Supplementary Material: A Comparison of Single and Multiple Changepoint Techniques for Time Series Data

Shi, X., Gallagher, C., Lund, R. and Killick, R.

1 Supplement for the Single Changepoint Section

1.1 The Asymptotic Distribution of a CUSUM Test

Under H_0, the asymptotic distribution of the CUSUM statistic follows the probability law

$$\mathbb{P}\left[\sup_{t\in[0,1]} |B(t)| > x\right] = 2\sum_{n=1}^{\infty}(-1)^{n+1}e^{-2n^2x^2}, \quad x > 0. \tag{1}$$

Critical values for CUSUM statistics can be obtained from (1); listed here are some common percentiles.

Table 1: Critical CUSUM Values

Percentile	Critical Value
90.0%	1.224
95.0%	1.358
97.5%	1.480
99.0%	1.628

1.2 Cropped CUSUM Statistics

CUSUM tests have relatively poor detection power when the changepoint occurs near the boundaries (times 1 or N). Conversely, false detection is more likely to be signaled at boundary locations. This is expected since few observations lie between the changepoint and the boundary and estimation of a segment mean may be less precise. Mathematically, [2] address this problem by applying a weight function $w(\cdot)$, denoted by $w(t)$ at time $t = k/N$.

For example, with $w(t) = \sqrt{t(1-t)}$ and

$$\lambda_X(k) = \frac{\text{CUSUM}_X^2(k)}{\frac{k}{N}(1-\frac{k}{N})} \text{ and } \lambda_Z(k) = \frac{\text{CUSUM}_Z^2(k)}{\frac{k}{N}(1-\frac{k}{N})}. \tag{2}$$

Theorem 4 from [3], states the following.

Theorem 1. *Given $0 < \ell < h < 1$ and suppose that $\hat{\sigma}^2$ and $\hat{\eta}^2$ are $\sqrt{N}$-consistent estimates of σ^2 and η^2 respectively. Under H_0,*

$$\frac{1}{\hat{\eta}^2} \max_{\ell \leq k/N \leq h} \lambda_X(k) \xrightarrow{\mathcal{D}} \sup_{\ell < t < h} \frac{B^2(t)}{t(1-t)}, \tag{3}$$

and

$$\frac{1}{\hat{\sigma}^2} \max_{\ell \leq k/N \leq h} \lambda_z(k) \xrightarrow{\mathcal{D}} \sup_{\ell < t < h} \frac{B^2(t)}{t(1-t)}. \tag{4}$$

One can approximate p-values for cropped CUSUM tests via

$$\mathbb{P}\left[\sup_{\ell \leq t \leq h} \frac{B^2(t)}{t(1-t)} > x\right] \approx \sqrt{\frac{xe^{-x}}{2\pi}} \left[\left(1 - \frac{1}{x}\right) \log\left(\frac{(1-\ell)h}{\ell(1-h)}\right) + \frac{4}{x}\right]. \tag{5}$$

1.3 Cropped LRT Tests

Similar to the cropped CUSUM tets, cropped LRTs simply truncate admissible times at the boundaries; for example,

$$U_{\text{crop}} = \max_{\ell \leq k/N \leq h} (-2\log(\Lambda_k)). \tag{6}$$

Based on [3], the following theorem holds:

Theorem 2. *Under H_0, the cropped LRT statistic obeys*

$$U_{\text{crop}} = \max_{\ell \leq k/N \leq h} (-2\log(\Lambda_k)) \xrightarrow{\mathcal{D}} \sup_{\ell \leq t \leq h} \frac{B^2(t)}{t(1-t)}. \tag{7}$$

Connections exist between U_{crop} and $\lambda_X(k)$:

$$\max_{\ell \leq k/N \leq h} (-2\log \Lambda_k) - \frac{1}{\hat{\eta}_2^2} \max_{\ell \leq k/N \leq h} (\lambda_X(k)) = o_p(1). \tag{8}$$

This identifies U_{crop}'s asymptotic null hypothesis distribution. Further, [3] shows that the cropped $-2\log(\Lambda_k)$ is related to the CUSUM $\lambda_Z(k)$ statistic through

$$\max_{\ell \leq k/N \leq h} (-2\log(\Lambda_k)) - \frac{1}{\hat{\sigma}^2} \max_{\ell \leq k/N \leq h} \lambda_Z(k) = o_p(1). \tag{9}$$

Thus, if $\ell \searrow 0$ and $h \nearrow 1$, $\text{CUSUM}_Z(k)$ and LRTs are linked by

$$T = \frac{1}{\hat{\sigma}^2} \max_{\ell \leq k/N \leq h} \lambda_Z(k), \tag{10}$$

$$W_T = \sqrt{2T \log\log(N)} - \left[2\log\log(N) + \frac{1}{2}\log\log\log(N) - \frac{1}{2}\log\pi\right]. \tag{11}$$

As $N \to \infty$, W_T converges to the Gumbel distribution. There is no need to crop boundaries here as extreme value scalings allow all admissible changepoint times to be considered.

1.4 The Sum of Squared CUSUM Test (SCUSUM_Z)

Note that CUSUM statistic of the one-step-ahead prediction residuals is

$$\text{CUSUM}_Z(k) = \frac{k}{N}\left(1 - \frac{k}{N}\right)\sqrt{N} \cdot \bar{Z}_k - \frac{k}{N}\left(1 - \frac{k}{N}\right)\sqrt{N} \cdot \bar{Z}_k^*, \tag{12}$$

where $\bar{Z}_k = \frac{1}{k}\sum_{t=1}^{k} \hat{Z}_t$ and $\bar{Z}_k^* = \frac{1}{N-k}\sum_{t=k+1}^{N} \hat{Z}_t$. Under the null hypothesis of no changepoints, the central limit theorem provides

$$\frac{\text{CUSUM}_Z(k)}{\hat{\sigma}\sqrt{k/N(1-k/N)}} \xrightarrow{\mathcal{D}} \mathbf{N}(0,1). \tag{13}$$

This holds for IID data and asymptotically for one-step-ahead prediction residuals when σ^2 and all ARMA parameters are estimated in a $\sqrt{N}$-consistent manner.

Let $t = k/N$. By the functional central limit theorem (See Section 8 of [1]), the process-based convergence

$$\left\{\frac{\text{CUSUM}_Z(k/N)}{\hat{\sigma}}\right\} \xrightarrow{\mathcal{D}} \{B(t)\}_{t=0}^{t=1} \tag{14}$$

can be shown to hold weakly, where $\{B(t)\}_{t=0}^{t=1}$ is a Brownian bridge. On the sample paths of Brownian bridges, sum of squared paths converge to integrals of squared paths; hence, application of the continuous mapping theorem provides

$$\text{SCUSUM}_Z = \frac{1}{N}\sum_{k=1}^{N}\left[\frac{\text{CUSUM}_Z(k/N)}{\hat{\sigma}}\right]^2 \xrightarrow{\mathcal{D}} \int_0^1 B^2(t)\, dt. \tag{15}$$

The distribution of $\int_0^1 B(t)^2 dt$ was thoroughly investigated in [4]. A simulation was conducted to obtain critical values; these are reported below for convenience.

Table 2: Critical Values for Sum of Squared CUSUM Statistics

Percentile	Critical Value
90.0%	0.3473046
95.0%	0.4613744
97.5%	0.5806168
99.0%	0.7434348

2 Single Changepoint Simulations: Type 1 Error and Power Tables

2.1 Type 1 Errors

An acronym list for our tests is as follows:

- CUSUM_X: CUSUM applied to the original series X_t
- CUSUM_Z: CUSUM applied to the one-step-ahead predicted residuals Z_t
- λ_X: Cropped CUSUM applied to the original series X_t
- λ_Z: Cropped CUSUM applied to the one-step-ahead predicted residuals Z_t
- SCUSUM_Z: Sum of squared CUSUM applied to the one-step-ahead predicted residuals Z_t
- $\text{LRT}(W_U)$: Likelihood ratio test defined in Theorem 5
- LRT(Crop): Likelihood ratio test with boundaries cropped
- $\text{LRT}(\text{CUSUM}_Z)$: Likelihood ratio test linked by CUSUM_Z

Simulation results for AR(1) series with varying autocorrelation levels ϕ and varying lengths N. The series has no changepoints.

Table 3: Empirical type I errors for AR(1) series with varying ϕ but no changepoints and $\sigma^2 = 1$. Bold entries are those that are inflated.

Test / ϕ	CUSUMx	λ_X	CUSUMz	λ_Z	SCUSUMz	$LRT(CUSUM)$	LRT(W_U)	$LRT(Crop)$
$\phi = .9, N = 100$	0	0	0.0142	0.0341	0.046	**0.1053**	**0.0717**	**0.2261**
$(N = 500)$	(0.0084)	(0.0021)	(0.0279)	(0.0281)	(0.0409)	**(0.1143)**	**(0.0759)**	**(0.2898)**
$[N = 1000]$	[0.0179]	[0.0092]	[0.0336]	[0.0298]	[0.0448]	**[0.1193]**	**[0.0749]**	**[0.2755]**
$\{N = 2500\}$	{0.0308}	{0.0247}	{0.0435}	{0.0421}	{0.0512}	**{0.1216}**	**{0.0795}**	**{0.2369}**
$\phi = .8$	0.0007	0	0.0112	0.0152	0.0352	0.0281	0.0538	**0.1824**
	(0.0192)	(0.0096)	(0.032)	(0.029)	(0.0431)	(0.0344)	(0.0421)	**(0.1546)**
	[0.0313]	[0.0213]	[0.0423]	[0.0379]	[0.0479]	[0.039]	[0.0382]	**[0.1407]**
	{0.0367}	{0.0321}	{0.0431}	{0.042}	{0.0456}	{0.0392}	{0.0349}	**{0.1088}**
$\phi = .7$	0.0035	0.0003	0.0144	0.0148	0.0373	0.0107	0.0381	**0.1344**
	(0.0255)	(0.016)	(0.0387)	(0.0312)	(0.0468)	(0.0164)	(0.0283)	**(0.101)**
	[0.0316]	[0.0252]	[0.0398]	[0.0377]	[0.0467]	[0.0221]	[0.0274]	**[0.0901]**
	{0.043}	{0.0391}	{0.0478}	{0.0478}	{0.0477}	{0.022}	{0.0274}	**{0.0791}**
$\phi = .6$	0.0081	0.0018	0.0195	0.0119	0.0388	0.0041	0.0261	**0.1017**
	(0.0281)	(0.02)	(0.0353)	(0.0322)	(0.0459)	(0.0101)	(0.0204)	**(0.0744)**
	[0.0369]	[0.0312]	[0.0437]	[0.0402]	[0.0476]	[0.0125]	[0.0209]	**[0.0717]**
	{0.0445}	{0.0401}	{0.0497}	{0.0474}	{0.0519}	{0.0172}	{0.0235}	**{0.069}**
$\phi = .5$	0.0101	0.0045	0.0206	0.0128	0.0389	0.0038	0.0237	**0.0829**
	(0.0334)	(0.0285)	(0.0414)	(0.0387)	(0.0487)	(0.0091)	(0.0181)	**(0.0671)**
	[0.0341]	[0.0308]	[0.0388]	[0.038]	[0.0445]	[0.0104]	[0.0156]	**[0.0572]**
	{0.0426}	{0.0424}	{0.0455}	{0.0491}	{0.0489}	{0.0137}	{0.0171}	**{0.0643}**
$\phi = .4$	0.0148	0.0055	0.0209	0.0158	0.03407	0.0021	0.0185	**0.0667**
	(0.0314)	(0.0247)	(0.0364)	(0.0309)	(0.0437)	(0.0084)	(0.0139)	(0.0515)
	[0.0417]	[0.0381]	[0.0457]	[0.0423]	[0.051]	[0.0099]	[0.0155]	**[0.0558]**
	{0.0468}	{0.0426}	{0.0499}	{0.047}	{0.0533}	{0.014}	{0.0183}	**{0.0571}**
$\phi = .3$	0.018	0.0084	0.0244	0.0149	0.041	0.0025	0.0138	**0.0597**
	(0.0392)	(0.0335)	(0.0425)	(0.0392)	(0.0529)	(0.0103)	(0.0144)	(0.0542)
	[0.0466]	[0.0366]	[0.0486]	[0.0405]	[0.0518]	[0.0117]	[0.0170]	[0.0511]
	{0.0436}	{0.0433}	{0.0455}	{0.0464}	{0.0482}	{0.0144}	{0.0164}	{0.0530}
$\phi = .2$	0.0277	0.0117	0.0268	0.0177	0.0424	0.0024	0.0127	0.0507
	(0.0383)	(0.0343)	(0.0402)	(0.0378)	(0.049)	(0.0118)	(0.0153)	(0.0508)
	[0.0401]	[0.0396]	[0.0418]	[0.0428]	[0.0487]	[0.0112]	[0.0132]	[0.0495]
	{0.0450}	{0.0425}	{0.0465}	{0.0438}	{0.0470}	{0.0109}	{0.0116}	{0.0489}
$\phi = .1$	0.0218	0.0146	0.0238	0.017	0.0455	0.0029	0.0126	0.0436
	(0.0432)	(0.0364)	(0.0443)	(0.0383)	(0.0533)	(0.0086)	(0.0117)	(0.0465)
	[0.0419]	[0.0407]	[0.0431]	[0.0423]	[0.0449]	[0.013]	[0.0139]	[0.0459]
	{0.0459}	{0.044}	{0.0466}	{0.0449}	{0.0483}	{0.0153}	{0.0155}	{0.0474}
$\phi = -.25$	0.035	0.0321	0.0286	0.0223	0.045	0.0044	0.0102	0.0388
	(0.0407)	(0.0394)	(0.0376)	(0.0348)	(0.044)	(0.0084)	(0.0088)	(0.0386)
	[0.0478]	[0.0443]	[0.0453]	[0.041]	[0.0509]	[0.0117]	[0.0118]	[0.0408]
	{0.0441}	{0.0455}	{0.0423}	{0.0438}	{0.047}	{0.0121}	{0.0118}	{0.0435}
$\phi = -.5$	0.0466	0.0481	0.0306	0.0212	0.0464	0.005	0.0086	0.033
	(0.0519)	(0.0506)	(0.0432)	(0.0386)	(0.0504)	(0.0132)	(0.0118)	(0.0382)
	[0.05]	[0.0498]	[0.0449]	[0.0416]	[0.0498]	[0.011]	[0.0089]	[0.0407]
	{0.0513}	{0.0482}	{0.048}	{0.0445}	{0.0536}	{0.0135}	{0.0108}	{0.0426}
$\phi = -.9$	**0.1372**	**0.2928**	0.0295	0.0337	0.052	**0.1017**	0.007	0.0291
	(0.0845)	**(0.1096)**	(0.044)	(0.0423)	(0.0549)	(0.1228)	(0.0101)	(0.0364)
	[0.0705]	**[0.0856]**	[0.0451]	[0.0417]	[0.051]	**[0.1187]**	[0.0103]	[0.0388]
	{0.0595}	{0.0675}	{0.0452}	{0.0438}	{0.0528}	**{0.1161}**	{0.0093}	{0.0406}
$\phi = -.95$	**0.2698**	0.5276	0.0262	0.0429	0.0484	**0.2385**	0.0047	0.022
	(0.107)	**(0.1941)**	(0.0424)	(0.0469)	(0.0531)	**(0.2597)**	(0.0075)	(0.0359)
	[0.0925]	**[0.1331]**	[0.0485]	[0.0499]	[0.0535]	**[0.2644]**	[0.0107]	[0.0431]
	{0.0682}	**{0.0861}**	{0.0472}	{0.0447}	{0.0486}	**{0.2604}**	{0.0102}	{0.0412}

The following table reports simulation results for MA(1), ARMA(1, 1), and ARMA(2, 2) series. The series is of length $N = 1000$ and has no changepoints.

AR(1) Models

ϕ	CUSUMz	λ_z	SCUSUMz	LRT(W_U)
-0.9	0.0451	0.0417	0.051	0.0103
-0.5	0.0449	0.0416	0.0498	0.0089
-0.25	0.0453	0.041	0.0509	0.0118
0.1	0.0431	0.0423	0.0449	0.0139
0.5	0.0388	0.038	0.0445	0.0156
0.7	0.0398	0.0377	0.0467	0.0274
0.9	0.0336	0.0298	0.0448	0.0749

ARMA(1, 1) Models

ϕ_1	θ_1	CUSUMz	λ_z	SCUSUMz	LRT(W_U)
0.5	-0.95	0.0324	0.1698	0.0609	0.2952
0.5	-0.9	0.0373	0.06	0.0521	0.1232
0.5	-0.1	0.0426	0.0364	0.0482	0.019
0.1	-0.5	0.0424	0.0436	0.0476	0.0145
0.9	-0.5	0.0351	0.0333	0.0475	0.0607
0.95	-0.5	0.0328	0.0261	0.043	0.1281

MA(1) Models

θ	CUSUMz	λ_z	SCUSUMz	LRT(W_U)
-0.95	0.0363	0.2257	0.0752	0.2733
-0.9	0.0393	0.0748	0.0567	0.1109
-0.5	0.0494	0.0466	0.0527	0.0123
0.1	0.0447	0.0435	0.0517	0.0123
0.5	0.0422	0.0438	0.0499	0.0171
0.9	0.0467	0.0439	0.0532	0.1133
0.95	0.0437	0.0433	0.0513	0.2469

ARMA(2, 2) Models

$\{\phi_1, \phi_2\}$	$\{\theta_1, \theta_2\}$	CUSUMz	λ_z	SCUSUMz	LRT(W_U)
$\{0.6, 0.35\}$	$\{0.6, 0.3\}$	0.0309	0.0216	0.042	0.4874
$\{0.6, 0.3\}$	$\{0.5, -0.2\}$	0.0346	0.0272	0.0441	0.4344
$\{0.6, -0.1\}$	$\{-0.6, 0.3\}$	0.043	0.038	0.05	0.3473
$\{0.5, -0.2\}$	$\{-0.45, -0.5\}$	0.024	0.2065	0.0622	0.5768
$\{0.5, -0.2\}$	$\{-0.4, -0.5\}$	0.0371	0.0742	0.0591	0.4224
$\{0.2, -0.5\}$	$\{-0.45, -0.05\}$	0.0435	0.0424	0.0506	0.3433

Table 4: Empirical type I errors for AR(1), MA(1), ARMA$(1, 1)$, and ARMA$(2, 2)$ models with $N = 1000$ and white noise variance $\sigma^2 = 1$. The type I error is $\alpha = 0.05$

Table 5: Powers for an AR(1) series with a single changepoint in the middle of the record. Here, $\sigma^2 = 1$ and the mean shift size is $\Delta = 0.15$.

Test / ϕ	CUSUMx	λ_X	CUSUMz	λ_Z	SCUSUMz	LRT(CUSUM)	LRT(W_U)	LRT(Crop)
$\phi = .9, N = 100$	0	0	0.0151	0.0349	0.0487	0.1047	0.0729	0.2348
$(N = 500)$	(0.0096)	(0.0034)	(0.0318)	(0.027)	(0.0481)	(0.1149)	(0.0795)	(0.291)
$[N = 1000]$	[0.021]	[0.0109]	[0.0397]	[0.0338]	[0.0514]	[0.1129]	[0.0862]	[0.2853]
$\{N = 2500\}$	{0.0371}	{0.0238}	{0.0504}	{0.0417}	{0.0561}	{0.1172}	{0.0755}	{0.2403}
$\phi = .8$	0.0009	0.0001	0.0124	0.0167	0.0375	0.0311	0.0549	0.1809
	(0.0248)	(0.0116)	(0.0427)	(0.0328)	(0.0559)	(0.037)	(0.0478)	(0.1685)
	[0.0391]	[0.027]	[0.0555]	[0.0458]	[0.0673]	[0.0377]	[0.0479]	[0.1591]
	{0.08}	{0.0554}	{0.0927}	{0.0726}	{0.0963}	{0.0496}	{0.055}	{0.1592}
$\phi = .7$	0.0034	0.0003	0.0153	0.0131	0.0378	0.0119	0.0375	0.1271
	(0.039)	(0.0228)	(0.0547)	(0.0394)	(0.0681)	(0.0179)	(0.0354)	(0.1267)
	[0.0677]	[0.0442]	[0.0804]	[0.0596]	[0.0878]	[0.0234]	[0.0379]	[0.1285]
	{0.1443}	{0.0983}	{0.1568}	{0.1104}	{0.1707}	{0.0418}	{0.0609}	{0.1701}
$\phi = .6$	0.0084	0.0012	0.0218	0.0117	0.0439	0.0041	0.0312	0.108
	(0.0548)	(0.0315)	(0.0672)	(0.0472)	(0.0825)	(0.0157)	(0.031)	(0.1062)
	[0.0968]	[0.0643]	[0.1111]	[0.076]	[0.1281]	[0.0243]	[0.0428]	[0.1274]
	{0.2413}	{0.1668}	{0.2528}	{0.1786}	{0.2682}	{0.0597}	{0.0816}	{0.2284}
$\phi = .5$	0.014	0.0041	0.0233	0.0149	0.0462	0.0032	0.0252	0.0872
	(0.0765)	(0.0484)	(0.0879)	(0.0603)	(0.1072)	(0.0161)	(0.0337)	(0.109)
	[0.1479]	[0.0929]	[0.1601]	[0.1042]	[0.1762]	[0.03]	[0.0488]	[0.146]
	{0.3565}	{0.254}	{0.3676}	{0.2642}	{0.385}	{0.0967}	{0.1196}	{0.3061}
$\phi = .4$	0.0189	0.0064	0.0302	0.0155	0.0529	0.0028	0.0214	0.0834
	(0.1035)	(0.0651)	(0.1131)	(0.075)	(0.133)	(0.02)	(0.0368)	(0.1108)
	[0.2087]	[0.1347]	[0.2177]	[0.1463]	[0.2403]	[0.0446]	[0.0627]	[0.1835]
	{0.5003}	{0.3724}	{0.5078}	{0.3817}	{0.5173}	{0.165}	{0.1908}	{0.4216}
$\phi = .3$	0.0284	0.0132	0.0366	0.0221	0.0619	0.0039	0.0216	0.0808
	(0.1416)	(0.0869)	(0.1481)	(0.0952)	(0.1726)	(0.0275)	(0.0442)	(0.1319)
	[0.2813]	[0.1884]	[0.2894]	[0.1973]	[0.3096]	[0.0702]	[0.0931]	[0.234]
	{0.6351}	{0.4962}	{0.6418}	{0.5042}	{0.6468}	{0.2581}	{0.2904}	{0.5354}
$\phi = .2$	0.0315	0.0153	0.0385	0.0213	0.0646	0.0023	0.0198	0.069
	(0.1874)	(0.1178)	(0.1933)	(0.1244)	(0.2149)	(0.0363)	(0.0551)	(0.1588)
	[0.3597]	[0.2503]	[0.365]	[0.2569]	[0.3846]	[0.0982]	[0.1204]	[0.29]
	{0.7573}	{0.6374}	{0.7607}	{0.6409}	{0.7653}	{0.3795}	{0.4058}	{0.6655}
$\phi = .1$	0.0434	0.0249	0.0478	0.0276	0.0743	0.0039	0.02	0.0702
	(0.2324)	(0.1538)	(0.2352)	(0.1565)	(0.2614)	(0.0492)	(0.0687)	(0.1864)
	[0.4419]	[0.3219]	[0.4445]	[0.3242]	[0.4632]	[0.1343]	[0.1575]	[0.3545]
	{0.8574}	{0.7611}	{0.8578}	{0.7621}	{0.8577}	{0.5196}	{0.5468}	{0.7788}
$\phi = -.25$	0.0904	0.0566	0.0768	0.0436	0.1122	0.008	0.023	0.0792
	(0.4366)	(0.3113)	(0.4235)	(0.2964)	(0.4516)	(0.1227)	(0.1419)	(0.3241)
	[0.7493]	[0.6301]	[0.7426]	[0.6198]	[0.7544]	[0.3608]	[0.3876]	[0.6402]
	{0.9903}	{0.975}	{0.9899}	{0.9741}	{0.9887}	{0.9008}	{0.9105}	{0.9766}
$\phi = -.5$	0.1449	0.1026	0.1114	0.0628	0.1499	0.0135	0.0292	0.0942
	(0.6111)	(0.4775)	(0.5858)	(0.447)	(0.6091)	(0.217)	(0.2393)	(0.4654)
	[0.8967]	[0.8202]	[0.8897]	[0.8062]	[0.8902]	[0.5918]	[0.6078]	[0.8137]
	{0.9994}	{0.998}	{0.9994}	{0.9978}	{0.9993}	{0.9859}	{0.9867}	{0.998}
$\phi = -.9$	0.365	0.4273	0.1753	0.115	0.2329	0.1272	0.0441	0.1254
	(0.8563)	(0.7737)	(0.7981)	(0.6824)	(0.8088)	(0.5085)	(0.4626)	(0.6877)
	[0.9885]	[0.972]	[0.9827]	[0.9592]	[0.9811]	[0.8838]	[0.8732]	[0.9605]
	{1}	{1}	{1}	{1}	{1}	{0.9999}	{0.9999}	{1}
$\phi = -.95$	0.4785	0.6186	0.1748	0.1268	0.2368	0.2554	0.0425	0.1248
	(0.8958)	(0.8369)	(0.821)	(0.7123)	(0.8306)	(0.6087)	(0.4875)	(0.7184)
	[0.9922]	[0.9811]	[0.9855]	[0.9649]	[0.9851]	[0.9224]	[0.898]	[0.9673]
	{1}	{1}	{1}	{1}	{1}	{1}	{1}	{1}

Table 6: Powers with a single changepoint in the middle of the record. Here, $\sigma^2 = 1$ and $\Delta = 1$.

Test / ϕ	CUSUMx	λ_X	CUSUMz	λ_Z	SCUSUMz	LRT(CUSUM)	LRT(W_U)	LRT(Crop)
$\phi = .9, N = 100$ $(N = 500)$ $[N = 1000]$ $\{N = 2500\}$	0 (0.022) [0.1164] {0.4578}	0 (0.0039) [0.0444] {0.2917}	0.018 (0.0703) [0.1795] {0.5102}	0.0312 (0.0353) [0.0935] {0.3503}	0.0526 (0.1043) [0.2093] {0.5377}	0.1026 (0.1118) [0.1179] {0.2144}	0.0804 (0.1662) [0.2569] {0.5076}	0.2584 (0.4607) [0.5563] {0.7689}
$\phi = .8$	0.0012 (0.2143) [0.622] {0.9862}	0 (0.0852) [0.4148] {0.9564}	0.0229 (0.297) [0.6762] {0.9884}	0.0133 (0.1467) [0.4817] {0.9637}	0.0644 (0.3598) [0.707] {0.9891}	0.0238 (0.0506) [0.2063] {0.8548}	0.1062 (0.3501) [0.6438] {0.9827}	0.2939 (0.6345) [0.8661] {0.9968}
$\phi = .7$	0.0072 (0.5939) [0.96] {1}	0.0005 (0.3635) [0.8855] {1}	0.0419 (0.6597) [0.969] {1}	0.0162 (0.4395) [0.907] {1}	0.0957 (0.7009) [0.9692] {1}	0.0098 (0.1536) [0.6914] {0.9995}	0.1397 (0.6802) [0.9637] {1}	0.3551 (0.8766) [0.9936] {1}
$\phi = .6$	0.026 (0.8994) [0.9994] {1}	0.0023 (0.7382) [0.9959] {1}	0.0806 (0.9209) [0.9997] {1}	0.0193 (0.781) [0.9966] {1}	0.157 (0.9308) [0.9992] {1}	0.0035 (0.4615) [0.9699] {1}	0.1995 (0.9353) [0.9997] {1}	0.4512 (0.9869) [1] {1}
$\phi = .5$	0.079 (0.9914) [1] {1}	0.0099 (0.9515) [0.9999] {1}	0.1531 (0.9934) [1] {1}	0.035 (0.9626) [0.9999] {1}	0.2607 (0.9913) [1] {1}	0.003 (0.8067) [0.9997] {1}	0.3147 (0.9968) [1] {1}	0.5849 (0.9998) [1] {1}
$\phi = .4$	0.1655 (0.9995) [1] {1}	0.0309 (0.9954) [1] {1}	0.2514 (0.9995) [1] {1}	0.0664 (0.9964) [1] {1}	0.3742 (0.9994) [1] {1}	0.0054 (0.9669) [1] {1}	0.4492 (1) [1] {1}	0.7107 (1) [1] {1}
$\phi = .3$	0.2935 (1) [1] {1}	0.076 (0.9999) [1] {1}	0.3797 (1) [1] {1}	0.1237 (0.9999) [1] {1}	0.5075 (1) [1] {1}	0.0116 (0.9977) [1] {1}	0.6175 (1) [1] {1}	0.8368 (1) [1] {1}
$\phi = .2$	0.4692 (1) [1] {1}	0.1671 (1) [1] {1}	0.5345 (1) [1] {1}	0.2272 (1) [1] {1}	0.6584 (1) [1] {1}	0.0297 (1) [1] {1}	0.7698 (1) [1] {1}	0.9289 (1) [1] {1}
$\phi = .1$	0.6592 (1) [1] {1}	0.3131 (1) [1] {1}	0.6986 (1) [1] {1}	0.3715 (1) [1] {1}	0.7882 (1) [1] {1}	0.0778 (1) [1] {1}	0.8914 (1) [1] {1}	0.9728 (1) [1] {1}
$\phi = -.25$	0.9872 (1) [1] {1}	0.913 (1) [1] {1}	0.9858 (1) [1] {1}	0.9121 (1) [1] {1}	0.9911 (1) [1] {1}	0.6183 (1) [1] {1}	0.9991 (1) [1] {1}	1 (1) [1] {1}
$\phi = -.5$	0.9997 (1) [1] {1}	0.9972 (1) [1] {1}	0.9998 (1) [1] {1}	0.997 (1) [1] {1}	0.9998 (1) [1] {1}	0.9569 (1) [1] {1}	0.9999 (1) [1] {1}	1 (1) [1] {1}
$\phi = -.9$	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}
$\phi = -.95$	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}	1 (1) [1] {1}

2.2 Simulated Power: Fixed and Effective Mean Shift Sizes

In the single changepoint simulation (Figure 2 in the main article), the shift size Δ is fixed at 0.15, thus the effective mean shift size varies with the level of autocorrelation. The power for the effective change size is also investigated in AR(1) series. The supplement includes both results and we compare the two in the two figures below. To do this, we simply keep the variance of $\{X_t\}$ as unity for all levels of correlation. That makes $|\Delta|/\mathrm{Var}(X_t)^{1/2}$, where Δ is the mean shift size, constant across different AR parameters. The conclusions stated in the text remain the same, whether the fixed or effective change size is used.

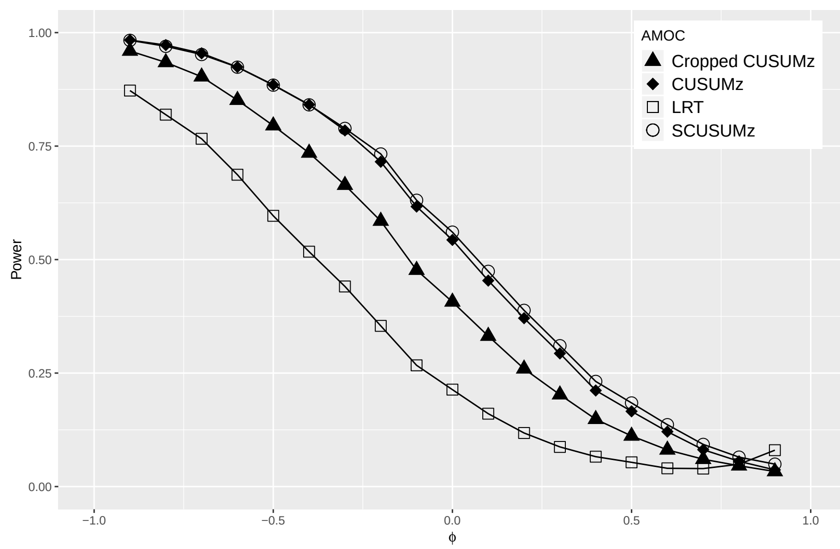

Figure 1: Mean shift size is not adjusted for ϕ. Original Figure 2 in the main article

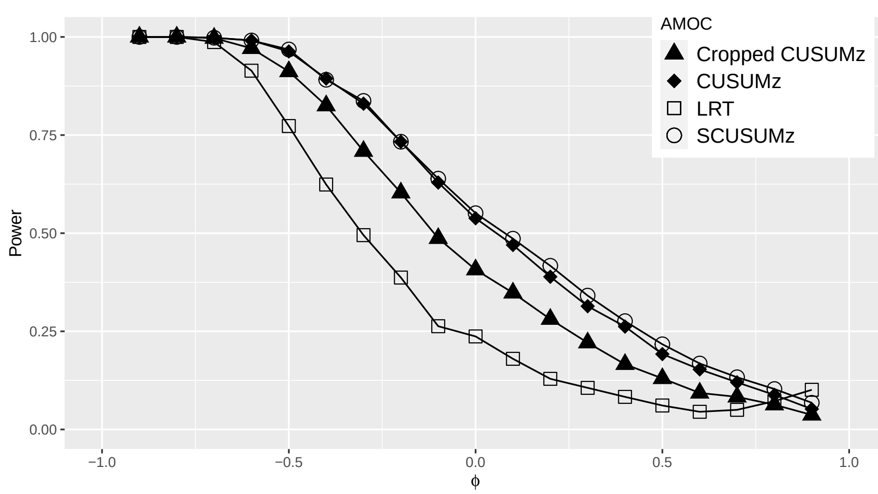

Figure 2: Mean shift size is adjusted according to ϕ.

3 Supplement for the Multiple Changepoint Section

3.1 The distance in comparing multiple changepoint techniques: an example

The distance computation between two changepoint configurations can be formulated as the following optimization problem:

$$\min \sum_{i=1}^{k} \sum_{j=1}^{m} c_{ij} x_{ij} \tag{16}$$

$$\sum_{i=1}^{k} x_{ij} = 1, \text{ for } j = 1, \cdots, m; \tag{17}$$

$$\sum_{j=1}^{m} x_{ij} \leq 1, \text{ for } i = 1, \cdots, k; \tag{18}$$

$$x_{ij} \in \{0, 1\}, \tag{19}$$

which turns out to be a **linear assignment problem** or particularly the **agent and task assignment problem**. Suppose that two changepoint configurations, $\mathcal{C}_1$ and $\mathcal{C}_2$ for a time series of length N are $\mathcal{C}_1(m; \tau_1, \cdots, \tau_m)$ and $\mathcal{C}_2(k; \eta_1, \cdots, \eta_k)$. Note that m and k may not be equal. The cost for assigning τ_i to η_j is defined by

$$c_{i,j} = \frac{|\tau_i - \eta_j|}{N}.$$

Then the cost matrix for matching the changepoints between two configurations can be formulated:

$$\text{cost matrix} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,k} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m,1} & c_{m,2} & \cdots & c_{m,k} \end{bmatrix}. \tag{20}$$

The distance between two changepoint configurations is defined to be

$$d(\mathcal{C}_1, \mathcal{C}_2) = \text{optimal solution of the assignment problem} + |m - k|. \tag{21}$$

The assignment problem can be balanced or unbalanced ($m \neq k$). We address the unbalanced case since it is more difficult. First, note that if $m >> k$, the distance can be approximated simply by $|m - k|$ since the first part of Equation (21) is less than 1. When $m > k$, the cost matrix is not square and we add some "virtual" nodes with zero cost to $\mathcal{C}_2$ to make the matrix square. For example, suppose that two changepoint configurations $\mathcal{C}_1$ and $\mathcal{C}_2$ for a time series of length 100 are

$$\mathcal{C}_1(3; 25, 78, 99);$$
$$\mathcal{C}_2(2; 26, 51).$$

The corresponding cost matrix is

$$\text{cost matrix} = \begin{bmatrix} 1 & 26 \\ 52 & 27 \\ 73 & 48 \end{bmatrix} /100.$$

We have to add a virtual "node" to $\mathcal{C}_2$ and make the cost matrix balanced:

$$\text{cost matrix} = \begin{bmatrix} 1 & 26 & 0 \\ 52 & 27 & 0 \\ 73 & 48 & 0 \end{bmatrix} /100.$$

The optimum solution of the assignment problem is $(1 + 27 + 0)/100 = 28/100$ and the distance is 1.28.

4 Multiple Changepoint Simulations: Figures and Tables for Various Settings

The general settings of these AR(1) simulations are $N = 500$ and $\sigma^2 = 1$.

4.1 AR(1) series without changepoints

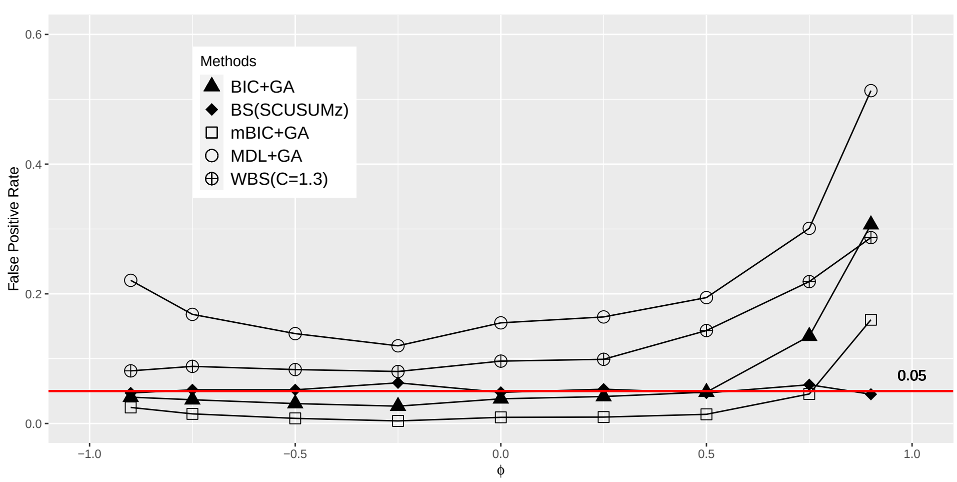

Figure 3: Empirical False Positive Detection Rates for an AR(1) Series with Various ϕ. Truth: No Changepoints.

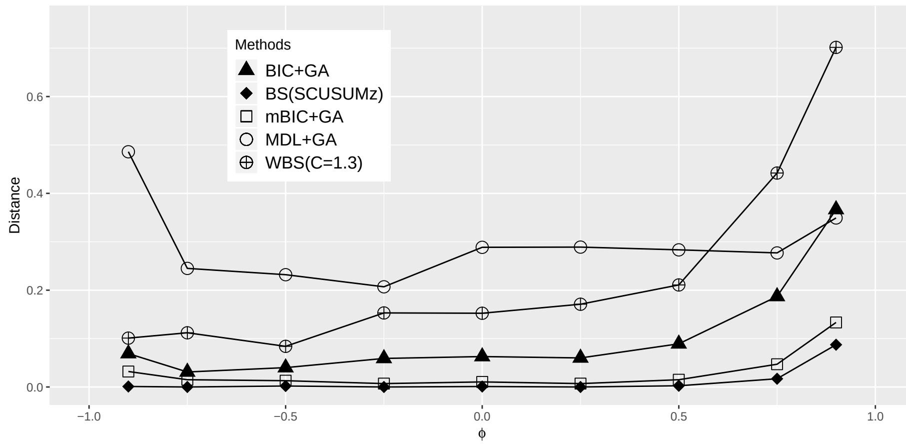

Figure 4: Average Distances for an AR(1) Series with Various ϕ. Truth: No Changepoints.

4.2 AR(1) series with one changepoint in the middle of the record

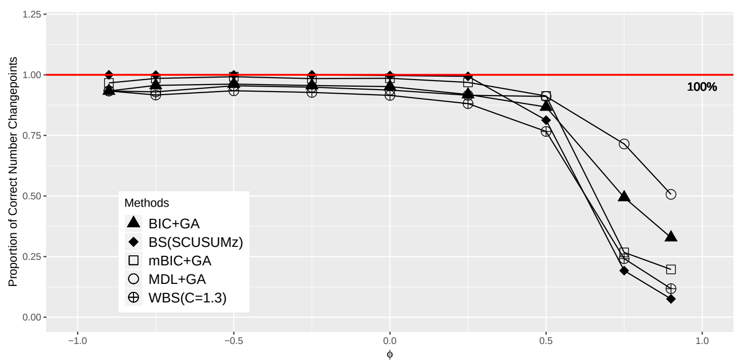

Figure 5: Proportion of Runs Correctly Estimating the Single Changepoint for an AR(1) Series with Varying ϕ. Truth: One Changepoint in the Middle Moving the Series Upwards.

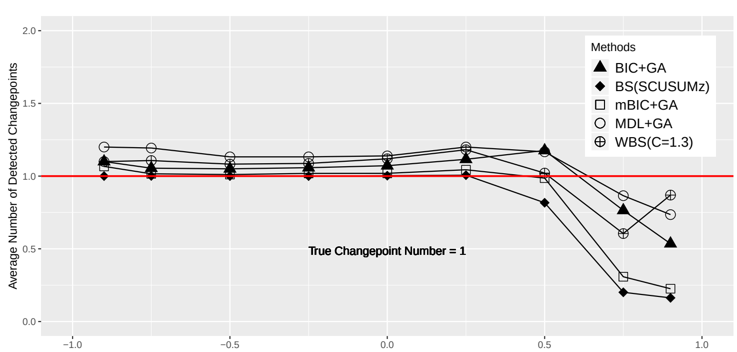

Figure 6: Average Number of Detected Changepoints for an AR(1) Series with Varying ϕ. Truth: One Changepoint in the Middle Moving the Series Upwards.

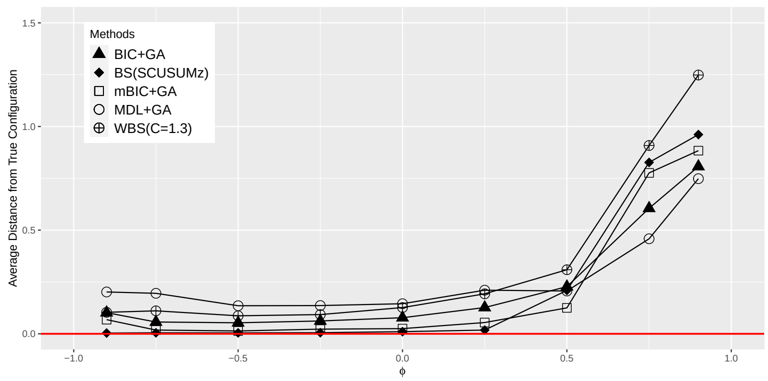

Figure 7: Average Distances for an AR(1) Series with Varying ϕ. Truth: One Changepoint in the Middle Moving the Series Upwards.

4.3 AR(1) series with three changepoints at the times 126, 251, and 376.

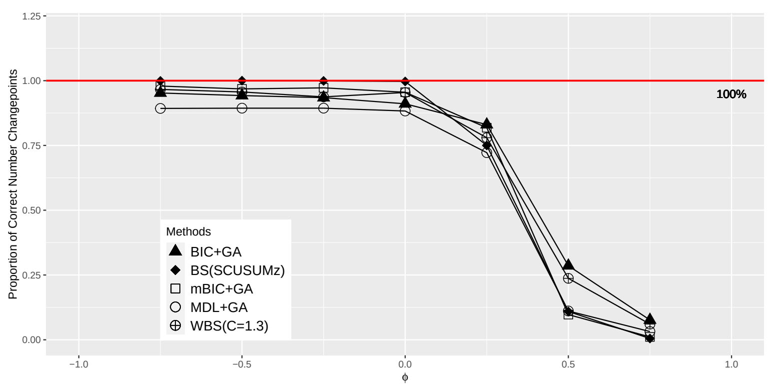

Figure 8: Proportion of Runs Correctly Detecting the Changepoint Numbers for AR(1) Series with Different ϕ. Truth: Three Equally Spaced Changepoints Moving the Series Up-Up-Up.

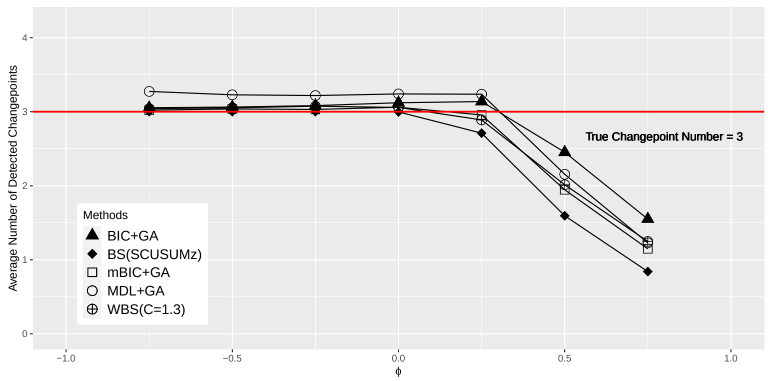

Figure 9: Average Number of Detected Changepoints for AR(1) Series with Different ϕ. Truth: Three Equally Spaced Changepoints Moving the Series Up-Up-Up.

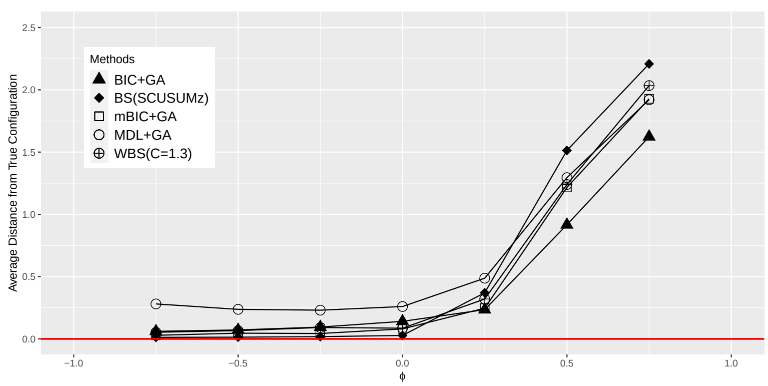

Figure 10: Average Distances for AR(1) Series with Different ϕ. Truth: Three Equally Spaced Changepoints Moving the Series Up-Up-Up.

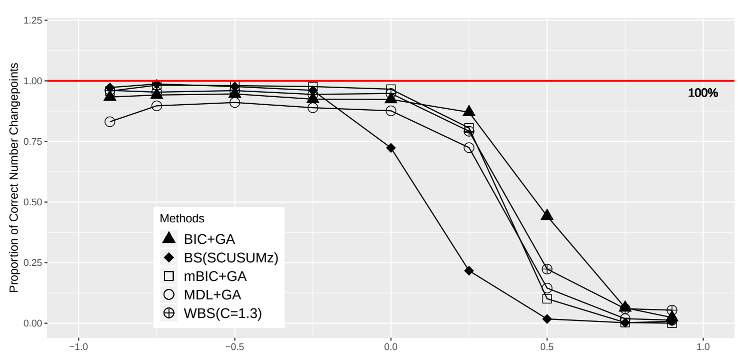

Figure 11: Proportion of Runs Correctly Estimating the Three Changepoints for an AR(1) Series with Varying ϕ. Truth: Three Equally Spaced Changepoints Moving the Series Up-Down-Up.

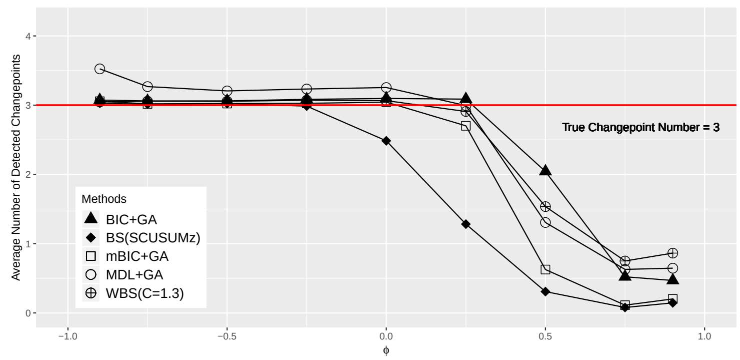

Figure 12: Average Number of Detected Changepoints for an AR(1) Series with Varying ϕ. Truth: Three Equally Spaced Changepoints Moving the Series Up-Down-Up.

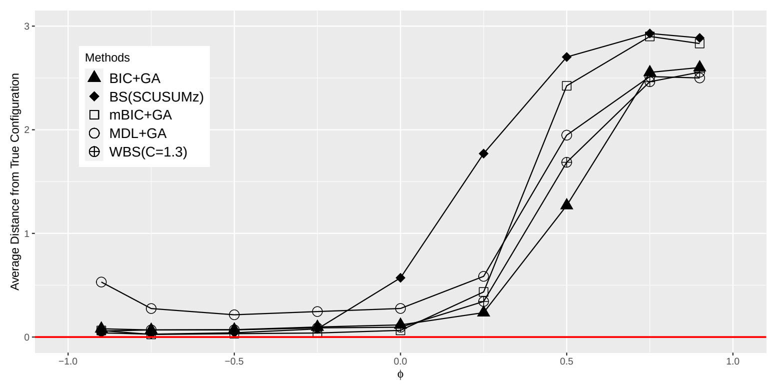

Figure 13: Average Distances for an AR(1) Series with Varying ϕ. Truth: Three Equally Spaced Changepoints Moving the Series Up-Down-Up.

4.4 AR(1) series of varying lengths N

Table 7: Performance of multiple changepoint techniques when a changepoint is at middle of an AR(1) series and the series length N varies. $\sigma^2 = 1$, $\phi = 0.5$, $\Delta = 1$.

Methods / Results	BIC+GA	mBIC+GA	MDL+GA	BS(SCUSUM)	WBS(C=1.3)
% Detect one changepoint, $N = 100$	40.9%	30.3%	56.9%	21.9%	27.9%
($N = 500$)	(86.7%)	(91.3%)	(91.1%)	(81.3%)	(76.6%)
[$N = 1000$]	[93.2%]	[96.7%]	[90.7%]	[99.6%]	[91.9%]
{$N = 2500$}	{94.16% }	{97.72%}	{91.39%}	{100%}	{97.52%}
Avg # of detected changepoints	1.436	0.530	3.291	0.252	0.738
	(1.176)	(0.986)	(1.167)	(0.817)	(1.021)
	[1.113]	[1.052]	[1.226]	[0.998]	[1.120]
	{1.115}	{1.042}	{1.209}	{1}	{1.027}
Avg distance to the true location	1.195	0.852	2.774	0.807	0.876
	(0.227)	(0.125)	(0.206)	(0.209)	(0.309)
	[0.126]	[0.066]	[0.239]	[0.017]	[0.135]
	{0.121}	{0.047}	{0.214}	{0.005}	{0.032}

Table 8: Performance of multiple changepoint techniques when three alternating changepoints are equally spaced on the sequence with different lengths N. $\sigma^2 = 1$, $\phi = 0.5$, $\Delta's = 1$.

Methods / Results	BIC+GA	mBIC+GA	MDL+GA	BS(SCUSUM)	WBS(C=1.3)
% Detect 3 changepoints, $N = 100$	10.0 %	1.96%	3.73%	0.78%	7.94%
($N = 500$)	(44.2%)	(10.1%)	(14.5%)	(1.8%)	(22.4%)
[$N = 1000$]	[82.8%]	[61.4%]	[56.7%]	[11.0%]	[67.0%]
{$N = 2500$}	{93.1% }	{97.1% }	{85.7% }	{76.3% }	{98.2% }
Avg # of detected changepoints	1.25	0.40	3.03	0.13	0.85
	(2.04)	(0.63)	(1.31)	(0.31)	(1.54)
	[3.00]	[2.17]	[2.70]	[0.81]	[2.58]
	{3.10}	{3.03}	{3.35}	{2.55}	{3.02}
Avg distance to the true locations	2.57	2.90	3.98	2.88	2.35
	(1.27)	(2.42)	(1.95)	(2.70)	(1.69)
	[0.311]	[0.921]	[0.899]	[2.201]	[0.588]
	{0.123}	{0.066}	{0.364}	{0.486}	{0.036}

Table 9: Performance of multiple changepoint techniques when nine alternating changepoints are equally spaced on an AR(1) series with different lengths N. $\sigma^2 = 1$, $\phi = 0.5$, $\Delta' s = 1$.

Methods / Results	BIC+GA	mBIC+GA	MDL+GA	BS(SCUSUM)	WBS(C=1.3)
% Detect 9 changepoints, [$N = 1000$] {$N = 2500$}	[1.49%] {3.20%}	[0.00%] {9.41%}	[0.00%] {7.41%}	[0.00%] {2.50%}	[1.39%] {3.70%}
Avg # of detected changepoints	[1.37] {8.06}	[0.10] {3.45}	[0.75] {5.55}	[0.08] {0.51}	[2.42] {6.99}
Avg distance to the true locations	[7.69] {1.29}	[8.91] {5.66}	[8.29] {3.96}	[8.93] {8.50}	[6.69] {2.31}

4.5 AR(1) series with nine equally-spaced changepoints

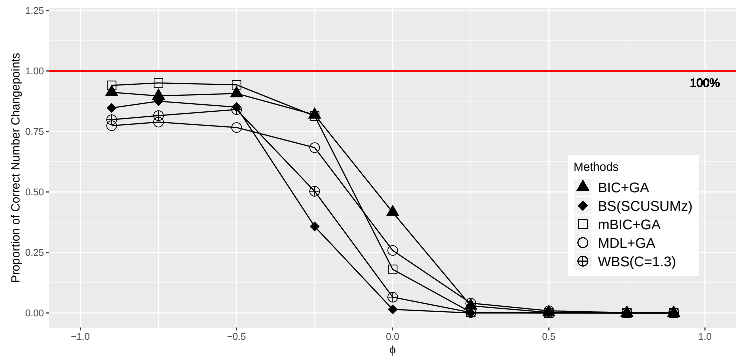

Figure 14: Proportion of Runs Detecting the Nine Changepoints for an AR(1) Series with Varying ϕ. Truth: Nine Changepoints, All Up, $\Delta = 2$

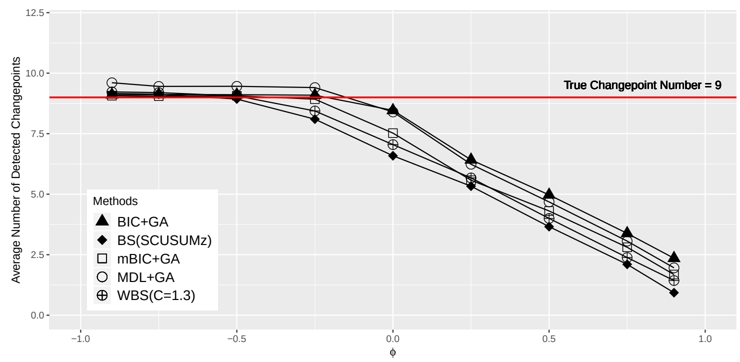

Figure 15: Average Number of Detected Changepoints for an AR(1) Series with Varying ϕ. Truth: Nine Changepoints, All Up, $\Delta = 2$.

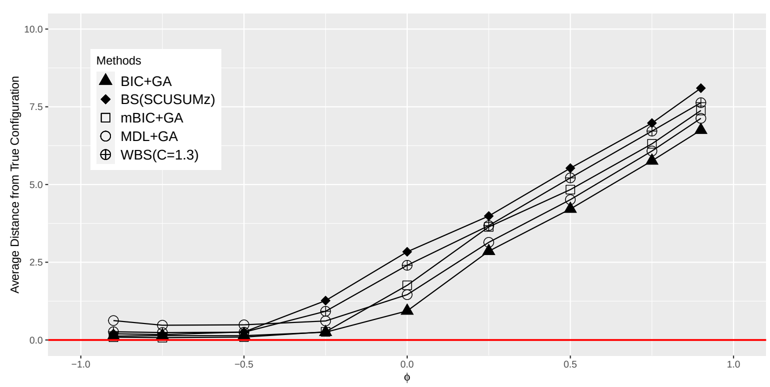

Figure 16: Average Distances for an AR(1) Series with Varying ϕ. Truth: Nine Changepoints, All Up, $\Delta = 2$.

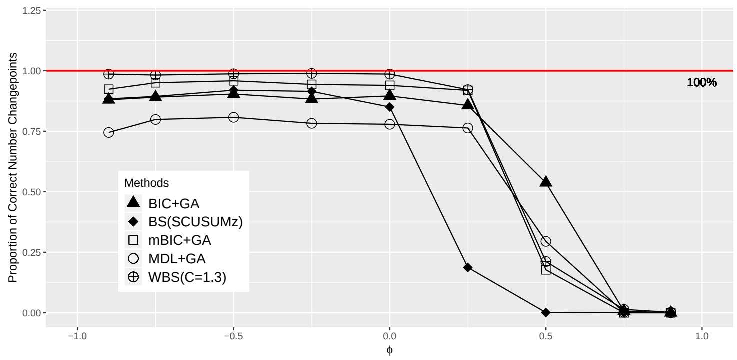

Figure 17: Proportion of Runs Correctly Detecting the Nine Changepoints for an AR(1) Series with Varying ϕ. Truth: Nine Alternating Changepoints, $\Delta = 2$.

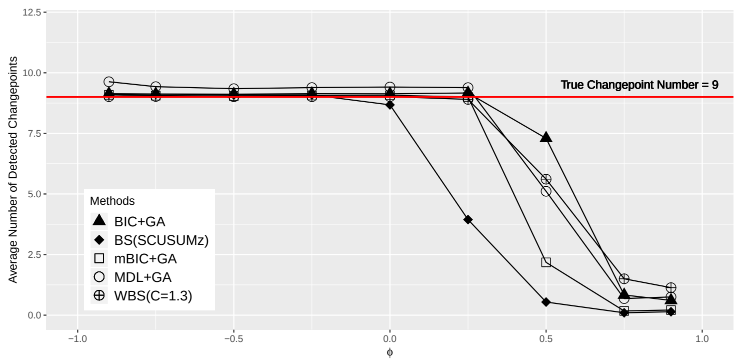

Figure 18: Average Number of Detected Changepoints for an AR(1) Series with Varying ϕ. Truth: Nine Alternating Changepoints, $\Delta = 2$.

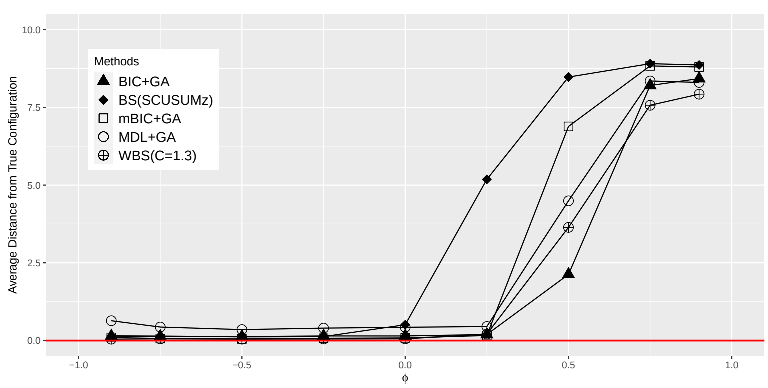

Figure 19: Average Distances for an AR(1) Series with Varying ϕ. Truth: Nine Alternating Changepoints, $\Delta = 2$.

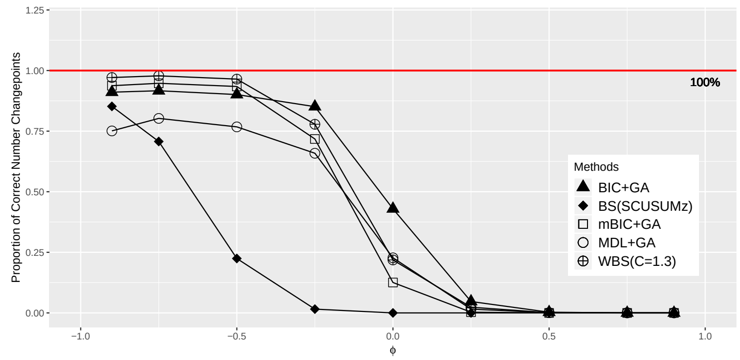

Figure 20: Proportion of Runs Correctly Detecting the Nine Changepoints for an AR(1) Series with Varying ϕ. Truth: Nine Alternating Changepoints, $\Delta = 1$.

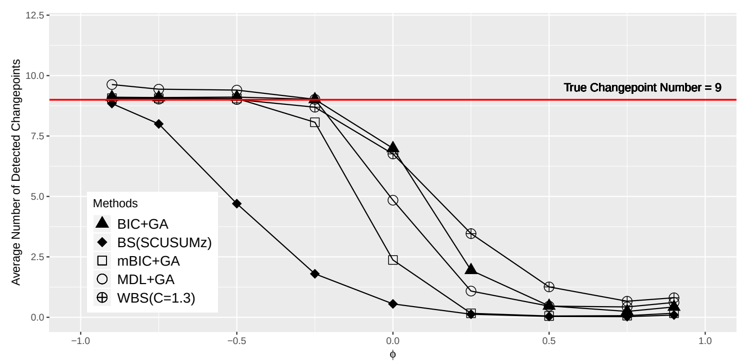

Figure 21: Average Number of Detected Changepoints for an AR(1) Series with Varying ϕ. Truth: Nine Alternating Changepoints, $\Delta = 1$.

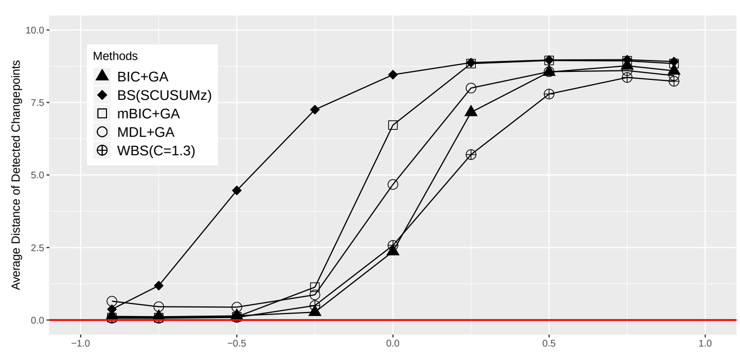

Figure 22: Average Distances for an AR(1) Series with Varying ϕ. Truth: Nine Alternating Changepoints, $\Delta = 1$.

4.6 Keyblade nine changepoint simulations

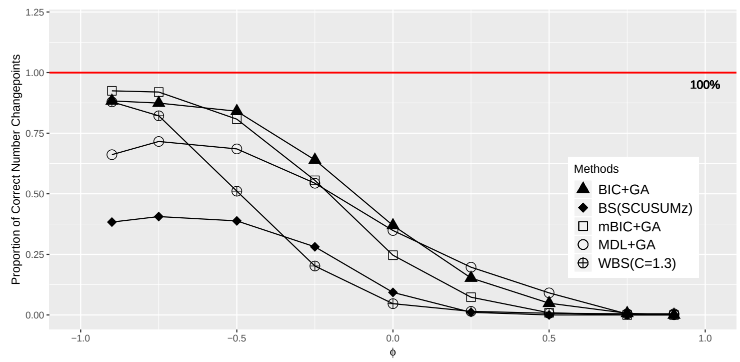

Figure 23: Proportion of Runs Correctly Detecting the Nine Changepoints for the Keyblade AR(1) Series with Varying ϕ. Truth: Nine Changepoints.

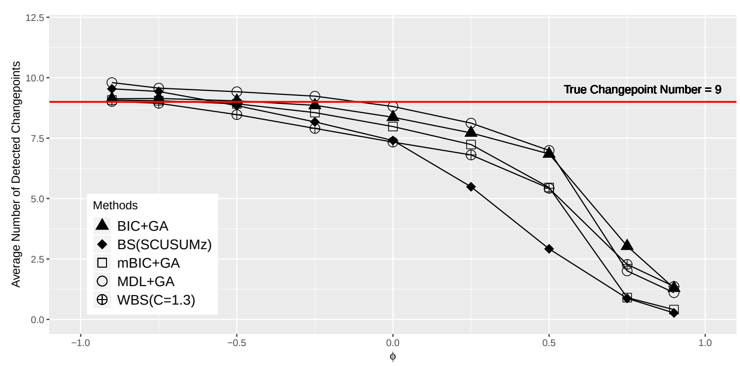

Figure 24: Average Number of Detected Changepoints for the Keyblade AR(1) Series with Varying ϕ. Truth: Nine Changepoints.

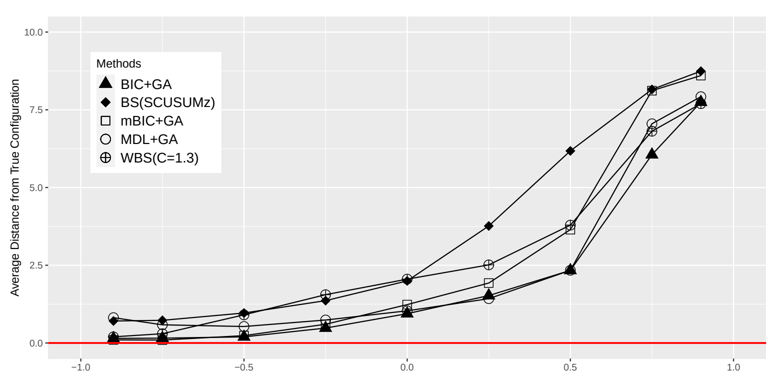

Figure 25: Average Distances for the Keyblade AR(1) Series with Varying ϕ. Truth: Nine Changepoints.

4.7 Random changepoint simulations

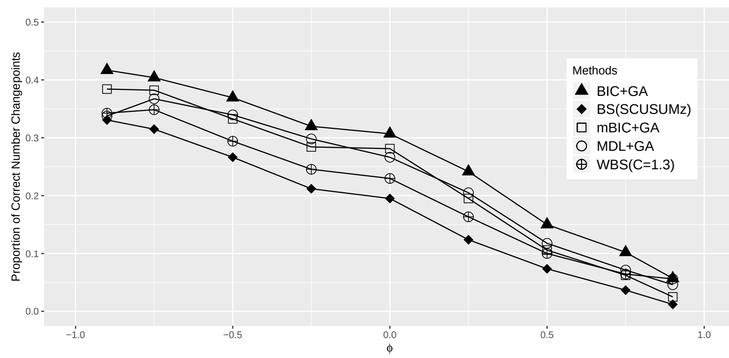

Figure 26: Proportion of Runs that Detect Correct Number of Changepoints.

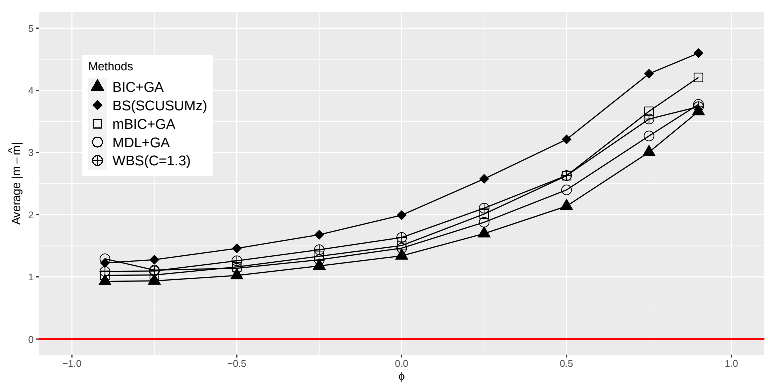

Figure 27: Average Absolute Difference between Correct Changepoint Number and Estimated Changepoint Number.

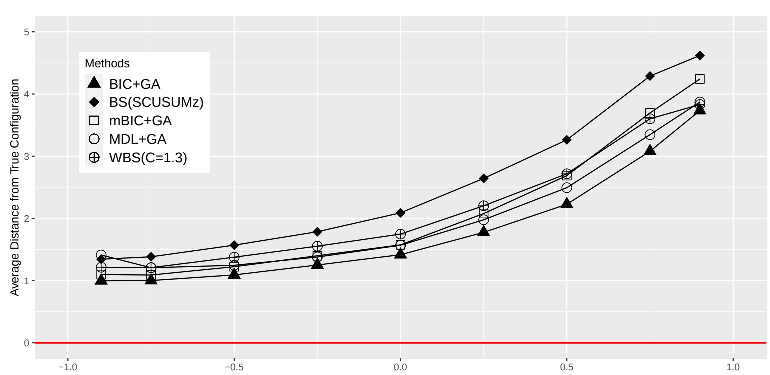

Figure 28: Average Distance between Correct Changepoint Locations and Estimated Changepoint Locations.

We examine method performance as the mean shift magnitudes increase. Here, we fix $N = 500, \phi = 0.5$, and $\sigma^2 = 1$ and consider three alternating changepoints placed at the times $126, 251$, and 376. Mean shift magnitudes are varied from 1 to 3. Average distances over $1,000$ simulations are reported in Table 10. As the mean shift magnitudes increases, all methods improve. BIC and MDL, two frequent winners of past scenarios, perform worst when the mean shift size is largest; moreover, WBS and binary segmentation, two frequent past losers, perform best. mBIC reports the smallest average distance when $\Delta \geq 2$.

Δ	BIC+GA	mBIC+GA	MDL+GA	BS(SCUSUMz)	WBS(C=1.3)
$\Delta = 1$	1.269	2.424	1.948	2.702	1.686
$\Delta = 2$	0.140	0.051	0.209	0.843	0.149
$\Delta = 3$	0.126	0.042	0.188	0.077	0.079

Table 10: Average Distance for an AR(1) Series with Varying Mean Shift Magnitudes

We now consider other autoregressive error structures. We begin with AR(2) errors and the case of no changepoints. Table 11 shows false positive rates of signaling one or more changepoints when in truth none exist for various AR(2) parameters ϕ_1 and ϕ_2. In this and all four tables below, 1,000 independent simulations are conducted, $N = 500$, $\sigma^2 = 1$, and all mean shift sizes are two units (this adds additional information to cases above where mean shifts were unity). All four tables are discussed below in tandem after they are presented.

$\{\phi_1, \phi_2\}$	BIC+GA	mBIC+GA	MDL+GA	BS(SCUSUMz)	WBS(C=1.3)
{0.6, 0.35}	21.5%	2.5%	38.8%	22.6%	50.0%
{0.6, 0.3}	17.5%	2.6%	33.2%	10.2%	36.6%
{0.6, -0.1}	5.9%	1.1%	15.6%	0.3%	17.4%
{0.5, -0.2}	4.1%	1.6%	13.6%	0.0%	11.7%
{0.2, -0.5}	3.0%	0.6%	9.4%	0.1%	9.1%

Table 11: False Positive Rates for an AR(2) Series with Varying $\{\phi_1, \phi_2\}$. Truth: No Changepoints

Table 12 reports average distances for the AR(2) scenario of the last table, but now with three changepoints. The three shifts induce four equal length regimes and shift the series mean in an up-down-up manner.

$\{\phi_1, \phi_2\}$	BIC+GA	mBIC+GA	MDL+GA	BS(SCUSUMz)	WBS(C=1.3)
{0.6, 0.35}	2.757	2.932	2.759	2.633	2.265
{0.6, 0.30}	2.484	2.895	2.510	2.742	2.337
{0.6, -0.1}	0.167	0.052	0.182	0.818	0.193
{0.5, -0.2}	0.131	0.032	0.163	0.072	0.101
{0.2, -0.5}	0.086	0.023	0.111	0.040	0.068

Table 12: Average Distances for an AR(2) Series with Varying $\{\phi_1, \phi_2\}$. Truth: Three Alternating Changepoints of Size $\Delta = 2$ Occurring at the time 126, 251, and 376.

Table 13 shows false positive rates of signaling one or more changepoints when in truth there are none for various parameter choices in an AR(4) series.

$\{\phi_1,\ \phi_2,\ \phi_3,\ \phi_4\}$	BIC+GA	mBIC+GA	MDL+GA	BS(SCUSUMz)	WBS(C=1.3)
{0.5, 0.25, 0.15, 0.05}	66.5 %	44.8%	76.4%	29.7%	54.4%
{0.6, 0.3, 0.1, -0.3}	16.7 %	8.5%	42.0%	0.6%	21.5%
{0.6, 0.3, -0.3, -0.1}	9.9%	4.9%	32.5%	0.1%	14.8%
{0.6, -0.4, -0.2, -0.1}	5.0%	2.5%	27.0%	0.2%	10.3%
{0.6, -0.4, 0.3, -0.2}	5.3%	1.6%	22.9%	0.2%	17.4%

Table 13: False Positive Rates for an AR(4) Series with Varying $\{\phi_1, \phi_2, \phi_3, \phi_4\}$. Truth: No Changepoints.

Finally, Table 14 reports average distances over 1,000 independent simulations for the same AR(4) scenario above. The mean shift specifications are repeated from Table 12.

$\{\phi_1,\ \phi_2,\ \phi_3,\ \phi_4\}$	BIC+GA	mBIC+GA	MDL+GA	BS(SCUSUMz)	WBS(C=1.3)
{0.5, 0.25, 0.15, 0.05}	2.723	2.420	3.360	2.516	2.151
{0.6, 0.3, 0.1, -0.3}	0.615	1.582	1.292	2.318	1.256
{0.6, 0.3, -0.3, -0.1}	0.205	0.107	0.251	0.834	0.211
{0.6, -0.4, -0.2, -0.1}	0.127	0.055	0.319	0.031	0.079
{0.6, -0.4, 0.3, -0.2}	0.161	0.066	0.246	0.228	0.101

Table 14: Average Distances for AR(4) Errors with Varying $\{\phi_1, \phi_2, \phi_3, \phi_4\}$. Truth: Three Alternating Changepoints of Size $\Delta = 2$ Occurring at the time 126, 251, and 376.

In the above tables, when there are no changepoints, binary segmentation appears best and MDL and WBS worst, repeating conclusions for AR(1) errors. In the tables with three changepoints and heavily positively correlated errors, MDL, BIC, and WBS all do comparatively well; when the correlation becomes negative, the situation reverses and mBIC and binary segmentation are best. These aspects were also seen for AR(1) series, although we did not remark about the negatively correlated results.

References

[1] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 2013.

[2] Miklos Csörgo and Lajos Horváth. *Limit Theorems in Change-point Analysis*. John Wiley & Sons Chichester, 1997.

[3] Michael Robbins, Colin Gallagher, Robert Lund, and Alexander Aue. Mean shift testing in correlated data. *Journal of Time Series Analysis*, 32(5):498–511, 2011.

[4] Leonid Tolmatz. On the distribution of the square integral of the Brownian bridge. *The Annals of Probability*, 30(1):253–269, 2002.