Multiscale and multilevel technique for consistent segmentation of nonstationary time series

Haeran Cho† Piotr Fryzlewicz*
University of Bristol†
London School of Economics*

INSPIRE 2009
Imperial College London
22 September 2009
Introduction

(Weak) stationarity assumption in time series analysis

- Autocovariance structure does not change over time.
- Appealing when analysing short time series.
- Often unrealistic for longer processes.
- Many naturally occurring phenomena cannot be modelled as stationary processes.
  - Signal processing, quality control, econometrics, etc.
Introduction (cont’d)

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.

Dow-Jones Industrial Average index between 8 January 2007 and 16 January 2009
Piecewise stationarity

Simplest departure from stationarity ⇒ piecewise stationarity.

- Estimating a piecewise stationary process by consistently detecting its structural breakpoints.
  - Each segment between two breakpoints can be modelled (approximately) stationary.
  - Retrospective (a posteriori) segmentation.
  - Theoretically tractable, fast, well-performing.

Segmentation ↔ multiple breakpoint detection.
Piecewise stationarity (cont’d)

Simplest departure from stationarity ⇒ piecewise stationarity.

How many breakpoints do you see in this process?
Outline

● Locally Stationary Wavelet (LSW) model
  – Wavelets: whitening, rapidly computable, multiscale.
  – Piecewise stationary, linear time series.
  – No further parametric model assumed.
  – Local periodograms at multiple scales encode the entire autocovariance structure ⇒ basic statistics for the segmentation.
● A binary segmentation based method ⇒ multilevel
  – Applied to the local periodograms of original time series at each scale separately.
● Within-scale post-processing.
● Across-scales post-processing.
● Consistent estimation of breakpoints in the second-order structure of the original process.
Preliminaries: wavelets

Wavelets can be seen as “short oscillations”.

- Rescaled ($i$) and shifted in its location ($k$).
- e.g. the finest scale Haar wavelet $\psi_{-1} = (1/\sqrt{2}, -1/\sqrt{2})$ rescaled (left) and shifted (right).
Piecewise stationary processes

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

- \( i \): scale parameter.
- \( k \): location parameter.
- \( W_i(z) : [0, 1] \rightarrow \mathcal{R} \): real-valued, scale- and location-dependent piecewise constant function with finite but unknown number of jumps ⇔ “time-varying transfer function”.
- \( \psi_i \): discrete, real-valued, compactly supported, non-decimated wavelet vectors ⇔ “Fourier exponentials”.
- \( \xi_{i,k} \): i.i.d. Gaussian variables ⇔ “orthonormal increment process”.
Wavelet periodograms

The wavelet periodogram of $X_{t,T}$ at scale $i$

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

- A sequence of squared wavelet coefficients of $X_{t,T}$.
- Scaled $\chi_1^2$ variables.
Wavelet periodograms

The wavelet periodogram of $X_{t,T}$ at scale $i$

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

- A sequence of squared wavelet coefficients of $X_{t,T}$.
- Scaled $\chi_1^2$ variables.
Wavelet periodograms (cont’d)

\[ I^{(i)}_{t,T} = \mathbb{E}I^{(i)}_{t,T} \cdot Z^2_{t,T}. \]

- \( \mathbb{E}I^{(i)}_{t,T} \): “almost” piecewise constant as \( W_i(z) \).
  - The entire piecewise-constant autocovariance structure is encoded in \( \mathbb{E}I^{(i)}_{t,T} \).
- \( \{Z_{t,T}\}_{t=0}^{T-1} \): standard Gaussian variables (autocorrelated).
- Sufficient to look at the wavelet periodogram sequences at the fixed number of scales \( i = -1, \ldots, -I^* \).
  - At coarser scales, the autocorrelation within each \( I^{(i)}_{t,T} \) becomes stronger \( \Rightarrow \) little information is lost by disregarding coarse scales.
  - Unknown \( I^* \) \( \Rightarrow \) choose \( I^* < \lfloor \log_2 T \rfloor \) such that \( I^* \) is allowed to increase with \( T \).
Binary segmentation based procedure

wavelet periodogram

first iteration

second iteration 1

second iteration 2
Binary segmentation based procedure (cont’d)

- Test the validity of a detected breakpoint.
  - CUSUM type test statistic
    \[
    \sqrt{\frac{T-b}{T-b}} \sum_{t=0}^{b-1} I_{t,T}^{(i)} - \sqrt{\frac{b}{T(T-b)}} \sum_{t=b}^{T-1} I_{t,T}^{(i)}.
    \]

  - Scaled by the local mean over the segment \( \sum_{t=s}^{e} I_{t,T}^{(i)}/(e-s+1) \) \( \Rightarrow \) variance stabilization for multiplicative wavelet periodogram sequences.
  - Test criterion depends only on the length of the time series \( O(T^\theta \sqrt{\log T}) \) \( \Rightarrow \) rapidly computable.

- Applicable to generic multiplicative sequences, allowing autocorrelation in the data.
Within-scale post-processing

- Further reduces the risk of overestimating the number of breakpoints at each scale.
- For a single wavelet periodogram sequence, each breakpoint is checked against the adjacent ones.
Across-scales post-processing

Combines the estimated breakpoints across the scales $i = -1, \ldots, -I^*$. 
Consistency of breakpoint detection

**Theorem 1.** *The segmentation method combined with within-scale and across-scales post-processing procedures detects breakpoints which are consistent estimates of true breakpoints in the second-order structure of the original nonstationary process, in terms of their total number and locations.*
Simulation: overview

Sim 1 Piecewise stationary AR(2) process with observable changes in the parameters.
Sim 2 PS AR(1) process with less observable changes in the parameters.
Sim 3 PS AR(1) process with a short segment in the parameters.
Sim 4 Random-walk-like PS AR(1) with changes only in the variance.
Simulation: examples

- dotted red lines: true breakpoints, dashed blue lines: detected breakpoints
Simulation: outcome

Table 1: Summary of the simulation study in total number of breakpoints detected

<table>
<thead>
<tr>
<th>number of breakpoints (%)</th>
<th>Sim 1</th>
<th>Sim 2</th>
<th>Sim 3</th>
<th>Sim 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ours</td>
<td>Ours</td>
<td>Ours</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td>Auto-PARM†</td>
<td>AP</td>
<td>AP</td>
<td>AP</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>94</td>
<td>99</td>
</tr>
<tr>
<td>2</td>
<td>90</td>
<td>97</td>
<td>100</td>
<td>94</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>1</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Simulation 4: the example from Introduction

Random-walk-like piecewise stationary AR(1) process with changes in variance only.

- Financial time series, such as stock indices, are often modelled as random walk with a time-varying variance for e.g. pricing of derivative instruments.
Simulation 4 (cont’d)

\[ X_t = \begin{cases} 
0.999X_{t-1} + \epsilon_t & \text{for } 1 \leq t \leq 400, \\
0.999X_{t-1} + 1.5\epsilon_t & \text{for } 401 \leq t \leq 750, \\
0.999X_{t-1} + \epsilon_t & \text{for } 751 \leq t \leq 1024
\end{cases} \]
Simulation 4 (cont’d)

Selection frequency for each $t$ over 100 realizations.

(a) The breakpoints detected from our method centre on $\eta_1 = 400$, $\eta_2 = 750$,

(b) while those from the Auto-PARM are scattered over time.
Daily closing values of Dow-Jones Industrial Average from 8 January 2007 to 16 January 2009 ($T = 512 = 2^9$).
The first breakpoint ($\hat{\eta}_1 = 135$) coincided with the outbreak of the worldwide “credit crunch” (the last week of July 2007).

The second breakpoint ($\hat{\eta}_2 = 424$) coincided with the bankruptcy of Lehman Brothers, a major financial services firm (September 14, 2008).
Conclusion

• LSW model ⇒ wavelet periodograms following a multiplicative statistical model.

• Our binary segmentation procedure allows correlated data ⇒ essential to work with wavelet periodograms.

• Test criterion depends only on the length of the time series and is thus fast to compute.

• Novel across-scales post-processing step.

• Good performance and consistency in probability.
Thank you!

  
  http://www.maths.bris.ac.uk/~mahrc/msml_technique.pdf