Multiscale and multilevel technique for consistent segmentation of nonstationary time series

Haeran Cho and Piotr Fryzlewicz

Department of Statistics, London School of Economics, UK.

1 The proof of Theorem 1

The consistency of our algorithm is first proved for the sequence below,

$$\tilde{Y}_{t,T}^2 = \sigma^2(t/T) \cdot Z_{t,T}^2, \ t = 0, \ldots, T-1. \tag{1.1}$$

Note that unlike in (3), the above model features the true piecewise constant $\sigma^2(t/T)$. Denote $n = e - s + 1$ and define

$$\tilde{\mathbb{Y}}_{s,e}^b \quad = \quad \frac{\sqrt{e-b}}{\sqrt{n}\sqrt{b-s+1}} \sum_{t=s}^{b} \tilde{Y}_{t,T}^2 - \frac{\sqrt{b-s+1}}{\sqrt{n}\sqrt{e-b}} \sum_{t=b+1}^{e} \tilde{Y}_{t,T}^2.$$

$\tilde{\mathbb{S}}_{s,e}^b$ and $\mathbb{S}_{s,e}^b$ are defined similarly, replacing $\tilde{Y}_{t,T}^2$ with $\sigma^2(t/T)$ and $\sigma_{t,T}^2$, respectively. Note that the above are simply inner products of the respective sequences and a vector whose support starts at s, is constant and positive until b, then constant negative until e, and normalised such that it sums to zero and sums to one when squared. Let s, e satisfy $\eta_{p_0} \le s < \eta_{p_0+1} < \ldots < \eta_{p_0+q} < e \le \eta_{p_0+q+1}$ for $0 \le p_0 \le B - q$, which will always be the case at all stages of the algorithm. In Lemmas 1–5 below, we impose at least one of following conditions:

$$s < \eta_{p_0+r} - C\delta_T < \eta_{p_0+r} + C\delta_T < e \text{ for some } 1 \le r \le q, \tag{1.2}$$

$$\{(\eta_{p_0+1} - s) \wedge (s - \eta_{p_0})\} \vee \{(\eta_{p_0+q+1} - e) \wedge (e - \eta_{p_0+q})\} \le C\epsilon_T, \tag{1.3}$$

where $\wedge$ and $\vee$ are the minimum and maximum operators, respectively and C denotes a generic positive constant. We remark that both conditions (1.2) and (1.3) hold throughout the algorithm for all those segments starting at s

and ending at e which contain previously undetected breakpoints. As Lemma 6 concerns the case when all breakpoint have already been detected, it does not use either of these conditions.

The proof of the theorem is constructed as follows. Lemma 1 is used in the proof of Lemma 2, which in turn is used alongside Lemma 3 in the proof of Lemma 4. From the result of Lemma 4, we derive Lemma 5 and finally, Lemmas 5 and 6 are used to prove Theorem 1.

Lemma 1. Let s and e satisfy (1.2), then there exists $1 \leq r^* \leq q$ such that

$$\left|\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r^*}}\right| = \max_{s<t<e} |\tilde{\mathbb{S}}_{s,e}^t| \geq C\delta_T/\sqrt{T}. \tag{1.4}$$

Proof. The equality in (1.4) is proved by Lemmas 2.2 and 2.3 of Venkatraman (1993). For the inequality part, we note that in the case of a single breakpoint in $\sigma^2(z)$, r in (1.2) coincides with r^* and we can use the constancy of $\sigma^2(z)$ to the left and to the right of the breakpoint to show that

$$\left|\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}}\right| = \left|\frac{\sqrt{\eta_{p_0+r} - s + 1}\sqrt{e - \eta_{p_0+r}}}{\sqrt{n}} \left(\sigma^2(\eta_{p_0+r}/T) - \sigma^2((\eta_{p_0+r}+1)/T)\right)\right|,$$

which is bounded from below by $C\delta_T/\sqrt{T}$. In the case of multiple breakpoints, we remark that for any r satisfying (1.2), the above order remains the same and thus (1.4) follows. □

Lemma 2. Suppose (1.2) holds, and further assume that $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} > 0$ for some $1 \leq r \leq q$. Then for b satisfying $|\eta_{p_0+r} - b| = C\epsilon_T$ and $\tilde{\mathbb{S}}_{s,e}^b < \tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}}$, we have $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} \geq \tilde{\mathbb{S}}_{s,e}^b + 2\log T$ for a large T.

Proof. Without loss of generality, assume $\eta_{p_0+r} < b$. As in Lemma 1, we first derive the result in the case of a single breakpoint in $\sigma^2(z)$. The following holds;

$$\tilde{\mathbb{S}}_{s,e}^b = \frac{\sqrt{\eta_{p_0+r} - s + 1}\sqrt{e - b}}{\sqrt{e - \eta_{p_0+r}}\sqrt{b - s + 1}}\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}}, \text{ and} \tag{1.5}$$

$$\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} - \tilde{\mathbb{S}}_{s,e}^{b} = \left(1 - \frac{\sqrt{\eta_{p_0+r} - s + 1}\sqrt{e-b}}{\sqrt{e-\eta_{p_0+r}}\sqrt{b-s+1}}\right)\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}}$$

$$= \frac{\sqrt{1+\frac{b-\eta_{p_0+r}}{\eta_{p_0+r}-s+1}} - \sqrt{1-\frac{b-\eta_{p_0+r}}{e-\eta_{p_0+r}}}}{\sqrt{1+\frac{b-\eta_{p_0+r}}{\eta_{p_0+r}-s+1}}}\cdot \tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} \tag{1.6}$$

$$\geq \frac{(1+\frac{c_1\epsilon_T}{2\delta_T}) - (1+\frac{c_2\epsilon_T}{2\delta_T}) + o(\frac{\epsilon_T}{\delta_T})}{\sqrt{2}}\cdot \tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} \geq C\frac{\epsilon_T}{\delta_T}\cdot\frac{\delta_T}{\sqrt{T}} \geq 2\log T$$

for a large T, where c_1 and c_2 are positive constants. The Taylor expansion is applied in the last but one step, and Lemma 1 in the last step. Similar arguments are applicable when $b < \eta_{p_0+r}$. Since the order of (1.5) remains the same in the case of multiple breakpoints, the lemma is proved. □

Lemma 3. $\left|\tilde{\mathbb{Y}}_{s,e}^{b} - \tilde{\mathbb{S}}_{s,e}^{b}\right| \leq \log T$ with probability converging to 1 with T uniformly over $(s,b,e) \in \mathcal{D}$, where, for $c \in [1/2, 1)$,

$$\mathcal{D} := \left\{1 \leq s < b < e \leq T;\ e-s+1 \geq C\delta_T,\ \max\left\{\sqrt{\frac{b-s+1}{e-b}}, \sqrt{\frac{e-b}{s-b+1}}\right\} \leq c\right\}.$$

Proof. We need to show that

$$\mathbf{P}\left(\max_{(s,b,e)\in\mathcal{D}} \frac{1}{\sqrt{n}}\left|\sum_{t=s}^{e}\sigma^2(t/T)(Z_{t,T}^2 - 1)\cdot c_t\right| > \log T\right) \longrightarrow 0, \tag{1.7}$$

where $c_t = \sqrt{e-b}/\sqrt{b-s+1}$ for $t \in [s,b]$ and $c_t = \sqrt{b-s+1}/\sqrt{e-b}$ otherwise. Let $\{U_t\}_{t=s}^{e}$ be i.i.d. standard normal variables, $\mathbf{V} = (v_{i,j})_{i,j=1}^{n}$ with $v_{i,j} = \mathrm{cor}\,(Z_{i,T}, Z_{j,T})$, and $\mathbf{W} = (w_{i,j})_{i,j=1}^{n}$ be a diagonal matrix with $w_{i,i} = \sigma^2(t/T)\cdot c_t$ where $i = t-s+1$. By standard results (see e.g. Johnson and Kotz (1970), page 151), showing (1.7) is equivalent to showing that $\left|\sum_{t=s}^{e}\lambda_{t-s+1}(U_t^2-1)\right|$ is bounded by $\sqrt{n}\log T$ with probability converging to 1, where λ_i are eigenvalues of the matrix $\mathbf{VW}$. Due to the Gaussianity of U_t, $\lambda_{t-s+1}(U_t^2-1)$ satisfy the Cramér's condition, i.e., there exists a constant $C > 0$ such that

$$\mathbb{E}\left|\lambda_{t-s+1}(U_t^2-1)\right|^p \leq C^{p-2}p!\,\mathbb{E}\left|\lambda_{t-s+1}(U_t^2-1)\right|^2,\ p = 3, 4, \ldots.$$

Therefore we can apply Bernstein's inequality (Bosq (1998)) and obtain

$$\mathbf{P}\left(\left|\sum_{t=s}^{e}\sigma^2(t/T)(Z_{t,T}^2-1)\cdot c_t\right| > \sqrt{n}\log T\right) \leq 2\exp\left(-\frac{n\log^2 T}{4\sum_{i=1}^n \lambda_i^2 + 2\max_i|\lambda_i|C\sqrt{n}\log T}\right).$$

Note that $\sum_{i=1}^n \lambda_i^2 = \mathrm{tr}\,(\mathbf{VW})^2 \leq c^2 \max_z \sigma^4(z) n\rho_\infty^2$. Also it follows that $\max_i|\lambda_i| \leq c\max_z \sigma^2(z)\|\mathbf{V}\|$ where $\|\cdot\|$ denotes the spectral norm of a matrix, and $\|\mathbf{V}\| \leq \rho_\infty^1$ since $\mathbf{V}$ is non-negative definite. Then (1.7) is bounded by

$$\begin{aligned}&\sum_{(s,b,e)\in\mathcal{D}} \exp(-n\log^2 T/(4c^2\max_z \sigma^4(z)n\rho_\infty^2 + 2c\max_z\sigma^2(z)\sqrt{n}\log T\rho_\infty^1))\\ &\leq CT^3\exp\left(-\log^2 T\right) \to 0,\end{aligned}$$

as $\rho_\infty^p \leq C2^{I^*}$, which can be made to be of order $\log T$, since the only requirement on I^* is that it converges to infinity but no particular speed is required. Thus the lemma follows. □

Lemma 4. Assume (1.2) and (1.3). For $b = \arg\max_{s<t<e}|\tilde{\mathbb{Y}}_{s,e}^t|$, there exists $1 \leq r \leq q$ such that $|b-\eta_{p_0+r}| \leq C\epsilon_T$ for a large T.

Proof. Let $\tilde{\mathbb{S}}_{s,e} = \max_{s<t<e}|\tilde{\mathbb{S}}_{s,e}^t|$. From Lemma 3, $\tilde{\mathbb{Y}}_{s,e}^b \geq \tilde{\mathbb{S}}_{s,e} - \log T$ and $\tilde{\mathbb{S}}_{s,e}^b \geq \tilde{\mathbb{Y}}_{s,e}^b - \log T$, hence $\tilde{\mathbb{S}}_{s,e}^b \geq \tilde{\mathbb{S}}_{s,e} - 2\log T$. Assume that $|b-\eta_{p_0+r}| > C\epsilon_T$ for any r. From Lemma 2.2 in Venkatraman (1993), $\tilde{\mathbb{S}}_{s,e}^t$ is either monotonic or decreasing and then increasing on $[\eta_{p_0+r},\eta_{p_0+r+1}]$ and $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} \vee \tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r+1}} > \tilde{\mathbb{S}}_{s,e}^b$. Suppose that $\tilde{\mathbb{S}}_{s,e}^t$ is decreasing and then increasing on the interval. Then from Lemma 2, we have $b' = \eta_{p_0+r} + C\epsilon_T$ satisfying $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} - 2\log T \geq \tilde{\mathbb{S}}_{s,e}^{b'}$. Since $\tilde{\mathbb{S}}_{s,e}^t$ is locally increasing at $t=b$ (for $\tilde{\mathbb{S}}_{s,e}^b > \tilde{\mathbb{S}}_{s,e}^{b'}$), we have $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r+1}} > \tilde{\mathbb{S}}_{s,e}^b$ and there will again be a $b'' = \eta_{p_0+r+1} - C\epsilon_T$ satisfying $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} - 2\log T \geq \tilde{\mathbb{S}}_{s,e}^{b''}$. As $b'' > b$, it contradicts that $\tilde{\mathbb{S}}_{s,e}^b \geq \tilde{\mathbb{S}}_{s,e} - 2\log T$. Similar arguments are applicable when $\tilde{\mathbb{S}}_{s,e}^t$ is monotonic and therefore the lemma follows. □

Lemma 5. Under (1.2) and (1.3), $\mathbf{P}\left(\left|\tilde{\mathbb{Y}}_{s,e}^b\right| < \tau T^\theta\sqrt{\log T}\cdot n^{-1}\sum_{t=s}^e \tilde{Y}_{t,T}^2\right) \longrightarrow 0$ for $b = \arg\max_{s<t<e}|\tilde{\mathbb{Y}}_{s,e}^t|$.

Proof. From Lemma 4, there exists some r such that $|b-\eta_{p_0+r}| < C\epsilon_T$. Denote

$\tilde{d} = \tilde{\mathbb{Y}}^b_{s,e} = \tilde{d}_1 - \tilde{d}_2$ and $\tilde{m} = n^{-1/2}\sum_{t=s}^{e} \tilde{Y}^2_{t,T} = c_1\tilde{d}_1 + c_2\tilde{d}_2$, where

$$\tilde{d}_1 = \frac{\sqrt{e-b}}{\sqrt{n}\sqrt{b-s+1}}\sum_{t=s}^{b}\tilde{Y}^2_{t,T}, \quad \tilde{d}_2 = \frac{\sqrt{b-s+1}}{\sqrt{n}\sqrt{e-b}}\sum_{t=b+1}^{e}\tilde{Y}^2_{t,T}, \quad \text{and } c_1 = c_2^{-1} = \sqrt{\frac{b-s+1}{e-b}}.$$

For simplicity, let $c_2 > c_1$. Further, let $\mu_i = \mathbb{E}\tilde{d}_i$ and $w_i = \mathrm{var}(\tilde{d}_i)$ for $i = 1, 2$, and define $\mu = \mathbb{E}\tilde{d}$ and $w = \mathrm{var}(\tilde{d})$. Finally, t_n denotes the threshold $\tau T^\theta\sqrt{\log T/n}$. We need to show $\mathbf{P}(|\tilde{d}| \le \tilde{m}\cdot t_n) \to 0$. Note that $w_i \le c^2\sup_z \sigma^4(z)\rho^2_\infty$. Using Markov's and the Cauchy-Schwarz inequalities, we bound $\mathbf{P}(\tilde{d} \le \tilde{m}\cdot t_n)$ by

$$\begin{aligned}
&\mathbf{P}\left\{(\tilde{d}_1 - \mu_1)(c_1 t_n - 1) + (\tilde{d}_2 - \mu_2)(c_2 t_n + 1) + 2c_1 t_n\mu_1 + (c_2 - c_1)t_n\mu_2 \ge (1 + c_1 t_n)\mu\right\}\\
&\le 4\mu^{-2}(1 + c_1 t_n)^{-2}\left\{(c_1 t_n - 1)^2 w_1 + (c_2 t_n + 1)^2 w_2 + 4c_1^2 t_n^2\mu_1^2 + (c_2 - c_1)^2 t_n^2\mu_2^2\right\}\\
&\le O\left\{\mu^{-2}\sup_z \sigma^4(z)\left(\rho^2_\infty + \tau^2 T^{2\theta}\log T\right)\right\},
\end{aligned}$$

and since $\mu = \tilde{\mathbb{S}}^b_{s,e} = O\left(\delta_T/\sqrt{T}\right) > T^\theta\sqrt{\log T}$, the conclusion follows. □

Lemma 6. For some positive constants $C,\ C'$, let s, e satisfy either

(i) $\exists\, 1 \le p \le B$ such that $s \le \eta_p \le e$ and $[\eta_p - s + 1] \wedge [e - \eta_p] \le C\epsilon_T$ or

(ii) $\exists\, 1 \le p \le B$ such that $s \le \eta_p < \eta_{p+1} \le e$ and $[\eta_p - s + 1] \vee [e - \eta_{p+1}] \le C'\epsilon_T$.

Then for a large T,

$$\mathbf{P}\left(\left|\tilde{\mathbb{Y}}^b_{s,e}\right| > \tau T^\theta\sqrt{\log T}\cdot n^{-1}\sum_{t=s}^{e}\tilde{Y}^2_{t,T}\right) \longrightarrow 0,$$

where $b = \arg\max_{s<t<e}|\tilde{\mathbb{Y}}^t_{s,e}|$.

Proof. First we assume (i). Let $\mathcal{A} = \left\{\left|\tilde{\mathbb{Y}}^b_{s,e}\right| > \tau T^\theta\sqrt{\log T}\cdot n^{-1}\sum_{t=s}^{e}\tilde{Y}^2_{t,T}\right\}$ and

$$\mathcal{B} = \left\{\frac{1}{n}\left|\sum_{t=s}^{e}\left(\tilde{Y}^2_{t,T} - \mathbb{E}\tilde{Y}^2_{t,T}\right)\right| < h = \frac{(\eta_p - s + 1)\sigma_1^2 + (e - \eta_p)\sigma_2^2}{2n}\right\},$$

where $\sigma_1^2 = \sigma^2(\eta_p/T)$ and $\sigma_2^2 = \sigma^2((\eta_p + 1)/T)$. We have $\mathbf{P}(\mathcal{A}) = \mathbf{P}(\mathcal{A}\cap\mathcal{B}) +$

$\mathbf{P}\left(\mathcal{A}\,|\mathcal{B}^c\right)\mathbf{P}\left(\mathcal{B}^c\right) \le \mathbf{P}\left(\mathcal{A}\cap\mathcal{B}\right) + \mathbf{P}\left(\mathcal{B}^c\right)$. The first part is bounded as

$$\mathbf{P}\left(\mathcal{A}\cap\mathcal{B}\right) \le \mathbf{P}\left(\left|\tilde{\mathbb{Y}}^b_{s,e}\right| > \tau T^\theta \sqrt{\log T}\cdot n^{-1}\sum_{t=s}^{e}\left(\mathbb{E}\tilde{Y}^2_{t,T} - h\right)\right). \tag{1.8}$$

From Lemma 3, we have $|\tilde{\mathbb{Y}}^b_{s,e} - \tilde{\mathbb{S}}^b_{s,e}| \le \log T$. Also Lemmas 2.2 and 2.3 of Venkatraman (1993) indicate that $\max_{s<t<e}|\tilde{\mathbb{S}}^t_{s,e}| = |\tilde{\mathbb{S}}^{\eta_p}| = O(\sqrt{n^{-1}\epsilon_T(n - C\epsilon_T)}) = O(\sqrt{\epsilon_T})$. Therefore $|\tilde{\mathbb{Y}}^b_{s,e}| \le |\tilde{\mathbb{S}}^{\eta_p}| + \log T = O(\sqrt{\epsilon_T})$ and (1.8) is bounded by $\mathbb{E}\left(\tilde{\mathbb{Y}}^b_{s,e}\right)^2/(\tau^2h^2T^{2\theta}\log T) \le O\left(T^{1/2-2\theta}\right) \longrightarrow 0$, by applying Markov's inequality. Turning our attention to $\mathbf{P}\left(\mathcal{B}^c\right)$, we need to show that

$$\mathbf{P}\left(\frac{1}{n}\left|\sum_{t=s}^{e}\sigma^2(t/T)(Z^2_{t,T} - 1)\right| > h\right) \longrightarrow 0.$$

This can be shown by applying Bernstein's inequality as in the proof of Lemma 3, and the lemma follows. Similar arguments are applied when (ii) holds. □

We now prove Theorem 1. At the start of the algorithm, as $s = 0$ and $e = T-1$, all conditions for Lemma 5 are met and it finds a breakpoint within the distance of $C\epsilon_T$ from the true breakpoint, by Lemma 4. Under Assumption 2, both (1.2) and (1.3) are satisfied within each segment until every breakpoint in $\sigma^2(t/T)$ is identified. Then, either of two conditions (i) or (ii) in Lemma 6 is met and therefore no further breakpoint is detected with probability converging to 1.

Next we study how the bias present in $\mathbb{E}I^{(i)}_{t,T}(=\sigma^2_{t,T})$ affects the consistency. First we define the autocorrelation wavelet $\Psi_i(\tau) = \sum_{k=-\infty}^{\infty}\psi_{i,k}\psi_{i,k+\tau}$, the autocorrelation wavelet inner product matrix $A_{i,j} = \sum_\tau \Psi_i(\tau)\Psi_j(\tau)$, and the across-scales autocorrelation wavelets $\Psi_{i,j}(\tau) = \sum_k \psi_{i,k}\psi_{j,k+\tau}$. Then it is shown in Fryzlewicz and Nason (2006) that the integrated bias between $\mathbb{E}I^{(i)}_{t,T}$ and $\beta_i(t/T)$ converges to zero.

Proposition 1 (Propositions 2.1-2.2 (Fryzlewicz and Nason (2006)))**.** Let $I^{(i)}_{t,T}$ be the wavelet periodogram at a fixed scale i. Under Assumption 1,

$$T^{-1}\sum_{t=0}^{T-1}\left|\mathbb{E}I^{(i)}_{t,T} - \beta_i(t/T)\right|^2 = O(T^{-1}2^{-i}) + b_{i,T}, \tag{1.9}$$

where $b_{i,T}$ depends on the sequence $\{L_i\}_i$. Further, each $\beta_i(z)$ is a piecewise constant function with at most B jumps, all of which occur in the set $\mathcal{B}$.

Suppose the interval $[s, e]$ includes a true breakpoint η_p as in (1.2), and denote $b = \arg\max_{t\in(s,e)} |\tilde{\mathbb{S}}^t_{s,e}|$ and $\hat{b} = \arg\max_{t\in(s,e)} \left|\mathbb{S}^t_{s,e}\right|$. Recall that $\mathbb{E}I^{(i)}_{t,T}$ remains constant within each stationary segment, apart from short (of length $C2^{-i}$) intervals around the discontinuities in $\beta_i(t/T)$. Suppose a jump occurs at η_p in $\beta_i(t/T)$ yet there is no change in $\mathbb{E}I^{(i)}_{t,T}$ for $t \in [\eta_p - C2^{-i}, \eta_p + C2^{-i}]$. Then the integrated bias is bounded from below by $C\delta_T/T$ from Assumption 2, and Proposition 1 is violated. Therefore there will be a change in $\mathbb{E}I^{(i)}_{t,T}$ as well on such intervals around η_p and $\mathbb{E}I^{(i)}_{t_1,T} \neq \mathbb{E}I^{(i)}_{t_2,T}$ for $t_1 \leq \eta_p - C2^{-i}$ and $t_2 \geq \eta_p + C2^{-i}$. Although the bias of $\mathbb{E}I^{(i)}_{t,T}$ in relation to $\beta_i(t/T)$ may cause some bias between $\hat{b}$ and b, we have that $|\hat{b} - b| \leq C2^{I^*} < \epsilon_T$ holds for $I^* = O(\log\log T)$, which is an admissible rate for I^*. Besides, once one breakpoint is detected in such intervals, the algorithm does not allow any more breakpoints to be detected within the distance of Δ_T from the detected breakpoint, by construction. Hence the bias in $\mathbb{E}I^{(i)}_{t,T}$ does not affect the results of Lemmas 1–6 for wavelet periodograms at finer scales and the consistency still holds for $Y^2_{t,T}$ in (3).

Finally, we note that the within-scale post-processing step in Section 3.2.1 is in line with the theoretical consistency of our procedure; (a) Lemma 5 implies that our test statistic exceeds the threshold when there is a breakpoint η within a segment $[s, e]$ which is of sufficient distance from both s and e, and (b) Lemma 6 shows that it does not exceed the threshold when (s, η, e) does not satisfy the condition in (a).

2 The proof of Theorem 2

From Assumption 1 and the invertibility of the autocorrelation wavelet inner product matrix A, there exists at least one sequence of wavelet periodograms among $I^{(i)}_{t,T}$, $i = -1, \ldots, -I^*$ in which any breakpoint in $\mathcal{B}$ is detected. Suppose there is only one such scale, i_0, for $\nu_q \in \mathcal{B}$ and denote the detected breakpoint as $\hat{\eta}^{(i_0)}_{p_0}$. After the across-scales post-processing, $\hat{\eta}^{(i_0)}_{p_0}$ is selected as $\hat{\nu}_q$ since no other $\hat{\eta}^{(i)}_p$, $i \neq i_0$, is within the distance of $\Lambda_T = C\epsilon_T$ from either $\hat{\nu}_q$ or $\hat{\eta}^{(i_0)}_{p_0}$, and $\left|\nu_q - \hat{\eta}^{(i_0)}_{p_0}\right| \leq \epsilon_T$ with probability converging to 1 from Theorem 1. If there are

$D(\leq I^*)$ breakpoints detected for ν_q, denote them as $\hat{\eta}_{p_1}^{(i_1)}, \ldots, \hat{\eta}_{p_D}^{(i_D)}$. Then for any $1 \leq a < b \leq D$, $\left|\hat{\eta}_{p_a}^{(i_a)} - \hat{\eta}_{p_b}^{(i_b)}\right| \leq \left|\hat{\eta}_{p_a}^{(i_a)} - \nu_q\right| + \left|\hat{\eta}_{p_b}^{(i_b)} - \nu_q\right| \leq C\epsilon_T$, and only the one from the finest scale is selected as $\hat{\nu}_q$ among them by the post-processing procedure. Hence the across-scales post-processing preserves the consistency for the breakpoints selected as its outcome.

References

Bosq, D. (1998). *Nonparametric statistics for stochastic process: estimation and prediction.* Springer.

Fryzlewicz, P. and Nason, G. (2006). Haar-Fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society, B.* **68**, 611–634.

Johnson, N. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions, Vol. 1.* Houghton Mifflin Company.

Venkatraman, E. S. (1993). Consistency results in multiple change-point problems. *PhD Thesis, Stanford University.*

Department of Statistics, London School of Economics, UK.

E-mail: `h.cho1@lse.ac.uk`

Department of Statistics, London School of Economics, UK.

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