Structural Break Detection in Time Series Models

Richard A. Davis

Thomas Lee

Gabriel Rodriguez-Yam

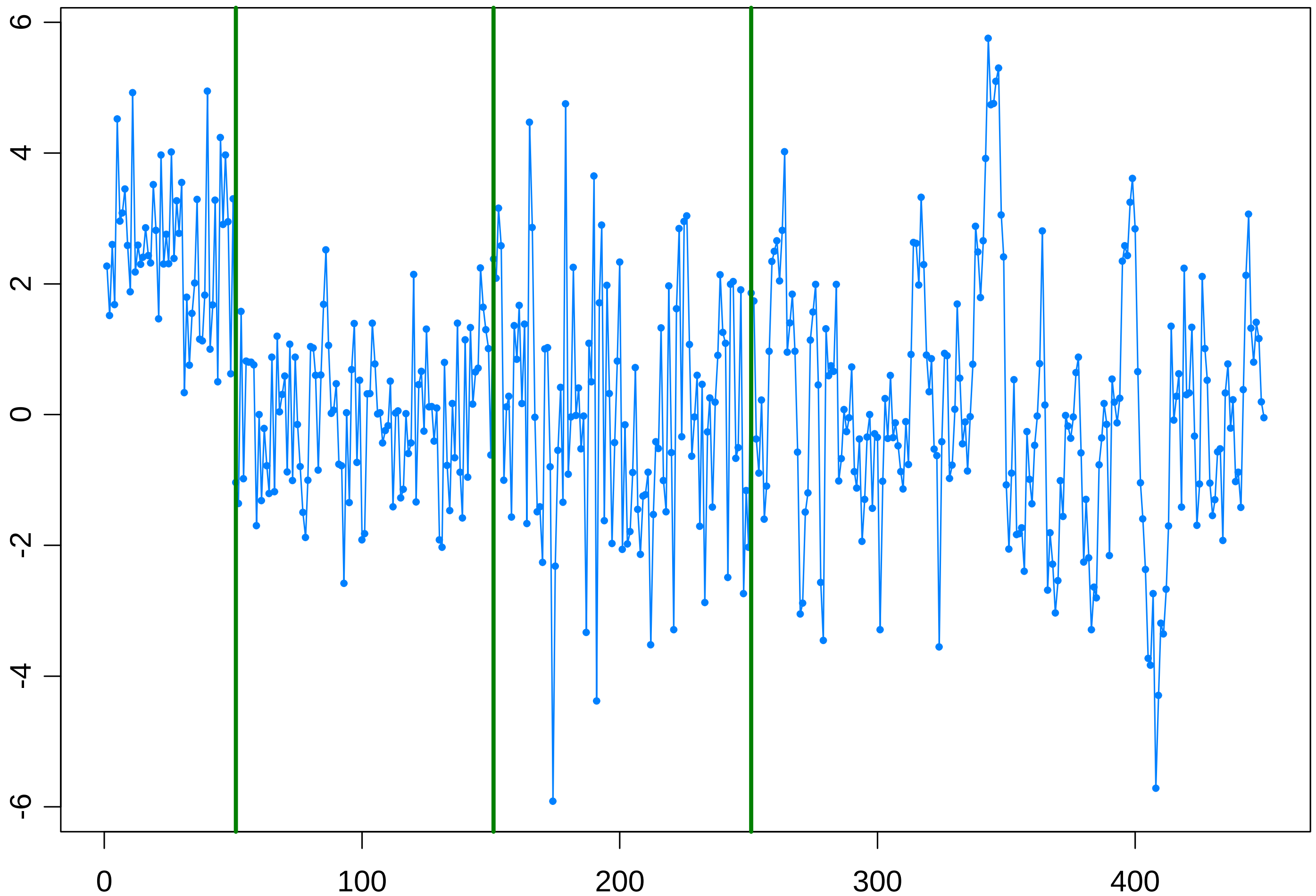
Colorado State University

(http://www.stat.colostate.edu/~rdavis/lectures)

This research supported in part by an IBM faculty award.

Illustrative Example

How many segments do you see?

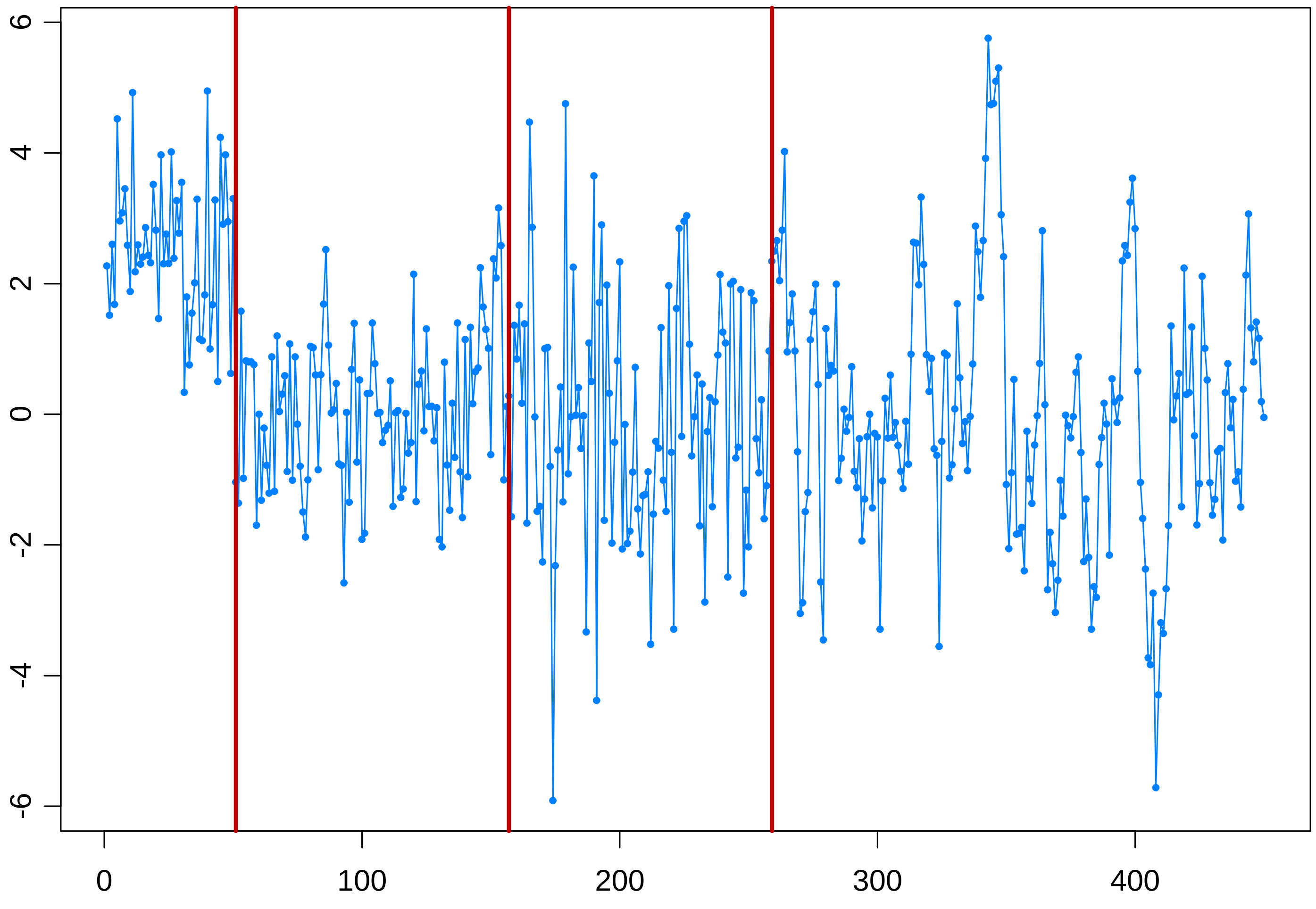


$\tau_1 = 51$ $\tau_2 = 151$ time $\tau_3 = 251$

Illustrative Example

Auto-PARM=Auto-Piecewise AutoRegressive Modeling

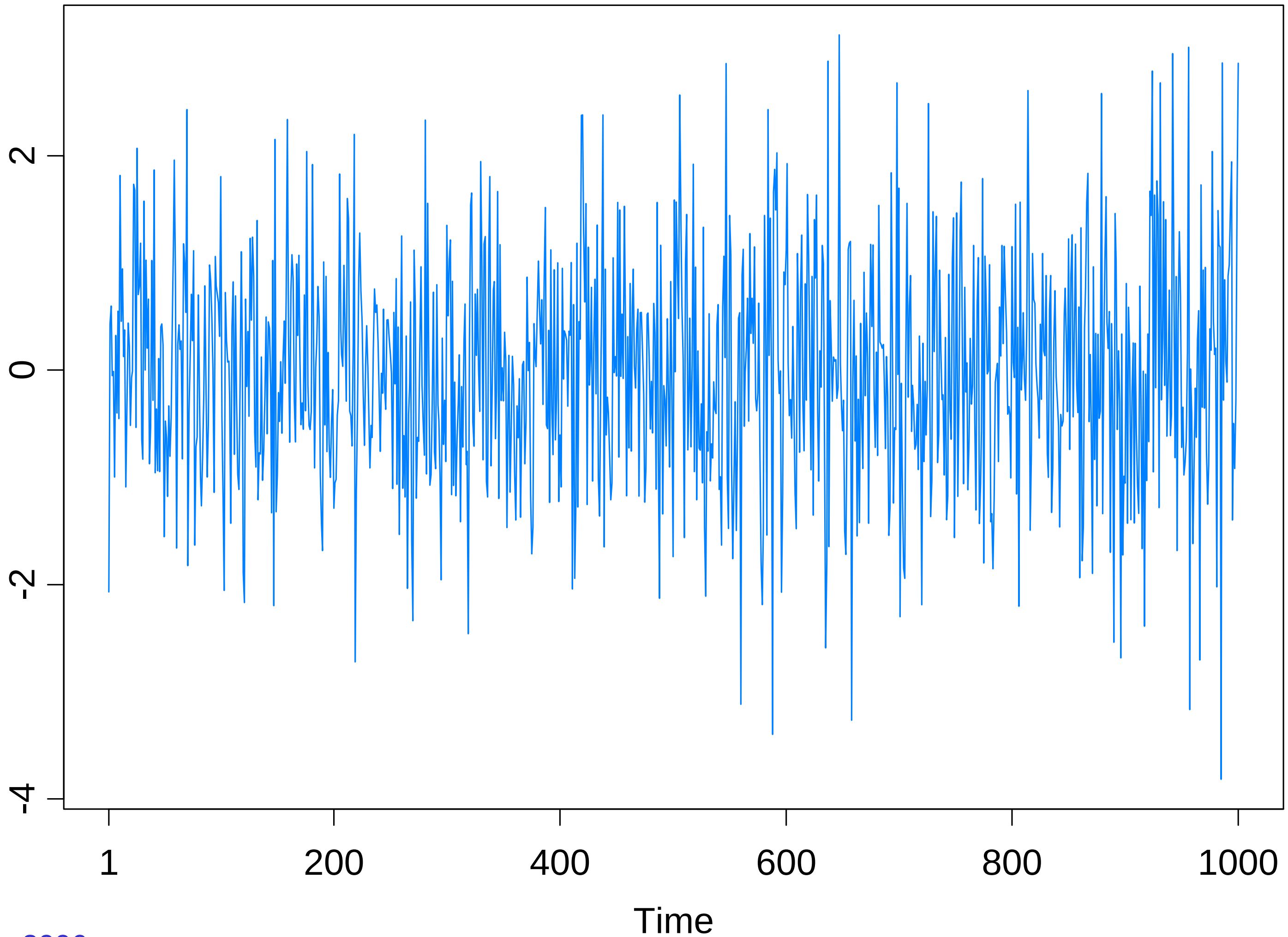
4 pieces, 2.58 seconds.



$\tau_1 = 51$ $\tau_2 = 157$ $\tau_3 = 259$

A Second Example

Any breaks in this series?



- Introduction
  - Examples
    - AR
    - GARCH
    - Stochastic volatility
    - State space models
- Model selection using Minimum Description Length (MDL)
  - General principles
  - Application to AR models with breaks
- Optimization using a Genetic Algorithm
  - Basics
  - Implementation for structural break estimation
- Simulation results
- Applications
- Simulation results for GARCH and SV models

Examples

1. Piecewise AR model:

$$Y_t = \gamma_j + \phi_{j1} Y_{t-1} + \cdots + \phi_{jp_j} Y_{t-p_j} + \sigma_j \varepsilon_t, \quad \text{if } \tau_{j\text{-}1} \le t < \tau_j,$$

where $\tau_0 = 1 < \tau_1 < \ldots < \tau_{m\text{-}1} < \tau_m = n + 1$, and $\{\varepsilon_t\}$ is IID(0,1).

Goal: Estimate

m = number of segments

τ_j = location of j^{th} break point

γ_j = level in j^{th} epoch

p_j = order of AR process in j^{th} epoch

$(\phi_{j1}, \ldots, \phi_{jp_j})$ = AR coefficients in j^{th} epoch

σ_j = scale in j^{th} epoch

Examples (cont)

2. Segmented GARCH model:

$$Y_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \omega_j + \alpha_{j1} Y_{t-1}^2 + \cdots + \alpha_{jp_j} Y_{t-p_j}^2 + \beta_{j1}\sigma_{t-1}^2 + \cdots + \beta_{jq_j}\sigma_{t-q_j}^2, \quad \text{if } \tau_{j\text{-}1} \le t < \tau_j,$$

where $\tau_0 = 1 < \tau_1 < \ldots < \tau_{m\text{-}1} < \tau_m = n + 1$, and $\{\varepsilon_t\}$ is IID(0,1).

3. Segmented stochastic volatility model:

$$Y_t = \sigma_t \varepsilon_t,$$
$$\log \sigma_t^2 = \gamma_j + \phi_{j1} \log \sigma_{t-1}^2 + \cdots + \phi_{jp_j} \log \sigma_{t-p_j}^2 + \nu_j \eta_t, \quad \text{if } \tau_{j\text{-}1} \le t < \tau_j.$$

4. Segmented state-space model (SVM a special case):

$$p(y_t \mid \alpha_t, \ldots, \alpha_1, y_{t-1}, \ldots, y_1) = p(y_t \mid \alpha_t) \text{ is specified}$$
$$\alpha_t = \gamma_j + \phi_{j1}\alpha_{t-1} + \cdots + \phi_{jp_j}\alpha_{t-p_j} + \sigma_j \eta_t, \quad \text{if } \tau_{j\text{-}1} \le t < \tau_j.$$

Model Selection Using Minimum Description Length

Basics of MDL:

Choose the model which *maximizes the compression* of the data or, equivalently, select the model that *minimizes the code length* of the data (i.e., amount of memory required to encode the data).

$\boldsymbol{M}$ = class of operating models for $y = (y_1, \ldots, y_n)$

$L_{\boldsymbol{F}}(y)$ = code length of y relative to $\boldsymbol{F} \in \boldsymbol{M}$

Typically, this term can be decomposed into two pieces (two-part code),

$$L_{\boldsymbol{F}}(y) = L(\hat{\boldsymbol{F}}|y) + L(\hat{e}\,|\,\hat{\boldsymbol{F}}),$$

where

$L(\hat{\boldsymbol{F}}|y)$ = code length of the fitted model for $\boldsymbol{F}$

$L(\hat{e}|\hat{\boldsymbol{F}})$ = code length of the residuals based on the fitted model

Model Selection Using Minimum Description Length (cont)

Applied to the segmented AR model:

$$Y_t = \gamma_j + \phi_{j1} Y_{t-1} + \cdots + \phi_{jp_j} Y_{t-p_j} + \sigma_j \varepsilon_t, \quad \text{if } \tau_{j\text{-}1} \le t < \tau_j,$$

First term $L(\hat{\mathcal{F}}|y)$:

$$\begin{aligned} L(\hat{\mathcal{F}}|y) &= L(m) + L(\tau_1,\ldots,\tau_m) + L(p_1,\ldots,p_m) + L(\hat{\psi}_1 \mid y) + \cdots + L(\hat{\psi}_m \mid y) \\ &= \log_2 m + m \log_2 n + \sum_{j=1}^{m} \log_2 p_j + \sum_{j=1}^{m} \frac{p_j + 2}{2} \log_2 n_j \end{aligned}$$

Second term $L(\hat{e} \mid \hat{\mathcal{F}})$:

$$L(\hat{e} \mid \hat{\mathcal{F}}) \approx -\sum_{j=1}^{m} \log_2 L(\hat{\psi}_j \mid y)$$

$$\begin{aligned} &MDL(m,(\tau_1,p_1),\ldots,(\tau_m,p_m)) \\ &\quad = \log_2 m + m \log_2 n + \sum_{j=1}^{m} \log_2 p_j + \sum_{j=1}^{m} \frac{p_j + 2}{2} \log_2 n_j + \sum_{j=1}^{m} (\log_2(2\pi\hat{\sigma}_j^2) + n_j) \end{aligned}$$

Optimization Using Genetic Algorithm

Basics of GA:

Class of optimization algorithms that mimic natural evolution.

- Start with an initial set of *chromosomes*, or population, of possible solutions to the optimization problem.
- Parent chromosomes are randomly selected (proportional to the rank of their objective function values), and produce offspring using *crossover* or *mutation* operations.
- After a sufficient number of offspring are produced to form a second generation, the process then *restarts to produce a third generation*.
- Based on Darwin's *theory of natural selection*, the process should produce future generations that give a *smaller (or larger)* objective function.

Optimization Using Genetic Algorithm

Genetic Algorithm: Chromosome consists of n genes, each taking the value of –1 (no break) or p (order of AR process). Use natural selection to find a *near* optimal solution.

Map the break points with a chromosome c via

$$(m,(\tau_1, p_1)\ldots,(\tau_m, p_m)) \longleftrightarrow c = (\delta_1,\ldots,\delta_n),$$

where

$$\delta_t = \begin{cases} -1, & \text{if no break point at } t, \\ p_j, & \text{if break point at time } t = \tau_{j-1} \text{ and AR order is } p_j. \end{cases}$$

For example,

c = (2, -1, -1, -1, -1, 0, -1, -1, -1, -1, 0, -1, -1, -1, 3, -1, -1, -1, -1,-1)
t: 1 6 11 15

would correspond to a process as follows:

AR(2), t=1:5; AR(0), t=6:10; AR(0), t=11:14; AR(3), t=15:20

Implementation of Genetic Algorithm—(cont)

Generation 0: Start with L (200) randomly generated chromosomes, $c_1, \ldots, c_L$ with associated MDL values, $MDL(c_1), \ldots, MDL(c_L)$.

Generation 1: A new child in the next generation is formed from the chromosomes $c_1, \ldots, c_L$ of the previous generation as follows:

- with probability π_c, *crossover* occurs.
  - two parent chromosomes c_i and c_j are selected at random with probabilities proportional to the ranks of $MDL(c_i)$.
  - k^{th} gene of child is $\delta_k = \delta_{i,k}$ w.p. ½ and $\delta_{j,k}$ w.p. ½
- with probability $1-\pi_c$, *mutation* occurs.
  - a parent chromosome c_i is selected
  - k^{th} gene of child is $\delta_k = \delta_{i,k}$ w.p. π_1 ; -1 w.p. π_2;and p w.p. $1-\pi_1-\pi_2$.

Implementation of Genetic Algorithm—(cont)

Execution of GA: Run GA until *convergence* or until a *maximum number of generations* has been reached. *.*

Various Strategies:

- include the *top ten* chromosomes from last generation in next generation.

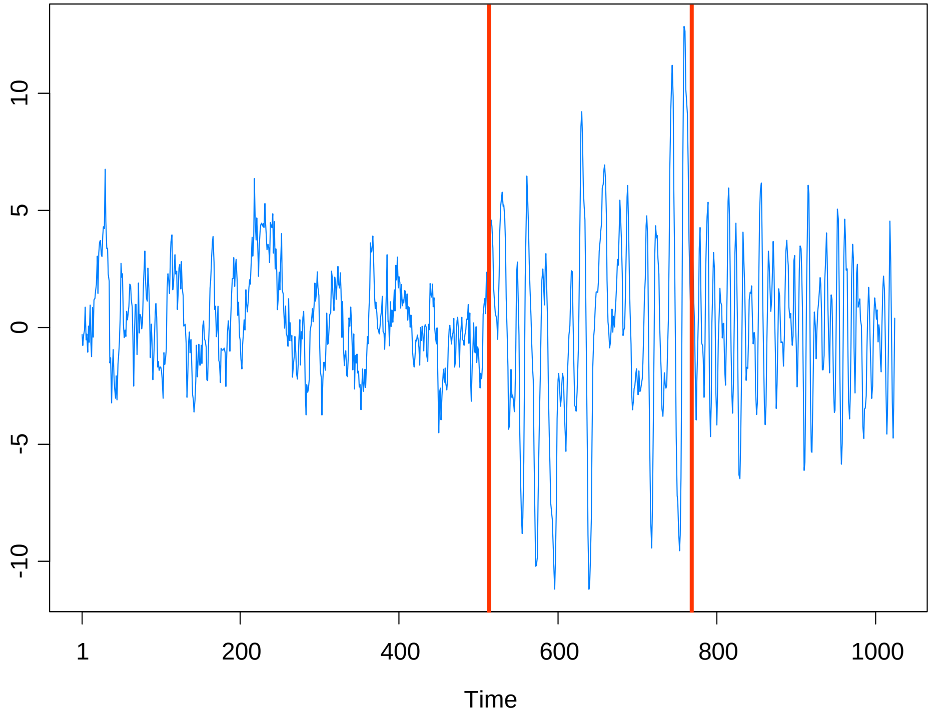
- use multiple *islands*, in which populations run independently, and then allow *migration* after a fixed number of generations. This implementation is amenable to *parallel computing*.

Simulation Examples-based on Ombao et al. (2001) test cases

1. Piecewise stationary with dyadic structure: Consider a time series following the model,

$$Y_t = \begin{cases} .9Y_{t-1} + \varepsilon_t, & \text{if } 1 \le t < 513, \\ 1.69Y_{t-1} - .81Y_{t-2} + \varepsilon_t, & \text{if } 513 \le t < 769, \\ 1.32Y_{t-1} - .81Y_{t-2} + \varepsilon_t, & \text{if } 769 \le t \le 1024, \end{cases}$$

where $\{\varepsilon_t\}$ ~ IID N(0,1).

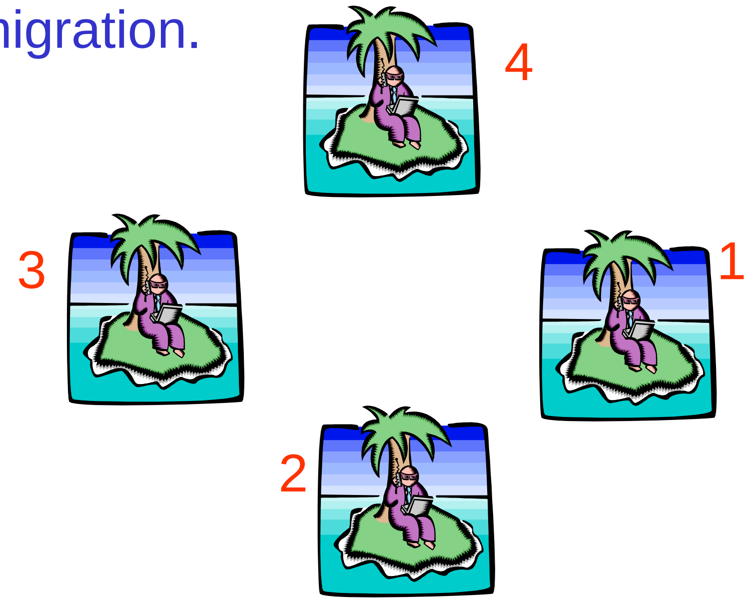


1. Piecewise stat (cont)

Implementation: Start with $NI = 50$ islands, each with population size $L = 200$.

After every $Mi = 5$ generations, allow migration.

Replace worst 2 in Island 1
with best 2 from Island 4.



Stopping rule: Stop when the max MDL does not change for 10 consecutive migrations or after 100 migrations.

Span configuration for model selection: Max AR order $K = 10$,

p	0	1	2	3	4	5	6	7-10	11-20
m_p	10	10	12	14	16	18	20	25	50
π_p	1/21	1/21	1/21	1/21	1/21	1/21	1/21	1/21	1/21

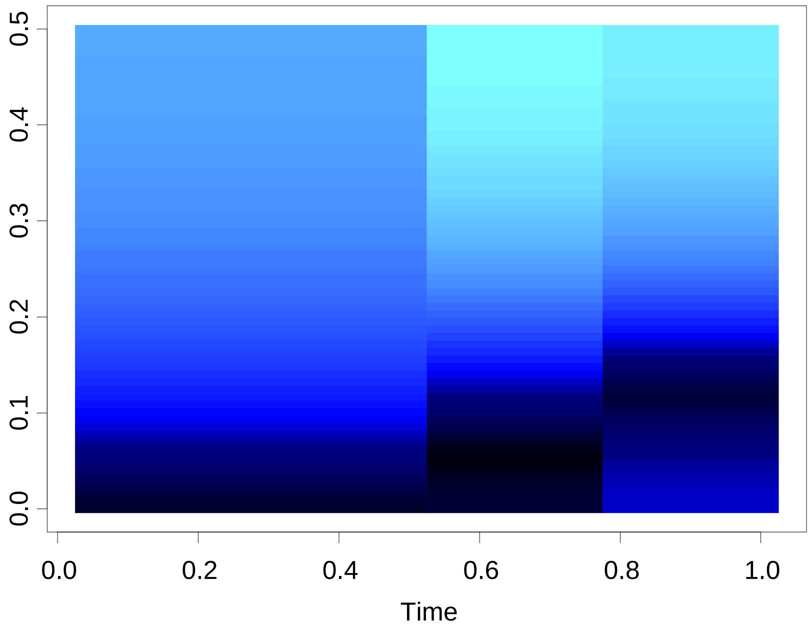
1. Piecewise stat (cont)

GA results: 3 pieces breaks at τ_1=513; τ_2=769. Total run time 16.31 secs

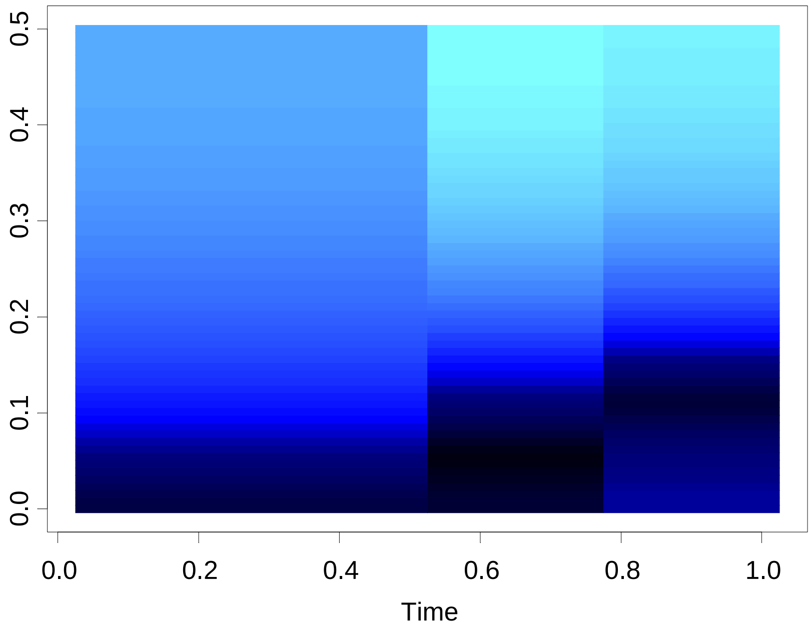
Fitted model:

	ϕ_1	ϕ_2	σ^2
1- 512:	.857		.9945
513- 768:	1.68	-0.801	1.1134
769-1024:	1.36	-0.801	1.1300

True Model



Fitted Model

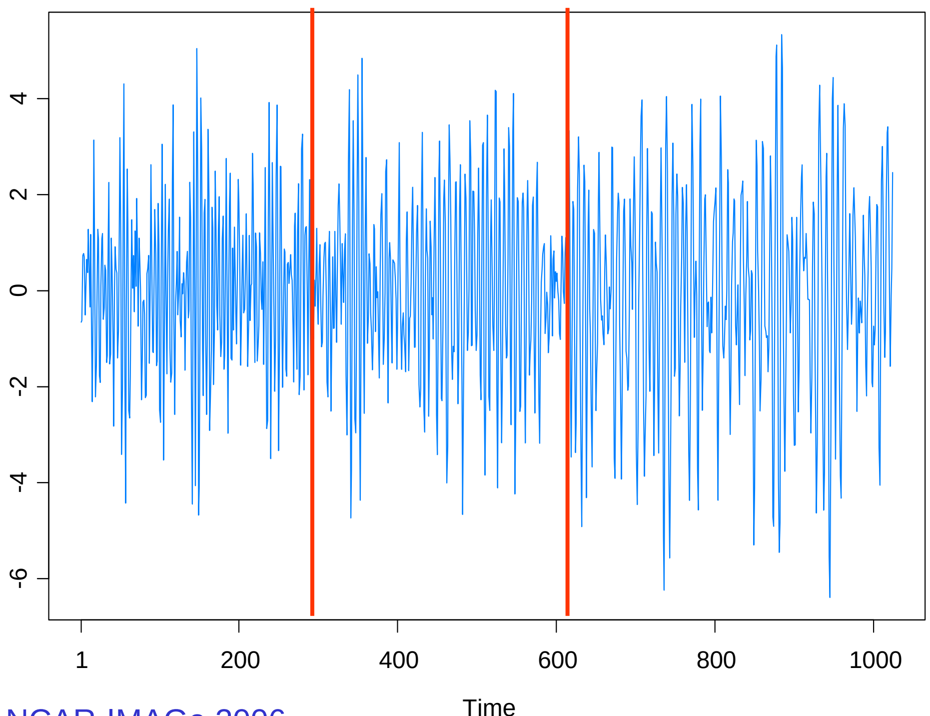


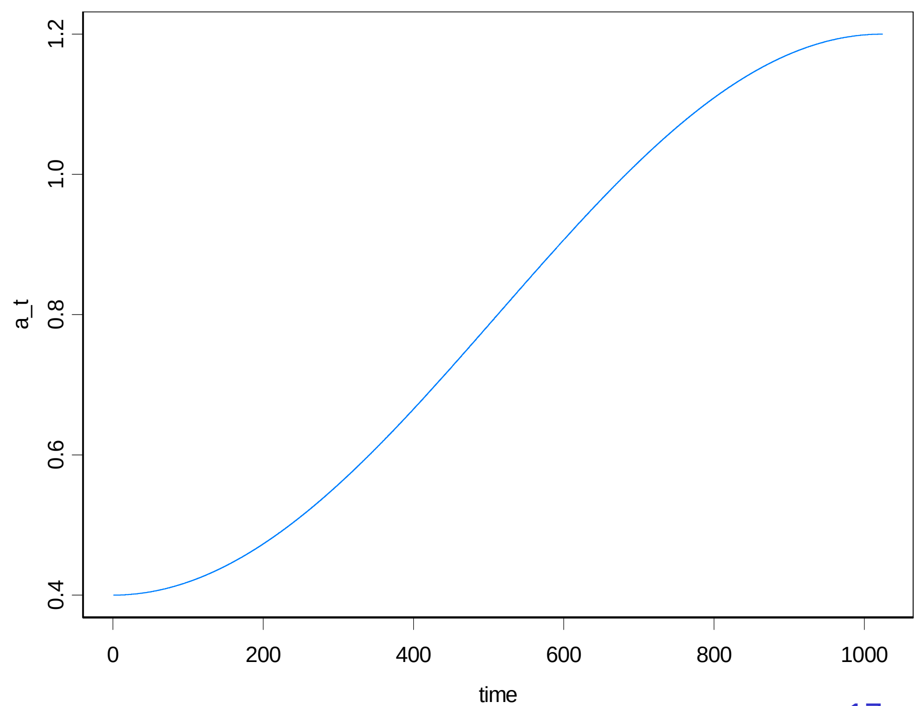
Simulation Examples (cont)

2. Slowly varying AR(2) model:

$$Y_t = a_t Y_{t-1} - .81 Y_{t-2} + \varepsilon_t \quad \text{if} \;\; 1 \le t \le 1024$$

where $a_t = .8[1 - 0.5\cos(\pi t / 1024)]$, and $\{\varepsilon_t\}$ ~ IID N(0,1).





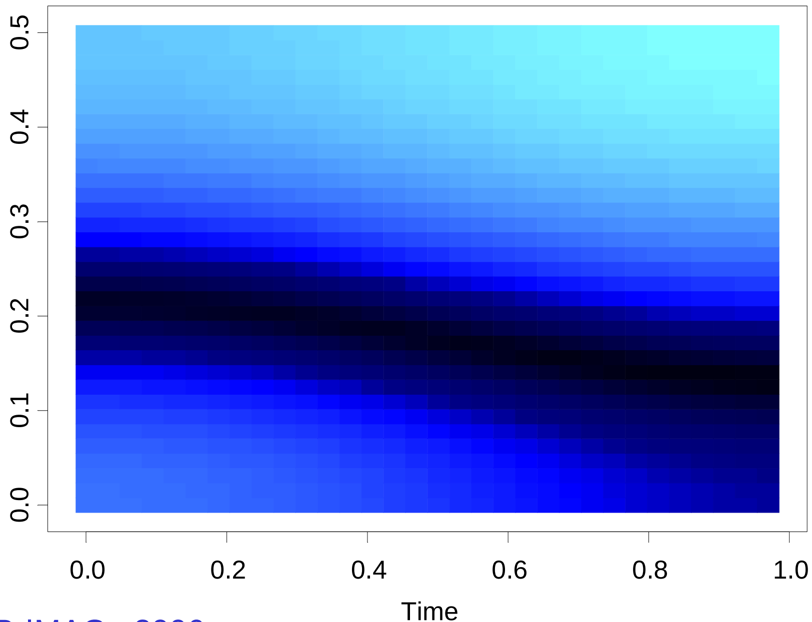
2. Slowly varying AR(2) (cont)

GA results: 3 pieces, breaks at τ_1=293, τ_2=615. Total run time 27.45 secs

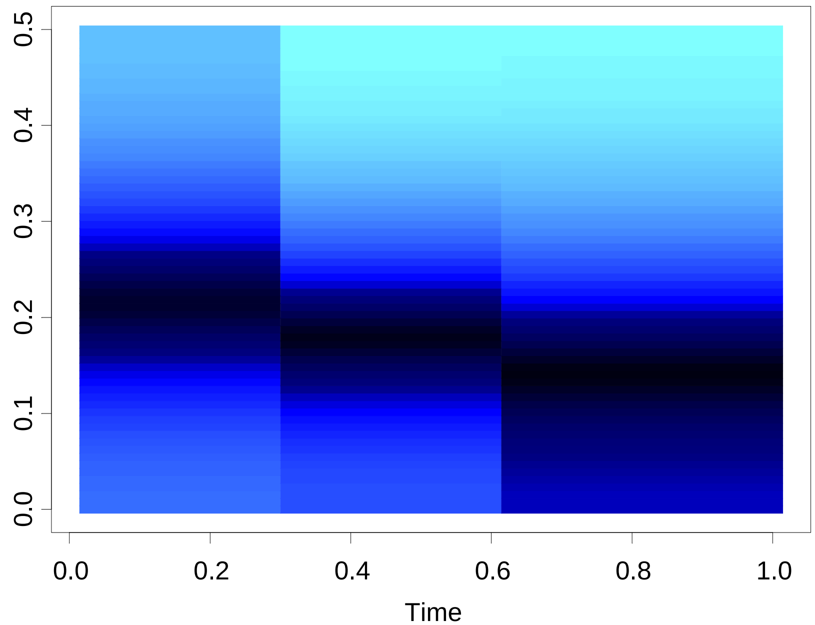
Fitted model:

	ϕ_1	ϕ_2	σ^2
1- 292:	.365	-0.753	1.149
293- 614:	.821	-0.790	1.176
615-1024:	1.084	-0.760	0.960

True Model



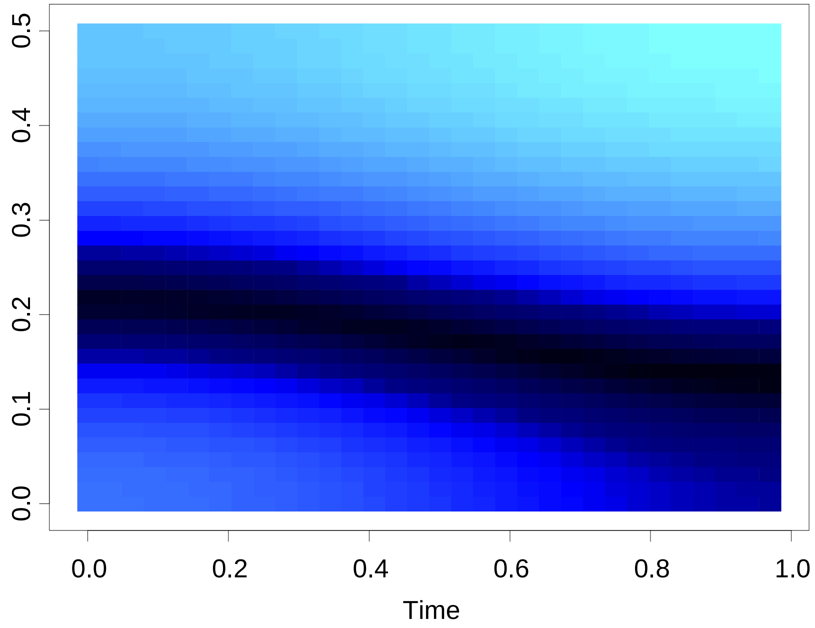
Fitted Model



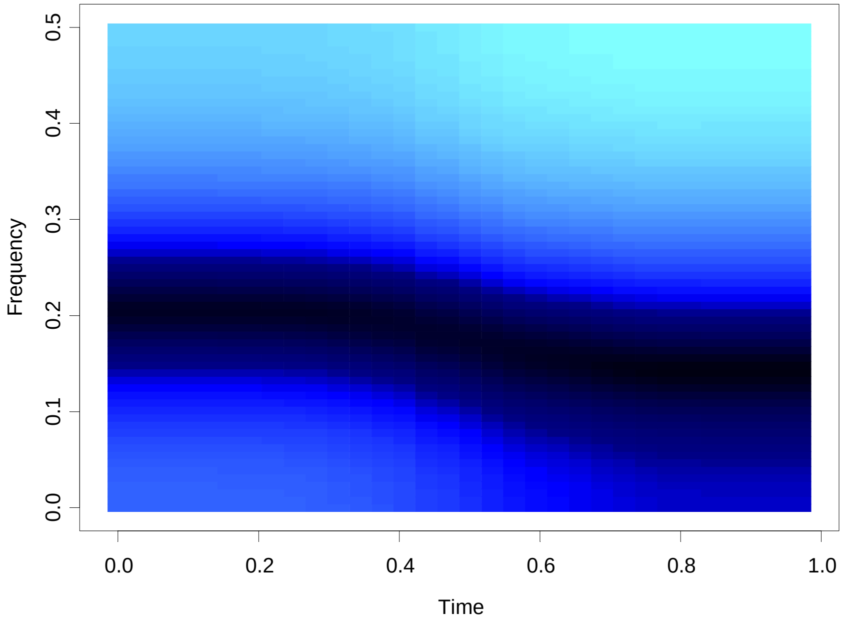
2. Slowly varying AR(2) (cont)

In the graph below right, we average the spectogram over the *GA fitted models* generated from each of the 200 simulated realizations.

True Model



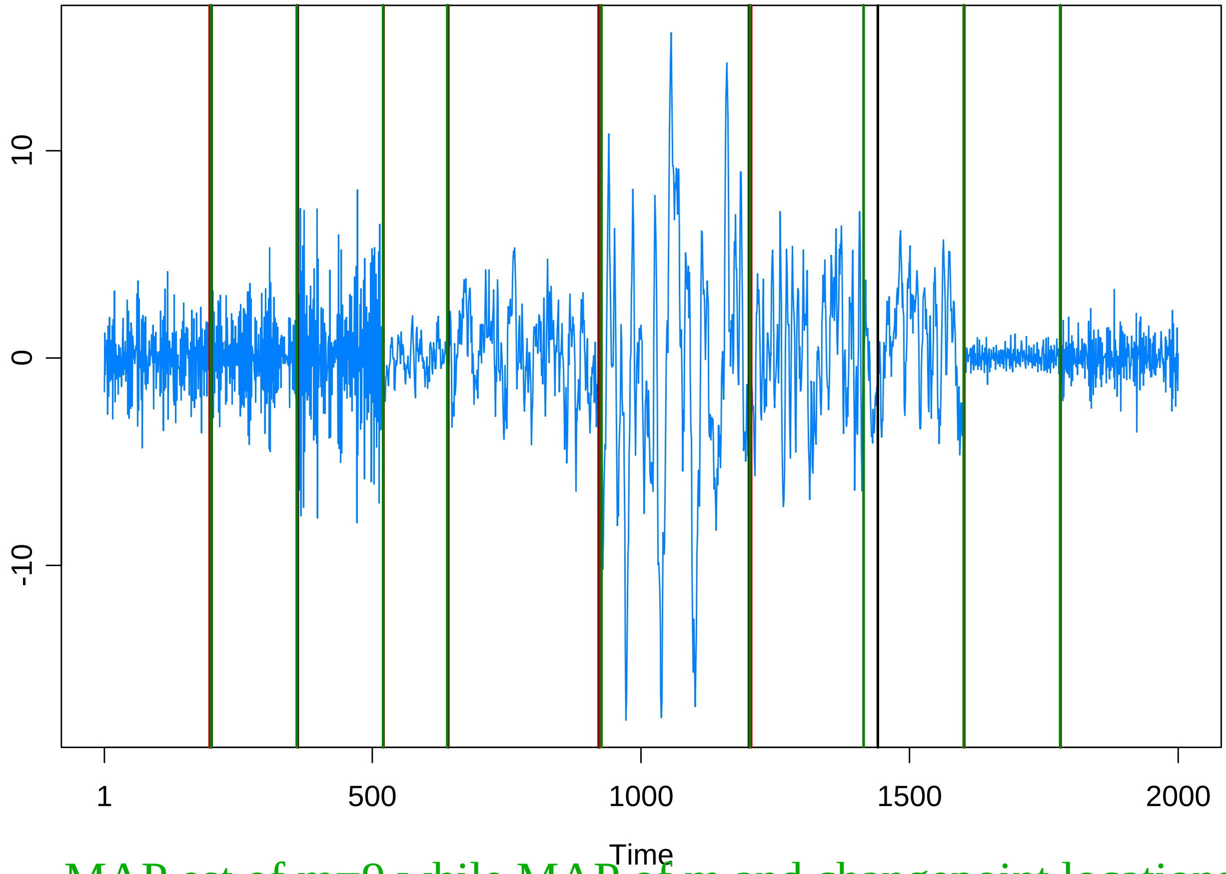
Average Model



Simulation Examples (cont)

3. Simulated data from Fearnhead (2005):

True model has 9 changepoints



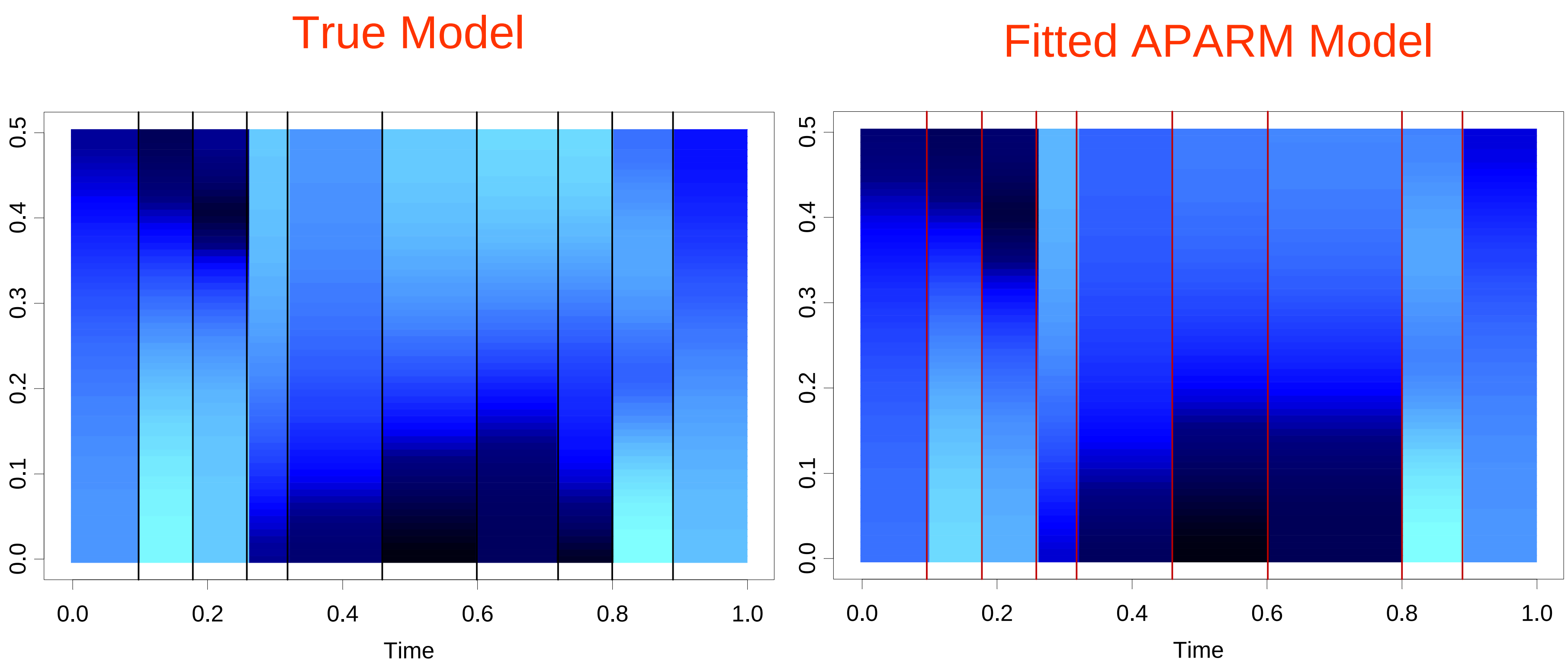
black=true

red=APARM

green=MAP

MAP est of m=9 while MAP of m and changepoint locations gives m= 8 changepts. Plot is conditional on 9 changepoints.

4. Fearnhead example



Theory

Consistency.

Suppose the number of change points m is known and let

$$\lambda_1=\tau_1/n, \; . \; . \; . \; , \; \lambda_m=\tau_m/n$$

be the relative (true) changepoints. Then

$$\hat{\lambda}_j \rightarrow \lambda_j \quad \text{a.s.}$$

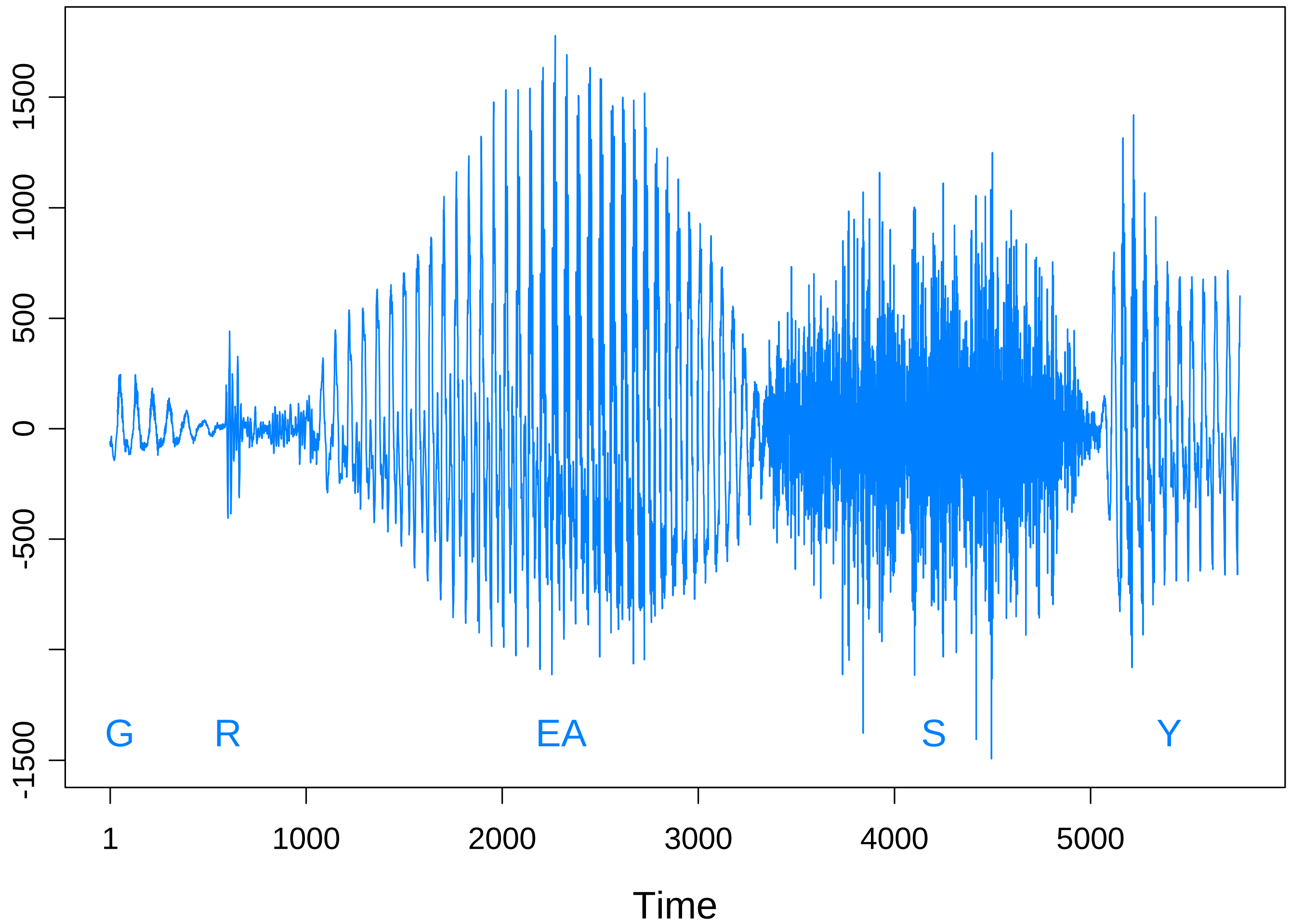
where $\hat{\lambda}_j = \hat{\tau}_j / n$ and $\hat{\tau}_j$ = Auto-PARM estimate of τ_j.

Consistency of the estimate of *m*?

- For *n* large, Auto-PARM estimate is $\geq$ *m*.
- Have not proved equality.

Examples

Speech signal: GREASY

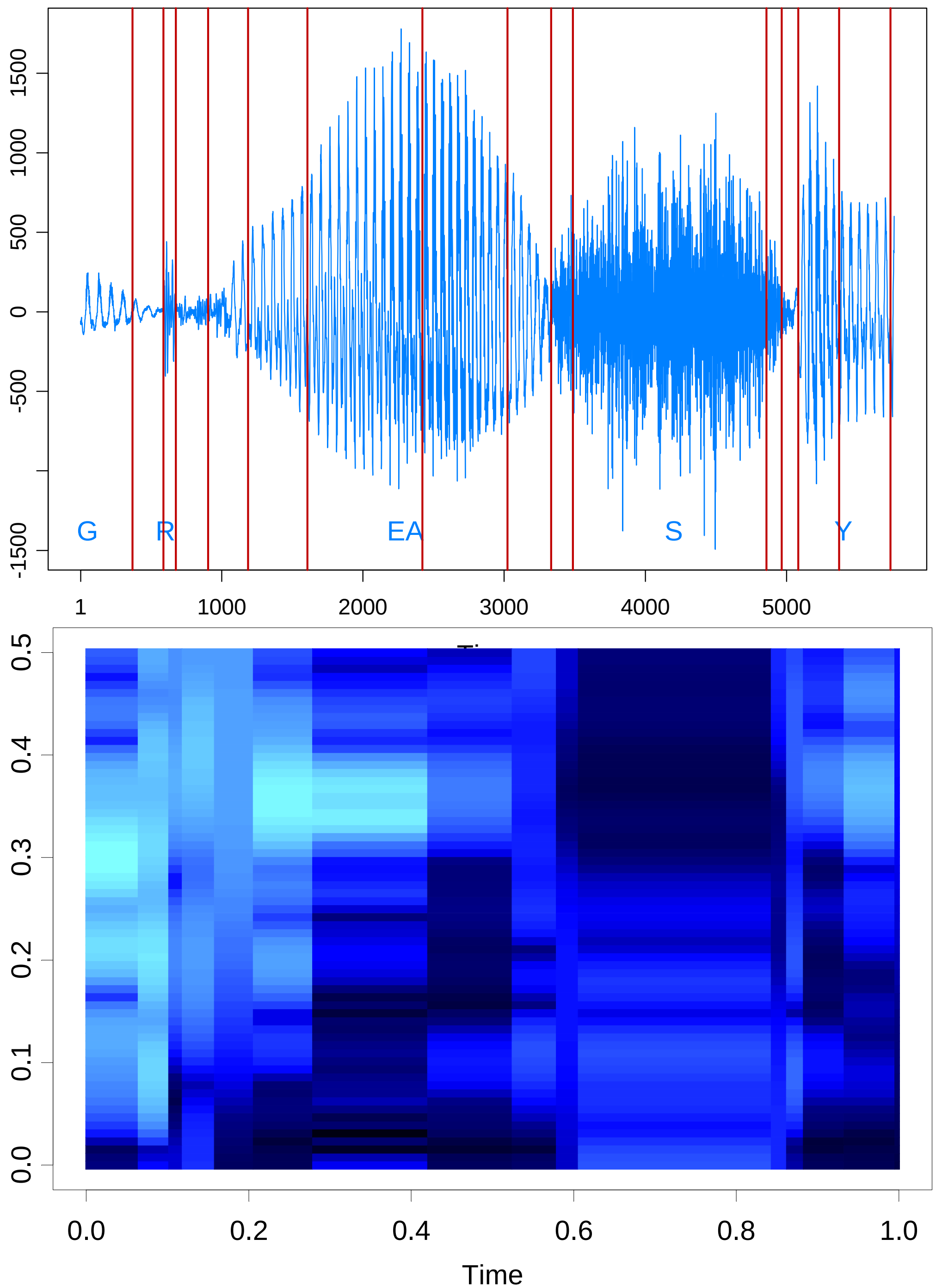


Speech signal: GREASY

n = 5762 observations

m = 15 break points

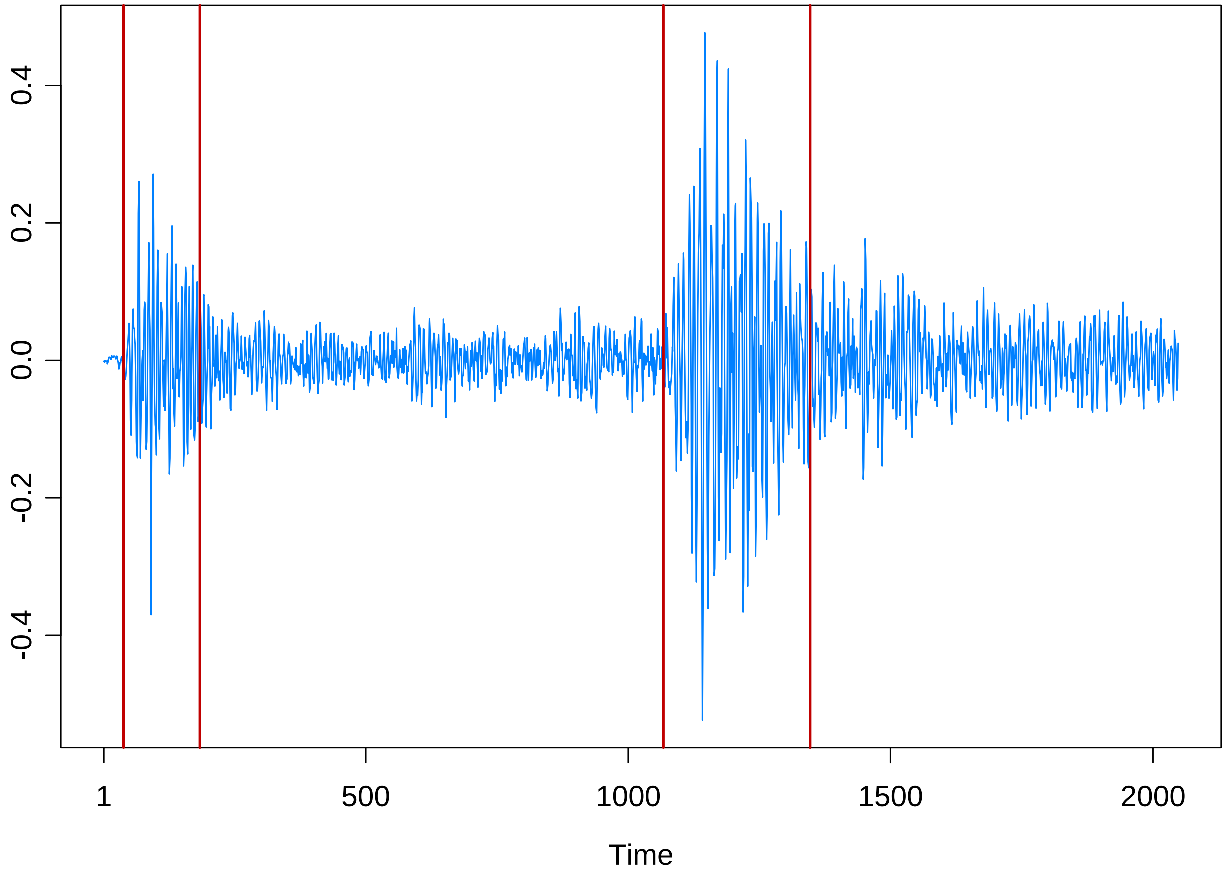
Run time = 18.02 secs



Examples

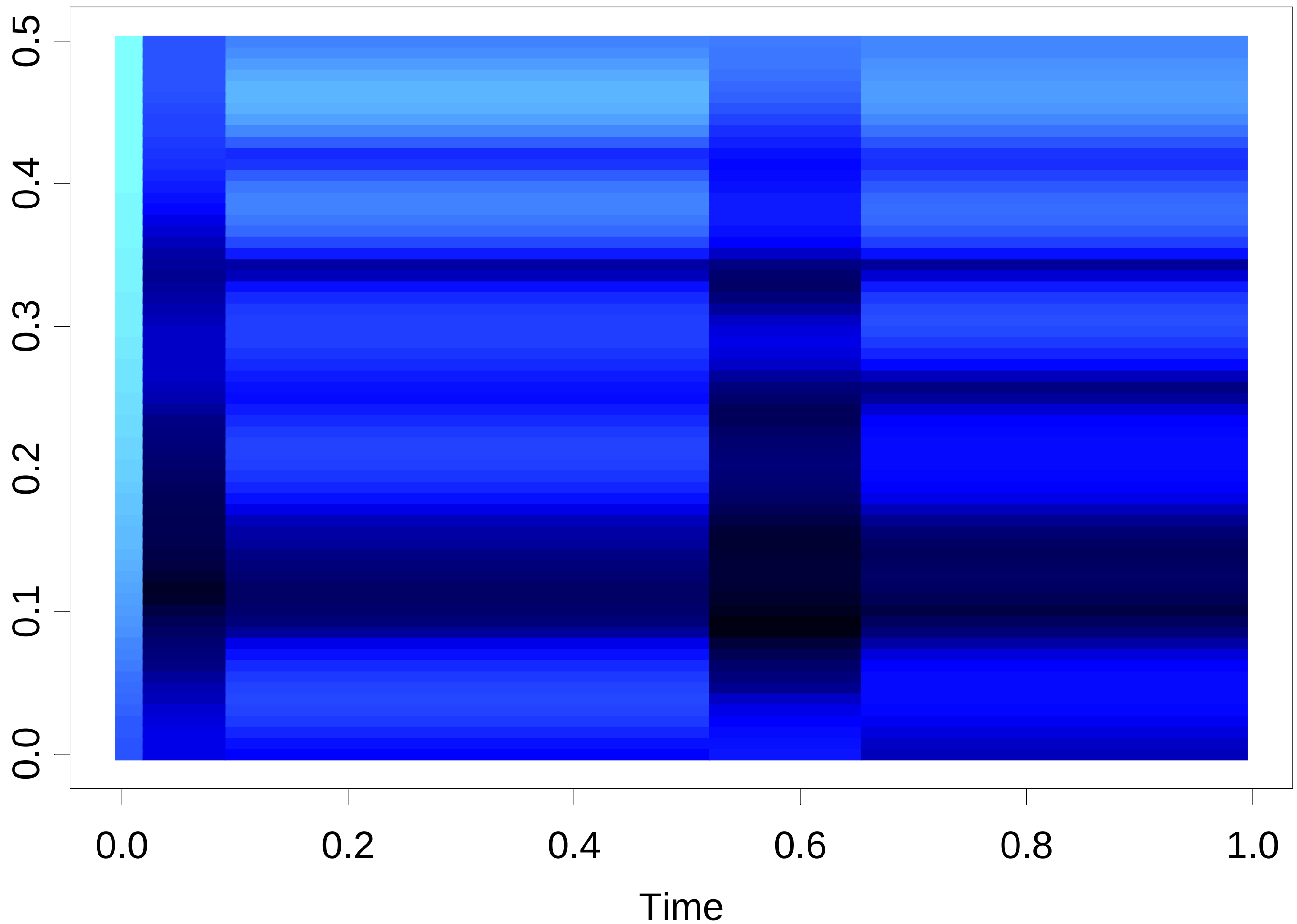
Mine explosion seismic trace in Scandinavia: (Shumway and Stoffer 2000, Stoffer et al. 2005)

Two waves: P (primary) compression wave and S (shear) wave



Examples

AR orders: 1 7 17 13 15



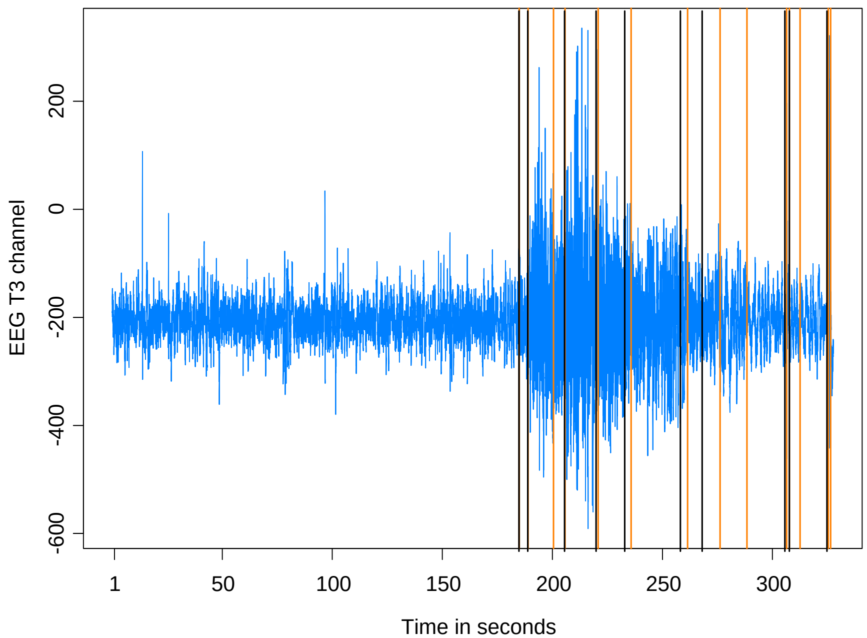
Example: EEG Time series

Data: Bivariate EEG time series at channels T3 (left temporal) and P3 (left parietal). Female subject was diagnosed with left temporal lobe epilepsy. Data collected by Dr. Beth Malow and analyzed in Ombao et al (2001). (n=32,768; sampling rate of 100H). Seizure started at about 1.85 seconds.

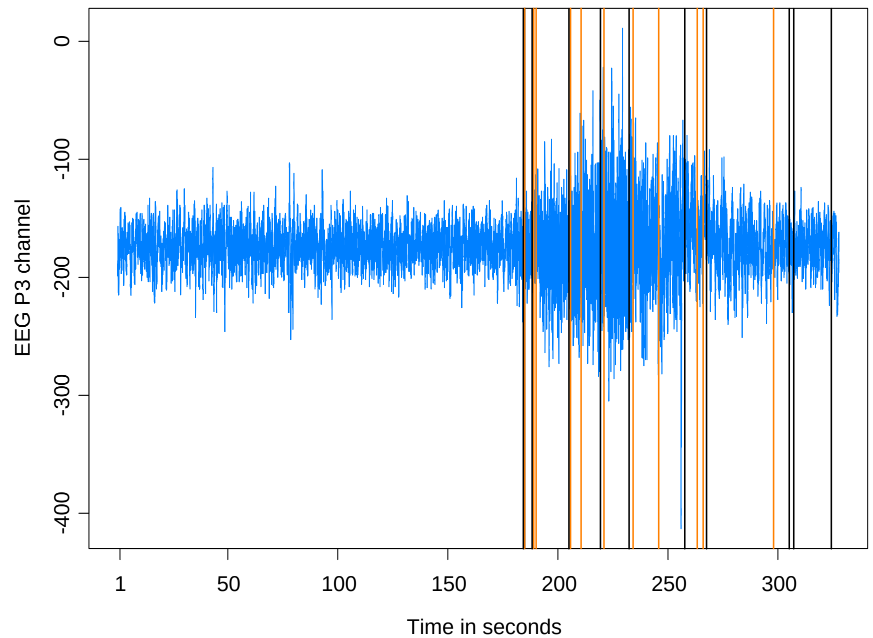
GA bivariate results: 11 pieces with AR orders 17, 2, 6, 15, 2, 3, 5, 9, 5, 4, 1

GA univariate results: 14 breakpoints for T3; 11 breakpoints for P3

T3 Channel



P3 Channel

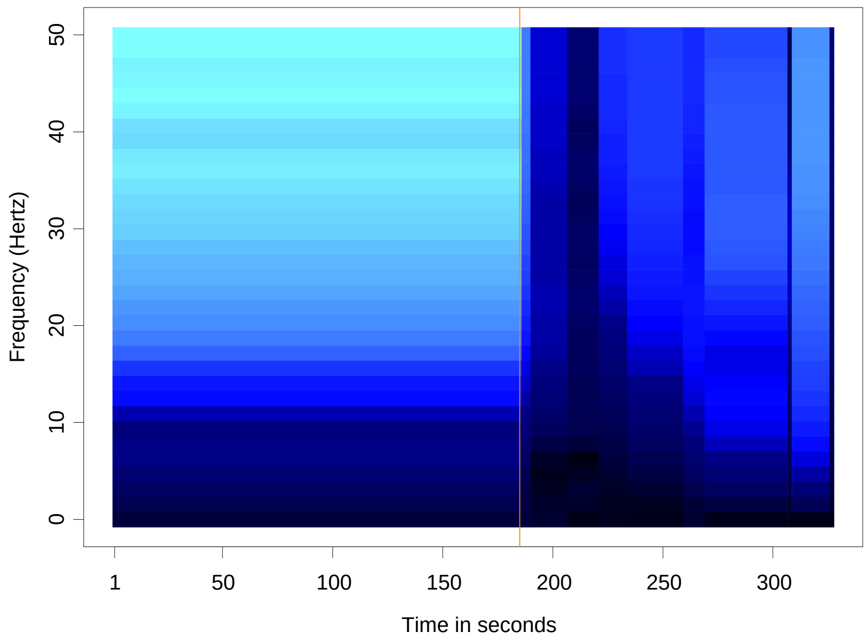


Example: EEG Time series (cont)

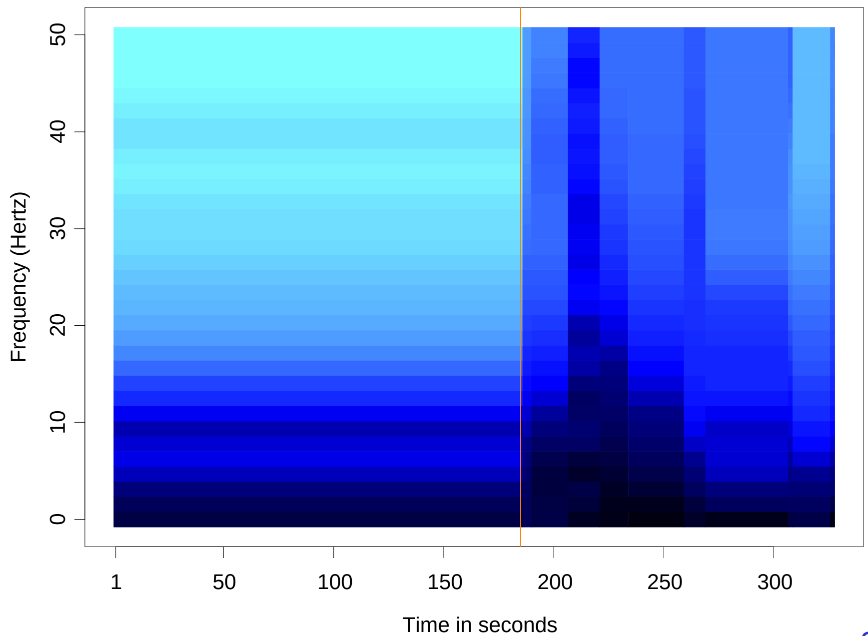
Remarks:

- the general conclusions of this analysis are similar to those reached in Ombao et al.
- prior to seizure, power concentrated at lower frequencies and then spread to high frequencies.
- power returned to the lower frequencies at conclusion of seizure.

T3 Channel



P3 Channel

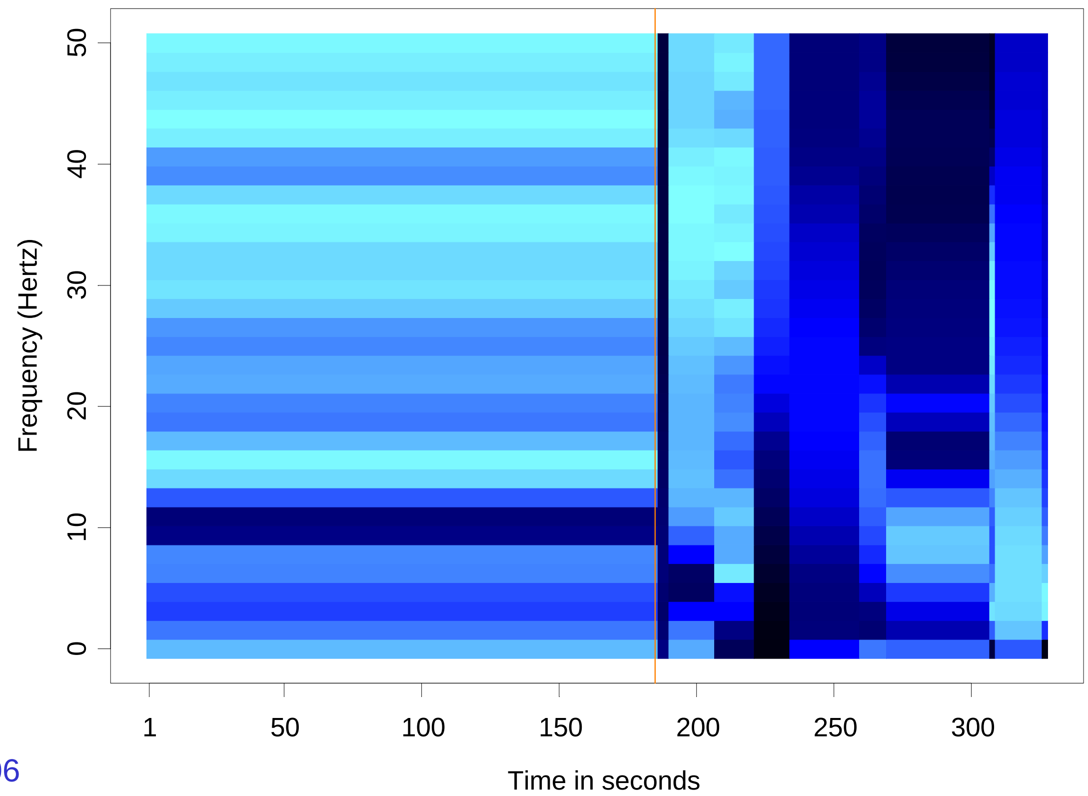


Example: EEG Time series (cont)

Remarks (cont):

- T3 and P3 strongly coherent at 9-12 Hz prior to seizure.
- strong coherence at low frequencies just after onset of seizure.
- strong coherence shifted to high frequencies during the seizure.

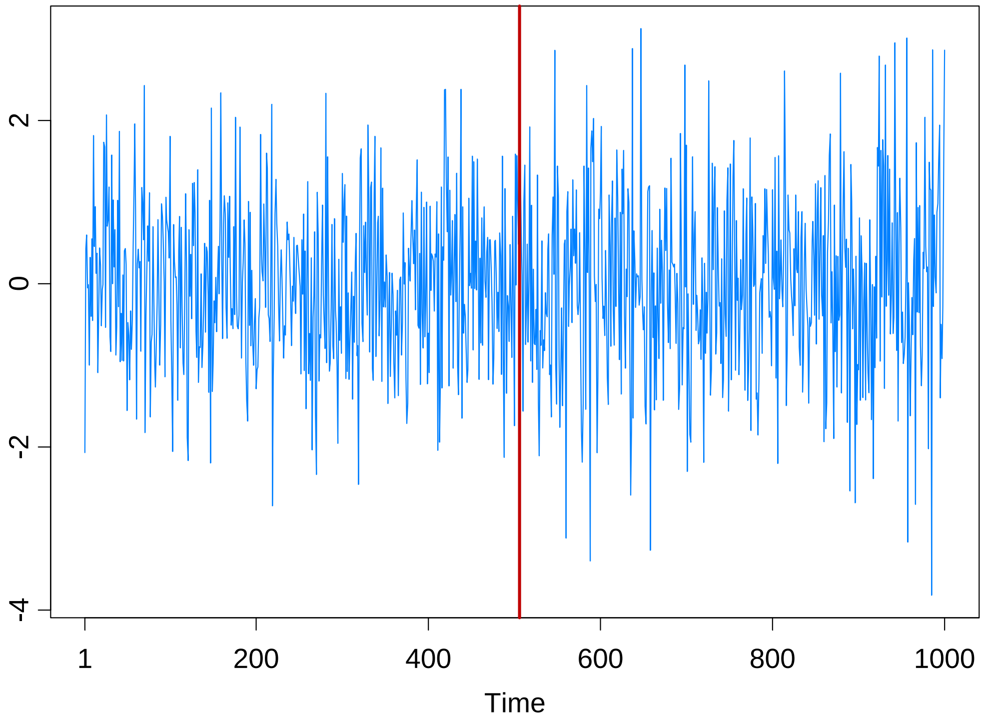
T3/P3 Coherency



Application to GARCH

Garch(1,1) model: $Y_t = \sigma_t \varepsilon_t, \qquad \{\varepsilon_t\} \sim \text{IID}(0,1)$

$$\sigma_t^2 = \omega_j + \alpha_j Y_{t-1}^2 + \beta_j \sigma_{t-1}^2, \qquad \text{if } \tau_{j\text{-}1} \le t < \tau_j.$$



$$\sigma_t^2 = \begin{cases} .4 + .1Y_{t-1}^2 + .5\sigma_{t-1}^2, & \text{if } 1 \le t < 501, \\ .4 + .1Y_{t-1}^2 + .6\sigma_{t-1}^2, & \text{if } 501 \le t < 1000 \end{cases}$$

# of CPs	GA %	AG %
0	80.4	72.0
1	19.2	24.0
≥ 2	0.4	0.4

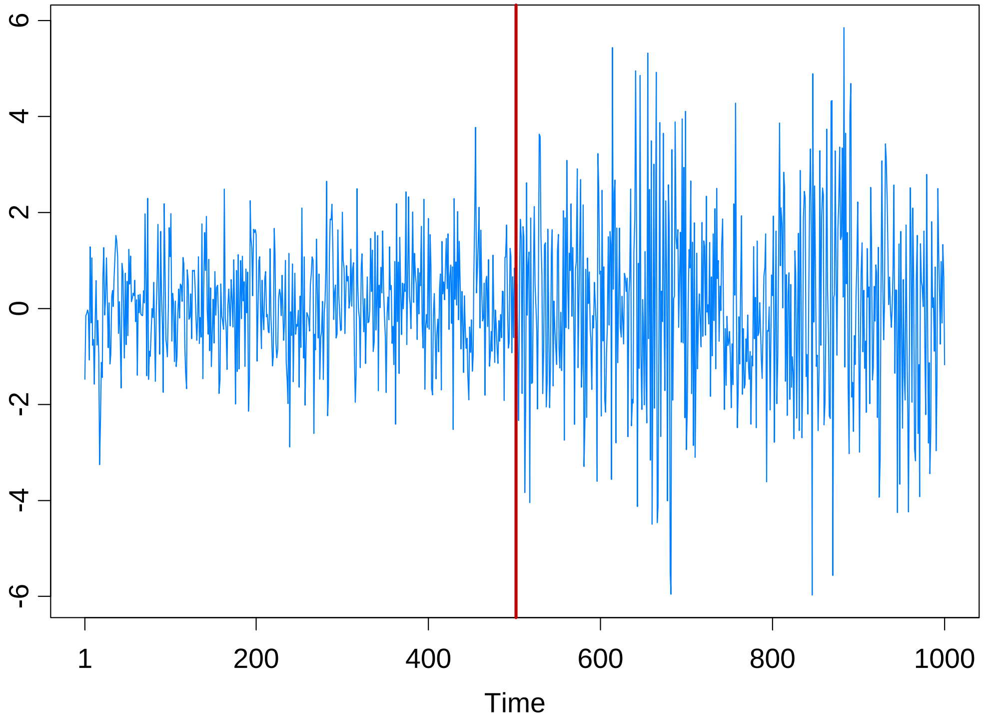
CP estimate = 506

AG = Andreou and Ghysels (2002)

Application to GARCH (cont)

Garch(1,1) model: $Y_t = \sigma_t \varepsilon_t, \qquad \{\varepsilon_t\} \sim \text{IID}(0,1)$

$$\sigma_t^2 = \omega_j + \alpha_j Y_{t-1}^2 + \beta_j \sigma_{t-1}^2, \quad \text{if } \tau_{j\text{-}1} \le t < \tau_j.$$



$$\sigma_t^2 = \begin{cases} .4 + .1Y_{t-1}^2 + .5\sigma_{t-1}^2, & \text{if } 1 \le t < 501, \\ .4 + .1Y_{t-1}^2 + .8\sigma_{t-1}^2, & \text{if } 501 \le t < 1000. \end{cases}$$

# of CPs	GA %	AG %
0	0.0	0.0
1	96.4	95.0
≥ 2	3.6	0.5

CP estimate = 502

AG = Andreou and Ghysels (2002)

Application to GARCH (cont)

More simulation results for Garch(1,1) : $Y_t = \sigma_t \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{IID}(0,1)$

$$\sigma_t^2 = \begin{cases} .05 + .4Y_{t-1}^2 + .3\sigma_{t-1}^2, & \text{if } 1 \le t < \tau_1, \\ 1.00 + .3Y_{t-1}^2 + .2\sigma_{t-1}^2, & \text{if } \tau_1 \le t < 1000. \end{cases}$$

τ_1		Mean	SE	Med	Freq
50	GA	**52.62**	11.70	**50**	.98
	Berkes	**71.40**	12.40	**71**	
250	GA	**251.18**	4.50	**250**	.99
	Berkes	**272.30**	18.10	**271**	
500	GA	**501.22**	4.76	**502**	.98
	Berkes	**516.40**	54.70	**538**	

Berkes = Berkes, Gombay, Horvath, and Kokoszka (2004).

Application to Parameter-Driven SS Models

State Space Model Setup:

Observation equation:

$$p(y_t \mid \alpha_t) = \exp\{\alpha_t y_t - b(\alpha_t) + c(y_t)\}.$$

State equation: $\{\alpha_t\}$ follows the piecewise AR(1) model given by

$$\alpha_t = \gamma_k + \phi_k \alpha_{t-1} + \sigma_k \varepsilon_t, \quad \text{if} \quad \tau_{k-1} \le t < \tau_k,$$

where $1 = \tau_0 < \tau_1 < \ldots < \tau_m < n$, and $\{\varepsilon_t\}$ ~ IID N(0,1).

Parameters:

m = number of break points

τ_k = location of break points

γ_k = level in k^{th} epoch

ϕ_k = AR coefficients k^{th} epoch

σ_k = scale in k^{th} epoch

Application to Structural Breaks—(cont)

Estimation: For (m, τ_1, . . . , τ_m) fixed, calculate the approximate likelihood evaluated at the "MLE", i.e.,

$$L_a(\hat{\psi}; \mathrm{y}_\mathrm{n}) = \frac{|G_n|^{1/2}}{(K+G_n)^{1/2}} \exp\{\mathrm{y}_n^T \alpha^* - 1^T\{b(\alpha^*) - c(\mathrm{y}_n)\} - (\alpha^* - \mu)^T G_n (\alpha^* - \mu)/2\},$$

where $\hat{\psi} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_m, \hat{\phi}_1, \ldots, \hat{\phi}_m, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2)$ is the MLE.

Remark: The exact likelihood is given by the following formula

$$L(\psi; \mathrm{y}_n) = L_a(\psi; \mathrm{y}_n) Er_a(\psi),$$

where

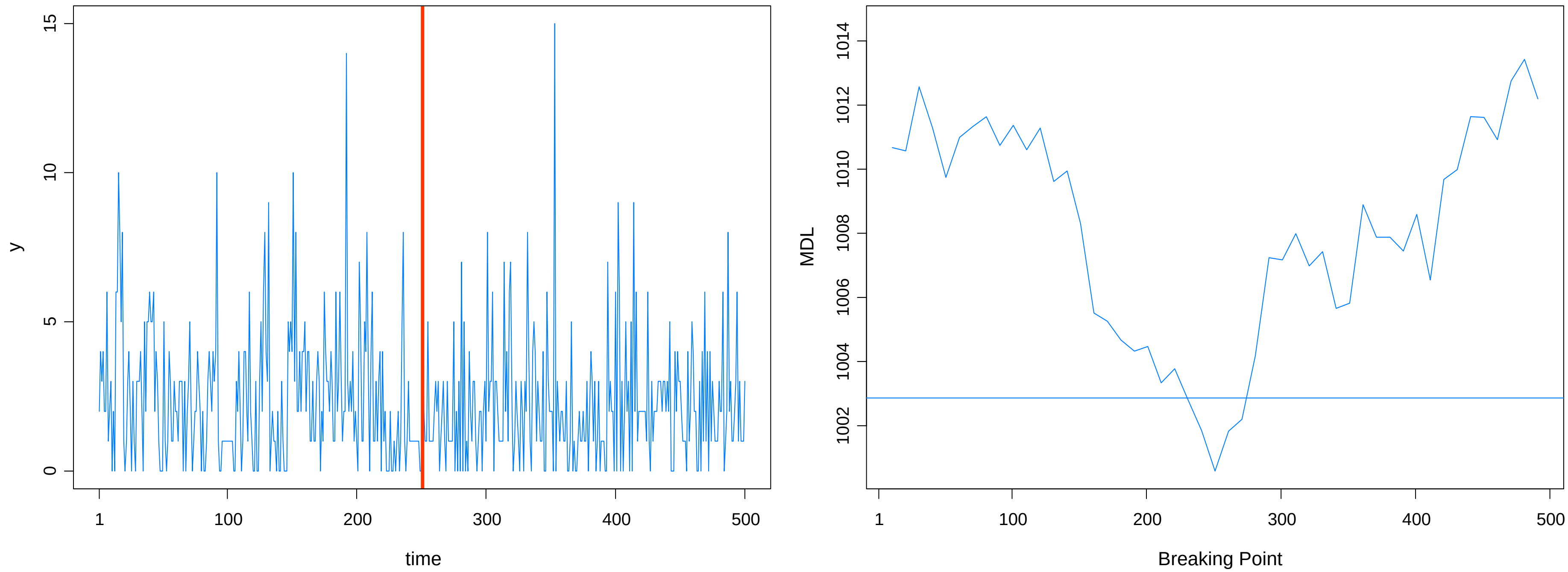
$$Er_a(\psi) = \int \exp\{R(\alpha_n; \alpha^*)\} p_a(\alpha_n \,|\, \mathrm{y}_n; \psi)\, \mathrm{d}\alpha_n.$$

It turns out that $\log(Er_a(\psi))$ is nearly linear and can be approximated by a linear function via importance sampling,

$$e(\psi) \sim e(\hat{\psi}_{AL}) + \dot{e}(\hat{\psi}_{AL})(\psi - \hat{\psi}_{AL})$$

Count Data Example

Model: $Y_t \mid \alpha_t \sim Pois(\exp\{\beta + \alpha_t\}),\ \ \alpha_t = \phi\alpha_{t-1} + \varepsilon_t,\ \ \{\varepsilon_t\} \sim \text{IID } N(0, \sigma^2)$

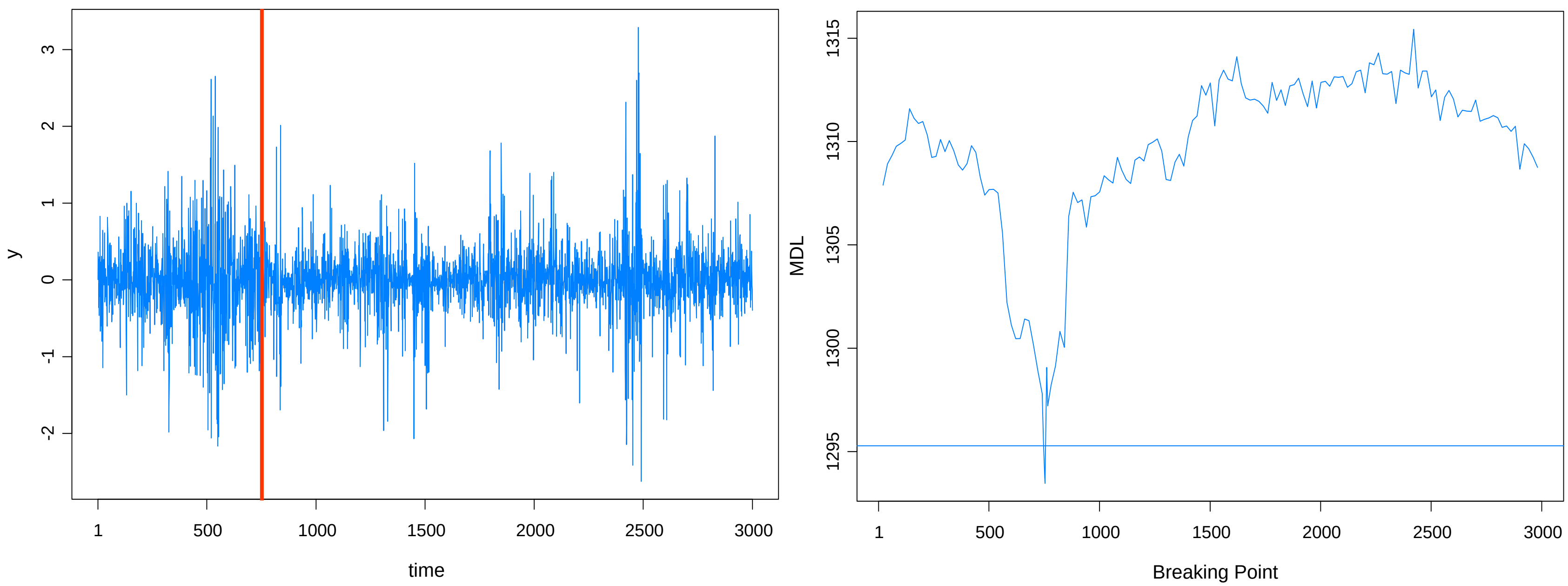


True model:

- $Y_t \mid \alpha_t \sim Pois(\exp\{.7 + \alpha_t\}),\ \alpha_t = .5\alpha_{t-1} + \varepsilon_t,\ \ \{\varepsilon_t\} \sim \text{IID } N(0, .3),\ \ t < 250$
- $Y_t \mid \alpha_t \sim Pois(\exp\{.7 + \alpha_t\}),\ \ \alpha_t = -.5\alpha_{t-1} + \varepsilon_t,\ \ \{\varepsilon_t\} \sim \text{IID } N(0, .3),\ \ t > 250.$
- GA estimate 251, time 267secs

SV Process Example

Model: $Y_t \mid \alpha_t \sim N(0, \exp\{\alpha_t\})$, $\alpha_t = \gamma + \phi\, \alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim$ IID $N(0, \sigma^2)$

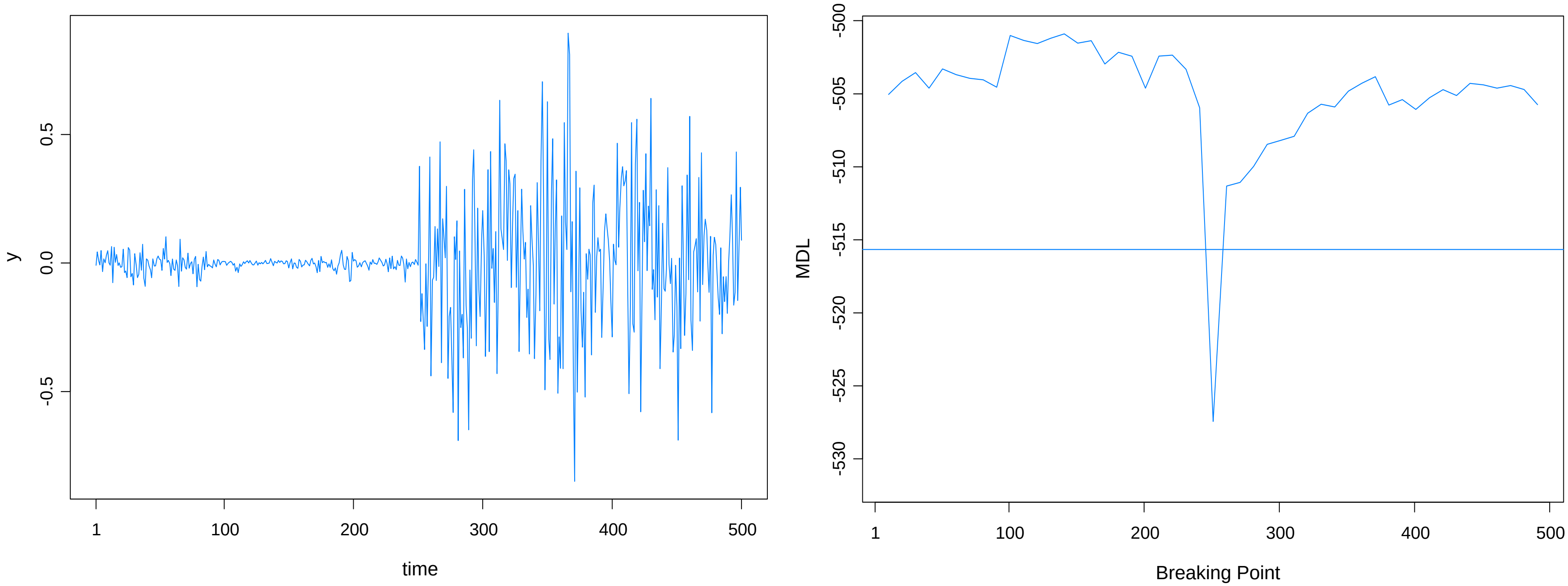


True model:

- $Y_t \mid \alpha_t \sim N(0, \exp\{\alpha_t\})$, $\alpha_t = -.05 + .975\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim$ IID $N(0, .05)$, $t \le 750$
- $Y_t \mid \alpha_t \sim N(0, \exp\{\alpha_t\})$, $\alpha_t = -.25 + .900\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim$ IID $N(0, .25)$, $t > 750$.
- GA estimate 754, time 1053 secs

SV Process Example

Model: $Y_t \mid \alpha_t \sim N(0,\exp\{\alpha_t\})$, $\alpha_t = \gamma + \phi\,\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim \text{IID } N(0, \sigma^2)$



True model:

- $Y_t \mid \alpha_t \sim N(0, \exp\{\alpha_t\})$, $\alpha_t = -.175 + .977\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim \text{IID N}(0, .1810)$, $t \le 250$
- $Y_t \mid \alpha_t \sim N(0, \exp\{\alpha_t\})$, $\alpha_t = -.010 + .996\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim \text{IID N}(0, .0089)$, $t > 250$.
- GA estimate 251, time 269s

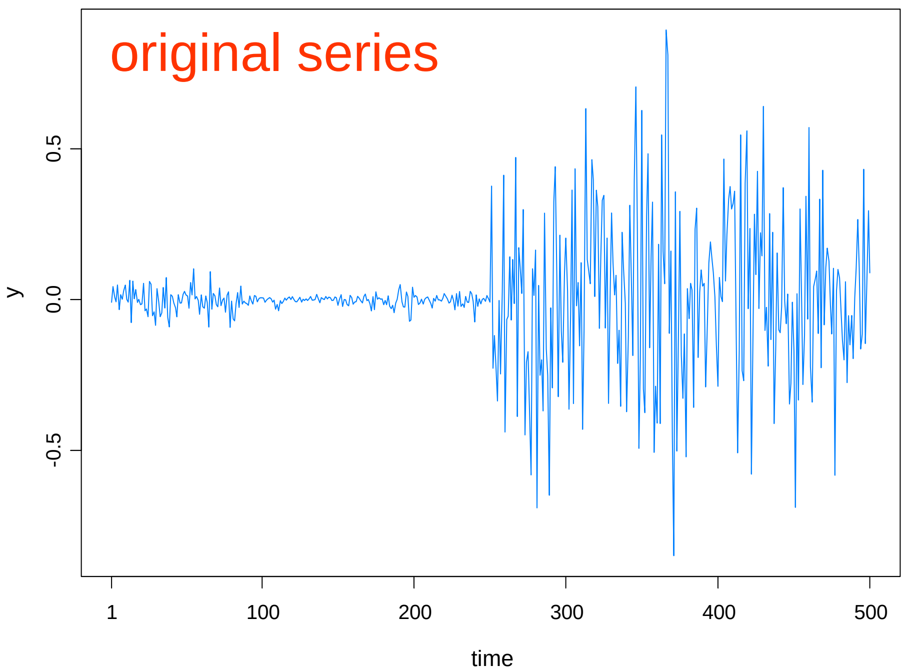
SV Process Example-(cont)

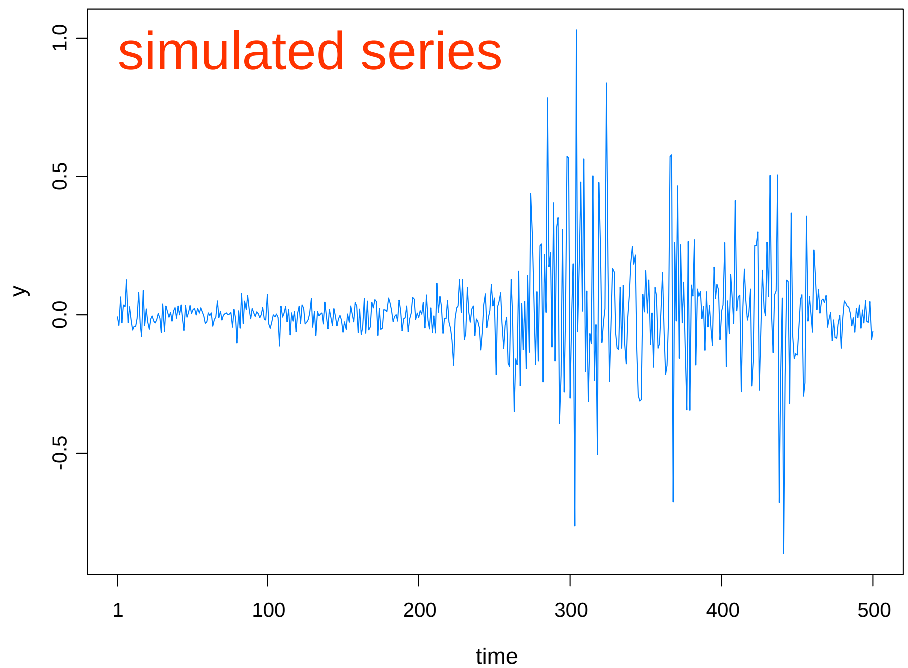
True model:

- $Y_t \mid \alpha_t \sim N(0, \exp\{a_t\})$, $\alpha_t = -.175 + .977\alpha_{t-1} + e_t$, $\{\varepsilon_t\} \sim \text{IID } N(0, .1810)$, $t \le 250$
- $Y_t \mid \alpha_t \sim N(0, \exp\{\alpha_t\})$, $\alpha_t = -.010 + .996\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim \text{IID } N(0, .0089)$, $t > 250$.

Fitted model based on no structural break:

- $Y_t \mid \alpha_t \sim N(0, \exp\{\alpha_t\})$, $\alpha_t = -.0645 + .9889\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim \text{IID } N(0, .0935)$





SV Process Example-(cont)

Fitted model based on no structural break:

- $Y_t \mid \alpha_t \sim N(0, \exp\{\alpha_t\})$, $\alpha_t = -.0645 + .9889\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\}$~IID N(0, .0935)