# Multiscale and multilevel technique for consistent segmentation of nonstationary time series

Haeran Cho

*University of Bristol, Bristol, UK.*

Piotr Fryzlewicz

*London School of Economics, London, UK.*

**Summary**. The importance of nonstationary modelling in time series analysis was borne out, for example, by the recent financial crisis, which, as some authors argue, has been partly caused by over-reliance on stationary time series models in financial risk management. Piecewise stationarity is the simplest form of departure from stationarity, and one basic task under this assumption is time series segmentation.

In this paper, we propose a fast, well-performing and consistent method for segmenting a piecewise-stationary, linear time series with an unknown number of breakpoints. The time series model we use is the nonparametric Locally Stationary Wavelet model, in which a complete description of the piecewise-stationary second-order structure is provided by wavelet periodograms computed at multiple scales and locations. The initial stage of our method is a new binary segmentation procedure, with a theoretically justified and rapidly computable test criterion, which detects breakpoints in wavelet periodograms at each scale separately. It is then followed by within-scale and across-scales post-processing steps, leading to consistent estimation of the number and locations of breakpoints in the second-order structure of the original process.

An extensive simulation study demonstrates good performance of our method in comparison to the state of the art, and its application to the Dow Jones index indicates two breakpoints, each corresponding to a significant event in the recent financial crisis. A complete R script, implementing our methodology, is provided.

*Keywords*: binary segmentation, breakpoint detection, locally stationary wavelet model, piecewise stationarity, post-processing, wavelet periodogram.

## 1. Introduction

Stationarity assumption is appealing when analysing short time series. However, it is often unrealistic in many circumstances, for example when observing time series evolving in naturally nonstationary environments. One such example can be found in econometrics, where price processes are considered to have time-varying variance in response to events taking place in the market. For example, given the explosion of market volatility during the recent financial crisis, it is unlikely that the same stationary time series model can accurately describe the evolution of market prices before and during the crisis. Indeed, Janeway (2009) goes further and argues that the (over)use of stationary time series models in financial risk management might have been a contributing factor in the crisis due to those models' lack of flexibility in reacting to rapid changes in the statistical properties of the markets. As a taster, we note that some interesting findings resulting from the application of our methodology to financial data observed before and during the crisis are presented in Section 5 of this paper.

Exact or approximate piecewise stationarity is a well studied and arguably the simplest form of departure from stationarity, and one task of interest when faced with data of this form is to detect their structural breakpoints so that each segment between two breakpoints can be modelled as approximately stationary.

Breakpoint detection has received considerable attention and many methods have been developed which can be broadly categorized into two: retrospective (a posteriori) methods and on-line methods. The on-line approach is employed in areas such as quality control, where one wishes to detect changes in the process whilst monitoring is still in progress. A survey of literature on the on-line segmentation of serially independent observations can be found in Bhattacharya (1994). Methods for on-line segmentation of sequentially observed time series include Ombao et al. (2004), who use approximation via autoregressive (AR) processes, and Choi et al. (2008), who analyse the changing spectral characteristics of the process.

On the other hand, the "a posteriori" approach takes into account the entire set of observations at once and detects breakpoints which occurred in the past. The resulting segmentation can be of interest e.g. from the point of view of forecasting, where information from the last (approximately) stationary segment can be useful in forecasting the future. Our interest in this article lies in the "a posteriori" segmentation category, and we propose a retrospective segmentation procedure which achieves consistency in identifying multiple breakpoints for a class of nonstationary processes.

Early segmentation literature was mostly devoted to testing the existence of a single breakpoint in the mean or variance of independent observations (Chernoff and Zacks, 1964; Sen and Srivastava, 1975; Hawkins, 1977; Hsu, 1977; Worsley, 1986). When the presence of more than one breakpoint is suspected, an algorithm for detecting multiple breakpoints is needed. In Vostrikova (1981), a "binary segmentation" procedure was introduced, a computationally efficient multilevel version of the CUSUM test, which recursively locates and tests for multiple breakpoints, producing consistent breakpoint estimators for a class of random processes with piecewise constant means. However, the critical values of the tests at each stage were difficult to compute in practice due to the stochasticity in previously selected breakpoints. Venkatraman (1993) employed the same procedure to find multiple breakpoints in the mean of independent and normally distributed variables and showed the consistency of the detected breakpoints with the tests depending on the sample size only, and thus being easier to compute. The binary segmentation procedure was also adopted to detect multiple shifts in the variance of independent observations (Inclán and Tiao, 1994; Chen and Gupta, 1997).

Various multiple breakpoint detection methods have been proposed for time series of dependent observations. In Lavielle and Moulines (2000), least squares estimators of breakpoint locations were developed for linear processes, extending the work of Bai and Perron (1998). In the Bayesian framework, a number of procedures were studied for the segmentation of signals using piecewise constant linear regression models (McCulloch and Tsay, 1993; Punskaya et al., 2002; Fearnhead, 2005). Adak (1998) and Ombao et al. (2001) proposed methods which divided the time series into dyadic blocks and chose the best segmentation according to suitably tailored criteria. Whitcher et al. (2000, 2002) and Gabbanini et al. (2004) suggested to segment long memory processes by applying the iterative cumulative sum of squares (ICSS) algorithm (proposed in Inclán and Tiao (1994)) to discrete wavelet coefficients of the process, which were shown to be approximately Gaussian and decorrelated. Davis et al. (2006) developed the Auto-PARM procedure which found the optimal segmentation of piecewise stationary AR processes via the minimum description

length principle, later extended to the segmentation of non-linear processes in Davis et al. (2008). Kokoszka and Leipus (2000) and Andreou and Ghysels (2002) studied the problem of detecting structural shifts in the parameters of ARCH or GARCH models.

The aim of our work is to propose a well-performing, theoretically tractable and fast procedure for segmenting piecewise stationary time series which are linear but otherwise do not follow any particular parametric model. (We note that we use the term "segmentation" interchangeably with "multiple breakpoint detection".) The nonparametric model we use for this purpose is the Locally Stationary Wavelet (LSW) model first proposed by Nason et al. (2000) and later studied by Fryzlewicz and Nason (2006) and Van Bellegem and von Sachs (2008). Our choice of model is motivated by several factors, which we explain in Section 2 below. In the LSW model, the piecewise constant second-order structure of the process is completely described by the local wavelet periodograms at multiple scales, and it is those basic statistics that we use as a basis of our segmentation procedure.

To achieve the multiple breakpoint detection, we propose a binary segmentation method, which we apply to wavelet periodograms at each scale separately, and then propose a within-scale and across-scales post-processing procedure to obtain consistent estimators of breakpoints in the second-order structure of the process. We note that wavelet periodograms follow a multiplicative statistical model, but our binary segmentation procedure is different from previously proposed binary segmentation methods for multiplicative models (Inclán and Tiao, 1994; Chen and Gupta, 1997) in that it allows correlated data, which is essential when working with wavelet periodograms. We also mention other unique ingredients of our breakpoint detection procedure which lead to its good performance and consistency in probability; these are: the theoretical derivation of our test criterion (which only depends on the length of the time series and is thus fast to compute); and the novel across-scales post-processing step, essential in combining the results of the binary segmentation procedures performed for each wavelet periodogram scale separately.

We note that our method can simultaneously be termed "multiscale" and "multilevel", as the basic time series model used for our purpose is a wavelet-based, and thus a "multiscale" model; and the core methodology to segment each scale of the wavelet periodogram in the model is based on binary segmentation and is thus a "multilevel" procedure.

The paper is organised as follows. Section 2 explains the LSW model and justifies its choice. Our breakpoint detection methodology (together with the post-processing steps) is introduced in Section 3, where we also demonstrate its theoretical consistency. In Section 4, we describe the outcome of an extensive simulation study which demonstrates good performance of our method in comparison with the state of the art. In Section 5, we apply our technique to the segmentation of the Dow Jones index, which results in an exciting discovery of two breakpoints: one coinciding with the initial period of the recent financial crisis, and the other coinciding with the recent collapse of Lehman Brothers, a major financial services firm. The proofs of our theoretical results are in the appendix.

Software (an R script) implementing our methodology is available from: `http://www.maths.bris.ac.uk/~mahrc/msml_technique.html`.

## 2. Locally stationary wavelet time series

In this section, we first define the Locally Stationary Wavelet (LSW) time series model (noting that our definition is a slight modification of Fryzlewicz and Nason (2006)), and then justify its choice as an attractive framework for the purpose of developing our methodology

for time series segmentation.

DEFINITION 1. *A triangular stochastic array* $\{X_{t,T}\}_{t=0}^{T-1}$ *for* $T = 1, 2, \ldots,$ *is in a class of Locally Stationary Wavelet (LSW) processes if there exists a mean-square representation*

$$X_{t,T} = \sum_{i=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_i(k/T)\psi_{i,t-k}\xi_{i,k} \tag{1}$$

*where* $i \in \{-1, -2, \ldots\}$ *and* $k \in \mathbb{Z}$ *are scale and location parameters respectively,* $\psi_i = (\psi_{i,0}, \ldots, \psi_{i,\mathcal{L}_i})$ *are discrete, real-valued, compactly supported, non-decimated wavelet vectors, and* $\xi_{i,k}$ *are zero-mean, orthonormal, identically distributed random variables. Also for each* $i \leq -1$, $W_i(z) : [0,1] \to \mathcal{R}$ *is a real-valued, piecewise constant function with a finite (but unknown) number of jumps. Let* $L_i$ *denote the total magnitude of jumps in* $W_i^2(z)$*. The functions* $W_i(z)$ *satisfy*

- $\sum_{i=-\infty}^{-1} W_i^2(z) < \infty$ *uniformly in* $z$,
- $\sum_{i=-I}^{-1} 2^{-i} L_i = O(\log T)$ *where* $I = \log_2 T$.

The reader unfamiliar with basic concepts in wavelet analysis is referred, at this point, to the excellent monograph by Vidakovic (1999). Throughout the paper, $\xi_{i,k}$ are assumed to follow the normal distribution; extensions to non-Gaussianity are possible but would cause substantial technical difficulties in deriving our theoretical results. Comparing the above definition with the Cramér's representation of stationary processes, $W_i(k/T)$ is a (scale- and location-dependent) transfer function, the wavelet vectors $\psi_i$ are analogous to the Fourier exponentials, and the innovations $\xi_{i,k}$ correspond to the orthonormal increment process. Small negative values of the scale parameter $i$ denote "fine" scales where the wavelet vectors are the most localised and oscillatory; large negative values denote "coarser" scales with longer, less oscillatory wavelet vectors. By assuming that $W_i(z)$ is piecewise constant, we are able to model processes with a piecewise constant second-order structure where, between any two breakpoints in $W_i(z)$, the autocovariance function remains constant. The Evolutionary Wavelet Spectrum (EWS) is defined as $S_i(z) = W_i(z)^2$, and is in one-to-one correspondence with the time-dependent autocovariance function of the process. Our primary objective is to develop a consistent method for detecting breakpoints in the EWS, and consequently provide a segmentation of the original time series $\{X_{t,T}\}_{t=0}^{T-1}$. We place the following technical assumption on the breakpoints present in the EWS.

ASSUMPTION 1. *The set of those locations* $z$ *where (possibly infinitely many) functions* $S_i(z)$ *contain a jump, is finite. That is, let* $\mathcal{B} = \{z; \quad \exists\, i \ \lim_{u \to z-} S_i(u) \neq \lim_{u \to z+} S_i(u)\}$*, then* $B = |\mathcal{B}| < \infty$.

We further define the wavelet periodogram of the LSW time series.

DEFINITION 2. *Let* $X_{t,T}$ *be an LSW process as in (1). The triangular stochastic array*

$$I_{t,T}^{(i)} = \left| \sum_s X_{s,T}\psi_{i,s-t} \right|^2 \tag{2}$$

*is called the wavelet periodogram of* $X_{t,T}$ *at scale* $i$.

Fryzlewicz and Nason (2006) show that the expectation $\mathbb{E}I^{(i)}_{t,T}$ of the wavelet periodogram is "close" (in the sense of the integrated squared bias converging to zero) to the function $\beta_i(z) = \sum_{j=-\infty}^{-1} S_j(z)A_{i,j}$, where $\mathbf{A}$ is an invertible matrix related to the wavelet vectors in (1) and (2). Further, each function $\beta_i(z)$ is piecewise constant with at most $B$ jumps, all of which occur in the set $\mathcal{B}$. Besides, $\mathbb{E}I^{(i)}_{t,T}$ themselves are piecewise constant by definition, except intervals of length $O(2^{-i})$ around the discontinuities occurring in $\mathcal{B}$.

The finiteness of $\mathcal{B}$ implies that there exists a fixed index $I^* < \lfloor \log_2 T \rfloor$ such that each breakpoint in $\mathcal{B}$ can be found in at least one of the functions $S_i(z)$ for $i = -1, \ldots, -I^*$, i.e., no further new breakpoints appear at scales coarser than $-I^*$. Thus, from the invertibility of $\mathbf{A}$ and the closeness of $\beta_i(z)$ and $\mathbb{E}I^{(i)}_{t,T}$, we conclude that every breakpoint is encoded in the wavelet periodogram sequences at scales $i = -1, \ldots, -I^*$. Therefore our procedure for detecting breakpoints in the autocorrelation structure of $X_{t,T}$ only considers the wavelet periodograms at these scales. Since $I^*$ is fixed but unknown, in our theoretical considerations we permit it to increase slowly to infinity with $T$.

A further reason for disregarding the coarse scales $i < -I^*$ is that the autocorrelation within each wavelet periodogram sequence becomes stronger at coarser scales; similarly, the intervals on which $\mathbb{E}I^{(i)}_{t,T}$ is not piecewise constant become longer. Thus, for coarse scales, wavelet periodograms provide little useful information about breakpoints and can safely be omitted.

We end this section by briefly summarising our reasons behind the choice of the LSW model as a suitable framework for developing our methodology:

(i) The entire second-order structure of the process, which varies over time in a piecewise-constant manner, is encoded in the (asymptotically) piecewise constant expectations of the wavelet periodogram sequences. Thus, any breakpoints in the second order structure must by definition be detectable by analysing the wavelet periodograms, which are relatively easy to handle as they follow a multiplicative model and are "localised" due to the compact support of the underlying wavelets.

(ii) Furthermore, due to the "whitening" properties of wavelets (see e.g. Vidakovic (1999), Chapter 9), the wavelet periodogram sequences are often much less autocorrelated than the original process.

(iii) The entire array of the wavelet periodograms at all scales is easily and rapidly computable via the non-decimated wavelet transform, which helps keep the computational load of our procedure to the minimum.

(iv) Last but not least, the use of the "rescaled time" $z = k/T$ in (1) and the associated regularity assumptions on the transfer functions $W_i(z)$ permit us to establish rigorous asymptotic properties of our procedure.

## 3. Binary segmentation algorithm

In this section, noting that each wavelet periodogram sequence follows a multiplicative model as described in Section 3.1 below, we first introduce a binary segmentation algorithm for such class of sequences. Binary segmentation is a computationally efficient tool which searches for multiple breakpoints in a recursive manner (and can be classed as a "greedy" and "multilevel" algorithm). Venkatraman (1993) applied the procedure to a sequence

of independent normal variables with multiple breakpoints in its mean and showed that the detected breakpoints were consistent in terms of their number and locations. In the following, we aim at extending these consistency results to the multiplicative model where, in addition, dependence between observations is permitted.

Our binary segmentation procedure is then applied to each wavelet periodogram sequence separately, and we establish a procedure which efficiently combines the breakpoint detection outcomes across the scales. We show that the consistency attained for each periodogram sequence carries over to the consistency in detecting breakpoints in the overall second-order structure of the process.

### 3.1. Generic multiplicative model

Recall that each wavelet periodogram ordinate is simply a squared wavelet coefficient of a zero-mean Gaussian time series, distributed as a scaled $\chi^2_1$ variable and satisfies $I^{(i)}_{t,T} = \mathbb{E}I^{(i)}_{t,T} \cdot Z^2_{t,T}$, where $\{Z_{t,T}\}_{t=0}^{T-1}$ are autocorrelated standard normal variables. Hence we first develop a generic breakpoint detection tool for multiplicative sequences defined by

$$Y^2_{t,T} = \sigma^2_{t,T} \cdot Z^2_{t,T},\ t = 0, \ldots, T-1, \tag{3}$$

where the following additional conditions hold.

(i) $\sigma^2_{t,T}$ is deterministic and "close" to a piecewise constant function $\sigma^2(t/T)$ in the sense that $\sigma^2_{t,T}$ is piecewise constant apart from intervals of length at most $O(2^{I^*})$ around the discontinuities in $\sigma^2(z)$, and $T^{-1}\sum_{t=0}^{T-1} |\sigma^2_{t,T} - \sigma^2(t/T)|^2 = o(\log^{-1} T)$, where the latter rate comes from the rate of convergence of the integrated squared bias between $\beta_i(t/T)$ and $\mathbb{E}I^{(i)}_{t,T}$ (see Fryzlewicz and Nason (2006) for details), and from the fact that our attention is limited to the $I^*$ finest scales only. Further, $\sigma^2(z)$ is bounded from above and away from zero, with a finite but unknown number of jumps.

(ii) The vector $\{Z_{t,T}\}_{t=0}^{T-1}$ is multivariate normal with mean zero and variance one, and its autocorrelation sequence is absolutely summable asymptotically; that is the function $\rho(\tau) = \sup_{t,T} |\mathrm{corr}(Z_{t,T}, Z_{t+\tau,T})|$ satisfies $\rho^1_\infty < \infty$ where $\rho^p_\infty = \sum_\tau \rho^p(\tau)$.

Once the breakpoint detection algorithm for the generic model (3) has been established, we apply it to the wavelet periodograms; thus, the reader is invited to relate $Y^2_{t,T}$, $\sigma^2(z)$ and $\sigma^2_{t,T}$ to $I^{(i)}_{t,T}$, $\beta_i(z)$ and $\mathbb{E}I^{(i)}_{t,T}$, respectively.

### 3.2. Algorithm

The first step of the binary segmentation procedure is to find the likely location of a breakpoint. We locate such a point in the interval $(0, T-1)$ as the one which maximizes the absolute value of

$$\mathbb{Y}^\nu_{0,T-1} = \sqrt{\frac{T-\nu}{T\cdot\nu}} \sum_{t=0}^{\nu-1} Y^2_{t,T} - \sqrt{\frac{\nu}{T\cdot(T-\nu)}} \sum_{t=\nu}^{T-1} Y^2_{t,T}. \tag{4}$$

$\mathbb{Y}^\nu_{0,T-1}$ can be interpreted as a scaled difference between the partial means of two segments $\{Y^2_{t,T}\}_{t=0}^{\nu-1}$ and $\{Y^2_{t,T}\}_{t=\nu}^{T-1}$, where the scaling is chosen so as to keep the variance $\mathbb{Y}^\nu_{0,T-1}$

constant over $\nu$ in the idealised case of $Y_{t,T}^2$ being i.i.d. Thus, it is a version of the well-known CUSUM statistic. Once such a $\nu$ has been found, we use $\mathbb{Y}_{0,T-1}^{\nu}$ (but not only this quantity; see below for details) to test the null hypothesis of $\sigma^2(t/T)$ being constant over $\{Y_{t,T}^2\}_{t=0}^{T-1}$. The test statistic and its critical value, or threshold, are established in such a way that when a breakpoint is present, the null hypothesis is rejected with probability converging to 1. If the null hypothesis is rejected, we continue the simultaneous locating and testing of breakpoints on the two segments to the left and to the right of $\nu$ in a recursive manner until no further breakpoints are detected. The algorithm is summarised below, where $j$ is the level index and $l$ is the location index of the node at each level.

**Algorithm**

**Step 1** Begin with $(j,l) = (1,1)$. Let $s_{j,l} = 0$ and $e_{j,l} = T-1$.

**Step 2** Iteratively compute $\mathbb{Y}_{s_{j,l},e_{j,l}}^{b}$ as in (4) for $b \in (s_{j,l}, e_{j,l})$, and find $b_{j,l}$ which maximizes its absolute value. Let $n_{j,l} = e_{j,l} - s_{j,l} + 1$, $d_{j,l} = \mathbb{Y}_{s_{j,l},e_{j,l}}^{b}$, and $m_{j,l} = \sum_{t=s_{j,l}}^{e_{j,l}} Y_{t,T}^2/\sqrt{n_{j,l}}$.

**Step 3** Perform hard thresholding on $|d_{j,l}|/m_{j,l}$ with the threshold defined as $t_{j,l} = \tau T^{\theta}\sqrt{\log T/n_{j,l}}$ so that $\hat{d}_{j,l} = d_{j,l}$ if $|d_{j,l}| > m_{j,l} \cdot t_{j,l}$, and $\hat{d}_{j,l} = 0$ otherwise. The choice of $\theta$ and $\tau$ is discussed in Section 3.4.

**Step 4** If either $\hat{d}_{j,l} = 0$ or $\max\{b_{j,l}-s_{j,l}+1, e_{j,l}-b_{j,l}\} < \Delta_T$ for $l$, stop the algorithm on the interval $[s_{j,l}, e_{j,l}]$. If not, let $(s_{j+1,2l-1}, e_{j+1,2l-1}) = (s_{j,l}, b_{j,l})$ and $(s_{j+1,2l}, e_{j+1,2l}) = (b_{j,l}+1, e_{j,l})$ and update the level $j$ as $j \to j+1$. Again, the choice of $\Delta_T$ is discussed in Section 3.4.

**Step 5** Repeat Steps 2–4.

The set of detected breakpoints is $\{b_{j,l}; \hat{d}_{j,l} \neq 0\}$. The test statistic $|d_{j,l}|/m_{j,l}$ is a scaled version of the test statistics in the ICSS algorithm (Inclán and Tiao, 1994; Whitcher et al., 2000). However, the test criteria in those papers are derived empirically under the assumption of independent observations and there is no guarantee that either algorithm produces consistent breakpoint estimates. On the other hand, our algorithm permits the sequences to be autocorrelated as in (3), and its test criterion enables the consistent identification of the total number and locations of breakpoints as shown in Section 3.3, provided that the true breakpoints are sufficiently scattered over time.

We also note the similarity between the statistic $|d_{j,l}|/m_{j,l}$ (with the convention $0/0 = 0$) and the Fisz transform of Fryzlewicz and Nason (2006) and Fryzlewicz et al. (2006), the difference being that the Fisz transform was only defined for the case $b = \frac{1}{2}(e_{j,l} + s_{j,l} + 1)$ (meaning the segments were split in half) and that it was not used for the purposes of breakpoint detection.

*3.2.1. Post-processing within a sequence*

We further equip the procedure with an extra step aimed at reducing the risk of overestimating the number of breakpoints. The ICSS algorithm in Inclán and Tiao (1994) has a "fine-tune" step whereby if more than one breakpoint is found, each breakpoint is checked against the adjacent ones to reduce the risk of overestimation. We propose a post-processing

procedure performing a similar task within the single-sequence multiplicative model (3). At each breakpoint, the test statistic is re-calculated over the interval between two neighbouring breakpoints and compared with the threshold again. Denote the breakpoint estimates as $\hat{\eta}_p, p = 1, \ldots, \hat{N}$ and $\hat{\eta}_0 = 0$, $\hat{\eta}_{\hat{N}+1} = T$. For each $\hat{\eta}_p$, we examine whether

$$\left| \mathbb{Y}^{\hat{\eta}_p}_{\hat{\eta}_{p-1}+1,\hat{\eta}_{p+1}} \right| > \tau T^{\theta} \sqrt{\log T} \cdot \frac{1}{(\hat{\eta}_{p+1} - \hat{\eta}_{p-1})} \sum_{t=\hat{\eta}_{p-1}+1}^{\hat{\eta}_{p+1}} Y^2_{t,T}.$$

If the above inequality does not hold, $\hat{\eta}_p$ is removed and the same procedure is repeated with the reduced set of breakpoints until the set does not change. Note that the fine-tune step of the ICSS algorithm re-calculates both the location and test statistic at each iteration and therefore the locations of breakpoints are subject to change after tuning. However in our post-processing procedure, only the test statistic is re-calculated at existing breakpoints, and thus their locations are preserved. We note that the post-processing does not affect the theoretical consistency of our procedure as (a) the extra checks above are of the same form as those done in the original algorithm of Section 3.2, (b) the locations of the breakpoints that survive the post-processing are unchanged. The next section provides details of our consistency result.

### 3.3. Consistency of detected breakpoints

In a breakpoint detection problem, it is desirable that the proposed procedure should (asymptotically) correctly identify the total number and locations of breakpoints. In this section, we first show the consistency of our algorithm for a multiplicative sequence as in (3), which corresponds to the wavelet periodogram sequence at a single scale. Later, Theorem 2 will show how this consistency result carries over to the consistency of our procedure in detecting breakpoints in the entire second-order structure of the input LSW process $X_{t,T}$.

Denote the number of breakpoints in $\sigma^2(t/T)$ by $N$ and the breakpoints themselves by $0 < \eta_1 < \ldots < \eta_N < T - 1$, with $\eta_0 = 0$, $\eta_{N+1} = T - 1$. The following assumption states that the breakpoints $\eta_p$ should be sufficiently scattered over time without being too close to each other.

ASSUMPTION 2. *For $\theta \in (1/4, 1/2)$ and $\Theta \in (\theta + 1/2, 1)$, the length of each segment in $\sigma^2(t/T)$ is bounded from below by $\delta_T = O(T^{\Theta})$. Further, the breakpoints cannot be too close to each other, i.e., there exists some constant $c > 0$ such that,*

$$\max_{1 \le p \le N} \left\{ \sqrt{\frac{\eta_p - \eta_{p-1}}{\eta_{p+1} - \eta_p}}, \sqrt{\frac{\eta_{p+1} - \eta_p}{\eta_p - \eta_{p-1}}} \right\} \le c.$$

THEOREM 1. *Suppose that $\{Y_{t,T}\}_{t=0}^{T-1}$ follows model (3). Assume that*

- $\inf_p \left| \sigma^2 \left( \eta_p / T \right) - \sigma^2 \left( \eta_{p+1} / T \right) \right| \ge \xi$ *for some* $\xi > 0$,
- $\sup_t |\sigma^2(t/T)| \le M$ *for some* $M > 0$.

*Under Assumption 2, the number and locations of detected breakpoints are consistent. That is,* $\Pr(\mathcal{A}) \to 1$ *as* $T \to \infty$ *where* $\mathcal{A} = \left\{ \hat{N} = N; |\hat{\eta}_p - \eta_p| \le O(\epsilon_T),\ 1 \le p \le N \right\}$; $\hat{\eta}_p,\ p = 1, \ldots, \hat{N}$ *are detected breakpoints and* $\epsilon_T = T^{1/2} \log T$. *(Interpreting this in the rescaled time interval* $[0, 1]$, $\epsilon_T / T = T^{-1/2} \log T$.*)*

*3.3.1. Post-processing across the scales*

As mentioned in Section 2, we only consider wavelet periodograms $I_{t,T}^{(i)}$ at scales $i = -1, \ldots, -I^*$, choosing $I^*$ to satisfy $2^{I^*} \ll \epsilon_T = T^{1/2} \log T$ so that the bias between $\sigma_{t,T}^2$ and $\sigma^2(t/T)$ does not cause us harm in deriving the results of Theorem 1. Recall from our previous discussion that any breakpoint in the second-order structure of the original process $X_{t,T}$ must be reflected in a breakpoint in at least one of the asymptotic wavelet periodogram expectations $\beta_i(z)$ for $i = -1, \ldots, -I^*$, and vice versa: a breakpoint in one of the $\beta_i(z)$'s implies a breakpoint in the second-order structure of $X_{t,T}$. Thus, it is sensible to combine the estimated breakpoints across the periodogram scales by, roughly speaking, selecting a breakpoint as significant if it appears in *any* of the wavelet periodogram sequences. This section describes a precise algorithm for doing this, and states a consistency result for the final set of breakpoints arising from combining them across scales.

We first provide a general description of the algorithm, and then make it more precise below. In the first stage, we produce a union of the sets of all breakpoints from scales $i = -1, \ldots, -I^*$. In the second stage, we cluster them into groups depending on their distance from one another. If there is a single scale containing representatives in each group, we choose those as our final breakpoint estimates. If not, then the representative breakpoint chosen from each group is the one that corresponds to the finest periodogram scale at which it appears.

The complete algorithm follows. Let $\hat{\mathcal{B}}_i = \left\{\hat{\eta}_p^{(i)},\ p = 1, \ldots, \hat{N}_i\right\}$ be the set of detected breakpoints from the sequence $I_{t,T}^{(i)}$. Then the post-processing finds a subset of $\cup_{i=-1}^{-I^*} \hat{\mathcal{B}}_i$, say $\hat{\mathcal{B}}$, as formulated below;

**Step 1** Arrange all breakpoints into groups so that those from different sequences and within the distance of $\Lambda_T$ from each other are classified as belonging to the same group; and denote the groups by $\mathcal{G}_1, \ldots, \mathcal{G}_{\hat{B}}$.

**Step 2** Find $i_0 = \max\left\{\arg\max_{-I^* \le k \le -1} \hat{N}_k\right\}$, i.e., the finest scale with the most breakpoints.

**Step 3** Check whether there exists $\hat{\eta}_{p_0}^{(i_0)}$ for every $\hat{\eta}_p^{(i)}$, $i \neq i_0,\ 1 \le p \le \hat{N}_i$, which satisfies $\left|\hat{\eta}_p^{(i)} - \hat{\eta}_{p_0}^{(i_0)}\right| < \Lambda_T$. If so, let $\hat{\mathcal{B}} = \hat{\mathcal{B}}_{i_0}$ and quit the post-processing.

**Step 4** Otherwise let $\hat{\mathcal{B}} = \left\{\hat{\nu}_p,\ p = 1, \ldots, \hat{B}\right\}$ where each $\hat{\nu}_p \in \mathcal{G}_p$ with the maximum $i$ (finest scale).

We set $\Lambda_T = \lfloor \epsilon_T/2 \rfloor$ in order to take into account the bias arising in deriving the results of Theorem 1. As argued previously, breakpoints detected at coarser scales are likely to be less accurate than those detected at finer scales; therefore, the above algorithm prefers the latter. The across-scales post-processing procedure preserves the number of "distinct" breakpoints and also their locations determined in the algorithm. Hence the breakpoints in set $\hat{\mathcal{B}}$ are still consistent estimates of true breakpoints in the second-order structure of the original nonstationary process $X_{t,T}$.

We acknowledge that the above is not the only possible way of combining the breakpoints across scales which is still permitted by the theory and preserves consistency; however, we advocate the above algorithm due to its good practical performance.

Denote the set of the true breakpoints in the second-order structure of $X_{t,T}$ by $\mathcal{B} = \{\nu_p,\ p = 1, \ldots, B\}$, and the estimated breakpoints by $\hat{\mathcal{B}} = \left\{\hat{\nu}_p,\ p = 1, \ldots, \hat{B}\right\}$.

THEOREM 2. *Suppose that $X_{t,T}$ satisfies Assumption 1 and that $\nu_p, 1 \le p \le B$, are scattered over time as in Assumption 2. Further assume that the conditions in Theorem 1 hold for each $\beta_i(z)$. Then our estimated number and locations of breakpoints are consistent, that is*

$$\Pr\left\{\hat{B} = B;\ |\hat{\nu}_p - \nu_p| \le O(\epsilon_T),\ 1 \le p \le B\right\} \to 1$$

*as $T \to \infty$.*

### 3.4. Choice of $\Delta_T$, $\theta$, $\tau$ and $I^*$

To ensure that each estimated segment is of sufficiently large length so as not to distort our theoretical results, $\Delta_T$ is chosen so that $\Delta_T \ge O(\epsilon_T)$. However, in practice our method works well for smaller values of $\Delta_T$ too; e.g. in the forthcoming simulation experiments, $\Delta_T = O(\sqrt{T})$ is used. As $\theta \in (1/4, 1/2)$, we use $\theta = 0.256$ (as we have found that the method works best when $\theta$ is close to the lower end of its permitted range) and elaborate on the choice of $\tau$ below (noting that our asymptotic theoretical results hold for any fixed positive $\tau$).

The selection of $\tau$ is not a straightforward task and to get some insight into the issue, a set of numerical experiments was conducted. A vector of random variables $\mathbf{X} \sim \mathcal{N}_T(0, \Sigma)$ was generated where $\mathbf{X} = (X_1, \ldots, X_T)^T$, then it was transformed into sequences of wavelet periodograms $I^{(i)}_{t,T}$. The covariance matrix satisfied $\Sigma = (\sigma_{i,j})^T_{i,j=1}$ where $\sigma_{i,j} = \rho^{|i-j|}$ such that with varying $\rho$, the variables $\{X_t\}^T_{t=1}$ were either independent or correlated. Then we found $b \in (1, T)$ which maximised

$$\mathbb{I}^b_i = \left|\sqrt{\frac{T-b}{T \cdot b}} \sum_{t=1}^{b} I^{(i)}_{t,T} - \sqrt{\frac{b}{T(T-b)}} \sum_{t=b+1}^{T} I^{(i)}_{t,T}\right|,$$

and computed $\mathbb{U}_{i,\rho,T} = \mathbb{I}^b_i \cdot \{T^{-1} \sum_{t=1}^{T} I^{(i)}_{t,T} \cdot T^{\theta}\sqrt{\log T}\}^{-1}$. This was repeated with a varying covariance matrix ($\rho = 0, 0.3, 0.6, 0.9$) and sample size ($T = 512, 1024, 2048$), 100 times for each combination.

The quantity $\mathbb{U}_{i,\rho,T}$ is the ratio between our test statistic $|d_{1,1}|/m_{1,1}$ and the time-dependent factor $T^{\theta}\sqrt{\log T}$ appearing in our threshold defined in the Algorithm of Section 3.2. $\mathbb{U}_{i,\rho,T}$ is computed under the "null hypothesis" of no breakpoints being present in the covariance structure of $X_t$, and its magnitude serves as a guideline as to how to select the value of $\tau$ for each scale $i$ to prevent spurious breakpoint detection in the null hypothesis case. The results showed that the values of $\mathbb{U}_{i,\rho,T}$ and their range tended to increase for coarser scales, due to the increasing dependence in the wavelet periodogram sequences. In comparison to the scale factor $i$, the parameters $\rho$ or $T$ had relatively little impact on $\mathbb{U}_{i,\rho,T}$.

Based on the above numerical experiments, we propose to use different values of $\tau$ in Step 3 of the Algorithm of Section 3.2 and in the within-scale post-processing procedure of Section 3.2.1. Denoting the former by $\tau_{i,1}$ and the latter by $\tau_{i,2}$, we choose $\tau_{i,1}$ differently for each $i$ as the 95% quantile of $\mathbb{U}_{i,\rho,T}$, and $\tau_{i,2}$ as its 97.5% quantile. The numerical values are summarised in Table 3.4.

**Table 1.** Values of $\tau$ for each scale $i = -1, \ldots, -6$.

| scale $i$ | $-1$ | $-2$ | $-3$ | $-4$ | $-5$ | $-6$ |
|---|---|---|---|---|---|---|
| $\tau_{i,1}$ | 0.40 | 0.50 | 0.65 | 0.80 | 0.95 | 1.25 |
| $\tau_{i,2}$ | 0.45 | 0.60 | 0.75 | 0.90 | 1.10 | 1.35 |

Finally, we discuss the choice of $I^*$, i.e. the coarsest wavelet periodogram scale at which we still apply our breakpoint detection procedure. Firstly, we detect breakpoints in wavelet periodograms at scales $i = -1, \ldots, -\lfloor \log_2 T/3 \rfloor$ and perform the across-scale post-processing as described in Section 3.3.1, obtaining the set of breakpoints $\hat{\mathcal{B}} = \left\{ \hat{\nu}_p,\ p = 1, \ldots, \hat{B} \right\}$. Subsequently, for the wavelet periodogram at the next finest scale, we compute the quantity $\mathbb{V}_p,\ p = 1, \ldots, \hat{B} + 1$ as

$$\mathbb{V}_p = \max_{\nu \in (\hat{\nu}_{p-1}, \hat{\nu}_p)} \left| \frac{\sqrt{\frac{\hat{\nu}_p - \nu}{(\hat{\nu}_p - \hat{\nu}_{p-1}) \cdot (\nu - \hat{\nu}_{p-1})}} \sum_{t=\hat{\nu}_{p-1}+1}^{\nu} I_{t,T}^{(i)} - \sqrt{\frac{\nu - \hat{\nu}_{p-1}}{(\hat{\nu}_p - \hat{\nu}_{p-1}) \cdot (\hat{\nu}_p - \nu)}} \sum_{t=\nu+1}^{\hat{\nu}_p} I_{t,T}^{(i)}}{\sum_{\hat{\nu}_{p-1}+1}^{\hat{\nu}_p} I_{t,T}^{(i)} / (\hat{\nu}_p - \hat{\nu}_{p-1})} \right|$$

where $\hat{\nu}_0 = -1$ and $\hat{\nu}_{\hat{B}+1} = T - 1$. Note that $\mathbb{V}_p$ is again of the same form as our basic test statistic. Then $\mathbb{V}_p$ is compared to $\tau_{i,1} \cdot T^{\theta} \sqrt{\log T}$ to see whether there are any further breakpoints yet to be detected in $I_{t,T}^{(i)}$ which have not been included in $\hat{\mathcal{B}}$. (This step is similar to our within-scale post-processing.) If there is an interval $[\hat{\nu}_{p-1} + 1, \hat{\nu}_p]$ where $\mathbb{V}_p$ exceeds the threshold, $I^*$ is updated as $I^* := I^* + 1$ and the above procedure is repeated to update $\hat{\mathcal{B}}$ until either no further changes are made or $I^* \geq \lfloor \log_2 T/2 \rfloor$. This approach is theoretically justified by Lemma 6 in the Appendix and shown to work well in our simulation study.

## 4. Simulation study

In Davis et al. (2006), the performance of the Auto-PARM was assessed and compared with the Auto-SLEX (Ombao et al., 2001) through simulation in various settings. The Auto-PARM was shown to be superior to Auto-SLEX in identifying both dyadic and non-dyadic breakpoints in piecewise stationary time series. Some examples from the paper were adopted for the comparative study between our method and the Auto-PARM, alongside some other new examples. In the simulations below, wavelet periodograms were computed using Haar wavelets and both post-processing procedures (Section 3.2.1 and Section 3.3.1) followed the application of the segmentation algorithm. In our examples, $T = 1024$ and therefore $I^*$ was set as 3 at the start of each application of the algorithm, then updated automatically if necessary, as described in Section 3.4. The algorithm was coded in R and took 0.32 seconds on average to run on a 2.40-GHz Intel Core 2 processor when analysing a realisation of length $T = 1024$.

### *4.1. Piecewise stationary AR process with clearly observable changes*

This example is taken from Davis et al. (2006). The target nonstationary process was generated from (5),

$$X_t = \begin{cases} 0.9X_{t-1} + \epsilon_t & \text{for } 1 \le t \le 512, \\ 1.68X_{t-1} - 0.81X_{t-2} + \epsilon_t & \text{for } 513 \le t \le 768, \\ 1.32X_{t-1} - 0.81X_{t-2} + \epsilon_t & \text{for } 769 \le t \le 1024 \end{cases} \tag{5}$$

where $\epsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$ in all examples. As seen in Figure 1 (a), there is a clear visual difference between the three segments in the model. Figure 1 (b) shows the wavelet periodogram at scale $-4$ and the estimation results, where the dotted lines indicate the true breakpoints ($\eta_1 = 512, \eta_2 = 768$) while the dashed lines indicate the detected ones ($\hat{\eta}_1 = 512, \hat{\eta}_2 = 758$). Note that although initially the procedure returned three breakpoints, the within-sequence post-processing (Section 3.2.1) successfully removed the false one. The experiment was repeated 100 times and the final counts of breakpoints are given in Table 4.4.

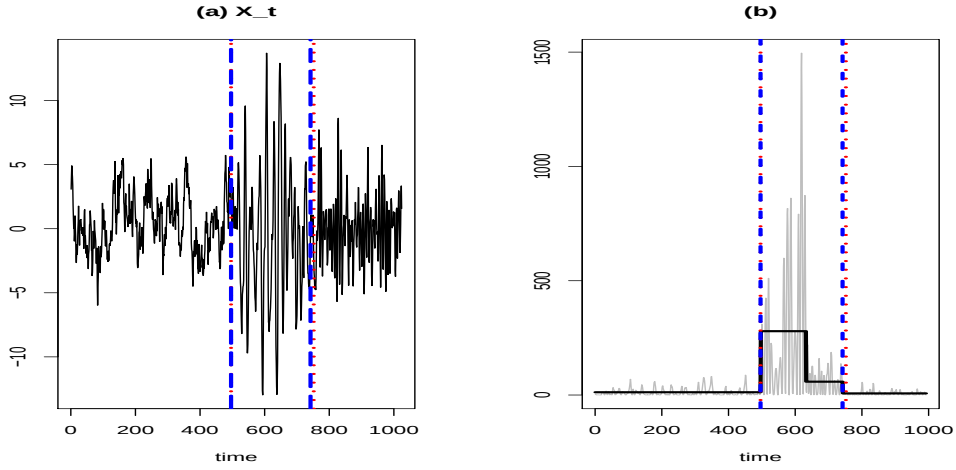


**Fig. 1.** (a) A realisation of model (5) and true (dotted) and detected (dashed) breakpoints; (b) Wavelet periodogram at scale $-4$ and the breakpoint detection outcome, showing also the spurious breakpoint which was later removed in post-processing.

### *4.2. Piecewise stationary AR process with less clearly observable changes*

In this example, the piecewise stationary AR model is revisited, but its breakpoints are less clear-cut, as seen in Figure 2 (a).

$$X_t = \begin{cases} 0.4X_{t-1} + \epsilon_t & \text{for } 1 \le t \le 400, \\ -0.6X_{t-1} + \epsilon_t & \text{for } 401 \le t \le 612, \\ 0.5X_{t-1} + \epsilon_t & \text{for } 613 \le t \le 1024 \end{cases} \tag{6}$$

Figure 2 (b) shows the wavelet periodogram at scale $-1$ for the realisation in the left panel and also its breakpoint estimates ($\hat{\eta}_1 = 372, \hat{\eta}_2 = 622$). Both procedures achieved a similarly good performance: our method accurately detected the two breakpoints in 97% of the cases while the Auto-PARM performed well for all cases.

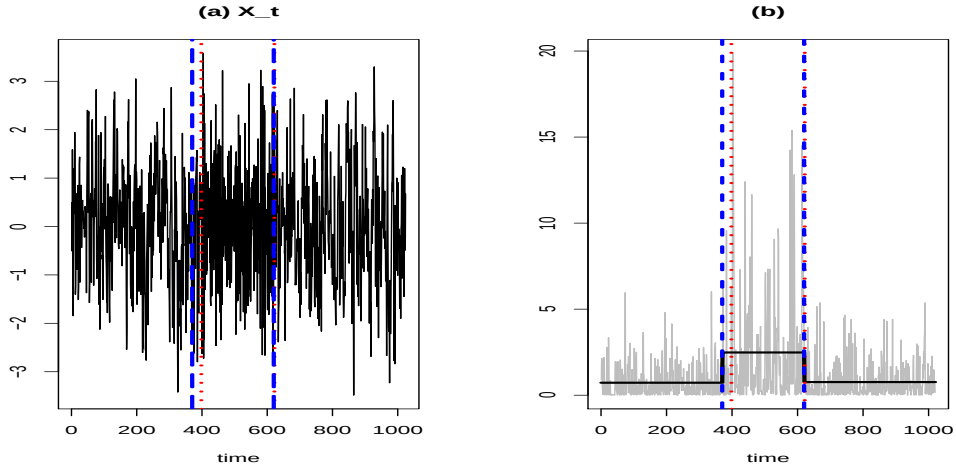


**Fig. 2.** (a) A realisation of model (6) and true (dotted) and detected (dashed) breakpoints; (b) Wavelet periodogram at scale $-1$ and the breakpoint detection outcome.

### *4.3. Piecewise stationary AR process with a short segment*

This example is again from Davis et al. (2006). A single breakpoint occurs and one segment is much shorter than the other.

$$X_t = \begin{cases} 0.75X_{t-1} + \epsilon_t & \text{for } 1 \le t \le 50, \\ -0.5X_{t-1} + \epsilon_t & \text{for } 51 \le t \le 1024. \end{cases} \tag{7}$$

A typical realisation of (7), its wavelet periodogram at scale $-3$, and the estimation outcome are shown in Figure 3, where the jump at $\eta_1 = 50$ was correctly identified as $\hat{\eta}_1 = 49$. Even though one segment is substantially shorter than the other, our procedure was able to detect exactly one breakpoint in 94% of the cases and underestimation did not occur even when it failed to detect exactly one.

### *4.4. Piecewise stationary unit-root-like process with changing variance*

Financial time series, such as stock indices, individual share or commodity prices, or currency exchange rates, are for certain purposes (such as e.g. pricing of derivative instruments) often modelled as random walk with a time-varying variance. Motivated by this, we generated a random-walk-like (but piecewise stationary) example from model (8) below, where the variance has two breakpoints over time and the AR parameter remains constant and very close to 1.

$$X_t = \begin{cases} 0.999X_{t-1} + \epsilon_t & \text{for } 1 \le t \le 400, \\ 0.999X_{t-1} + 1.5\epsilon_t & \text{for } 401 \le t \le 750, \\ 0.999X_{t-1} + \epsilon_t & \text{for } 751 \le t \le 1024. \end{cases} \tag{8}$$

Recall that the Auto-PARM is designed to find the "best" combination of the total number and locations of breakpoints and adopts a genetic algorithm to traverse the vast

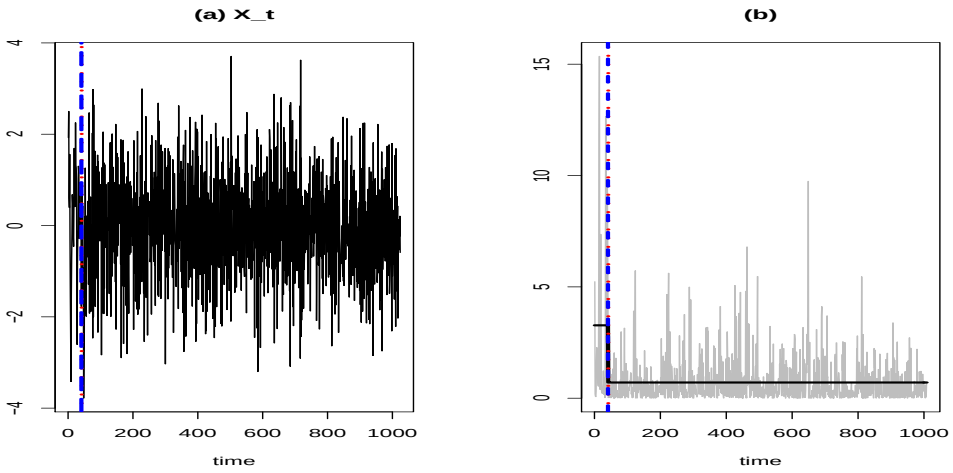


**Fig. 3.** (a) A realisation of model (7) and true (dotted) and detected (dashed) breakpoints; (b) Wavelet periodogram at scale $-3$ and the breakpoint detection outcome.

**Table 2.** Summary of breakpoint detection from simulations; Our method (CF) and the Auto-PARM (AP).

| | number of breakpoints | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | model (5) | | model (6) | | model (7) | | model (8) | |
| | CF | AP | CF | AP | CF | AP | CF | AP |
| 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 41 |
| 1 | 0 | 0 | 1 | 0 | **94** | **99** | 1 | 29 |
| 2 | **90** | **99** | **97** | **100** | 6 | 0 | **94** | **18** |
| 3 | 10 | 1 | 1 | 0 | 0 | 0 | 4 | 10 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

parameter space. However, due to the stochastic nature of the algorithm, the Auto-PARM occasionally fails to return consistent estimates, which is not a desirable feature. This instability was emphasised in this example, each run often returning different breakpoints. For one typical realisation, it detected $t = 21, 797$ as breakpoints, and then only $t = 741$ in the next run on the same sample path. Besides, from Figure 4, which summarises and compares the empirical distribution of the detected breakpoints by both methods over 100 repetitions, it is clear that the performance of Auto-PARM leaves much to be desired for this particular example, whereas our method performs very well. We emphasise that this is not to be taken as a criticism of Auto-PARM in general, which performed very well in our other examples. Note that it was at scale $-1$ of the wavelet periodogram that both breakpoints were consistently identified the most frequently. The computation of the wavelet periodogram at scale $-1$ with Haar wavelets is a differencing operation and naturally "whitens" the almost-unit-root process (8), clearly revealing any changes of variance in the sequence. Finally, though inferior to Haar wavelets, performance achieved by our method using other wavelets was also superior to Auto-PARM in this example.

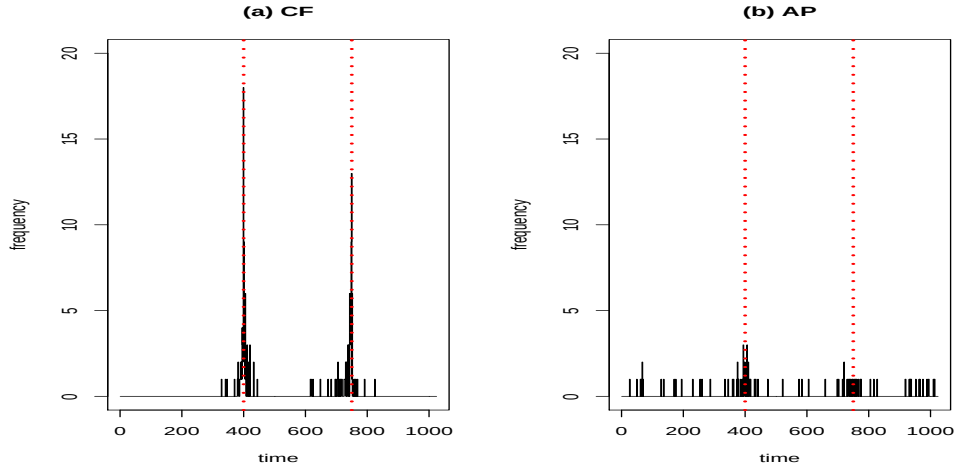


**Fig. 4.** Selection frequency for each $t$; (a) Our method (CF); (b) Auto-PARM (AP).

## 5. U.S stock market data analysis

### *5.1. Dow Jones weekly closing values 1970–1975*

The time series of weekly closing values of the Dow Jones Industrial Average index between July 1971 and August 1974 was studied in Hsu (1979) and revisited in Chen and Gupta (1997). Historical data are available on `www.google.com/finance/historical?q=INDEXDJX:.DJI`, where daily and weekly prices can be extracted for any time period. Both papers concluded that there was a change in the variance of the index around the third week of March 1973. For the ease of computation of the wavelet periodogram, we chose the same weekly index between 1 July 1970 and 19 May 1975 so that the data size was $T = 256$ and the above-mentioned time period was contained in this interval. The third week of March 1973 corresponds to $\eta = 141$ and our procedure detected $\hat{\eta} = 142$ as a breakpoint, which is illustrated in Figure 5 (b).

### *5.2. Dow Jones daily closing values 2007–2009*

We further investigated more recent *daily* data from the same source, between 8 January 2007 and 16 January 2009. Over this period, the global financial market experienced one of the worst crises in history, covered extensively in the media. Our breakpoint detection algorithm estimated two breakpoints (see the illustration in Figure 6), one occurring in the last week of July 2007, and the other in mid-September 2008. The first breakpoint ($\hat{\eta}_1 = 135$) coincided with the outbreak of the worldwide "credit crunch" as subprime mortgage backed securities were discovered in portfolios of banks and hedge funds around the world. The second breakpoint ($\hat{\eta}_2 = 424$) coincided with the bankruptcy of Lehman Brothers, a major financial services firm, (September 14, 2008), an event which brought even more volatility to the market.

We note that Wikipedia (`http://en.wikipedia.org/wiki/Financial_crisis_of_2007-2009`) also mentions these two dates as milestones of the crisis:

> "The financial crisis of 2007–2009 began in July 2007 when a loss of confidence by investors in the value of securitized mortgages in the United States resulted

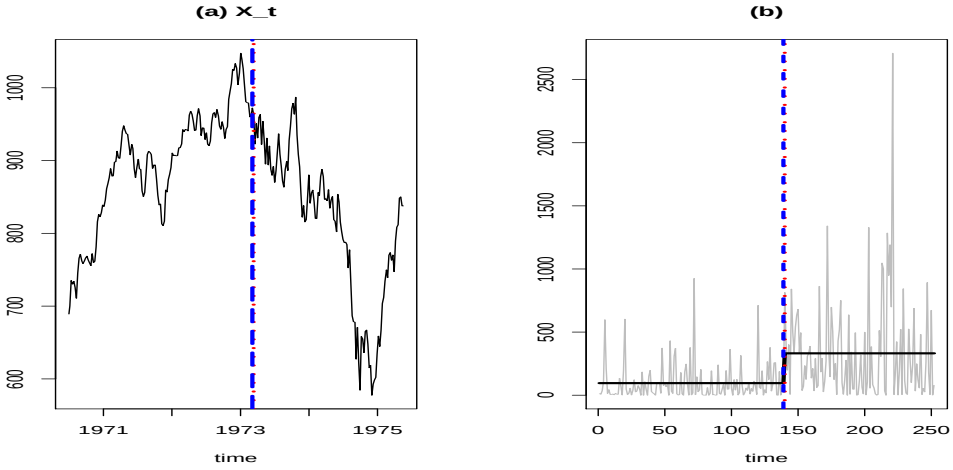


**Fig. 5.** (a) Weekly average values of the Dow Jones IA index (July 1970–May 1975); (b) Wavelet periodogram at scale $-1$ and the breakpoint detection outcome.

in a liquidity crisis that prompted a substantial injection of capital into financial markets by the United States Federal Reserve, Bank of England and the European Central Bank. The TED spread, an indicator of perceived credit risk in the general economy, spiked up in July 2007, remained volatile for a year, then spiked even higher in September 2008, reaching a record 4.65% on October 10, 2008. In September 2008, the crisis deepened, as stock markets worldwide crashed and entered a period of high volatility, and a considerable number of banks, mortgage lenders and insurance companies failed in the following weeks."

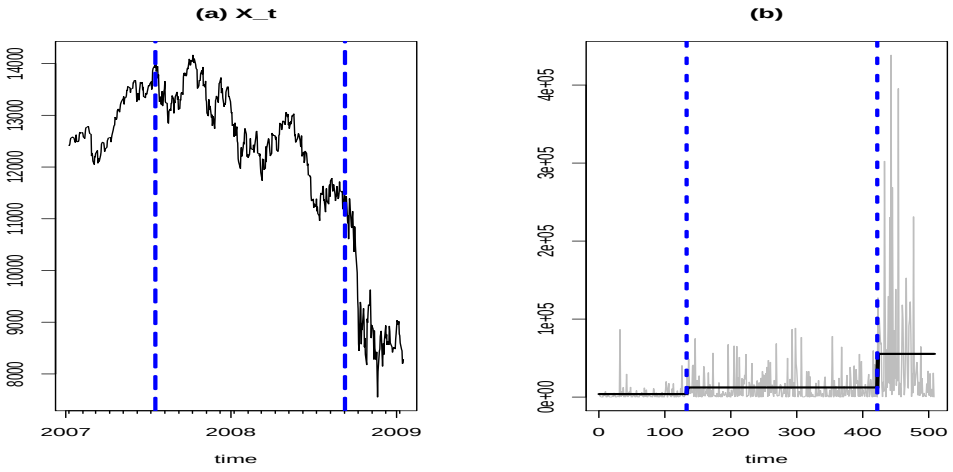


**Fig. 6.** (a) Daily average values of the Dow Jones IA index (Jan 2007–Jan 2009); (b) Wavelet periodogram at scale $-1$ and the breakpoint detection outcome.

## A. The proof of Theorem 1

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

$$\tilde{Y}^2_{t,T} = \sigma^2(t/T) \cdot Z^2_{t,T},\ t = 0, \ldots, T-1. \tag{9}$$

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

$$\begin{aligned}
\tilde{\mathbb{Y}}^b_{s,e} &= \frac{\sqrt{e-b}}{\sqrt{n}\sqrt{b-s+1}} \sum_{t=s}^{b} \tilde{Y}^2_{t,T} - \frac{\sqrt{b-s+1}}{\sqrt{n}\sqrt{e-b}} \sum_{t=b+1}^{e} \tilde{Y}^2_{t,T}, \\
\tilde{\mathbb{S}}^b_{s,e} &= \frac{\sqrt{e-b}}{\sqrt{n}\sqrt{b-s+1}} \sum_{t=s}^{b} \sigma^2(t/T) - \frac{\sqrt{b-s+1}}{\sqrt{n}\sqrt{e-b}} \sum_{t=b+1}^{e} \sigma^2(t/T), \\
\mathbb{S}^b_{s,e} &= \frac{\sqrt{e-b}}{\sqrt{n}\sqrt{b-s+1}} \sum_{t=s}^{b} \sigma^2_{t,T} - \frac{\sqrt{b-s+1}}{\sqrt{n}\sqrt{e-b}} \sum_{t=b+1}^{e} \sigma^2_{t,T}.
\end{aligned}$$

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 $\eta_{p_0} \leq s < \eta_{p_0+1} < \ldots < \eta_{p_0+q} < e \leq \eta_{p_0+q+1}$ for $0 \leq p_0 \leq B - q$, 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:

$$s < \eta_{p_0+r} - C\delta_T < \eta_{p_0+r} + C\delta_T < e \text{ for some } 1 \leq r \leq q, \tag{10}$$

$$\{(\eta_{p_0+1} - s) \wedge (s - \eta_{p_0})\} \vee \{(\eta_{p_0+q+1} - e) \wedge (e - \eta_{p_0+q})\} \leq \epsilon_T, \tag{11}$$

where (as elsewhere in the paper) $C$ is an arbitrary positive constant and $\wedge$ and $\vee$ are the minimum and maximum operators, respectively. We remark that both conditions (10) and (11) 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 (10), then there exists $1 \leq r \leq q$ such that*

$$\left|\tilde{\mathbb{S}}^{\eta_{p_0+r}}_{s,e}\right| = \max_{s<t<e} |\tilde{\mathbb{S}}^t_{s,e}| \geq O\left(\delta_T/\sqrt{T}\right). \tag{12}$$

**Proof.** The equality in (12) is proved by Lemmas 2.2 and 2.3 in Venkatraman (1993). For the inequality, note first that in the case of a single breakpoint in $\sigma^2(z)$ we can use the constancy of $\sigma^2(z)$ to the left and to the right of the breakpoint to show that

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

which is bounded from below by $O\left(\delta_T/\sqrt{T}\right)$. We remark that the order remains the same in the case of multiple breakpoints. □

LEMMA 2. *Suppose (10) holds, and further assume that* $\tilde{\mathbb{S}}^{\eta_{p_0+r}}_{s,e} > 0$ *for some* $1 \le r \le q$. *Then for any* $b$ *satisfying* $|\eta_{p_0+r} - b| = O(\epsilon_T)$ *we have, for a large* $T$, $\tilde{\mathbb{S}}^{\eta_{p_0+r}}_{s,e} \ge \tilde{\mathbb{S}}^{b}_{s,e} + 2\log T$.

**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;

$$\begin{aligned}
\tilde{\mathbb{S}}^{b}_{s,e} &= \frac{\sqrt{\eta_{p_0+r}-s+1}\sqrt{e-b}}{\sqrt{e-\eta_{p_0+r}}\sqrt{b-s+1}}\tilde{\mathbb{S}}^{\eta_{p_0+r}}_{s,e}, \text{ and} \qquad (13)\\
\tilde{\mathbb{S}}^{\eta_{p_0+r}}_{s,e} - \tilde{\mathbb{S}}^{b}_{s,e} &= \left(1 - \frac{\sqrt{\eta_{p_0+r}-s+1}\sqrt{e-b}}{\sqrt{e-\eta_{p_0+r}}\sqrt{b-s+1}}\right)\tilde{\mathbb{S}}^{\eta_{p_0+r}}_{s,e}\\
&= \frac{\sqrt{1+\frac{b-\eta_{p_0+r}}{\eta_{p_0+r}-s+1}} - \sqrt{1-\frac{b-\eta_{p_0+r}}{e-\eta_{p_0+r}}}}{\sqrt{1+\frac{b-\eta_{p_0+r}}{\eta_{p_0+r}-s+1}}}\tilde{\mathbb{S}}^{\eta_{p_0+r}}_{s,e}\\
&\ge \frac{\frac{1}{2}\left(\frac{b-\eta_{p_0+r}}{\eta_{p_0+r}-s+1} + \frac{b-\eta_{p_0+r}}{e-\eta_{p_0+r}}\right)}{\sqrt{2}}\tilde{\mathbb{S}}^{\eta_{p_0+r}}_{s,e} \ge 2\log T
\end{aligned}$$

for large $T$, applying the Taylor expansion 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 (13) remains the same in the case of multiple breakpoints, the lemma is proved. □

LEMMA 3. *Let* $n \ge O(\delta_T)$. *Assume* $\exists\, b \in (s,e)$ *satisfying*

$$\max\left\{\sqrt{\frac{b-s+1}{e-b}}, \sqrt{\frac{e-b}{s-b+1}}\right\} \le c.$$

*Then* $\left|\tilde{\mathbb{Y}}^{b}_{s,e} - \tilde{\mathbb{S}}^{b}_{s,e}\right| \le \log T$ *with probability converging to* 1 *with* $T$, *uniformly over* $s$, $b$, $e$.

**Proof.** We need to show that

$$\Pr\left(\frac{1}{\sqrt{n}}\left|\sum_{t=s}^{e}\sigma^2(t/T)(Z^2_{t,T}-1)\cdot c_t\right| > \log T\right) \longrightarrow 0, \qquad (14)$$

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 $\{U_t\}^{e}_{t=s}$ be i.i.d. standard normal variables, $\mathbf{V} = (v_{i,j})^{n}_{i,j=1}$ with $v_{i,j} = \mathrm{corr}\,(Z_{i,T}, Z_{j,T})$, and $\mathbf{W} = (w_{i,j})^{n}_{i,j=1}$ 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 (14) 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 $\lambda_i$ 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

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

Therefore we can apply Bernstein's inequality (Bosq, 1998) and bound the probability in (14) by

$$2\exp\left(-\frac{n\log^2 T}{4\sum_{i=1}^{n}\lambda_i^2 + 2\max_i|\lambda_i|C\sqrt{n}\log T}\right). \qquad (15)$$

Note that $\sum_{i=1}^{n}\lambda_i^2 = \mathrm{tr}\,(\mathbf{VW})^2 \le c^2 \max_z \sigma^4(z) n\rho_\infty^2$. Also it follows that $\max_i|\lambda_i| \le c\max_z\sigma^2(z)\|\mathbf{V}\|$ where $\|\cdot\|$ denotes the spectral norm of a matrix, and $\|\mathbf{V}\| \le \rho_\infty^1$ since $\mathbf{V}$ is non-negative definite. Then the lemma follows as (15) is bounded from above by

$$\exp\left(-\frac{n\log^2 T}{4c^2\max_z\sigma^4(z)n\rho_\infty^2 + 2c\max_z\sigma^2(z)\sqrt{n}\log T\rho_\infty^1}\right) \longrightarrow 0,$$

as $\rho_\infty^p \le O(2^{I^*})$, which can be made to be of order $O(\log T)$, since the only requirement on $I^*$ is that it converges to infinity but no particular speed is required. □

LEMMA 4. *Assume (10) and (11). For* $b = \arg\max_{s<t<e}|\tilde{\mathbb{Y}}_{s,e}^t|$, *there exists* $1 \le r \le q$ *such that, for large* $T$,

$$|b - \eta_{p_0+r}| \le \epsilon_T. \tag{16}$$

**Proof.** Let $\tilde{\mathbb{S}}_{s,e} = \max_{s<t<e}|\tilde{\mathbb{S}}_{s,e}^t|$. From Lemma 3, $\tilde{\mathbb{Y}}_{s,e}^b \ge \tilde{\mathbb{S}}_{s,e} - \log T$ and $\tilde{\mathbb{S}}_{s,e}^b \ge \tilde{\mathbb{Y}}_{s,e}^b - \log T$, hence $\tilde{\mathbb{S}}_{s,e}^b \ge \tilde{\mathbb{S}}_{s,e} - 2\log T$. Assume (16) does not hold that $b \in (\eta_{p_0+r}+\epsilon_T, \eta_{p_0+r+1}-\epsilon_T)$ for some $r$. From Lemma 2.2 in Venkatraman (1993), $\tilde{\mathbb{S}}_{s,e}^t$ is either monotonic or decreasing and then increasing between two breakpoints (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 $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} > \tilde{\mathbb{S}}_{s,e}^b$. Then there exists $b' \in (\eta_{p_0+r}, \eta_{p_0+r}+\epsilon_T]$ satisfying $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} - 2\log T \ge \tilde{\mathbb{S}}_{s,e}^{b'}$ from Lemma 2. Since $b > b'$, we also get $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r+1}} > \tilde{\mathbb{S}}_{s,e}^b$ (as $\tilde{\mathbb{S}}_{s,e}^t$ is locally increasing at $t = b$), and there will again be a $b'' \in [\eta_{p_0+r+1}-\epsilon_T, \eta_{p_0+r+1})$ satisfying $\tilde{\mathbb{S}}_{s,e}^{\eta_{p_0+r}} - 2\log T \ge \tilde{\mathbb{S}}_{s,e}^{b''}$. Since $b'' > b$, it contradicts that $\tilde{\mathbb{S}}_{s,e}^b \ge \tilde{\mathbb{S}}_{s,e} - 2\log T$. Similar arguments are applicable when $b < \eta_{p_0+r}$ and therefore the lemma follows. □

LEMMA 5. *Under (10) and (11), for* $b = \arg\max_{s<t<e}|\tilde{\mathbb{Y}}_{s,e}^t|$,

$$\Pr\left(\left|\tilde{\mathbb{Y}}_{s,e}^b\right| < \tau T^\theta\sqrt{\log T}\cdot\frac{1}{n}\sum_{t=s}^{e}\tilde{Y}_{t,T}^2\right) \longrightarrow 0.$$

**Proof.** From Lemma 4, there exists some $r$ such that $|b - \eta_{p_0+r}| < \epsilon_T$. Denote $\tilde{m} = \sum_{t=s}^{e}\tilde{Y}_{t,T}^2/n$ and $\tilde{d} = \tilde{\mathbb{Y}}_{s,e}^b = \tilde{d}_1 - \tilde{d}_2$ where

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

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 $\Pr(|\tilde{d}| \le \tilde{m}\cdot t_n) \to 0$.

We first note that $w_i \le c^2\sup_z\sigma^4(z)\rho_\infty^2$. Using Markov's and the Cauchy-Schwarz inequalities, $\mu^2\Pr(\tilde{d} \le \tilde{m}\cdot t_n)$ is bounded by

$$\begin{aligned}
& \mu^2\Pr\left\{(\tilde{d}_1-\mu_1)(ct_n-1) + (\tilde{d}_2-\mu_2)(ct_n+1) + 2ct_n\mu_1 \ge (1+ct_n)\mu\right\} \\
\le\ & 3(1+ct_n)^{-2}\left\{(ct_n-1)^2w_1 + (ct_n+1)^2w_2 + 4c^2t_n^2\mu_1^2\right\} \\
\le\ & O\left\{\sup_z\sigma^4(z)\left(\rho_\infty^2 + \tau^2T^{2\theta}\log T\right)\right\},
\end{aligned}$$

and since $\mu = \tilde{\mathbb{S}}_{s,e}^b = O\left(\delta_T/\sqrt{T}\right) > T^\theta\sqrt{\log T}$, 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* $b = \arg\max_{s<t<e} |\tilde{\mathbb{Y}}^t_{s,e}|$ *and large* $T$,

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

**Proof.** Let $\mathcal{A} = \left\{\left|\tilde{\mathbb{Y}}^b_{s,e}\right| > \tau T^\theta \sqrt{\log T} \cdot \sum_{t=s}^{e} \tilde{Y}^2_{t,T}/n\right\}$ 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 $\sigma_2^2 = \sigma^2((\eta_p + 1)/T)$. First we assume (i) holds. We have

$$\Pr(\mathcal{A}) = \Pr(\mathcal{A}\,|\mathcal{B})\Pr(\mathcal{B}) + \Pr(\mathcal{A}\,|\mathcal{B}^c)\Pr(\mathcal{B}^c) \le \Pr(\mathcal{A}\,|\mathcal{B}) + \Pr(\mathcal{B}^c).$$

The first part is bounded as

$$\Pr(\mathcal{A}\,|\mathcal{B}) \le \frac{\Pr\left(\left|\tilde{\mathbb{Y}}^b_{s,e}\right| > \tau T^\theta \sqrt{\log T} \cdot \frac{1}{n}\sum_{t=s}^{e}\left(\mathbb{E}\tilde{Y}^2_{t,T} - h\right)\right)}{\Pr(\mathcal{B})}. \tag{17}$$

We have

$$\mathbb{E}\left(\tilde{\mathbb{Y}}^b_{s,e}\right)^2 \le n\left(\tilde{\mathbb{S}}^b_{s,e}\right)^2 \le \frac{n \cdot C\epsilon_T}{n - C\epsilon_T}\max_z \sigma^4(z)$$

and $n \ge O(\delta_T)$. Therefore by applying Markov's inequality, the numerator is bounded by

$$\mathbb{E}\left(\tilde{\mathbb{Y}}^b_{s,e}\right)^2 / (\tau h^2 T^{2\theta}\log T) \le O\left(T^{1/2-2\theta}\log T\right) \longrightarrow 0.$$

Turning our attention to the denominator of (17), we need to show that

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

This can be shown by applying Bernstein's inequality as in the proof of Lemma 2, and the lemma follows. Similar arguments are applied to prove the lemma 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 $O(\epsilon_T)$ from the true breakpoint, by Lemma 4. Under Assumption 2, both (10) and (11) 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.

Finally 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,*

$$T^{-1}\sum_{t=0}^{T-1}\left|\mathbb{E}I^{(i)}_{t,T}-\beta_i(t/T)\right|^2 = O(T^{-1}2^{-i})+b_{i,T}, \tag{18}$$

*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 $\eta_p$ as in (10), 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 $O(2^{-i})$) intervals around the discontinuities in $\beta_i(t/T)$. Suppose a jump occurs at $\eta_p$ in $\beta_i(t/T)$ yet there is no change in $\mathbb{E}I^{(i)}_{t,T}$ for $t\in[\eta_p-O(2^{-i}),\eta_p+O(2^{-i})]$. Then the integrated bias is bounded from below by $O(\delta_T/T)$ 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 $\eta_p$ and $\mathbb{E}I^{(i)}_{t_1,T}\neq\mathbb{E}I^{(i)}_{t_2,T}$ for $t_1\le\eta_p-O(2^{-i})$ and $t_2\ge\eta_p+O(2^{-i})$. 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 $|\hat{b}-b|\le O(2^{I^*})<\epsilon_T$ 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 $\Delta_T$ 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 $Y^2_{t,T}=\sigma^2_{t,T}\cdot Z^2_{t,T}$ as in (3).

## B. 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 $I^{(i)}_{t,T}$, $i=-1,\dots,-I^*$ 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 $\Lambda_T=O(\epsilon_T)$ from either $\hat{\nu}_q$ or $\hat{\eta}^{(i_0)}_{p_0}$, and $\left|\nu_q-\hat{\eta}^{(i_0)}_{p_0}\right|\le\epsilon_T$ with probability converging to 1 from Theorem 1. If there are $D(\le I^*)$ breakpoints detected for $\nu_q$, denote them as $\hat{\eta}^{(i_1)}_{p_1},\dots,\hat{\eta}^{(i_D)}_{p_D}$. Then for any $1\le a<b\le D$, $\left|\hat{\eta}^{(i_a)}_{p_a}-\hat{\eta}^{(i_b)}_{p_b}\right|\le\left|\hat{\eta}^{(i_a)}_{p_a}-\nu_q\right|+\left|\hat{\eta}^{(i_b)}_{p_b}-\nu_q\right|\le O(\epsilon_T)$, 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

Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *J. Am. Stat. Assoc. 93*, 1488–1501.

Andreou, A. and E. Ghysels (2002). Detecting multiple breaks in financial market volatility dynamics. *J. Appl. Econ. 17*, 579–600.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Bhattacharya, P. (1994). Some aspects of change-point analysis. *Change-point problems (eds E. Carlstein, H.-G. Müller and D. Siegmund) 23*, 28–56.

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

Chen, J. and A. K. Gupta (1997). Testing and locating variance change-points with application to stock prices. *J. Am. Stat. Assoc. 92*, 739–747.

Chernoff, H. and S. Zacks (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Ann. Math. Statist. 35*, 999–1028.

Choi, H., H. Ombao, and B. Ray (2008). Sequential change-point detection method in time series. *Technometrics 50*, 40–52.

Davis, R. A., T. C. M. Lee, and G. A. Rodriguez-Yam (2006). Structural break estimation for non-stationary time series. *J. Am. Stat. Assoc. 101*, 223–239.

Davis, R. A., T. C. M. Lee, and G. A. Rodriguez-Yam (2008). Break detection for a class of nonlinear time series models. *J. Time Ser. Anal. 29*, 834–867.

Fearnhead, P. (2005). Exact Bayesian curve fitting and signal segmentation. *IEEE Trans. Signal Process. 53*, 2160–2166.

Fryzlewicz, P. and G. Nason (2006). Haar-Fisz estimation of evolutionary wavelet spectra. *J. Roy. Stat. Soc. B 68*, 611–634.

Fryzlewicz, P., T. Sapatinas, and S. Subba Rao (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika 93*, 687–704.

Gabbanini, F., M. Vannucci, G. Bartoli, and A. Moro (2004). Wavelet packet methods for the analysis of variance of time series with application to crack widths on the Brunelleschi Dome. *J. Comput. Graph. Stat. 13*, 639–658.

Hawkins, D. M. (1977). Testing a sequence of observations for a shift in location. *J. Am. Stat. Assoc. 72*, 180–186.

Hsu, D. A. (1977). Tests for variance shifts at an unknown time point. *Appl. Statist. 26*, 179–184.

Hsu, D. A. (1979). Detecting shifts of parameters in gamma sequences with applications to stock price and air traffic flow analysis. *J. Am. Stat. Assoc. 74*, 31–40.

Inclán, C. and G. C. Tiao (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Am. Stat. Assoc. 89*, 913–923.

Janeway, W. (2009). Six impossible things before breakfast: Lessons from the crisis. *Significance 6*, 28–31.

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

Kokoszka, P. and R. Leipus (2000). Change-point estimation in ARCH models. *Bernoulli 6*, 513–539.

Lavielle, M. and E. Moulines (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal. 21*, 33–59.

McCulloch, R. E. and R. S. Tsay (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *J. Am. Stat. Assoc. 88*, 968–978.

Nason, G. P., R. von Sachs, and G. Kroisandt (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. Roy. Stat. Soc. B 62*, 271–292.

Ombao, H., J. Heo, and D. Stoffer (2004). *Online analysis of seismic signals: An almost-real-time approach*. New York: Springer.

Ombao, H. C., J. A. Raz, R. von Sachs, and B. A. Malow (2001). Automatic statistical analysis of bivariate nonstationary time series. *J. Am. Stat. Assoc. 96*, 543–560.

Punskaya, E., C. Andrieu, A. Doucet, and W. J. Fitzgerald (2002). Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Transactions on Signal Processing 50*, 747–758.

Sen, A. and M. S. Srivastava (1975). On tests for detecting change in mean. *Ann. Statist. 3*, 98–108.

Van Bellegem, S. and R. von Sachs (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Ann. Statist. 36*, 1879–1924.

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

Vidakovic, B. (1999). *Statistical modeling by wavelets*. Wiley.

Vostrikova, L. J. (1981). Detecting 'disorder' in multidimensional random processes. *Sov. Math. Doklady. 24*, 55–59.

Whitcher, B., S. D. Byers, P. Guttorp, and D. B. Percival (2002). Testing for homogeneity of variance in time series: Long memory, wavelets, and the Nile River. *Water Resour. Res. 38*, 10–1029.

Whitcher, B., P. Guttorp, and D. B. Percival (2000). Multiscale detection and location of multiple variance changes in the presence of long memory. *J. Statist. Comput. Simul. 68*, 65–87.

Worsley, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika 73*, 91–104.