift towards 0, though the correct number of change-points is still rarely recovered. For the high-frequency scenario with 14 closely spaced change-points (Signal 3), several methods (`IsolateDetect`, `notic`, `wbs2`, `wbsth`, `dais`) regain stability and concentrate nearly all mass at $\widehat{K} - K = 0$, whereas others (`mosum`, `wbsbic`, `seedBS`) display non-negligible over-segmentation. `pelt` performs poorly in this dense-change setting, with heavy under-estimation and the worst `MSE` and `DH` values.

Across the non-Gaussian i.i.d. noise settings, where all distributions are normalized to have unit variance, method performance varies substantially, with skewness and tail-heaviness emerging as the dominant drivers of false detections rather than scale effects. Under positively skewed noise such as Gamma and Chi-square distributions, `wbsth`, `IsolateDetect_th`, `wbs2`, `dais`, `seedBS_th` consistently over-segment, frequently placing most of their mass in the "≥ 3" category and exhibiting elevated `MSE` values, despite the common variance. In contrast, methods such as `notic` and `mosum` demonstrate greater resilience to skewness, with a large proportion of realizations concentrated at $\widehat{K} - K = 0$, while `pelt`, `wbsbic`, `seedBS_ic`, and `IsolateDetect_sic` show intermediate behavior, displaying reduced but still noticeable over-segmentation under asymmetric noise. For heavy-tailed Student-t noise, performance improves progressively as the degrees of freedom increase from $df = 3$ to $df = 15$. At low degrees of freedom, heavy tails induce substantial over-segmentation and localisation errors, particularly for threshold-based methods, whereas IC-based approaches and `pelt` become increasingly stable as the distribution approaches normality under fixed variance. In contrast, centered Uniform noise provides a symmetric, light-tailed setting in which nearly all methods accurately recover the true number of change-points, achieving low `MSE` and `DH` values. Overall, these results indicate that even under variance normalization, non-Gaussian noise, especially when skewed or heavy-tailed, poses significant challenges for change-point detection, with IC-based methods, `mosum`, and `notic` exhibiting the most consistent robustness across these difficult scenarios.

When the noise exhibits serial dependence rather than being i.i.d., the introduction of temporal dependence has a substantial impact on change-point recovery, with performance degrading as correlation becomes stronger or more persistent. Under mild AR(1) dependence ($\alpha = 0.1$), nearly all procedures perform well, with low `MSE` and `DH` values and strong concentration of $\widehat{K} - K$ at zero. The IC-based methods (`wbssbic`, `seedBS_ic`, `notic`,`IsolateDetect_sic`), and `pelt` remain the most stable and accurate, while threshold-based procedures (`wbsth`, `seedBS_th`, `IsolateDetect_th`) already display noticeable over-segmentation. As autocorrelation increases ($\alpha = 0.3$ and $\alpha = 0.5$), dependence causes persistent noise patterns that lead to inflated estimation error, with `wbsth`, `wbs2`, `dais`, `IsolateDetect_th`, `mosum`, and `seedBS_th` deteriorating rapidly. In contrast, `wbssbic`, `seedBS_ic`, `notic`, `pelt`, and `IsolateDetect_sic` retain moderate robustness, although accuracy naturally declines. When dependence becomes extreme ($\alpha = 0.80$), all procedures suffer substantial loss of precision, with very large `MSE` and `DH` values, though `pelt` remains the only method that preserves some meaningful localisation accuracy. Similar behaviour appears under ARMA(1,1) dependence, where both short- and long-range correlation strongly degrade performance: threshold-based methods heavily over-segment, while IC-based procedures and `pelt` retain comparatively better, though still imperfect accuracy. In contrast, MA(1) noise introduces only short-range dependence, producing much more favourable conditions: most procedures, especially `wbssbic`, `notic`, `IsolateDetect_sic`, `seedBS_ic`, and `pelt`, achieve highly accurate recovery with low `MSE` and `DH`. Threshold-based methods, however, again consistently overestimate even under this mild dependence. Overall, persistent correlation (as in AR and ARMA models) significantly increases the difficulty of the change-point problem, while short-range dependence (MA) has comparatively little effect; across all correlated noise structures, `pelt` emerges as the most robust method, followed closely by the IC-based procedures.

References

A. Anastasiou and P. Fryzlewicz (2022). Detecting multiple generalized change-points by isolating single ones. *Metrika*, 84(2):141–174.

A. Anastasiou and S. Loizidou (2025). Data-adaptive structural change-point detection via isolation. *Statistics and Computing*, 35(5):117.

A.B. Olshen, E. S. Venkatraman, R. Lucito, and M. Wigler (2004). Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics*, 5(4):557–572.

B. Eichinger and C. Kirch (2018). A mosum procedure for the estimation of multiple random change points. *Bernoulli*, 24:526–564.

C.-J Kim, J. C. Morley, and C.R. Nelson (2005). The structural break in the equity premium. *Journal of Business & Economic Statistics*, 23(2):181–191.

J. Reeves, J. Chen, X.L. Wang, R. Lund, and Q.Q. Lu (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46:900–915.

M. Londschien, S. Kovács, and P. Bühlmann (2019). Change point detection for graphical models in presence of missing values. *arXiv preprint arXiv:1907.05409*.

P. Fryzlewicz (2014). Wild binary segmentation for multiple change-point detection. *Annals of Statistics*, 42:2243–2281.

P. Fryzlewicz (2020). Detecting possibly frequent change-points: Wild binary segmentation 2 and steepest-drop model selection. *Journal of the Korean Statistical Society*, 49:1027–1070.

R. Baranowski, Y. Chen, and P. Fryzlewicz (2019). Narrowest-over-threshold detection of multiple change-points and change-point-like features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81:649–672.

R. Killick, P. Fearnhead, and I. A. Eckley. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107:1590–1598.

S. Kovács, P. Bühlmann, H. Li, and A. Munk (2020). Seeded binary segmentation: A general methodology for fast and optimal change-point detection. *Biometrika*, 110(1):249–256.

T. Hotz, O.M. Schütte, H. Sieling, T. Polupanow, U. Diederichsen, C. Steinem, and A. Munk (2013). Idealizing ion channel recordings by a jump segmentation multiresolution filter. *IEEE Transactions on NanoBioscience*, 12(4):376–386.