

**Supplementary material:**

**MULTIPLE CHANGE-POINT DETECTION FOR NON-STATIONARY  
TIME SERIES USING WILD BINARY SEGMENTATION**

Karolos K. Korkas and Piotr Fryzlewicz

*London School of Economics*

**1 Parameter selection simulation study**

**1.1 Bootstrap threshold selection**

In Section 5.1 of Korkas and Fryzlewicz (2014) we discuss an alternative data-driven approach in selecting an appropriate threshold  $C^{(i)}$  by fitting an AR(p) model to the data. We refer to this bootstrap method as Bsp1. A referee has shown us a way that utilises the wavelet periodogram which is also our main change-point detection statistic. A similar procedure has appeared in Nunes et al. (2014), but in the context of stationarity testing for textured images. The algorithm for obtaining the thresholds is described below.

**Bootstrap algorithm for obtaining  $C^{(i)}$  (Bsp2)**

- Input: compute the spectrum  $\hat{S}_i(z)$  for  $i = -1, -2, \dots, -I^*$ .
- 1. Compute the average  $\bar{S}_i(z)$  by taking the average of spectrum values for each  $i = -1, -2, \dots, -I^*$ .
- 2. For  $g = 1, \dots, B$  repeat the following
  - Simulate a stationary model  $X_{t,T}^{(g)}$  using squared amplitudes given by  $\bar{S}_i(z)$  and with  $\mathcal{N}(0, 1)$  innovations.
  - Compute the following ratio

$$C_g^{(i)} = \mathbb{Y}_v^{(i)} (\log T)^{-1} \left( \sum_{t=1}^T I_{t,T}^{(i)} \right)^{-1} \quad T \text{ for } i = -1, -2, \dots, -I^*$$

where  $v$  maximises the absolute value of (3.1) from Korkas and Fryzlewicz (2014).

- 3. Output: the  $q$ th quantile of  $C_g^{(i)}$  (we use the 95% quantile in this work).

For a comparison study between the two bootstrap algorithms we repeat the simulation study from the main text using the models S1-S7 (no change-points) and A-I. The results are given in Table 1.1. Even though Bsp2 shows a slightly better performance compared with Bsp1 in terms of the hit ratio for models A-I, it produces more spurious change-points in models S3, S6 and S7, similarly to the universal thresholds discussed in the main text. Hence, we favour Bsp1 which appears to be more robust in most cases examined here.

Table 1.1: Table compares the performance of the two bootstrap approaches Bsp1 and Bsp2 for selecting  $C^{(i)}$ . Panel I shows the number of occasions either Bsp1 or Bsp2 detected at least one change-point for models without change-points (taken from the main text i.e. S1-S7). Panel II shows the hit ratios achieved by either Bsp1 or Bsp2 for the models A-I from the main text. All the models have size 1024 and the hit ratios are calculated assuming  $d_{\max} = 25$ . In all the cases we use WBS2.

Panel I			Panel II		
Model	Bsp1	Bsp2	Model	Bsp1	Bsp2
S1	0	17	A	0.673	0.700
S2	1	14	B	0.765	0.758
S3	5	93	C	0.816	0.784
S4	0	22	D	0.695	0.680
S5	0	13	E	0.460	0.483
S6	0	29	F	0.634	0.635
S7	5	100	G	0.679	0.729
			H	0.552	0.573
			I	0.582	0.635

## 1.2 Other parameters

Another important parameter in our method is the unbalanceness condition (1.5) from the main text which is controlled by  $c_* \in (0.67, 1)$ . To choose an appropriate value that will work in most cases we conduct the following (simple) experiment. For sample sizes  $T = 500, 550, \dots, 1500$  we estimate the following model P1

$$y_t = \begin{cases} -0.6y_{t-1} + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, 1) & \text{for } 1 \leq t \leq \lfloor 0.12T \rfloor \\ 0.4y_{t-1} + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, 1) & \text{for } \lfloor 0.12T \rfloor + 1 \leq t \leq \lfloor 0.22T \rfloor \\ -0.6y_{t-1} + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, 1) & \text{for } \lfloor 0.22T \rfloor + 1 \leq t \leq \lfloor 0.52T \rfloor \\ 0.5y_{t-1} + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, 1) & \text{for } \lfloor 0.52T \rfloor + 1 \leq t \leq \lfloor 0.60T \rfloor \\ -0.6y_{t-1} + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, 1) & \text{for } \lfloor 0.60T \rfloor + 1 \leq t \leq \lfloor 0.72T \rfloor \\ 0.4y_{t-1} + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, 1) & \text{for } \lfloor 0.72T \rfloor + 1 \leq t \leq T \end{cases}$$

On a grid of  $c_*$  and  $T$  values we calculate the mean hit ratio over 50 repetitions for each value  $c_* = 0.67, 0.695, \dots, 1$  and  $T = 500, 550, \dots, 2000$ . For robustness we repeat the experiment

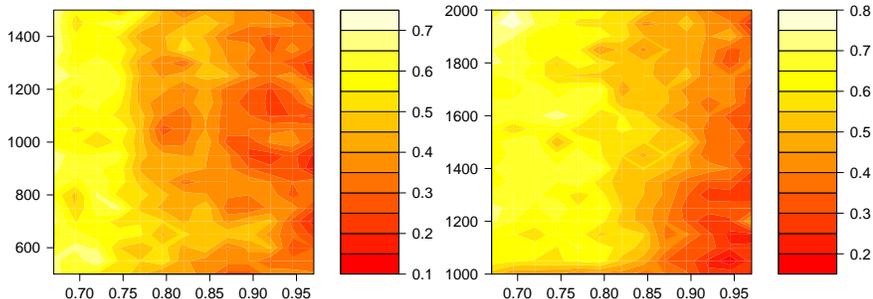


Figure 1.1: The hit ratio heat map for the model P1 for increasing  $c_*$  and  $T$ .

Figure 1.2: The hit ratio heat map for the model P2 for increasing  $c_*$  and  $T$ .

for a different model and for  $T = 1000, 1050, \dots, 2000$ . We choose model G from the main text but we keep the change-points  $\eta = 200, 400, 600, 800$  the same for all sample sizes. We call this model P2. In the colour map of Figures 1.1 and 1.2 the lighter (yellow) colours indicate high hit ratios (good performance).

It appears that our method performs better when  $c_*$  is less than 0.75, hence we set this value as the default for this parameter, even though the lower bound 0.67 shows an equally strong performance.

For selecting the parameter  $\lambda$ , which defines the finest scale to be used by our methods, we conduct a similar experiment. We allow  $\lambda = 0.25, 0.30, \dots, 1.1$  and  $T = 500, 550, \dots, 2000$  and choose Model A from the main text, but with change-points at  $\eta = 200, 300$ . The reason for selecting this model is due to our observation that the change-points were more likely to be found in coarser scales and, hence, the performance of our methods are more sensitive to the  $\lambda$  parameter. Similarly with the experiment described above, for every pair  $(c_*, T)$  we repeat the estimation 50 times. Figure 1.3 indicates that the range  $[0.7, 0.9]$  provides the best results and we choose  $\lambda = 0.7$  as our default value for this parameter.

## 2 Additional analysis to the simulation study of Section 5.2

In this section we provide more evidence for the performance of our methods. First, we repeat the simulation study described in the main text, but with  $d_{\max} = 25$  (2.5%) and not 50 (5%), and we measure the performance by the obtained hit ratio. The results are given in Table 1.2. Again the WBS methods do well compared with BS1, BS2 or CF methods and it is interesting to note that there have been changes in the ranking of the methods in certain methods. CF method is not within 10% of the best in Models B, C and F. The same holds for BS1 and WBS1 in Models A and C, while WBS2 is the best or within 10% of the best in all models except E for both  $d_{\max} = 25$  and 50. In general, WBS1 and WBS2 perform similarly and

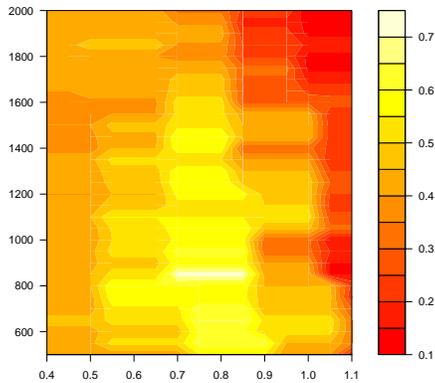


Figure 1.3: Hit ratio map for the model A to select the  $\lambda$  parameter.

the same holds for CF and BS1 (when  $d_{\max} = 50$ , CF does better). In addition, we generate empirical histograms for the estimated locations of the change-points using BS2 and WBS2. The histograms confirm the fact that WBS2 reveals the majority of the change-points more frequently, which is especially evident for models D, G, H and I.

### 3 Additional simulation studies

#### 3.1 Small sample size simulation study

We assess the performance of our methods for small sample cases, i.e.  $T = 200, 300$  using the models As, Bs and Cs. These models are modifications of the models discussed in the main text except Bs which is a modified version of Model G from Killick et al. (2013). All the results are shown in Table 1.3 and  $d_{\max} = 10$  for all cases.

**Model As:** *A non-stationary process similar to model D*

In this model the change-points occur in positions (90, 135) and the sample size is  $T = 200$ . The WBS1 and WBS2 methods achieve a high hit ratio, more than double of that of BS1/BS2.

**Model Bs:** *A piecewise constant MA process*

$$y_t = \begin{cases} \varepsilon_t + 0.8\varepsilon_{t-1}, \varepsilon_t \sim \mathcal{N}(0, 1) & \text{for } 1 \leq t \leq 85 \\ \varepsilon_t + 1.68\varepsilon_{t-1} - 0.81\varepsilon_{t-2}, \varepsilon_t \sim \mathcal{N}(0, 1) & \text{for } 86 \leq t \leq 120 \\ \varepsilon_t + 0.8\varepsilon_{t-1}, \varepsilon_t \sim \mathcal{N}(0, 1) & \text{for } 121 \leq t \leq 200 \end{cases}$$

In this model WBS1 and WBS2 revealed the two change-points in 84% and 85% of the

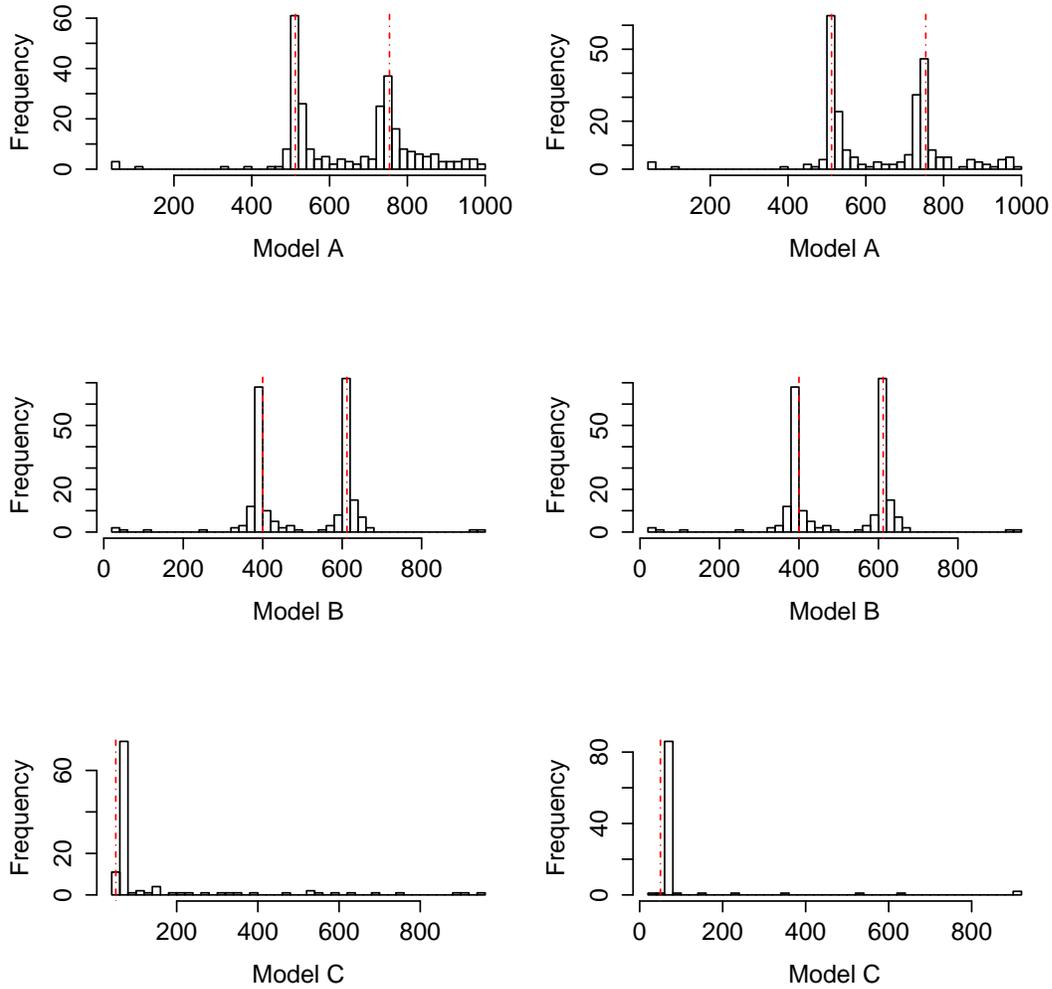


Figure 1.4: Empirical change-point densities for the Models A (top), B (middle) and C (bottom). Left: estimated locations frequency estimated by WBS2 method. Right: estimated locations frequency estimated by BS2 method. The red vertical dotted lines are the locations of the real change-points.

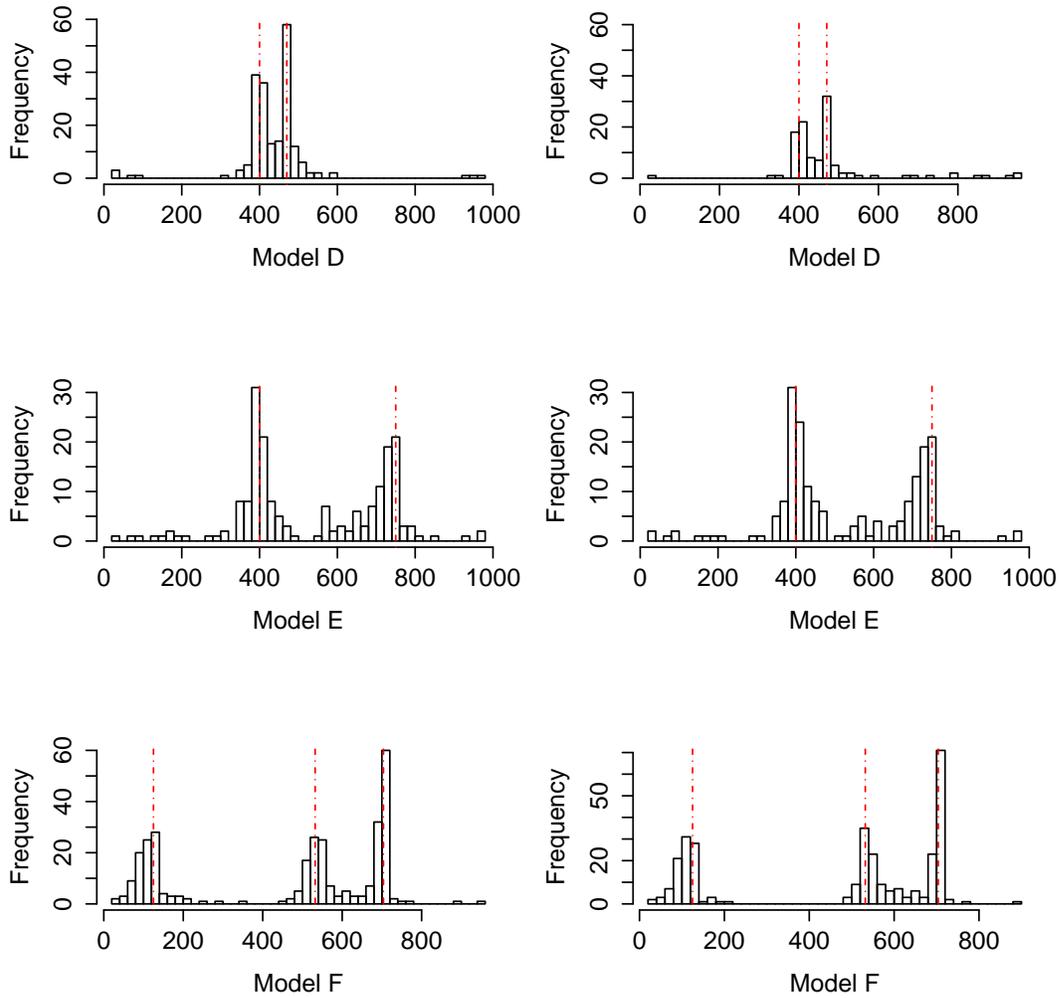


Figure 1.5: Empirical change-point densities for the Models D (top), E (middle) and F (bottom). Left: estimated locations frequency estimated by WBS2 method. Right: estimated locations frequency estimated by BS2 method. The red vertical dotted lines are the locations of the real change-points.

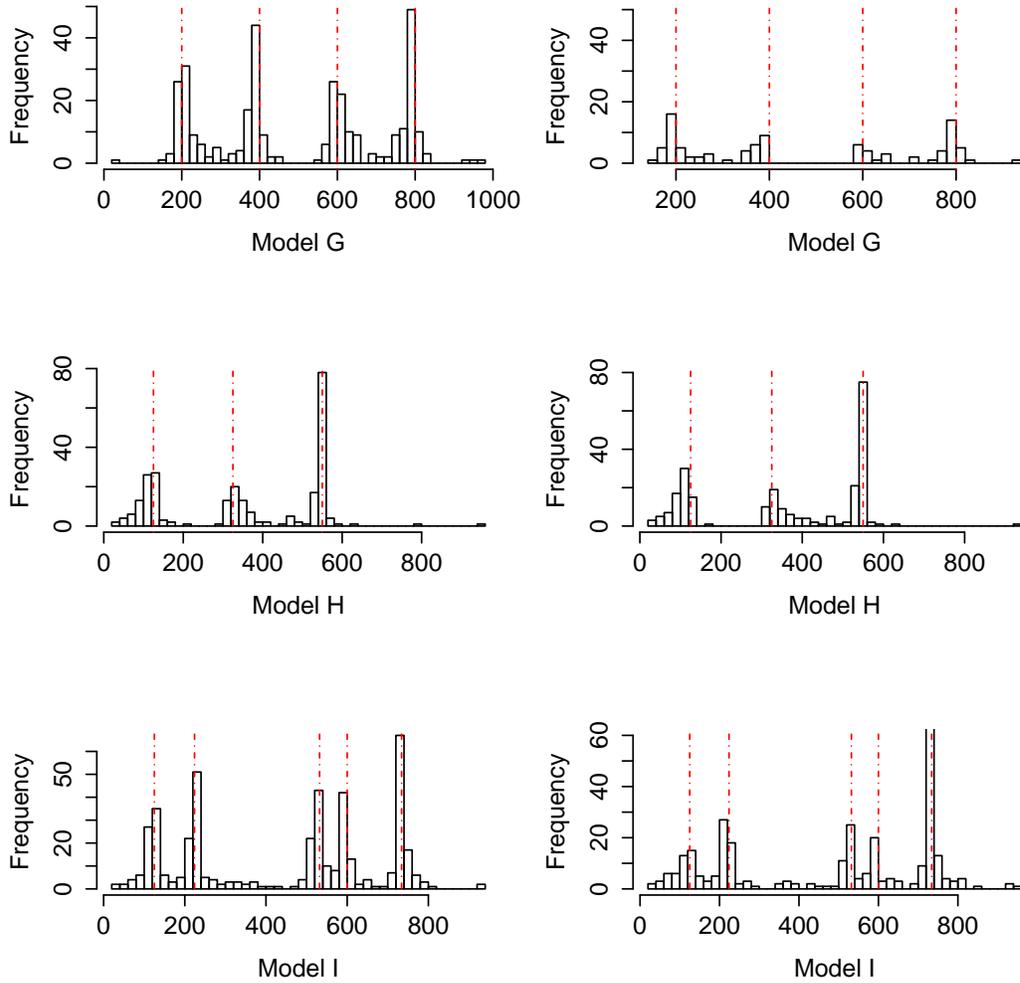


Figure 1.6: Empirical change-point densities for the Models G (top), H (middle) and I (bottom). Left: estimated locations frequency estimated by WBS2 method. Right: estimated locations frequency estimated by BS2 method. The red vertical dotted lines are the locations of the real change-points.

Table 1.2: Non-stationary processes results for the simulation described in the main text. Table shows the number of occasions a method detected that number of change-points within 25 data points from the real ones. Bold: the method with the highest hit ratio or within 10% from the highest.

Number of Change-points															
Model	A					B					C				
	BS1	BS2	WBS1	WBS2	CF	BS1	BS2	WBS1	WBS2	CF	BS1	BS2	WBS1	WBS2	CF
<b>0</b>	9	6	7	4	16	1	2	0	1	2	41	28	19	32	29
<b>1</b>	42	33	41	46	45	17	24	19	20	23	59	72	81	68	71
<b>2</b>	49	61	52	50	39	82	74	81	79	75	-	-	-	-	-
Hit ratio	0.619	<b>0.690</b>	0.607	0.599	0.530	<b>0.836</b>	<b>0.835</b>	<b>0.838</b>	<b>0.845</b>	0.775	0.570	<b>0.705</b>	0.753	0.628	0.648
Model	D					E					F				
	BS1	BS2	WBS1	WBS2	CF	BS1	BS2	WBS1	WBS2	CF	BS1	BS2	WBS1	WBS2	CF
<b>0</b>	52	52	11	8	50	11	22	20	28	6	10	0	2	1	5
<b>1</b>	11	14	29	37	16	58	54	57	50	55	25	13	25	12	22
<b>2</b>	37	34	60	55	34	31	24	23	22	39	40	50	41	53	50
<b>3</b>	-	-	-	-	-	-	-	-	-	-	25	37	32	34	23
Hit ratio	0.416	0.406	<b>0.733</b>	<b>0.723</b>	0.398	<b>0.570</b>	0.500	0.497	0.456	<b>0.631</b>	0.587	<b>0.737</b>	<b>0.663</b>	<b>0.720</b>	0.613
Model	G					H					I				
	BS1	BS2	WBS1	WBS2	CF	BS1	BS2	WBS1	WBS2	CF	BS1	BS2	WBS1	WBS2	CF
<b>0</b>	61	63	9	8	43	6	5	10	1	4	5	10	0	0	5
<b>1</b>	19	16	18	15	24	48	39	26	29	45	43	35	10	1	45
<b>2</b>	10	13	18	22	24	34	39	39	44	45	15	16	16	17	30
<b>3</b>	5	5	31	29	5	12	17	25	26	6	17	19	23	20	15
<b>4</b>	5	3	24	26	4	-	-	-	-	-	10	10	23	31	3
<b>5</b>	-	-	-	-	-	-	-	-	-	-	10	10	28	31	2
Hit ratio	0.183	0.170	<b>0.598</b>	<b>0.618</b>	0.257	0.501	0.552	<b>0.577</b>	<b>0.644</b>	0.498	0.425	0.428	<b>0.669</b>	<b>0.727</b>	0.343

occasions respectively achieving a high hit ratio. On the other hand, the BS methods did not detect any change-point at 60% of the occasions.

**Model Cs:** *A non-stationary process similar to model G*

In this model the change-points occur in positions (100, 150, 225) and the sample size is  $T = 300$ . Again, WBS1 and WBS2 do well in this example achieving a hit ratio almost 50% better than that of the BS methods. In more than 60% of the occasions they detected three change-points without over-segmenting the series and with a good accuracy.

### 3.2 Large sample size simulation study

We assess the performance of our methods for large sample cases, i.e.  $T = 5000, 10000$  using the models Ab, Bb and Cb. All the results are shown in Table 1.4 and  $d_{\max} = 25$  for all cases.

**Model Ab:** *A non-stationary process similar to model G, but with many change-points*

In this model the change-points occur in positions (1000, 1200, ..., 2800), i.e. there are 10 change-points in total. The sample size is  $T = 5000$ . The WBS1 and WBS2 methods achieve a high hit ratio, more than double of that of BS1/BS2.

Table 1.3: Non-stationary processes results for small sample sizes. Panel I shows the number of occasions a method detected that number of change-points within 10 data points from the real ones. Bold: the method with the highest hit ratio or within 10% from the highest. Panel II shows the percentage of occasions a method detected that number of change-points. True number of change-points is in bold.

Panel I												
Number of Change-points within 10 data points from the real ones												
Model	As				Bs				Cs			
	BS1	BS2	WBS1	WBS2	BS1	BS2	WBS1	WBS2	BS1	BS2	WBS1	WBS2
<b>0</b>	70	69	36	33	61	61	10	10	16	16	10	6
<b>1</b>	16	17	23	23	13	14	26	29	54	52	33	25
<b>2</b>	14	14	41	44	26	25	64	61	14	12	14	25
<b>3</b>	-	-	-	-	-	-	-	-	16	20	43	44
Hit ratio	0.216	0.225	<b>0.521</b>	<b>0.551</b>	0.325	0.429	<b>0.770</b>	<b>0.755</b>	0.433	0.453	<b>0.625</b>	<b>0.683</b>

Panel II												
Number of Change-points												
Model	As				Bs				Cs			
	BS1	BS2	WBS1	WBS2	BS1	BS2	WBS1	WBS2	BS1	BS2	WBS1	WBS2
<b>0</b>	64	63	33	30	60	60	10	9	5	5	3	1
<b>1</b>	16	16	9	9	5	6	6	8	59	55	29	21
<b>2</b>	<b>19</b>	<b>21</b>	<b>57</b>	<b>60</b>	<b>35</b>	<b>34</b>	<b>84</b>	<b>83</b>	7	7	8	10
$\geq$ <b>3</b>	1	0	1	1	0	0	0	0	<b>29</b>	<b>33</b>	<b>60</b>	<b>68</b>
Total	100	100	100	100	100	100	100	100	100	100	100	100

**Model Bb:** *A non-stationary process similar to model B, but with many change-points*

In this model the change-points occur in positions (1000, 1200, ..., 2800), i.e. there are 10 change-points in total. The sample size is  $T = 5000$ . WBS2 outperforms all other methods when measured by the hit ratio while it reveals ten change-points in 65% of the occasions and this is confirmed by the histogram in Figure 1.7.

**Model Bb:** *A non-stationary process similar to model B, but with three change-points*

In this model the change-points occur in positions (500, 700, 900) and the sample size is  $T = 10000$ . The reason for choosing this model is to examine the performance of the methods in large datasets where there are a few change-points occurring in short distances at the start of the series. Even though all methods perform well in that they reveal the three change-points in most occasions the WBS methods are significantly more accurate in detecting their locations. This is confirmed by the higher hit ratio achieved as well as the histogram of the estimated change-points (Figure 1.7) where it is clearly shown that the middle change-point is more concentrated around the real change-point 700.

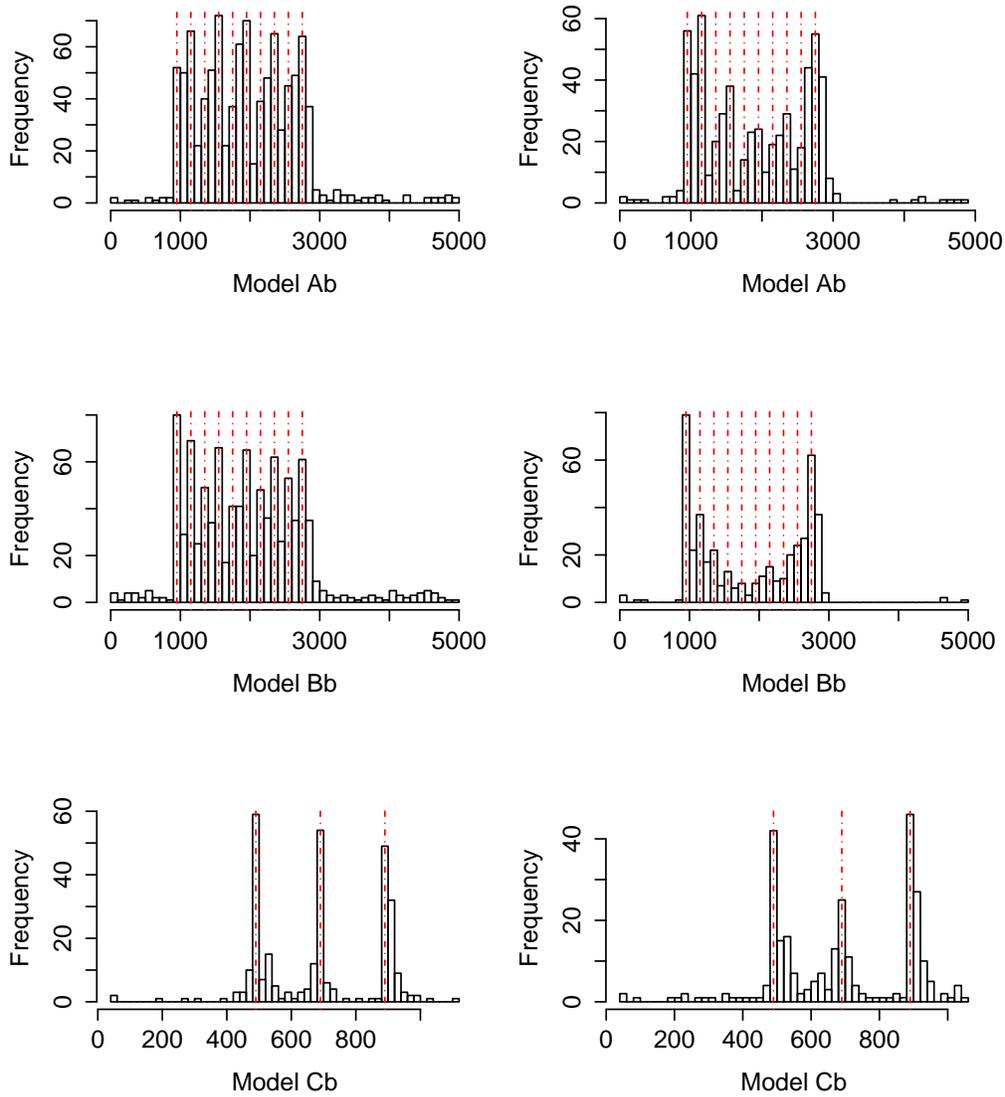


Figure 1.7: Empirical change-point densities for the Models Ab (top), Bb (middle) and Cb (bottom). Left: estimated locations frequency estimated by WBS2 method. Right: estimated locations frequency estimated by BS2 method. The red vertical dotted lines are the locations of the real change-points. Note: density plot of Cb is a zoomed in version.

Table 1.4: Non-stationary processes results for large sample sizes. Panel I shows the number of occasions a method detected that number of change-points within 25 data points from the real ones. Bold: the method with the highest hit ratio or within 10% from the highest. Panel II shows the percentage of occasions a method detected that number of change-points. True number of change-points is in bold.

Panel I												
Number of Change-points within 25 data points from the real ones												
Model	Ab				Bb				Cb			
	BS1	BS2	WBS1	WBS2	BS1	BS2	WBS1	WBS2	BS1	BS2	WBS1	WBS2
<b>0</b>	1	1	0	0	16	2	0	0	13	5	4	3
<b>1</b>	10	13	0	0	28	15	7	0	35	31	21	11
<b>2</b>	25	20	1	1	27	27	3	1	32	38	46	37
<b>3</b>	19	16	1	2	16	26	12	5	20	26	29	49
<b>4</b>	12	18	4	5	4	8	15	9	-	-	-	-
<b>5</b>	9	4	4	13	6	10	20	12	-	-	-	-
<b>6</b>	13	9	13	11	2	8	15	18	-	-	-	-
<b>7</b>	4	5	24	21	1	3	16	21	-	-	-	-
<b>8</b>	3	9	22	23	0	1	6	20	-	-	-	-
<b>9</b>	3	3	23	18	0	0	6	11	-	-	-	-
<b>10</b>	1	2	8	6	0	0	0	3	-	-	-	-
Hit ratio	0.375	0.392	<b>0.719</b>	<b>0.685</b>	0.195	0.306	0.495	<b>0.613</b>	0.515	0.602	<b>0.623</b>	<b>0.730</b>

Panel II												
Number of Change-points												
Model	Ab				Bb				Cb			
	BS1	BS2	WBS1	WBS2	BS1	BS2	WBS1	WBS2	BS1	BS2	WBS1	WBS2
<b>0</b>	0	0	0	0	1	0	0	0	0	0	0	0
<b>1</b>	0	0	0	0	0	0	0	0	9	10	3	2
<b>2</b>	15	8	0	0	27	21	3	0	17	10	9	4
<b>3</b>	12	16	0	1	23	17	4	0	<b>63</b>	<b>71</b>	<b>67</b>	<b>77</b>
<b>4</b>	23	12	0	0	18	22	5	3	11	7	12	7
<b>5</b>	11	16	1	0	12	11	4	2	0	2	9	10
<b>[6-9]</b>	31	30	24	26	16	25	50	30	0	0	0	0
<b>10</b>	<b>8</b>	<b>18</b>	<b>75</b>	<b>73</b>	<b>3</b>	<b>4</b>	<b>34</b>	<b>65</b>	0	0	0	0
Total	100	100	100	100	100	100	100	100	100	100	100	100

#### 4 Additional material on the variance stabilization

The “variance stabilization” refers to Stage II of our Algorithm of Section 4 in the main paper, in which we divide our CUSUM statistic

$$\mathbb{Y}_{s_m, e_m}^{b(i)} = \sqrt{\frac{e_m - b}{n(b - s_m + 1)}} \sum_{t=s_m}^b I_{t,T}^{(i)} - \sqrt{\frac{b - s_m + 1}{n(e_m - b)}} \sum_{t=b+1}^{e_m} I_{t,T}^{(i)} \quad (1.2)$$

by the local sample mean of  $I_{t,T}^{(i)}$ , given by

$$q_{s_m, e_m} = \sum_{t=s_m}^{e_m} I_{t,T}^{(i)} / n_m,$$

before testing the thus-rescaled CUSUM against the threshold  $\omega_T$ .

The basic reason for performing the variance stabilization is that if it is not done, the variance of  $\mathbb{Y}_{s_m, e_m}^{b(i)}$  will depend quadratically on  $E(I_{t,T}^{(i)})$ ,  $t = s_m, \dots, e_m$ .

This means that for two processes that are both stationary over the segment  $[s_m, e_m]$  but have two different values of  $E(I_{t,T}^{(i)})$ , the corresponding CUSUM statistics will be of different magnitudes. Therefore, one cannot possibly envisage a single threshold against which to compare the (unscaled) CUSUM statistics  $\mathbb{Y}_{s_m, e_m}^{b(i)}$  to test the stationarity of both these processes; the reason is that either process will require its own, different threshold, whose magnitude will have to depend on the unknown  $E(I_{t,T}^{(i)})$ . This would make threshold selection challenging.

To fix this, we proposed to first divide our CUSUM statistic by the local sample mean of  $I_{t,T}^{(i)}$  before performing the test against the threshold. This partly remedies the issue mentioned above in the sense that now the variance of the (scaled) CUSUM does not substantially depend on  $E(I_{t,T}^{(i)})$ , which allows  $\omega_T$  to be a *constant* multiple of  $\log(T)$ . This makes it relatively “easy” to select so that it works well for a range of processes, which we demonstrate empirically in the paper. It works in theory too, in the sense that it enables the result in our Theorem 1.

However, even after the division by  $q_{s_m, e_m}$ , the variance stabilization is not exact. This is because of the autocorrelation in the sequence  $\{I_{t,T}^{(i)}\}_t$ , which the variance of the rescaled statistic  $\mathbb{Y}_{s_m, e_m}^{b(i)} / q_{s_m, e_m}$  will still depend on, albeit (much) less strongly than that of the unscaled CUSUM  $\mathbb{Y}_{s_m, e_m}^{b(i)}$ .

To illustrate the essence of the problem and appreciate that the variance stabilization as described above is still desirable despite this problem, we consider the following simplified illustrative example. We take  $s_m = 1$ ,  $e_m = T$ ,  $b = T/2$ , and postulate that the autocorrelation in the periodogram sequence  $I_{t,T}^{(i)}$  resembles that of (squared) autoregression of order 1. To be more precise, we set

$$I_{t,T}^{(i)} = D_t^2; \quad D_t = a D_{t-1} + \varepsilon_t,$$

where  $a \in (-1, 1)$  and  $\varepsilon_t$  is a i.i.d. sequence distributed as  $N(0, \sigma^2)$ . While this is a very specific choice (and does not necessarily correspond to any particular original process on which the periodogram may have been computed), it will serve our illustrative purposes well, since it describes the entire autocorrelation structure by the single parameter  $a$ , so it is a convenient device for demonstrating *how the variance stabilization is impacted by increasing autocorrelation in the periodogram sequence*. Even with this simple choice, the demonstration will have to proceed via approximate arguments, as exact distributional arguments would be challenging to obtain.

With this example, we have

$$\mathbb{Y}_{1,T}^{T/2} / q_{1,T} = \sqrt{T} \frac{D_1^2 + \dots + D_{T/2}^2 - (D_{T/2+1}^2 + \dots + D_T^2)}{D_1^2 + \dots + D_T^2} = \sqrt{T} \frac{Z_1 - Z_2}{Z_1 + Z_2},$$

where

$$Z_i = D_{1+(i-1)T/2}^2 + \dots + D_{T/2+(i-1)T/2}^2.$$

Our aim is to examine the effect of the parameter  $a$  on  $\text{Var}(\mathbb{Y}_{1,T}^{T/2}/q_{1,T})$ . Some of the arguments below are approximate in nature. As suggested in Brown (1975), we approximate the distribution of  $Z_1$  (identical to the distribution of  $Z_2$ , of course) as  $\alpha\chi_\beta^2$ , and note

$$\begin{aligned} E(\alpha\chi_\beta^2) &= \alpha\beta \\ \text{Var}(\alpha\chi_\beta^2) &= 2\alpha^2\beta. \end{aligned}$$

Using simple algebra, we have

$$\begin{aligned} E(Z_1) &= \frac{T\sigma^2}{2(1-a^2)} \\ \text{Var}(Z_1) &\approx \frac{T\sigma^4(1+a^2)}{(1-a^2)^3}. \end{aligned}$$

Matching the respective moments gives

$$\begin{aligned} \alpha &= \frac{\sigma^2(1+a^2)}{(1-a^2)^2} \\ \beta &= \frac{T(1-a^2)}{2(1+a^2)}. \end{aligned}$$

For larger values of  $T$  (longer intervals are particularly interesting to us as it is unlikely that change-points detected by the WBS algorithm will come from the examination of short intervals), the variables  $Z_1$  and  $Z_2$  are approximately independent. Therefore, by Lemma 1 in Fryzlewicz et al. (2006), the distribution of  $\mathbb{Y}_{1,T}^{T/2}/q_{1,T}$  can be approximated as

$$\sqrt{T} \left( 2 \text{Beta} \left\{ \frac{T(1-a^2)}{4(1+a^2)}, \frac{T(1-a^2)}{4(1+a^2)} \right\} - 1 \right),$$

which leads (by standard results on the beta distribution) to

$$\text{Var}(\mathbb{Y}_{1,T}^{T/2}/q_{1,T}) \approx \frac{2(1+a^2)}{1-a^2}. \quad (1.3)$$

While this expression is independent of the scale parameter  $\sigma^2$ , it does depend to some extent on the autocorrelation parameter  $a$ . However, we now compare it to the variance of the *unscaled* CUSUM, approximated by similar arguments as

$$\text{Var}(\mathbb{Y}_{1,T}^{T/2}) \approx \frac{2\sigma^4(1+a^2)}{(1-a^2)^3}.$$

Unsurprisingly, it depends also on  $\sigma$ . However, even if  $\sigma$  is kept constant, the interesting aspect of this expression is that is (much) more dependent on  $a$  (in the sense of being further away from a constant function) than  $\text{Var}(\mathbb{Y}_{1,T}^{T/2}/q_{1,T})$ , which is illustrated in Figure 1.8. Therefore, even abstracting from the scale parameter  $\sigma$ , the division by  $q_{1,T}$  provides a substantial degree of variance stabilization in the sense of making  $\text{Var}(\mathbb{Y}_{1,T}^{T/2}/q_{1,T})$  “less dependent” on the degree of autocorrelation in  $I_{i,T}^{(i)}$  than  $\text{Var}(\mathbb{Y}_{1,T}^{T/2})$ .

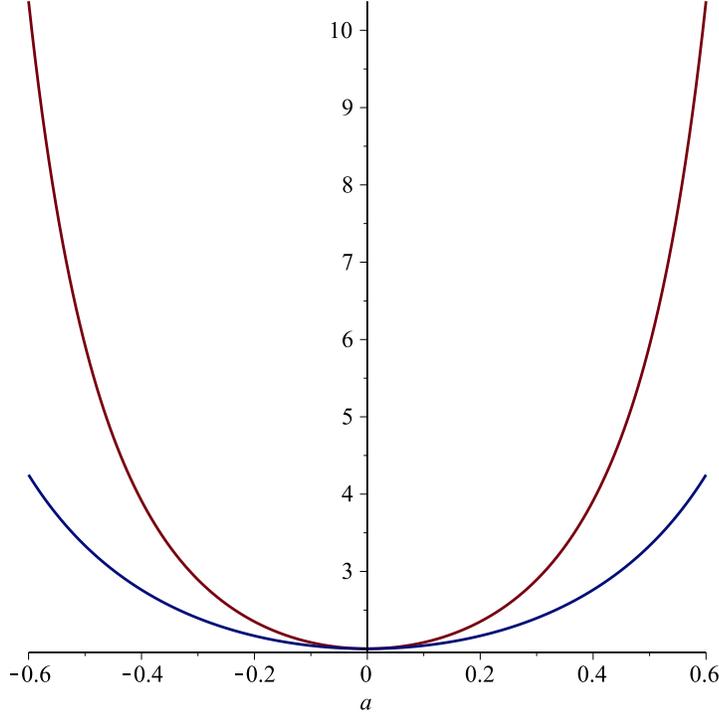


Figure 1.8: Red line:  $\frac{2(1+a^2)}{(1-a^2)^3}$ ; blue line:  $\frac{2(1+a^2)}{1-a^2}$ .

## 5 Proofs of theoretical results

### Proof of Theorem 1

The proof of consistency is based on the following multiplicative model

$$\tilde{Y}_{t,T} = \sigma(t/T)^2 Z_{t,T}^2, \quad t = 0, \dots, T-1.$$

We define the following two CUSUM statistics

$$\mathbb{Y}_{s,e}^b = \sqrt{\frac{e-b}{n(b-s+1)}} \sum_{t=s}^b \tilde{Y}_{t,T} - \sqrt{\frac{b-s+1}{n(e-b)}} \sum_{t=b+1}^e \tilde{Y}_{t,T}$$

and

$$\mathbb{S}_{s,e}^b = \sqrt{\frac{e-b}{n(b-s+1)}} \sum_{t=s}^b \sigma^2(t/T) - \sqrt{\frac{b-s+1}{n(e-b)}} \sum_{t=b+1}^e \sigma^2(t/T)$$

where  $n = e - s + 1$ , the size of the segment defined by  $(s, e)$ .

$\mathbb{Y}_{s,e}^b$  can be seen as the inner product between sequence  $\{\tilde{Y}_{t,T}\}_{t=s,\dots,e}$  and a vector  $\psi_{s,e}^b$  whose elements  $\psi_{s,e,t}^b$  are constant and positive for  $t \leq b$  and constant and negative for  $t > b$  such that they sum to zero and sum to one when squared. Similarly for  $\mathbb{S}_{s,e}^b$ .

Let  $s, e$  satisfy  $\eta_{p_0} \leq s < \eta_{p_0+1} < \dots < \eta_{p_0+q} < e \leq \eta_{p_0+q+1}$  for  $0 \leq p_0 \leq N - q$ . The inequality will hold at all stages of the algorithm until no undetected change-points are remained. We impose at least one of the following conditions

$$s < \eta_{p_0+r'} - C\delta_T < \eta_{p_0+r'} + C\delta_T < e, \quad \text{for some } 1 \leq r' \leq q \quad (1.4)$$

$$\{(\eta_{p_0+1} - s) \wedge (s - \eta_{p_0})\} \vee \{(\eta_{p_0+q+1} - e) \wedge (e - \eta_{p_0+q})\} \leq C\epsilon_T \quad (1.5)$$

where  $\wedge$  and  $\vee$  denote the minimum and maximum operators, respectively. These inequalities will hold throughout the algorithm until no further change-points are detected.

We define symmetric intervals  $\mathcal{I}_r^L$  and  $\mathcal{I}_r^R$  around change-points such that for every triplet  $\{\eta_{r-1}, \eta_r, \eta_{r+1}\}$

$$\mathcal{I}_r^L = \left[ \eta_r - \frac{2}{3}\delta_{\min}^r, \eta_r - \frac{1}{3}\delta_{\min}^r (1 + \bar{c}) \right]$$

and

$$\mathcal{I}_r^R = \left[ \eta_r + \frac{1}{3}\delta_{\min}^r (1 + \bar{c}), \eta_r + \frac{2}{3}\delta_{\min}^r \right] \quad \text{for } r = 1, \dots, N + 1$$

where  $\delta_{\min}^r = \min\{\eta_r - \eta_{r-1}, \eta_{r+1} - \eta_r\}$  and  $\bar{c} = 3 - \frac{2}{c_\star}$  for  $c_\star$  as in formula (1.5) of the main paper. We recall that at every stage of the WBS algorithm  $M$  intervals  $(s_m, e_m)$ ,  $m = 1, \dots, M$  are drawn from a discrete uniform distribution over the set  $\{(s, e) : s < e, 0 \leq s \leq T - 2, 1 \leq e \leq T - 1\}$ .

We define the event  $D_T^M$  as

$$D_T^M = \{\forall r = 1, \dots, N \exists m = 1, \dots, M (s_m, e_m) \in \mathcal{I}_r^L \times \mathcal{I}_r^R\}.$$

Also, note that

$$P((D_T^M)^c) \leq \sum_{r=1}^N \prod_{m=1}^M (1 - P((s_m, e_m) \in \mathcal{I}_r^L \times \mathcal{I}_r^R)) \leq \frac{T}{\delta_T} (1 - \delta_T^2 (1 - \bar{c})^2 T^{-2} / 9)^M. \quad (1.6)$$

On a generic interval satisfying (1.4) and (1.5) we consider

$$(m_0, b) = \arg \max_{(m,t): m \in \mathcal{M}_{s,e}, s_m \leq t \leq e_m} |\tilde{Y}_{s_m, e_m}^t| \quad (1.7)$$

where  $\mathcal{M}_{s,e} = \{m : (s_m, e_m) \subseteq (s, e), 1 \leq m \leq M\}$ .

**Lemma 1**

$$P \left( \max_{(s,b,e)} \left| \mathbb{Y}_{s,e}^b - \mathbb{S}_{s,e}^b \right| > \lambda_1 \right) \rightarrow 0 \quad (1.8)$$

for

$$\lambda_1 \geq \log T.$$

**Proof:** We start by studying the following event

$$\left| \sum_{t=s}^e c_t \sigma(t/T)^2 (Z_{t,T}^2 - 1) \right| > \sqrt{n} \lambda_1$$

where  $c_t = \sqrt{(e-b)/(b-s+1)}$  and  $c_t = \sqrt{(b-s+1)/(e-b)}$  for  $t \leq b$  and  $b+1 \leq t$  respectively. From formula (1.5) of the main paper, we have that  $c_t \leq c_\star \equiv \sqrt{\frac{c_\star}{1-c_\star}} < \infty$ . The proof proceeds as in Cho and Fryzlewicz (2015) and we have that (1.8) is bounded by

$$\sum_{(s,b,e)} 2 \exp\left(-\frac{n\lambda_1^2}{4c_\star^2 \max_z \sigma^2(z)n\rho_\infty^2 + 2c_\star \max_z \sigma(z)\sqrt{n}\lambda_1\rho_\infty^1}\right) \leq 2T^3 \exp(-C'_1(c_\star^{-2}) \log^2 T)$$

which converges to 0 since  $n \geq \delta_T = \mathcal{O}(\log^2 T)$  and  $\rho_\infty^1 < \infty$  from (A2).

**Lemma 2** *Assuming that (1.4) holds, then there exists  $C_2 > 0$  such that for  $b$  satisfying  $|b - \eta_{p_0+r'}| = C_2\gamma_T$  for some  $r'$ , we have  $|\mathbb{S}_{s_{m_0}, e_{m_0}}^{\eta_{p_0+r'}}| \geq |\mathbb{S}_{s_{m_0}, e_{m_0}}^b| + C\gamma_T\delta_T^{-1/2} \geq |\mathbb{S}_{s_{m_0}, e_{m_0}}^b| + 2\lambda_1$ , where  $\gamma_T = \sqrt{\delta_T}\lambda_1$ .*

**Proof:** From the proof of Theorem 3.2 in Fryzlewicz (2014) and Lemma 1 in Cho and Fryzlewicz (2012) we have the following result

$$|\mathbb{S}_{s_{m_0}, e_{m_0}}^b| \geq |\mathbb{Y}_{s_{m_0}, e_{m_0}}^b| - \lambda_1 \geq C_3\sqrt{\delta_T} \quad (1.9)$$

provided that  $\delta_T \geq C_4\lambda_1^2$ .

By Lemma 2.2 in Venkatraman (1992) there exists a change-point  $\eta_{p_0+r'}$  immediately to the left or right of  $b$  such that

$$|\mathbb{S}_{s_{m_0}, e_{m_0}}^{\eta_{p_0+r'}}| > |\mathbb{S}_{s_{m_0}, e_{m_0}}^b| \geq C_3\sqrt{\delta_T}.$$

Now, the following three cases are not possible:

1.  $(s_{m_0}, e_{m_0})$  contains a single change-point,  $\eta_{p_0+r'}$ , and both  $\eta_{p_0+r'} - s_{m_0}$  and  $e_{m_0} - \eta_{p_0+r'}$  are not bounded from below by  $c_1\delta_T$ .
2.  $(s_{m_0}, e_{m_0})$  contains a single change-point,  $\eta_{p_0+r'}$ , and either  $\eta_{p_0+r'} - s_{m_0}$  or  $e_{m_0} - \eta_{p_0+r'}$  are not bounded from below by  $c_1\delta_T$ .
3.  $(s_{m_0}, e_{m_0})$  contains two change-points,  $\eta_{p_0+r'}$  and  $\eta_{p_0+r'+1}$ , and both  $\eta_{p_0+r'} - s_{m_0}$  and  $e_{m_0} - \eta_{p_0+r'+1}$  are not bounded from below by  $c_1\delta_T$ .

The first case is not permitted by (A5). For the last two, if either case were true, then following the arguments as in Lemma A.5 of Fryzlewicz (2014), we would obtain that  $\max_{t: s_{m_0} \leq t \leq e_{m_0}} |\mathbb{S}_{s_{m_0}, e_{m_0}}^t|$  was not bounded from below by  $C_3\sqrt{\delta_T}$  which contradicted (1.9). Hence, interval  $(s_{m_0}, e_{m_0})$  satisfies condition (1.4) and following a similar argument to the proof of Lemma 2 in Cho and Fryzlewicz (2012) we can show that for any  $b$  satisfying  $|b - \eta_{p_0+r'}| = C_2\gamma_T$ , we have  $|\mathbb{S}_{s_{m_0}, e_{m_0}}^{\eta_{p_0+r'}}| \geq |\mathbb{S}_{s_{m_0}, e_{m_0}}^b| + C\gamma_T\delta_T^{-1/2}$ .

**Lemma 3** *Under conditions (1.4) and (1.5) there exists  $1 \leq r' \leq q$  such that  $|b - \eta_{p_0+r'}| \leq \epsilon_T$ , where  $b$  is given in (1.7) and  $\epsilon_T = C \log^2 T$  for a positive constant  $C$ .*

**Proof:** First, we mention that the model from formula (1.1) of the main paper can be written as  $\tilde{Y}_{t,T} = \sigma(t/T)^2 + \sigma(t/T)^2(Z_{t,T}^2 - 1)$  which has the form of a signal+noise model i.e.  $Y_t = f_t + \varepsilon_t$ . Now, let  $\bar{f}_{s_{m_0}, e_{m_0}}^d$  define the best function approximation to  $f_t$  such that  $\arg \max_d |\langle \psi_{s_{m_0}, e_{m_0}}^d, f \rangle| = \arg \min_d \sum_{t=s_{m_0}}^{e_{m_0}} (f_t - \bar{f}_{s_{m_0}, e_{m_0}}^d)$  where  $\bar{f}_{s_{m_0}, e_{m_0}}^d = \bar{f} + \langle f, \psi_{s_{m_0}, e_{m_0}}^d \rangle \psi_{s_{m_0}, e_{m_0}}^d$ ,  $\bar{f}$  is the mean of  $f$  and  $\psi_{s_{m_0}, e_{m_0}}^d$  is a set of vectors that are constant and positive until  $d$  and then constant and negative from  $d+1$  until  $e_{m_0}$ .

If it can be shown that for a certain  $\epsilon_T < C_2 \gamma_T$ , we have

$$\sum_{t=s_{m_0}}^{e_{m_0}} (Y_t - \bar{Y}_{s_{m_0}, e_{m_0}, t}^d)^2 > \sum_{t=s_{m_0}}^{e_{m_0}} (Y_t - \bar{f}_{s_{m_0}, e_{m_0}, t}^{\eta_{p_0+r'}})^2 \quad (1.10)$$

as long as

$$\epsilon_T \leq |d - \eta_{p_0+r'}|$$

then this would prove necessarily that  $|b - \eta_{p_0+r'}| \leq \epsilon_T$ .

By Lemma 2 and Lemma A.3 in Fryzlewicz (2014), we have the same triplet of inequalities as in the argument in the proof of Theorem 3.2 in Fryzlewicz (2014) i.e.

$$|d - \eta_{p_0+r'}| \geq C(\lambda_2 |d - \eta_{p_0+r'}| \delta_T^{-1/2}) \vee (\lambda_2 |d - \eta_{p_0+r'}|^{-1/2}) \vee (\lambda_2^2). \quad (1.11)$$

Hence, with the requirement that  $|d - \eta_{p_0+r'}| \leq C_2 \gamma_T = C_2 \lambda_1 \sqrt{\delta_T}$  we obtain

$$\delta_T > C^2 \lambda_2^2 \max(C^2 C_2^{-2} \lambda_1^{-2} \lambda_2^2, 1)$$

and  $\epsilon_T = \max(1, C^2) \lambda_2^2$ . From Lemma 1  $\lambda_1$  is of order  $\mathcal{O}(\log T)$ . For  $\lambda_2$ , which appears in the following two terms of the decomposition of (1.10)

$$I = \frac{1}{d - s_{m_0} + 1} \left( \sum_{t=s_{m_0}}^d \varepsilon_t \right)^2 \quad \text{and} \quad II = \frac{1}{e_{m_0} - d + 1} \left( \sum_{t=d+1}^{e_{m_0}} \varepsilon_t \right)^2$$

we show below that with probability tending to 1,  $I \leq \lambda_2^2 = \log^2 T$ . From Lemma 1 we have that  $c_t = 1$  for  $t = s_{m_0}, \dots, d$  and thus

$$P \left( \frac{1}{\sqrt{d - s_{m_0} + 1}} \left| \sum_{t=s_{m_0}}^d \varepsilon_t \right| > \lambda_2 \right) \rightarrow 0$$

since by the Bernstein inequality the probability is bounded by

$$2T^2 \exp \left( - \frac{(d - s_{m_0} + 1) \lambda_2^2}{4 \max_z \sigma^2(z) (d - s_{m_0} + 1) \rho_\infty^2 + 2c' \max_z \sigma(z) \sqrt{d - s_{m_0} + 1} \lambda_2 \rho_\infty^1} \right) \leq 2T^2 \exp(-C'_3 \lambda_2^2)$$

which converges to 0 due to  $(d - s_{m_0} + 1) = \mathcal{O}(\delta_T)$  from formula (1.5) of the main paper. Note that  $II$  has similar order and we omit the details. This concludes the lemma.

**Lemma 4** *Under conditions (1.4) and (1.5)*

$$P\left(\left|\mathbb{Y}_{s_{m_0}, e_{m_0}}^b\right| > \omega_T \frac{\sum_{t=s_{m_0}}^{e_{m_0}} \tilde{Y}_t}{n_{m_0}}\right) \rightarrow 1$$

where  $b$  is given in (1.7).

**Proof:** We define the following two events  $\mathcal{A} = \left\{ \left| \mathbb{Y}_{s_{m_0}, e_{m_0}}^b \right| < \omega_T \frac{1}{n_{m_0}} \sum_{t=s_{m_0}}^{e_{m_0}} \tilde{Y}_{t,T} \right\}$  and  $\mathcal{B} = \left\{ \frac{1}{n_{m_0}} \left| \sum_{t=s_{m_0}}^{e_{m_0}} \tilde{Y}_{t,T} - \sum_{t=s_{m_0}}^{e_{m_0}} \sigma(t/T)^2 \right| < \bar{\sigma} = \frac{1}{2n_{m_0}} \sum_{t=s_{m_0}}^{e_{m_0}} \sigma^2(t/T) \right\}$ .

Since  $P(\mathcal{A}) \leq P(\mathcal{A} \cap \mathcal{B}) + P(\mathcal{B}^c)$  we need to show that  $P(\mathcal{B}) \rightarrow 1$  and  $P(\mathcal{A} \cap \mathcal{B}) \rightarrow 0$ . To show that  $P(\mathcal{B}) = P\left(\frac{1}{n_{m_0}} \sum_{t=s_{m_0}}^{e_{m_0}} \tilde{Y}_{t,T} \in (\bar{\sigma}/2, 3\bar{\sigma}/2)\right) \rightarrow 1$  we apply the Bernstein inequality as in Lemma 1 and we have that

$$P(\mathcal{B}') = P\left(\frac{1}{n_{m_0}} \left| \sum_{t=s_{m_0}}^{e_{m_0}} \tilde{Y}_{t,T} - \sum_{t=s_{m_0}}^{e_{m_0}} \sigma(t/T)^2 \right| > \bar{\sigma}\right) = P\left(\left| \sum_{t=s_{m_0}}^{e_{m_0}} \sigma(t/T)^2 (Z_{t,T}^2 - 1) \right| > n_{m_0} \bar{\sigma}\right).$$

Hence,

$$P(\mathcal{B}') \leq 2 \exp\left(-\frac{n_{m_0}^2 \bar{\sigma}^2}{4 \max_z \sigma^2(z) n_{m_0} \rho_\infty^2 + 2C' \max_z \sigma(z) n_{m_0} \bar{\sigma} \rho_\infty^1}\right) \leq 2T^2 \exp(-C'_4 \log^2 T)$$

which converges to 0 since  $n_{m_0} \geq \delta_T = \mathcal{O}(\log^2 T)$  and  $\rho_\infty^1 < \infty$  from (A2).

Now, from Lemma (3), we have some  $\eta \equiv \eta_{p_0+r'}$  satisfying  $|b - \eta| \leq C\epsilon_T$ . Turning to  $P(\mathcal{A} \cap \mathcal{B})$  we have from conditions (1.4) and (1.5)

$$\begin{aligned} |\mathbb{Y}_{s_{m_0}, e_{m_0}}^b| \geq |\mathbb{Y}_{s_{m_0}, e_{m_0}}^\eta| &\geq |\mathbb{S}_{s_{m_0}, e_{m_0}}^\eta| - \log T \\ &= \left| \sqrt{\frac{(\eta - s_{m_0} + 1)(e_{m_0} - \eta)}{n_{m_0}}} \left( \sigma\left(\frac{\eta}{T}\right)^2 - \sigma\left(\frac{\eta+1}{T}\right)^2 \right) \right| - \log T \\ &= \sqrt{\frac{e_{m_0} - \eta}{n_{m_0}(\eta - s_{m_0} + 1)}} (\eta - s_{m_0} + 1) \sigma_* - \log T \geq C\sqrt{\delta_T} - \log T > \omega_T 3\bar{\sigma}/2, \end{aligned}$$

which concludes the Lemma.

**Lemma 5** *For some positive constants  $C, C'$ , let  $s, e$  satisfy either*

- $\exists 1 \leq p \leq N$  such that  $s \leq \eta_p \leq e$  and  $(\eta_p - s + 1) \wedge (e - \eta_p) \leq C\epsilon_T$  or
- $\exists 1 \leq p \leq N$  such that  $s \leq \eta_{p+1} \leq e$  and  $(\eta_p - s + 1) \vee (e - \eta_{p+1}) \leq C'\epsilon_T$ .

Then,

$$P\left(\left|\mathbb{Y}_{s_{m_0}, e_{m_0}}^b\right| < \omega_T \frac{\sum_{t=s_{m_0}}^{e_{m_0}} Y_t}{n_{m_0}}\right) \rightarrow 1$$

where  $b$  is given in (1.7).

**Proof:** A similar argument to the proof of Lemma 5 is applied here. We only need to show that  $P(\mathcal{A} \cap \mathcal{B}) \rightarrow 0$  where now event  $\mathcal{A} = \left\{ |\mathbb{Y}_{s_{m_0}, b, e_{m_0}}^b| > \omega_T \frac{1}{n_{m_0}} \sum_{t=s_{m_0}}^{e_{m_0}} \tilde{Y}_{t,T} \right\}$ . Using condition (i) or (ii) we have that

$$\begin{aligned} |\mathbb{Y}_{s_{m_0}, e_{m_0}}^b| &\leq |\mathbb{S}_{s_{m_0}, e_{m_0}}^b| + \log T \\ &= \left| \frac{\sqrt{b - s_{m_0} + 1} \sqrt{e_{m_0} - b}}{\sqrt{n_{m_0}}} (\sigma^2(b/T) - \sigma^2((b+1)/T)) \right| + \log T \\ &\leq \sigma^* C \sqrt{\epsilon_T} + \log T < \omega_T \bar{\sigma} / 2. \end{aligned}$$

The proof of Theorem 1 proceeds as follows: at the start of the algorithm when  $s = 0$  and  $e = T - 1$  all the conditions of (1.4) & (1.5) required by Lemma 4 are met and thus it detects a change-point on that interval defined by formula (1.7) within the distance of  $C\epsilon_T$  (by Lemma 3). The conditions of Lemma 4 are satisfied until all change-points have been identified. Then, every random interval  $(s_m, e_m)$  does not contain a change-point or the conditions of Lemma 5 are met; hence no more change-points are detected and the algorithm stops.

Finally, we examine whether the bias present in  $\mathbb{E}I_{t,T}^{(i)}$  (see condition (A0)) will affect the above result. We define  $\tilde{\mathbb{S}}_{s,e}^t$  similarly to  $\mathbb{S}_{s,e}^t$  by replacing  $\sigma(t/T)^2$  with  $\sigma_{t,T}^2$ . Assume that  $\eta_r$  is a change-point within the interval  $[s_{m_0}, e_{m_0}]$  and  $b = \arg \max_{t \in (s_{m_0}, e_{m_0})} |\mathbb{S}_{s_{m_0}, e_{m_0}}^b|$  and  $\hat{b} = \arg \max_{t \in (s_{m_0}, e_{m_0})} |\tilde{\mathbb{S}}_{s_{m_0}, e_{m_0}}^b|$ . Recall that  $\mathbb{E}I_{t,T}^{(i)}$  is constant within each segment apart from short intervals around true change-point  $\eta_r$  i.e.  $[\eta_r - K2^{-i}, \eta_r + K2^{-i}]$ . In addition, from Theorem 2 in Cho and Fryzlewicz (2015) the finest scale should satisfy  $i \geq I^* = -\lfloor \alpha \log \log T \rfloor$  in order for (A4) to hold. Then,  $|\hat{b} - b| \leq K2^{I^*} < \epsilon_T$  holds since  $I^* = \mathcal{O}(\log \log T)$ . Therefore, bias does not affect the above result and the consistency is preserved.

### Proof of Theorem 2

We start by the first method of aggregation. From the invertibility of the autocorrelation wavelet inner product matrix  $A$ , there exists at least one ordinate of wavelet periodogram in which a change-point  $\theta_r$  is detected. From Theorem 1 it holds that  $|\theta_r - \hat{\theta}_r| \leq C\epsilon_T$  with probability converging to 1 regardless of the scale  $i$ . Since the algorithm begins its search from the finest scale and only proceeds to the next one if no change-point is detected (until scale  $I^*$ ) then consistency is preserved.

We now turn to the second method of aggregation. We note that  $\mathbb{Y}_t^{thr}$  has the same functional form with each of  $\mathcal{Y}_t^{(i)}$  i.e.  $h^{(i)}(x) = (x(1-x))^{-1/2}(c_x^{(i)}x + d_x^{(i)}x)$  for  $x = (t - s_m + 1)/n \in (0, 1)$ , where  $c_x^{(i)}, d_x^{(i)}$  are determined by the location and the magnitude of the change-points of  $I_{t,T}^{(i)}$ . Let  $b = \arg \max_{s_{m_0} < t < e_{m_0}} \mathbb{Y}_t^{thr}$ ; then following a similar argument to Lemma 2 of Fryzlewicz (2014) we can show that  $\mathbb{Y}_t^{thr}$  must have a local maximum at  $t = \theta_{p_0+r'}$  and that  $|b - \theta_{p_0+r'}| \leq C_5\gamma_T$ . With this result, we can show that  $|b - \theta_{p_0+r}| \leq C'\epsilon_T$  for some  $1 \leq r' \leq q$  as in Lemma 3 above by constructing a signal+noise model  $y_t = f_t + \varepsilon_t$  and substituting  $f_t$  with  $\sum_{i=-I^*}^{-1} \mathbb{E}I_{t,T}^{(i)} \mathbb{I}(\mathcal{Y}_t^{(i)} > \omega_T^{(i)})/q_{s_m, e_m}^{(i)}$ . Then, conditions (1.4) and (1.5) are satisfied within each segment for at least one scale  $i \in \{-1, \dots, -I^*\}$ . When all change-points have been detected every subsequent random interval  $(s_m, e_m)$  will satisfy the conditions of Lemma 5 for every  $i \in \{-1, \dots, -I^*\}$  and the algorithm stops.

## References

- M. Brown. A method for combining non-independent, one-sided tests of significance. *Biometrics*, 31:987–992, 1975.
- H. Cho and P. Fryzlewicz. Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, 22:207–229, 2012.
- H. Cho and P. Fryzlewicz. Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. *Journal of the Royal Statistical Society Series B*, 77: 475–507, 2015.
- P. Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42:2243–2281, 2014.
- P. Fryzlewicz, T. Sapatinas, and S. Rao. A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, 93(3):687–704, 2006.
- R. Killick, I. Eckley, and P. Jonathan. A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electronic Journal of Statistics*, 7: 1167–1183, 2013.
- K. Korkas and P. Fryzlewicz. Multiple change-point detection for non-stationary time series using wild binary segmentation. *Preprint*, 2014.
- M. Nunes, S. Taylor, and I. Eckley. A multiscale test of spatial stationarity for textured images in  $r$ . *The R Journal*, 6:20–30, 2014.
- E. Venkatraman. *Consistency results in multiple change-point problems*. PhD thesis, Department of Statistics. Stanford University, 1992.

Department of Statistics, London School of Economics, Columbia House, LSE, Houghton Street, London, WC2A 2AE, UK

E-mail: k.k.korkas@lse.ac.uk

Department of Statistics, London School of Economics, Columbia House, LSE, Houghton Street, London, WC2A 2AE, UK

E-mail: p.fryzlewicz@lse.ac.uk