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.

- Introduction
  - Setup
  - Examples
    - AR
    - GARCH
    - Stochastic volatility
    - State space model
- Model Selection Using Minimum Description Length (MDL)
  - General principles
  - Application to AR models with breaks
- Optimization using Genetic Algorithms
  - Basics
  - Implementation for structural break estimation
- Simulation results
- Applications
- Simulation results for GARCH and SSM

Introduction

The Premise (in a general framework):

Base model: P_θ family or probability models for a stationary time series.

Observations: $y_1, \ldots, y_n$

Segmented model: there exist an integer $m \geq 0$ and locations

$$\tau_0 = 1 < \tau_1 < \ldots < \tau_{m-1} < \tau_m = n + 1$$

such that

$$Y_t = X_{t,j}, \quad \text{if } \tau_{j\text{-}1} \leq t < \tau_j,$$

where $\{X_{t,j}\}$ is a stationary time series with probability distr P_{θ_j} and $\theta_j \neq \theta_{j+1}$.

Objective: estimate

m = number of segments
τ_j = location of j^{th} break point
θ_j = parameter vector in j^{th} epoch

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

Piecewise AR models (cont)

Structural breaks:

Kitagawa and Akaike (1978)

- fitting locally stationary autoregressive models using AIC
- computations facilitated by the use of the Householder transformation

Davis, Huang, and Yao (1995)

- likelihood ratio test for testing a change in the parameters and/or order of an AR process.

Kitagawa, Takanami, and Matsumoto (2001)

- signal extraction in seismology-estimate the arrival time of a seismic signal.

Ombao, Raz, von Sachs, and Malow (2001)

- orthogonal complex-valued transforms that are localized in time and frequency- smooth localized complex exponential (SLEX) transform.
- applications to EEG time series and speech data.

Motivation for using piecewise AR models:

Piecewise AR is a special case of a *piecewise stationary process* (see Adak 1998),

$$\tilde{Y}_{t,n} = \sum_{j=1}^{m} Y_t^j I_{[\tau_{j-1},\tau_j)}(t/n),$$

where $\{Y_t^j\}$, $j = 1, \ldots, m$ is a sequence of stationary processes. It is argued in Ombao et al. (2001), that if $\{Y_{t,n}\}$ is a locally stationary process (in the sense of Dahlhaus), then there exists a piecewise stationary process $\{\tilde{Y}_{t,n}\}$ with

$$m_n \to \infty \quad \text{with } m_n / n \to 0, \text{ as } n \to \infty,$$

that approximates $\{Y_{t,n}\}$ (in average mean square).

Roughly speaking: $\{Y_{t,n}\}$ is a locally stationary process if it has a time-varying spectrum that is approximately $|A(t/n,\omega)|^2$, where $A(u,\omega)$ is a continuous function in u.

Examples (cont)

2. Segmented GARCH model:

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

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:

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

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

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

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 $y = (y_1, \ldots, y_n)$

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

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

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

where

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

$L(\hat{e}|\hat{F})$ = 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 $L(\hat{\boldsymbol{F}}|y)$: Let $n_j = \tau_j - \tau_{j\text{-}1}$; $\psi_j = (\gamma_j, \phi_{j1}, \ldots, \phi_{jp_j}, \sigma_j)$ denote the length of the jth segment and the parameter vector of the jth AR process, respectively. Then

$$\begin{aligned} L(\hat{\boldsymbol{F}}|y) &= L(m) + L(\tau_1, \ldots, \tau_m) + L(p_1, \ldots, p_m) + L(\hat{\psi}_1 \mid y) + \cdots + L(\hat{\psi}_m \mid y) \\ &= L(m) + L(n_1, \ldots, n_m) + L(p_1, \ldots, p_m) + L(\hat{\psi}_1 \mid y) + \cdots + L(\hat{\psi}_m \mid y) \end{aligned}$$

Encoding:

integer I : $\log_2 I$ bits (if I unbounded)

$\log_2 I_U$ bits (if I bounded by I_U)

MLE $\hat{\theta}$: ½ $\log_2 N$ bits (where N = number of observations used to compute $\hat{\theta}$; Rissanen (1989))

So,
$$L(\hat{\boldsymbol{F}}|y) = \log_2 m + m\log_2 n + \sum_{j=1}^{m} \log_2 p_j + \sum_{j=1}^{m} \frac{p_j + 2}{2} \log_2 n_j$$

Second term $L(\hat{e}\,|\,\hat{\boldsymbol{F}})$: Using Shannon's classical results on information theory, Rissanen demonstrates that the code length of $\hat{e}$ can be approximated by the negative of the log-likelihood of the fitted model, i.e., by

$$L(\hat{e}\,|\,\hat{\boldsymbol{F}}) \approx \sum_{j=1}^{m} \frac{n_j}{2} (\log_2(2\pi\hat{\sigma}_j^2) + 1)$$

For fixed values of m, $(\tau_1,p_1),\ldots,(\tau_m,p_m)$, we define the MDL as

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

The strategy is to find the best segmentation that minimizes $MDL(m,\tau_1,p_1,\ldots,\tau_m,p_m)$. To speed things up, we use Y-W estimates of AR parameters.

Optimization Using Genetic Algorithms

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.

Application to Structural Breaks—(cont)

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

$$(m, (\tau_1, p_1) \ldots, (\tau_m, p_m)) \longleftrightarrow c = (\delta_1, \ldots, \delta_{\mathrm{n}}),$$

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:

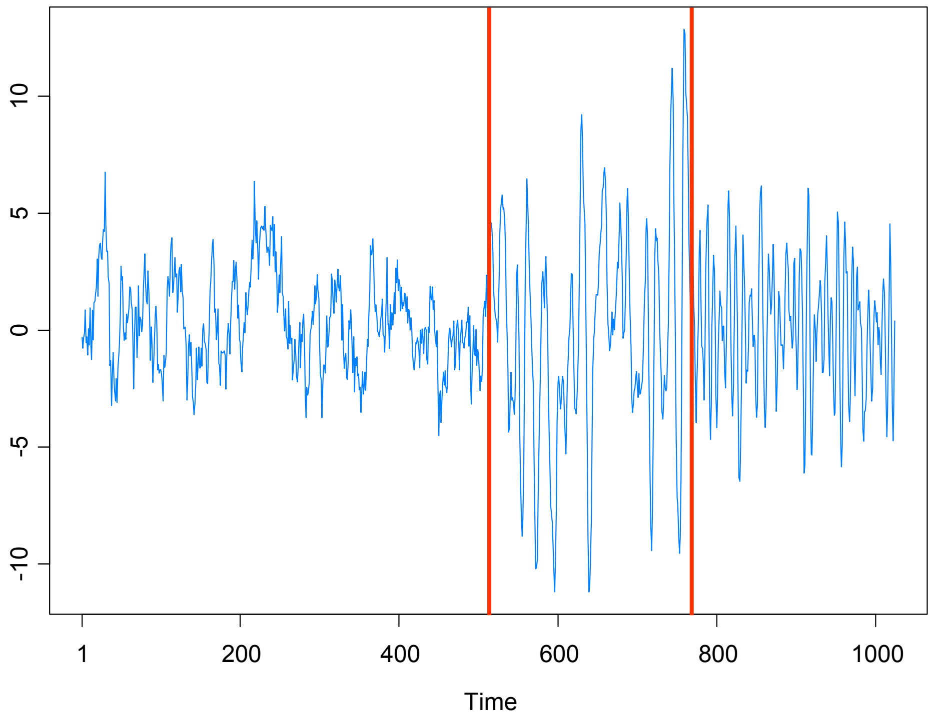
AR(2), t=1:5; AR(0), t=6:10; AR(0), t=11:14; AR(3), t=15:20

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).



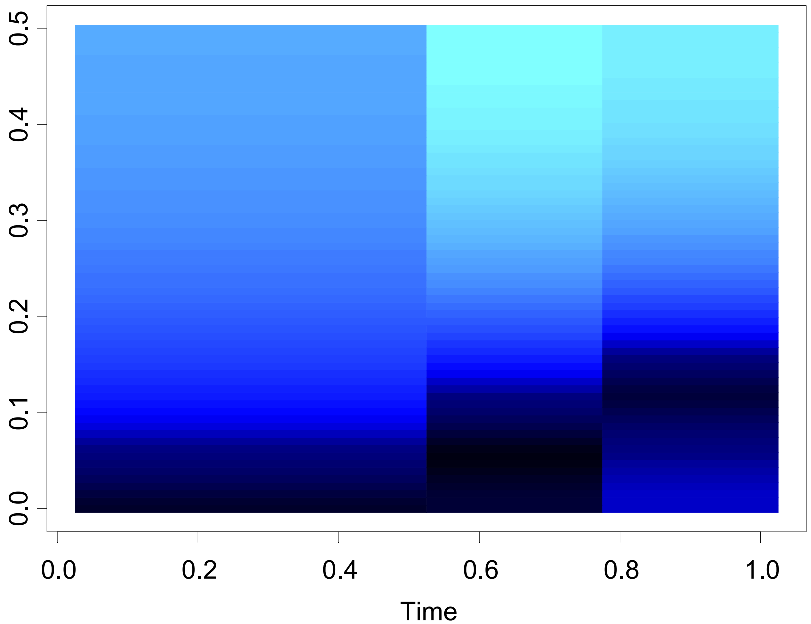
1. Piecewise stat (cont)

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

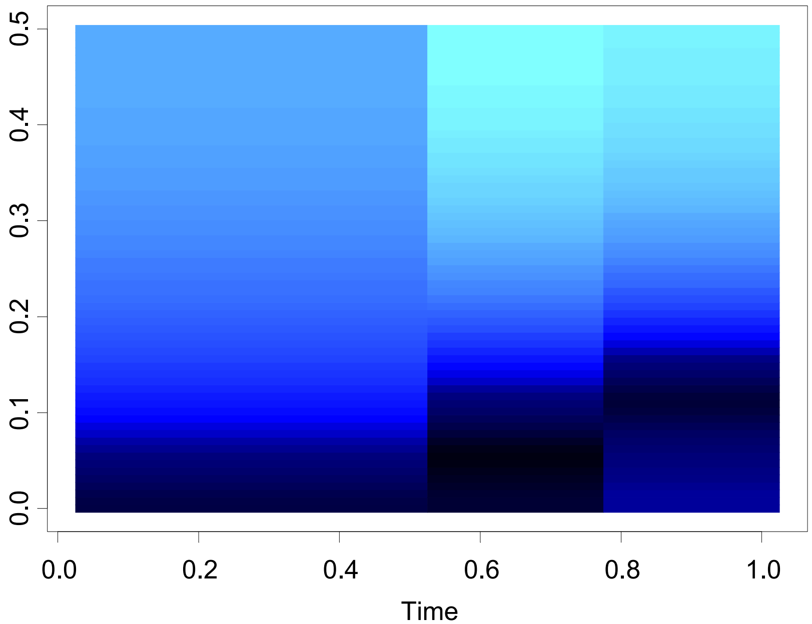
Fitted model:

	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

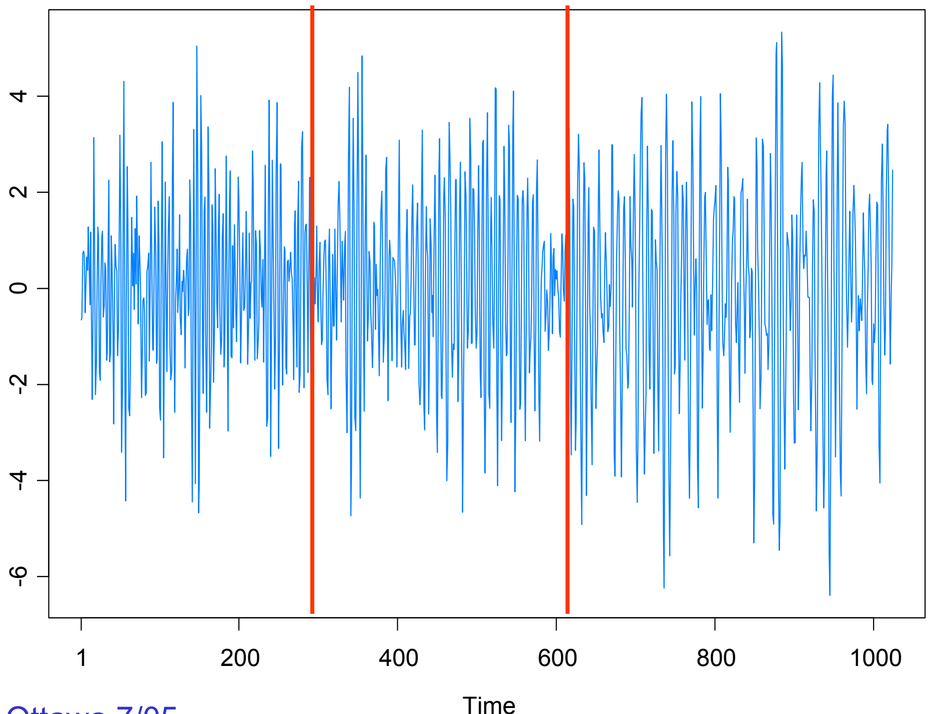


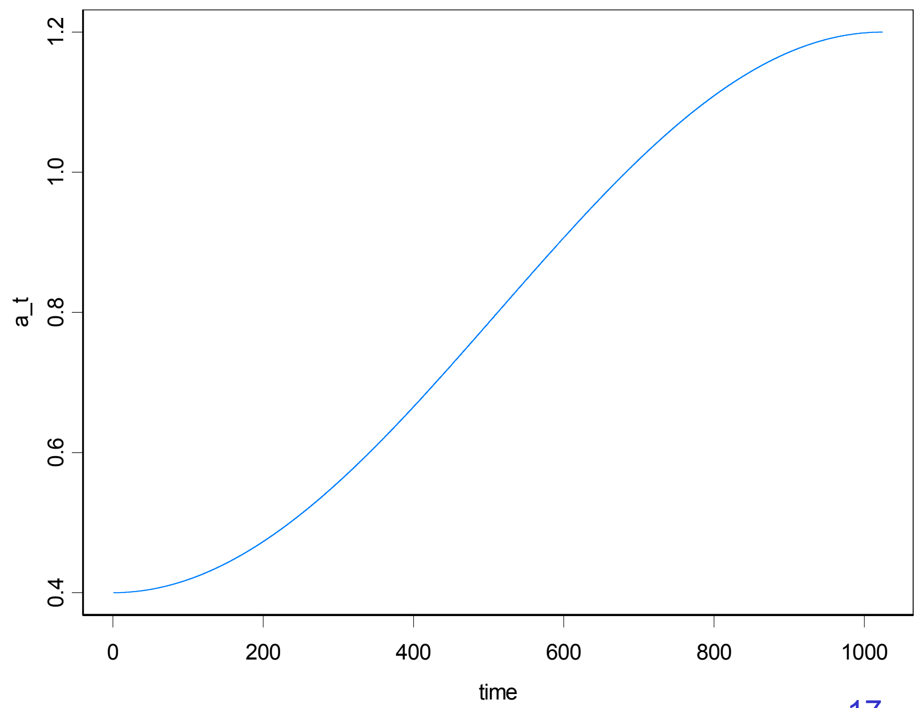
Simulation Examples (cont)

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).





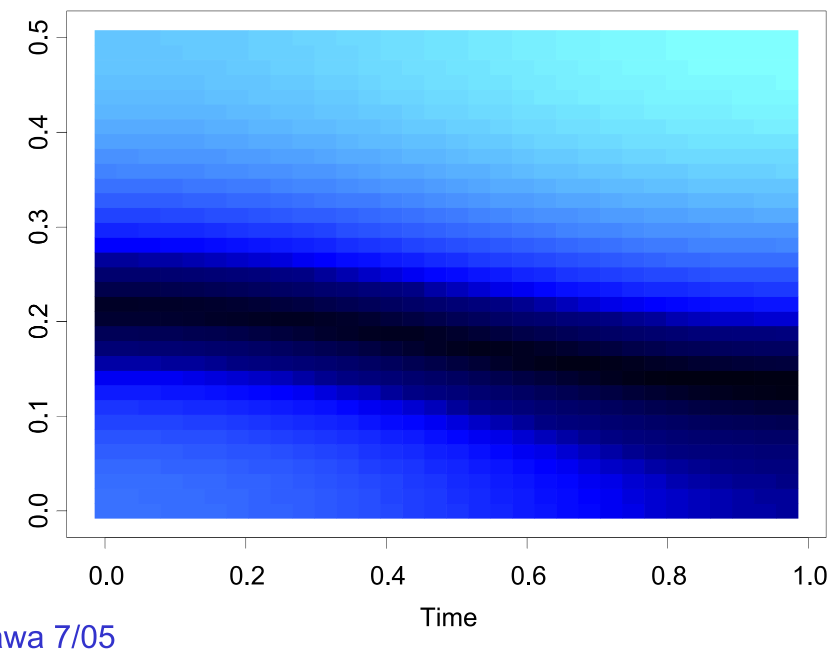
2. Slowly varying AR(2) (cont)

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

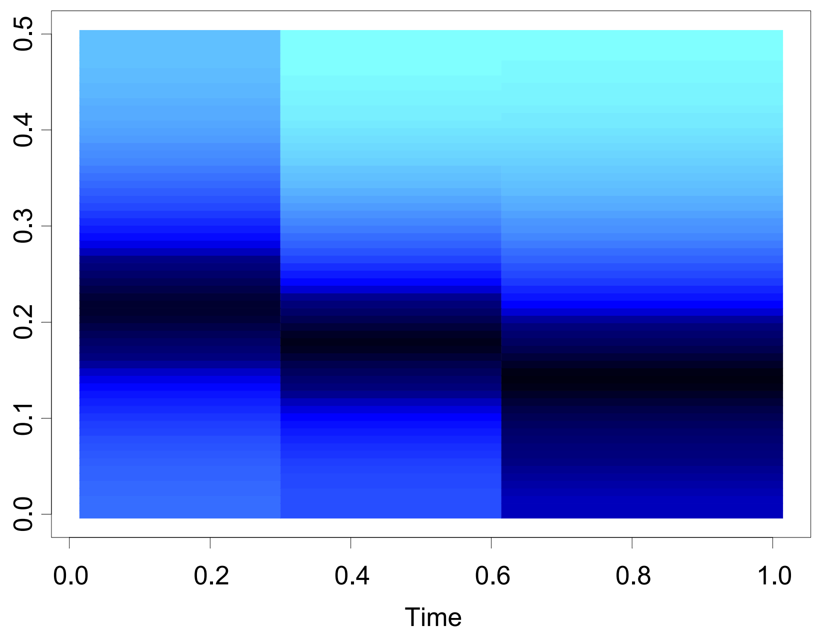
Fitted model:

	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

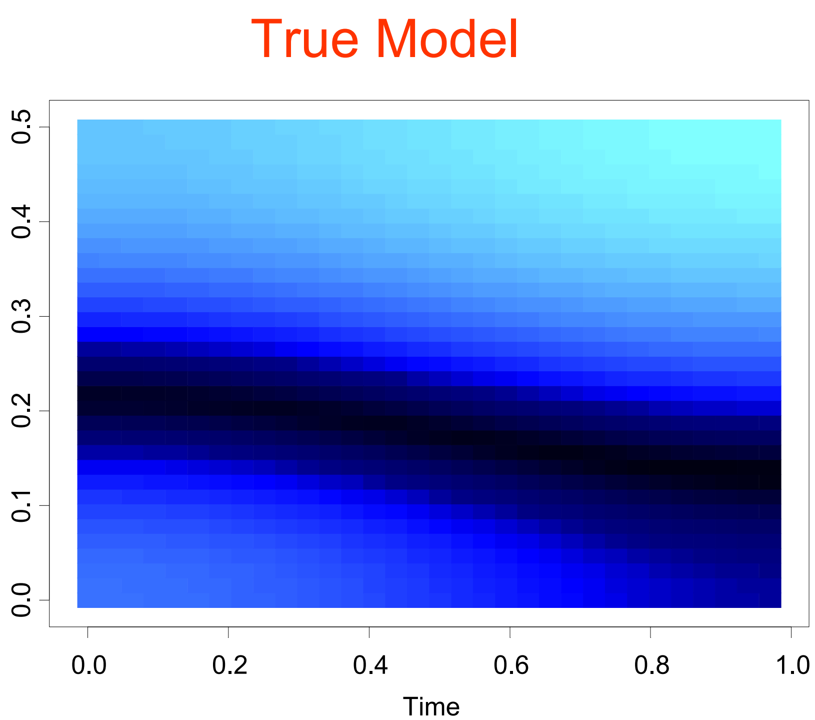
True Model

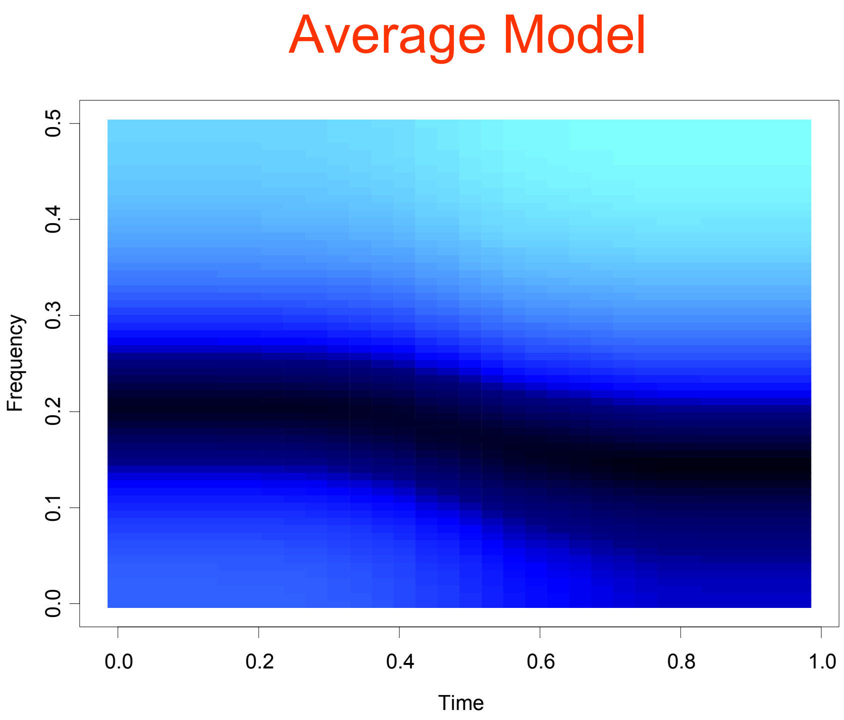


Fitted Model



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

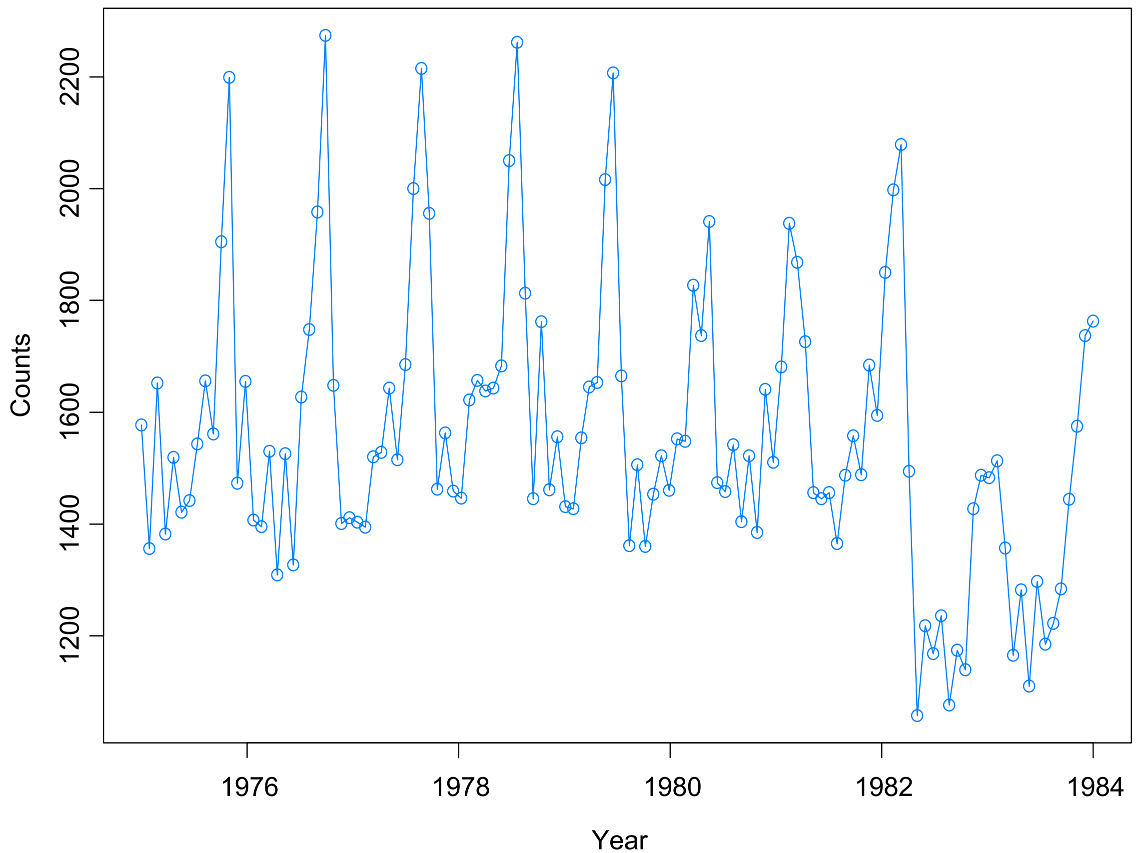




Example: Monthly Deaths & Serious Injuries, UK

Data: Y_t = number of monthly deaths and serious injuries in UK, Jan \`75 – Dec \`84, (t = 1,…, 120)

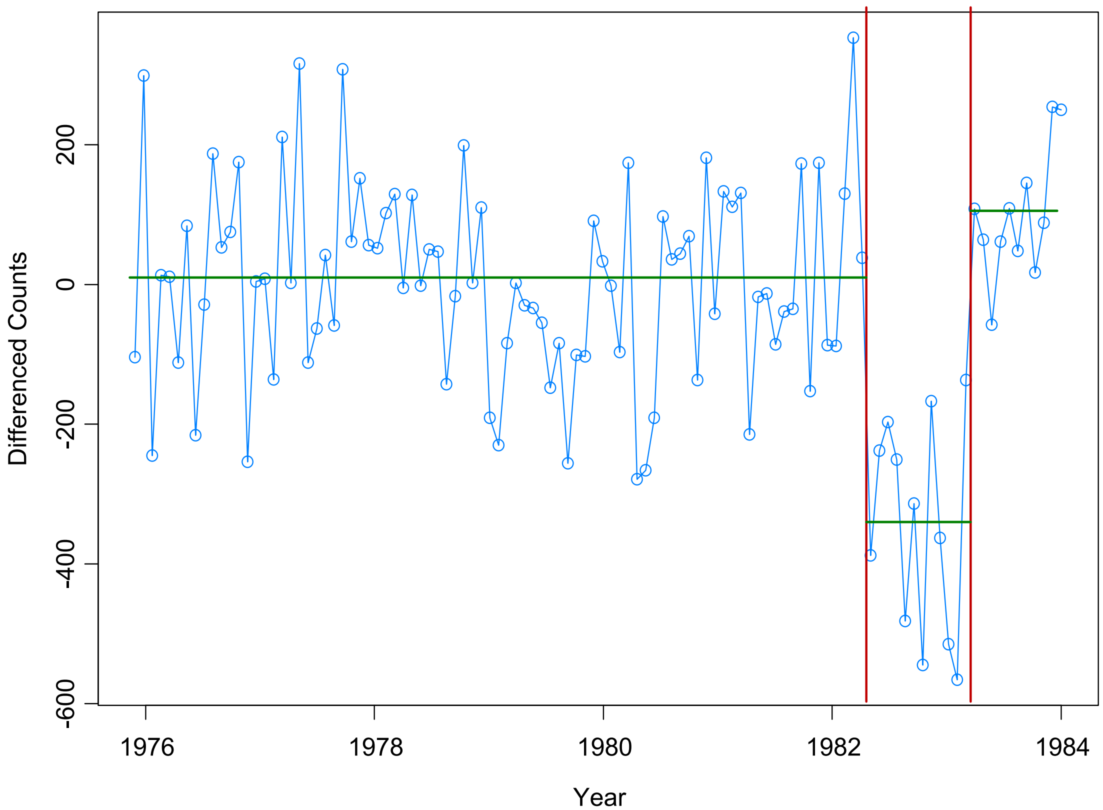
Remark: Seat belt legislation introduced in Feb \`83 (t = 99).



Example: Monthly Deaths & Serious Injuries, UK

Data: Y_t = number of monthly deaths and serious injuries in UK, Jan \`75 – Dec \`84, (t = 1,…, 120)

Remark: Seat belt legislation introduced in Feb \`83 (t = 99).

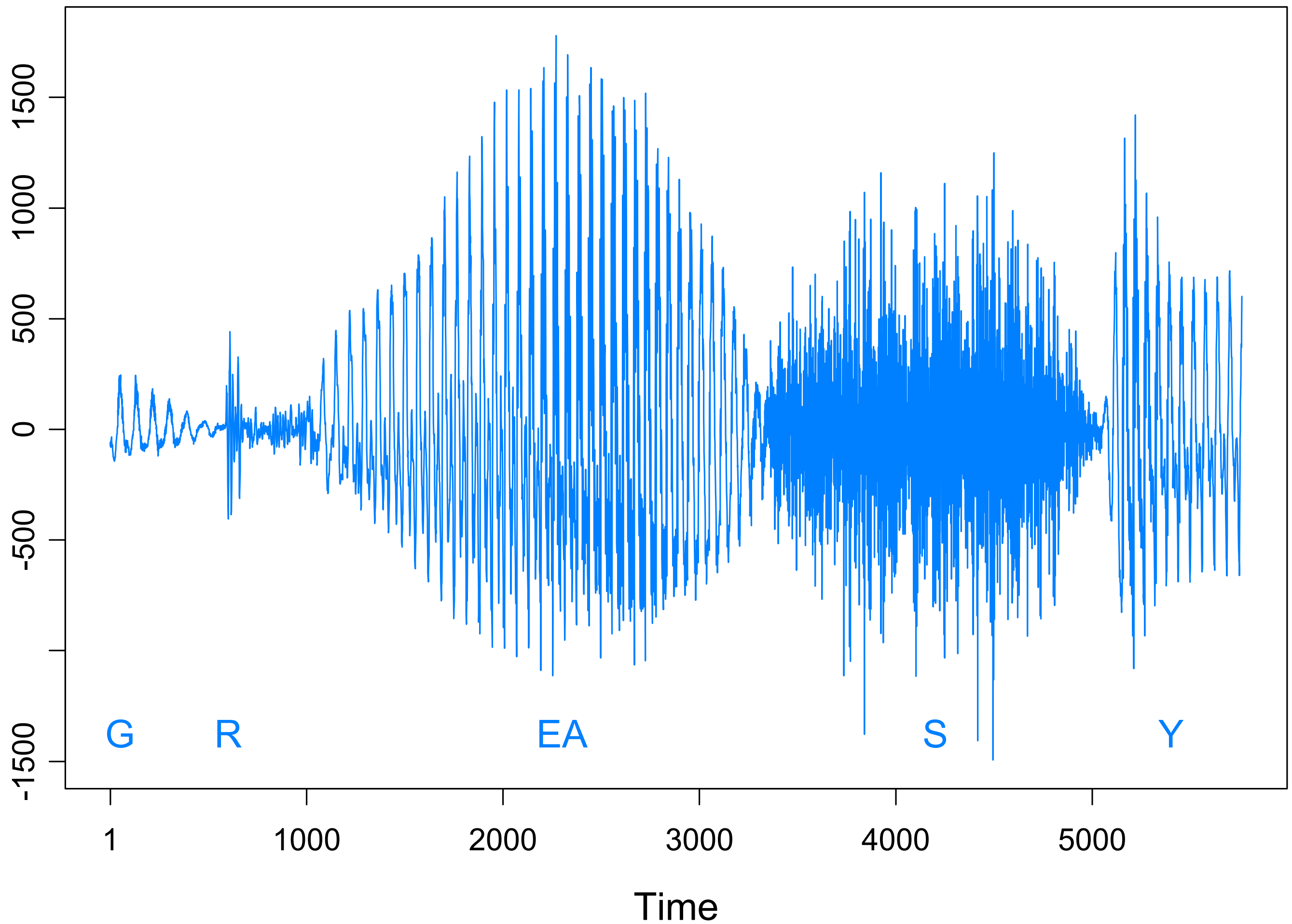


Results from GA: 3 pieces; time = 4.4secs

Piece 1: (t=1,…, 98) IID; Piece 2: (t=99,…108) IID; Piece 3: t=109,…,120 AR(1)

Examples

Speech signal: GREASY

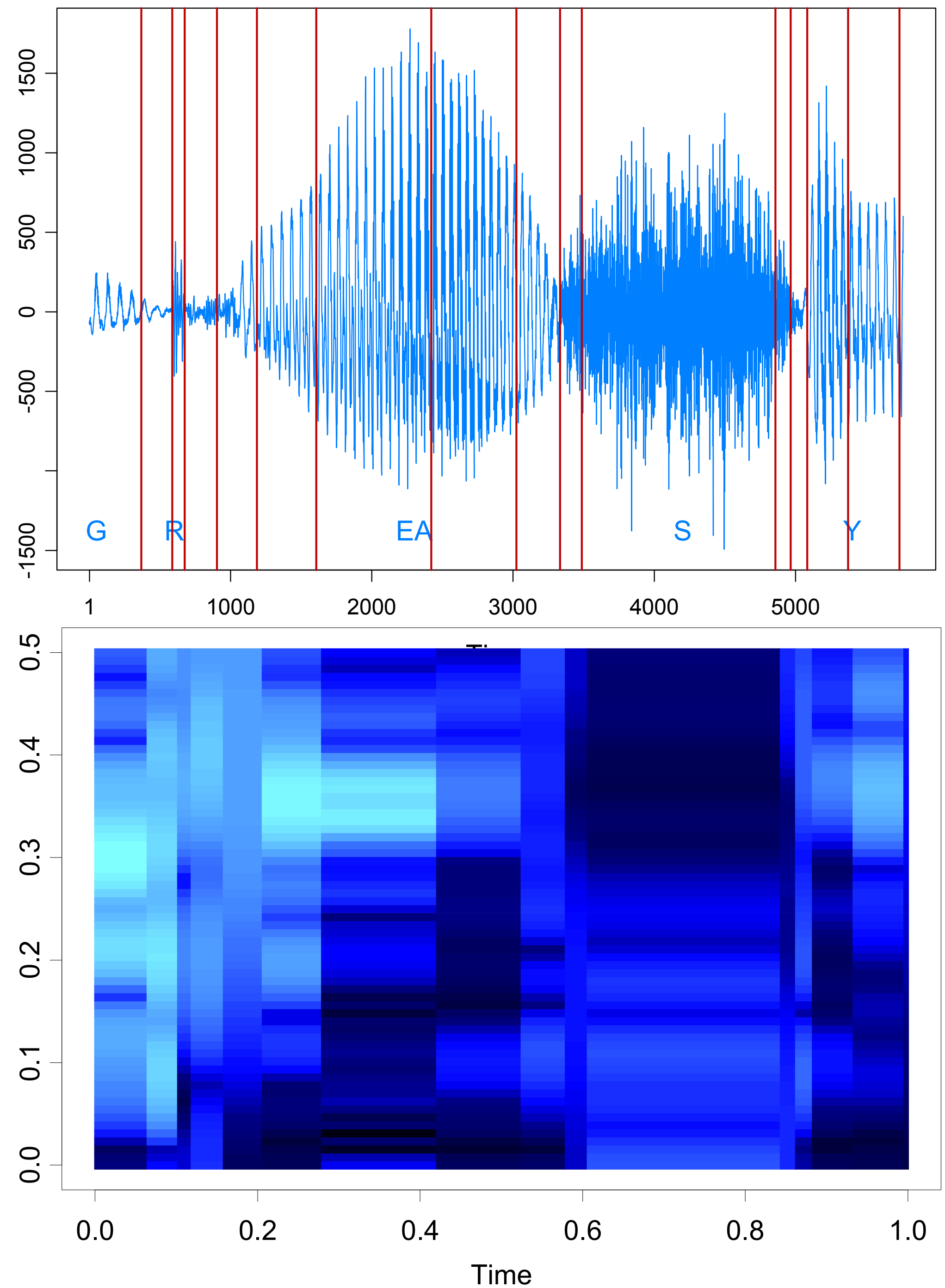


Speech signal: GREASY

n = 5762 observations

m = 15 break points

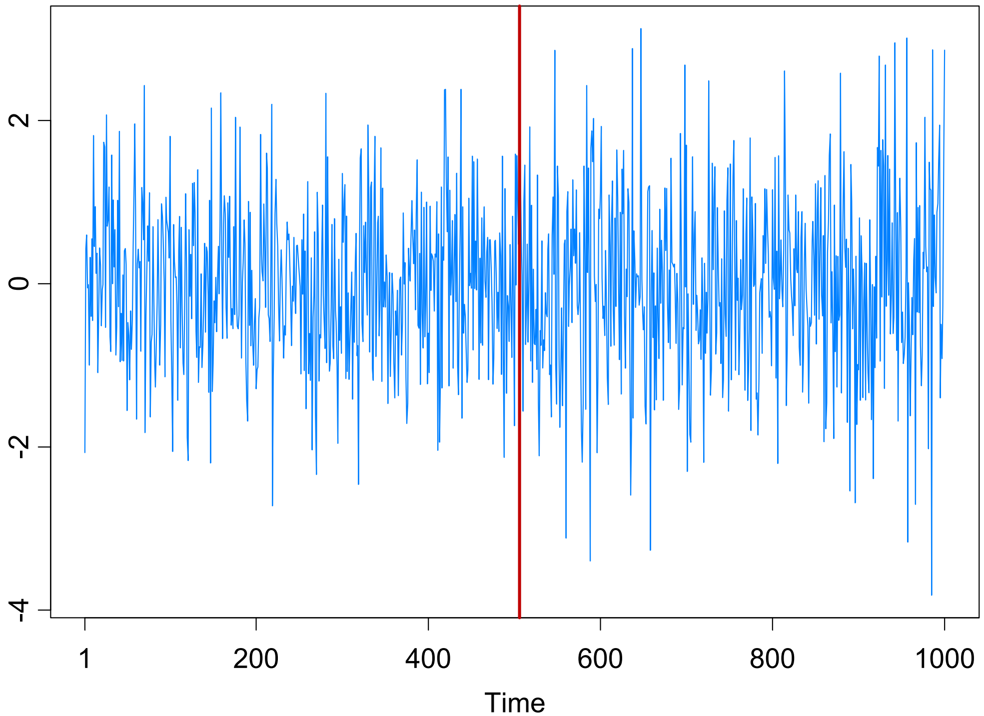
Run time = 18.02 secs



Application to GARCH (cont)

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

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



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

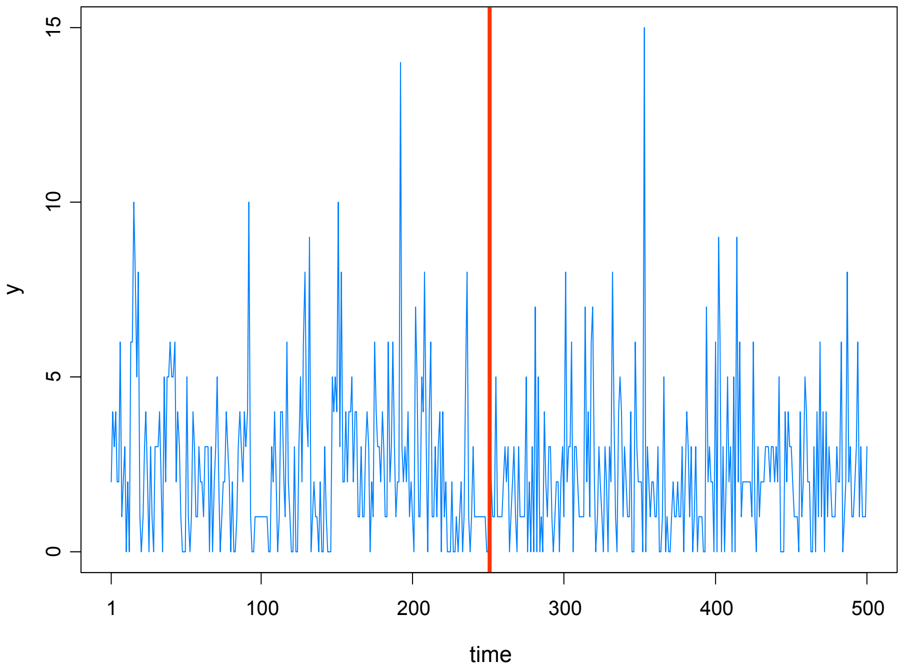
# of CPs	GA %	AG %
0	80.4	72.0
1	19.2	24.0
≥ 2	0.4	0.4

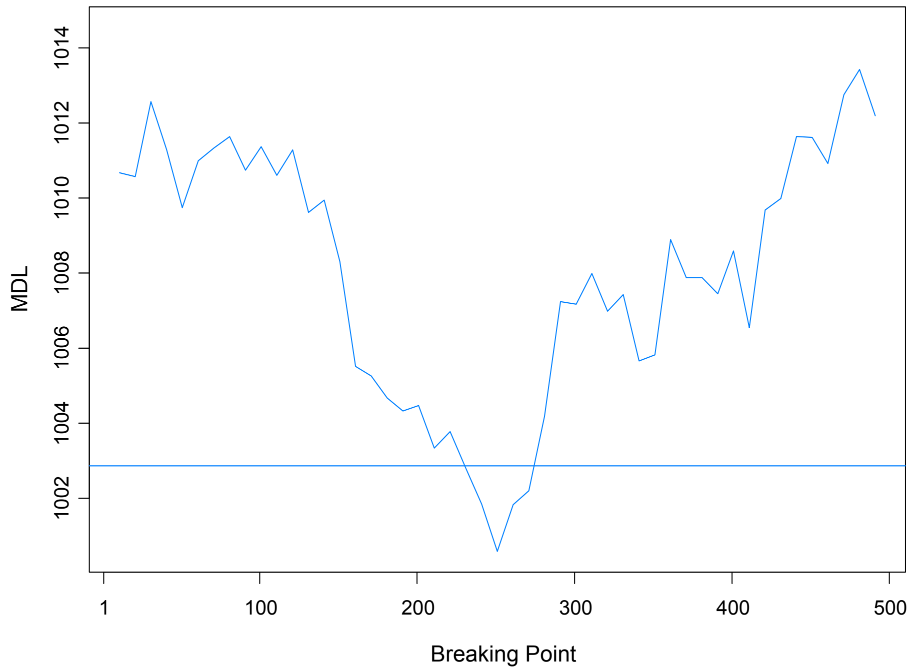
CP estimate = 506

AG = Andreou and Ghysels (2002)

Count Data Example

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



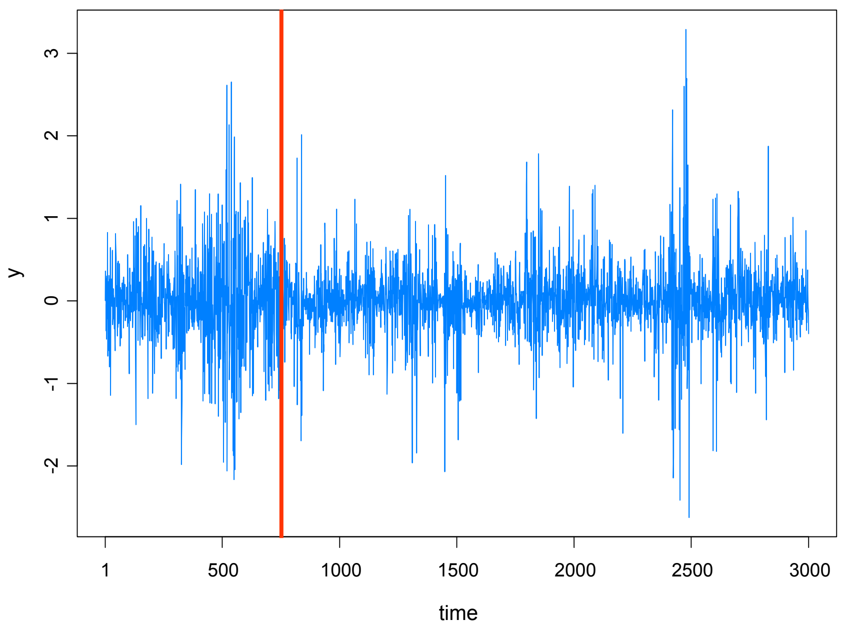


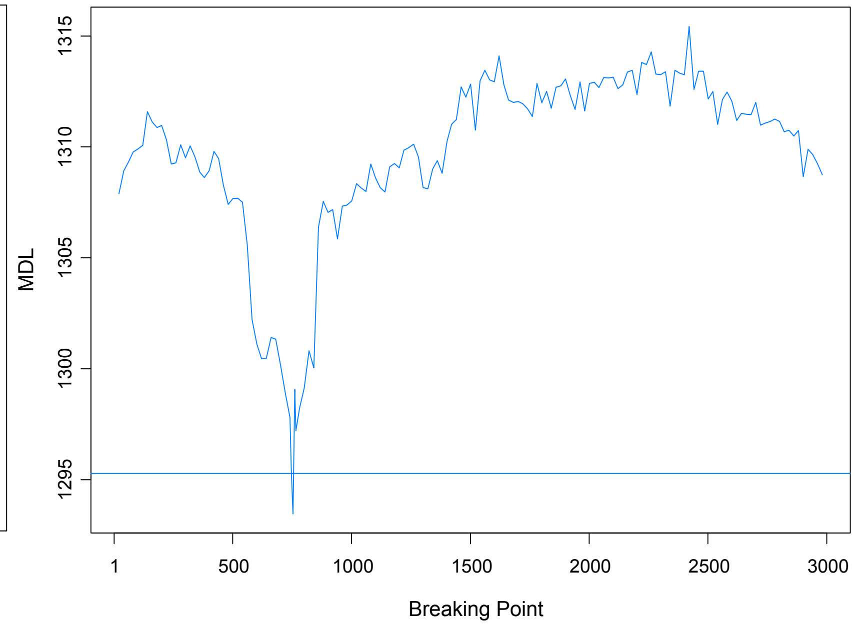
True model:

- $Y_t \mid \alpha_t \sim Pois(\exp\{.7 + \alpha_t\}),\ \alpha_t = .5\alpha_{t-1} + \varepsilon_t\ ,\ \ \{\varepsilon_t\} \sim \text{IID N}(0, .3),\ \ t < 250$
- $Y_t \mid \alpha_t \sim Pois(\exp\{.7 + \alpha_t\}),\ \ \alpha_t = -.5\alpha_{t-1} + \varepsilon_t\ ,\ \ \{\varepsilon_t\} \sim \text{IID N}(0, .3),\ \ t > 250.$
- 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 $N(0, \sigma^2)$



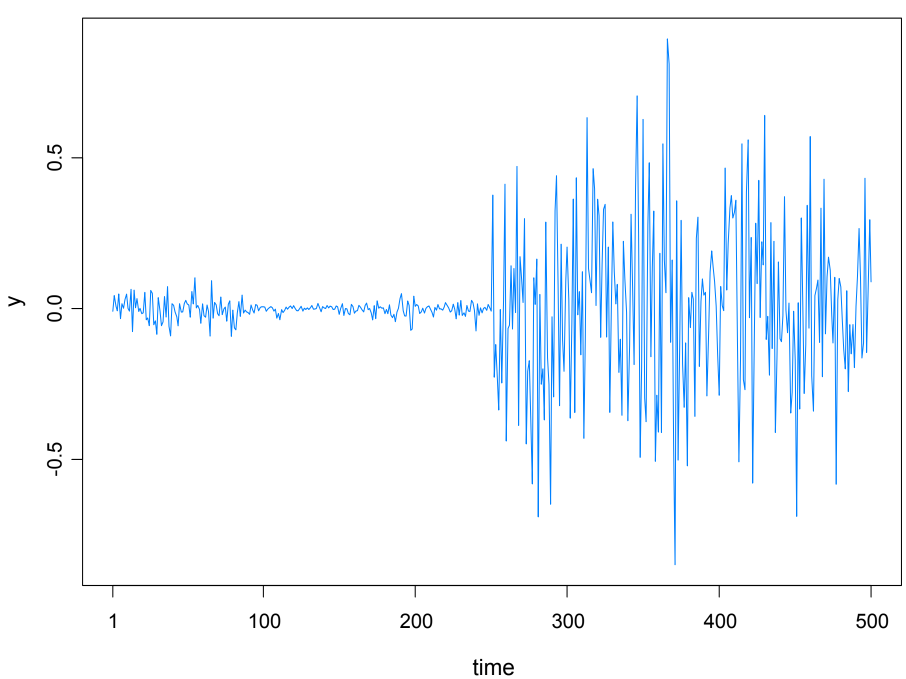


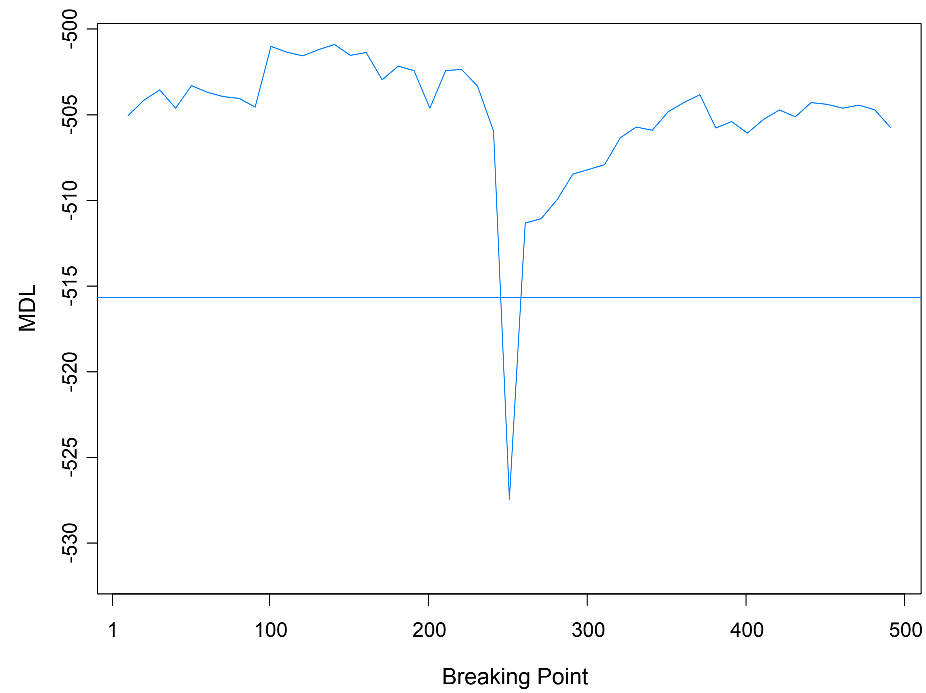
True model:

- $Y_t \mid \alpha_t \sim N(0, \exp\{\alpha_t\})$, $\alpha_t = -.05 + .975\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim$ 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

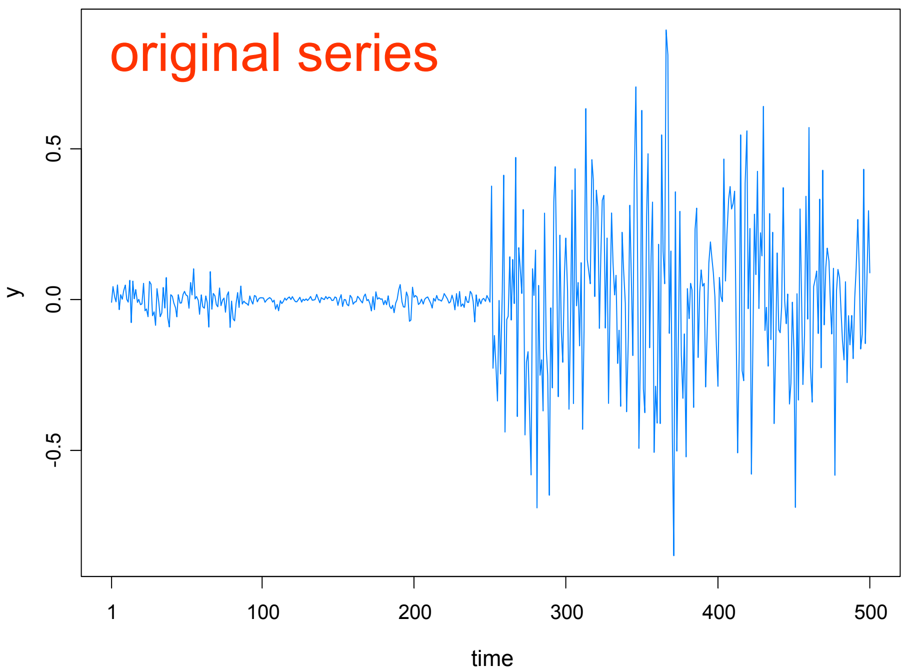
SV Process Example-(cont)

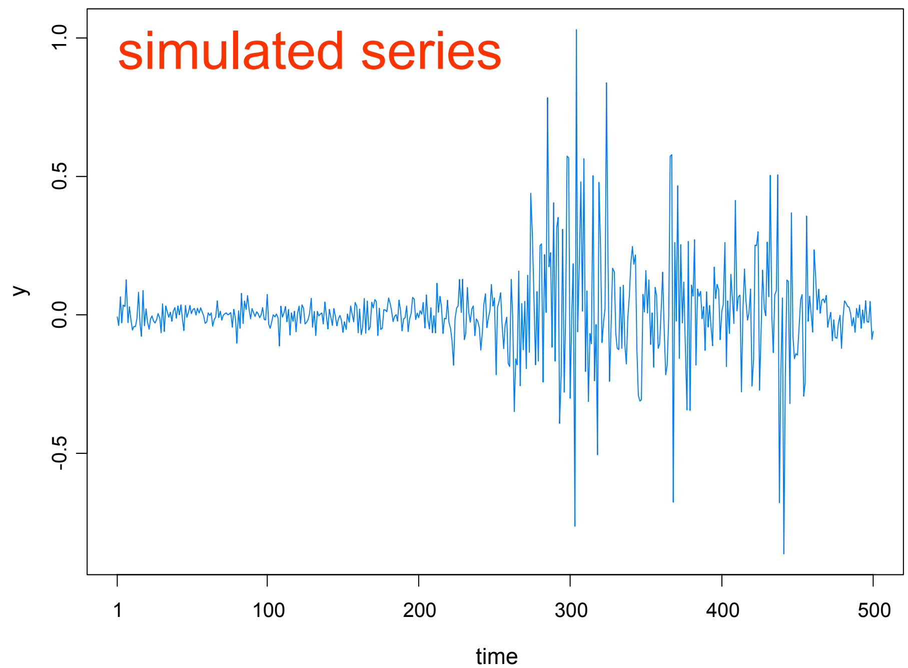
True model:

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

Fitted model based on no structural break:

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

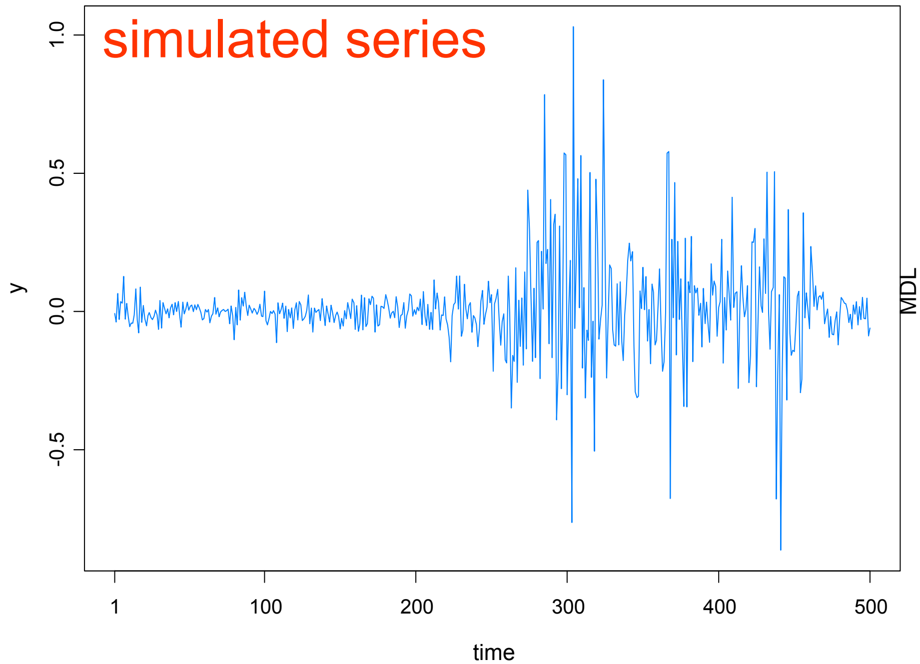


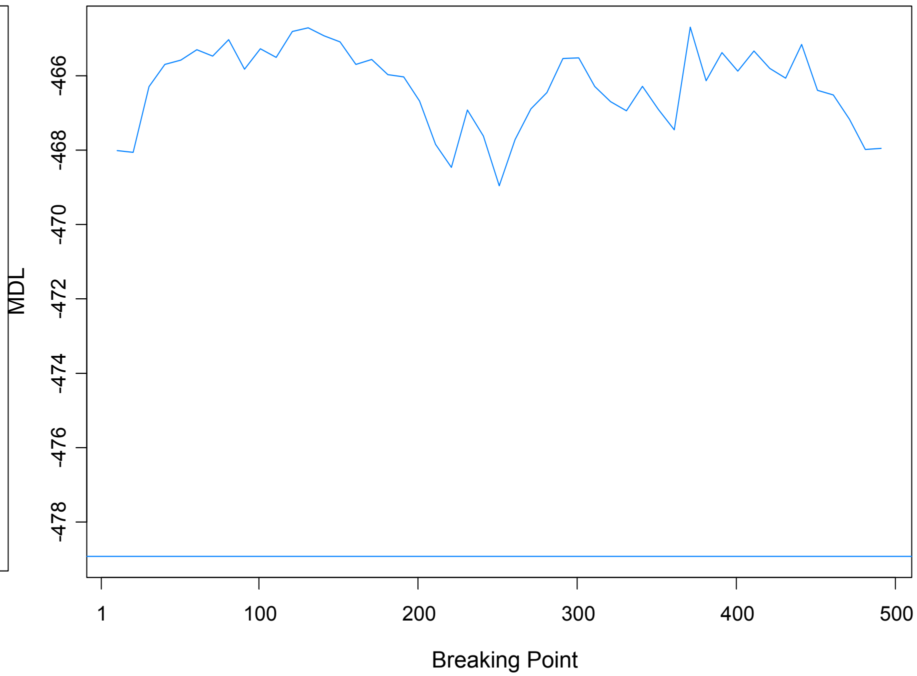


SV Process Example-(cont)

Fitted model based on no structural break:

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





Summary Remarks

1. *MDL* appears to be a good criterion for detecting structural breaks.

2. Optimization using a *genetic algorithm* is well suited to find a near optimal value of MDL.

3. This procedure extends easily to *multivariate* problems.

4. While estimating structural breaks for nonlinear time series models is *more challenging*, this paradigm of using *MDL together GA* holds promise for break detection in *parameter-driven* models and other nonlinear models.