Detecting breakpoints in piecewise stationary AR processes

Haeran Cho    Piotr Fryzlewicz

University of Bristol, UK

Young European Statisticians Workshop (YES-II), EURANDOM, Eindhoven, 7 October 2008
Outline

1. Introduction

2. Model description

3. Breakpoint detection procedure
   - Process transformation
   - Piecewise constant function estimation

4. Simulation
Introduction

- Stationarity assumption
  - Common for short time series.
  - Often unrealistic for longer processes.
  - Many naturally occurring phenomena cannot be modelled as stationary processes.
    - Speech processing, biomedical signal processing, seismology, etc.
Introduction

- Piecewise stationary AR processes
  - The simplest approach to model nonstationary time series.
  - Locally stationary time series can be approximated by piecewise stationary processes. (Adak, 1998)
  - Weakly stationary processes can be approximated by AR processes. (Brockwell and Davis, 1996)
Our method

Main objectives
- A posteriori segmentation of piecewise stationary AR processes.
- Approximation of locally stationary processes.

Our breakpoint detection method proceeds in two steps.
- Transform $X_t$ into pre-estimate sequences.
- Estimate piecewisely changing structure from each pre-estimate.
- Under the assumption that breakpoints are sufficiently “sparse”.

Haeran Cho
haeran.cho@bristol.ac.uk
(University of Bristol, UK)

Detecting breakpoints in piecewise stationary processes
7 October 2008
Piecewise stationary AR processes

A piecewise stationary AR($p$) process $X_t$ with $m$ breakpoints and $(m + 1)$ independent stationary segments

$$\Rightarrow X_t = \begin{cases} 
X_t^{(1)}, & \text{for } \eta_0 = 1 \leq t \leq \eta_1 \\
\vdots \\
X_t^{(m+1)}, & \text{for } \eta_m + 1 \leq t \leq \eta_{m+1} = n
\end{cases}$$

where each $X_t^{(j)}$ is an AR process of order $p$, i.e.,

$$X_t^{(j)} = \beta_{t,1}X_{t-1}^{(j)} + \beta_{t,2}X_{t-2}^{(j)} + \cdots + \beta_{t,p}X_{t-p}^{(j)} + \epsilon_t.$$ 

- $p$: the maximum among the orders of all stationary segments.
- At each breakpoint $\eta_j$, there exists at least one parameter $\beta_{t,i}$ with a “jump” for $i = 1, \ldots, p$. 

Haeran Cho haeran.cho@bristol.ac.uk (University of Bristol, UK) 

Detecting breakpoints in piecewise stationarity
Outline

1. Introduction

2. Model description

3. Breakpoint detection procedure
   - Process transformation
   - Piecewise constant function estimation

4. Simulation
Autocovariance and AR parameters

- $\gamma_t(r) := \mathbb{E} \left( X_t^{(j)} X_{t-r}^{(j)} \right)$: time-varying, piecewise constant autocovariance function of $X_t$.

- Within each stationary segment, AR parameters $\beta_{t,1}, \cdots, \beta_{t,p}$ satisfy

$$
\begin{pmatrix}
\gamma_t(0) & \gamma_t(1) & \cdots & \gamma_t(p-1) \\
\gamma_t(1) & \gamma_t(0) & \cdots & \gamma_t(p-2) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_t(p-1) & \gamma_t(p-2) & \cdots & \gamma_t(0)
\end{pmatrix}
\begin{pmatrix}
\beta_{t,1} \\
\beta_{t,2} \\
\vdots \\
\beta_{t,p}
\end{pmatrix} = 
\begin{pmatrix}
\gamma_t(1) \\
\gamma_t(2) \\
\vdots \\
\gamma_t(p)
\end{pmatrix}.
$$

$\Rightarrow$ Any break in $\beta_{t,i}, i = 1, \cdots, p$ is reflected as a break in at least one of $\gamma_t(r), r = 0, \cdots, p$. 

Haeran Cho

Detecting breakpoints in piecewise stationarity

7 October 2008 8 / 20
Pre-estimates for the autocovariance function

- Localised estimates of $\gamma_t(r)$ at lags $r = 0, \cdots, p$ as our pre-estimate sequences

\[ \tilde{\gamma}_{t,r} := X_t X_{t-r} = \mathbb{E}(X_t X_{t-r}) + \{X_t X_{t-r} - \mathbb{E}(X_t X_{t-r})\}. \]

- Pre-estimate sequences of $X_t$ at lags 0, 1, and 2.
Gaussian-like noise process

Let \( \nu_{t,r} := X_tX_{t-r} - \mathbb{E}(X_tX_{t-r}) \).

- Within each stationary segment, the noise process \( \nu_{t,r} \) “behaves like” the Gaussian process.
- Itself and its all partial sums can be bounded by a term of order \( O(\sqrt{\log n}) \) with probability converging to 1.
Outline

1 Introduction

2 Model description

3 Breakpoint detection procedure
   - Process transformation
   - Piecewise constant function estimation

4 Simulation
Estimation procedure

- **Aim:** Estimate piecewise constant functions with sparse breakpoints based on the pre-estimates.

- Let $x_t,r := \gamma_t(r) - \gamma_{t-1}(r)$, then only few $x_t,r$ are nonzero and

\[
\begin{pmatrix}
\tilde{\gamma}_1,r \\
\tilde{\gamma}_2,r \\
\vdots \\
\tilde{\gamma}_n,r
\end{pmatrix}
\begin{pmatrix}
1 & 0 & \cdots & 0 \\
1 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
x_1,r \\
x_2,r \\
\vdots \\
x_n,r
\end{pmatrix}
+ \begin{pmatrix}
v_1,r \\
v_2,r \\
\vdots \\
v_n,r
\end{pmatrix} := Kx + v
(Donoho, 2004) If there exists any sufficiently sparse solution (\(\equiv\) minimum \(l_0\)-norm) for a linear regression problem:

- the solution with minimum \(l_1\)-norm is a good approximation in \(l_2\) sense.
- while the procedure of looking for a solution with minimum \(l_0\)-norm is intractable, minimising \(l_1\)-norm changes the problem into a computationally feasible convex problem.
Therefore our estimate $\hat{\gamma}_{t,r}$ for each pre-estimate $\tilde{\gamma}_{t,r}$

- has the minimum total variation $\sum_{t=1}^{n-1} |\hat{\gamma}_{t+1,r} - \hat{\gamma}_{t,r}|$, and
- produces residuals which also behaves like the Gaussian process as $\nu_{t,r}$, i.e., the estimated residuals and all their partial sums are bounded by $\sigma \sqrt{\tau \log n}$.

Implementation: minimize $\|x\|_1 = \sum_t |x_t|$ subject to

$$\|K' (\tilde{\gamma} - Kx)\|_\infty \leq \sigma \sqrt{\tau \log n}$$

where $K'$ is column-wisely normalised $K$. 

Haeran Cho
haeran.cho@bristol.ac.uk
(University of Bristol, UK)
Comparison with the Dantzig selector

The Dantzig selector (Candès and Tao, 2005) solves a linear regression problem \( \mathbf{y} = \mathbf{X}\beta + \epsilon \) by

\[
\text{minimising } \|\beta\|_1 \text{ subject to } \|\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta)\|_\infty \leq \sigma \sqrt{\tau \log p}.
\]

\( \mathbf{K} \) does not satisfy the “uniform uncertainty principle” conditions imposed on \( \mathbf{X} \) in the paper.

- Any submatrix with \( S \) (sparsity) or fewer columns of \( \mathbf{X} \) should behave as if it were almost orthogonal.

Our interest is on locating breakpoints and highly correlated columns of \( \mathbf{K} \) are adjacent to each other.
Theoretical property

Assumption

The length of each segment between two adjacent breakpoints, say $d := d_n$, satisfies $\log n / d_n \to 0$.

With this assumption,

- the distance between true and detected breakpoints is bounded by $O(\log n)$
- any fluctuation between two neighbouring breakpoints is bounded by $o(1)$

with probability converging to 1.
Simulation study

The piecewise stationary AR(2) process below was repeatedly generated 50 times.

\[
X_t = \begin{cases} 
0.4X_{t-1} - 0.6X_{t-2} + \epsilon_t, & 1 \leq t \leq 150 \\
-0.2X_{t-1} + \epsilon_t, & 151 \leq t \leq 300 \\
0.5X_{t-1} + \epsilon_t, & 301 \leq t \leq 450 
\end{cases}
\]
Simulation study

Figure: (a) An example of $X_t$; (b) $\tilde{\gamma}_{t,1}$ and $\hat{\gamma}_{t,1}$; (c) $\tilde{\gamma}_{t,2}$ and $\hat{\gamma}_{t,2}$.
Simulation study

Table: Summary of the estimated breakpoints from the simulation

<table>
<thead>
<tr>
<th>Number of segments</th>
<th>Breakpoints (%)</th>
<th>Distance from true breakpoint</th>
<th>Breakpoints (%)</th>
<th>t = 150</th>
<th>t = 300</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>[log(n)] = 6</td>
<td>56</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>[2 log(n)] = 12</td>
<td>70</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>88</td>
<td>[3 log(n)] = 18</td>
<td>76</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>[4 log(n)] = 24</td>
<td>78</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>&gt; 3</td>
<td>0</td>
<td>[5 log(n)] = 30</td>
<td>86</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>