Break Detection for a Class of Nonlinear Time Series Models

Richard A. Davis
Columbia University

Thomas Lee
Chinese University of Hong Kong
Colorado State University

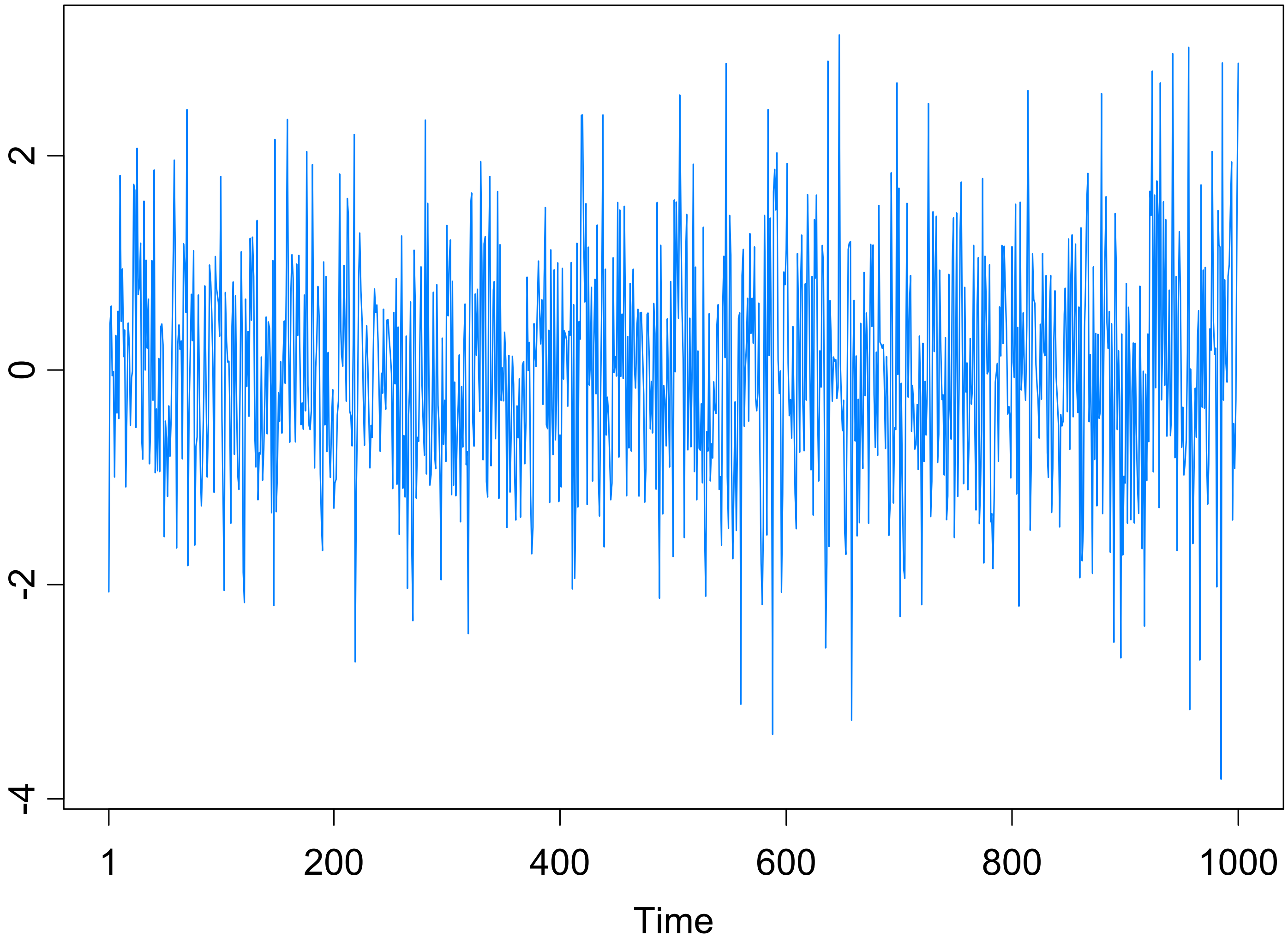
Gabriel Rodriguez-Yam
Universidad Autónoma Chapingo

http://www.stat.columbia.edu/~rdavis/lectures

This research supported in part by an IBM faculty award.

An Example

Any breaks in this series?



Game Plan

- Introduction
  - Examples of base models
    - AR
    - GARCH
    - Stochastic volatility
    - State space models
- Model selection using Minimum Description Length (MDL)
  - General principles
- Optimization using a Genetic Algorithm
  - Basics
- Simulation results for GARCH and SV models

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

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

such that

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

where the pieces $\{X_{t,j}\}$, $j=1,\ldots,m+1$ are independent and the j th piece is a stationary time series with distr P_{θ_j} and $\theta_j \neq \theta_{j+1}$.

Objective: estimate

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

Introduction—Examples

1. Piecewise AR model:

$$X_{t,j} = \phi_{j0} + \phi_{j1} X_{t-1,j} + \cdots + \phi_{jp_j} X_{t-p_j,j} + \sigma_j \varepsilon_{t,j}, \quad t = \ldots,\text{-}1,0,1,\ldots,$$

where and $\{\varepsilon_{t,j}\}$ is IID(0,1).

Goal: Estimate

m = number of breakpoints

τ_j = location of j^{th} break point

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

$(\phi_{j0}, \phi_{j1}, \ldots, \phi_{jp_j})$ = AR coefficients in j^{th} epoch

σ_j = scale in j^{th} epoch

Introduction—Examples

2. Segmented GARCH model: $Y_t = X_{t+1-\tau_{j-1},j}, \quad \text{if } \tau_{j\text{-}1} \le t < \tau_j,$

$$X_{t,j} = \sigma_{t,j}\varepsilon_{t,j},$$

$$\sigma^2_{t,j} = \alpha_{j0} + \alpha_{j1}X^2_{t-1,j} + \cdots + \alpha_{jp_j}X^2_{t-p_j,j} + \beta_{j1}\sigma^2_{t-1,j} + \cdots + \beta_{jq_j}\sigma^2_{t-q_j,j}, \qquad t = \ldots,\text{-}1,0,1,\ldots,$$

where $\{\varepsilon_{t,j}\}$ is IID(0,1).

3. Segmented stochastic volatility model:

$$Y_t = \sigma_t\varepsilon_t,$$

$$\log\sigma^2_t = X_{t+1-\tau_{j\text{-}1},j}, \qquad \text{if } \tau_{j\text{-}1} \le t < \tau_j.$$

where $\{X_{t,j}\}$ are the piecewise AR processes described in 1.

Introduction—Examples

4. Segmented state-space model (SVM a special case): Let $\{\alpha_t\}$ be a latent process (state-process). Then it is assumed that $\{\alpha_t\}$ follows the piecewise AR model in Eample 1. That is,

$$p(y_t \mid \alpha_t,\ldots,\alpha_1, y_{t-1},\ldots,y_1) = p(y_t \mid \alpha_t) \text{ is specified}$$
$$\alpha_t = X_{t+1-\tau_{j-1}}, \quad \text{if } \tau_{j\text{-}1} \le t < \tau_j.$$

For example, consider an observation eqn that belongs to the ***exponential family*** given by

$$p(y_t \mid \alpha_t) = \exp\{(z_t^T\beta + \alpha_t)y_t - b(z_t^T\beta + \alpha_t) + c(y_t)\},$$

where z_t is a vector of covariates, β a parameter vector, and b(·) and c(·) are known functions.

Remark: While the assumption of independence may seem restrictive, it can viewed as an approximating model in which dependence is allowed across segments.

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

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

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

$$CL_{\boldsymbol{F}}(y) = CL(\hat{\boldsymbol{F}}|y) + CL(\hat{e}\,|\,\hat{\boldsymbol{F}}),$$

where

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

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

Model Selection Using Minimum Description Length (cont)

First term $CL(\hat{\mathbf{F}}|y)$: For the j^{th} segment, let

$n_j = \tau_j - \tau_j$ sample size;

ζ_j = integer-valued parameters (e.g., model order) with dim c_j;

$\hat{\psi}_j$ = MLE of real-valued parameters (with dim d_j) given n_j and ζ_j.

Then

$$CL(\hat{\mathbf{F}}|y) = CL(m) + CL(n_1) + \cdots + CL(n_{m+1}) + CL(\zeta_1) + \cdots + CL(\zeta_{m+1}) + CL(\hat{\psi}_1) + \cdots + CL(\hat{\psi}_m)$$

$$= \log_2 m + (m+1)\log_2 n + \sum_{j=1}^{m+1}\sum_{k=1}^{c_j}\log_2 \zeta_{j,k} + \sum_{j=1}^{m+1}\frac{d_j}{2}\log_2 n_j$$

Second term $L(\hat{e}\,|\,\hat{\mathbf{F}})$: Using results by Rissanen

$$CL(\hat{e}\,|\,\hat{\mathbf{F}}) \approx -\sum_{j=1}^{m+1}\log_2 L(\hat{\psi}_j\,|\,y_j)$$

Model Selection Using Minimum Description Length (cont)

Putting the two terms together we obtain

$$MDL(m,\tau_1,\ldots,\tau_m,\zeta_1,\ldots,\zeta_{m+1}) = CL(y)$$
$$= \log_2 m + (m+1)\log_2 n + \sum_{j=1}^{m+1}\sum_{k=1}^{c_j}\log_2 \zeta_{j,k} + \sum_{j=1}^{m+1}\frac{d_j}{2}\log_2 n_j - \sum_{j=1}^{m+1}\log_2 L(\hat{\psi}_j \mid y_j)$$

Piecewise GARCH(1,1) model:

$$Y_t = \sigma_{t,j}\varepsilon_{t,j},$$
$$\sigma_{t,j}^2 = \alpha_{j0} + \alpha_{j1}Y_{t-1}^2 + \beta_{j1}\sigma_{t-1,j}^2, \qquad \tau_j \le t < \tau_{j+1}.$$

$$MDL(m,\tau_1,\ldots,\tau_m)$$
$$= \log_2 m + (m+1)\log_2 n + \sum_{j=1}^{m+1}\frac{3}{2}\log_2 n_j - \sum_{j=1}^{m+1}\log_2 L(\hat{\psi}_j \mid y)$$

Remark: For the GARCH, we replace 3/2 by 1 in MDL, i.e., reduce model complexity, due to the high correlation between α_0 and (α_1,α_2).

Optimization Using Genetic Algorithm

Basics of GA:

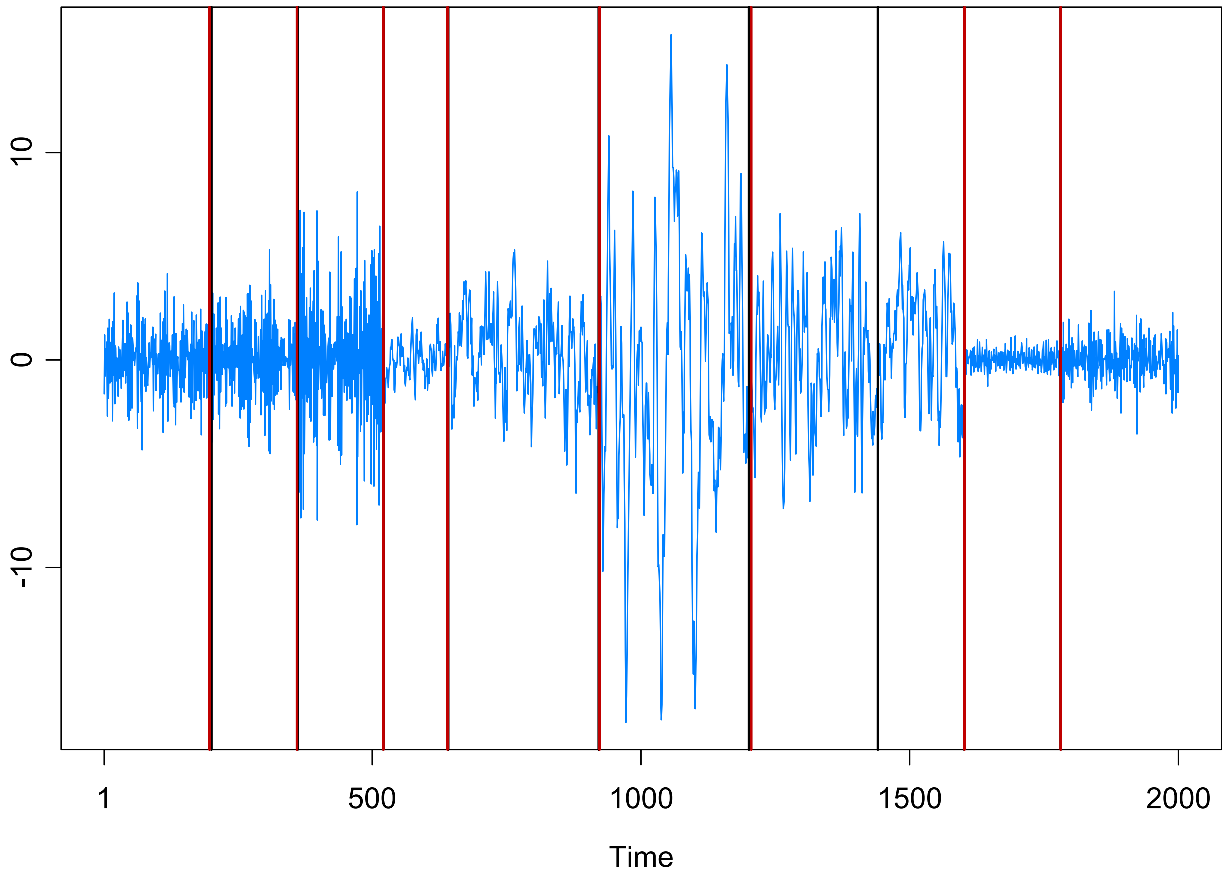
Class of optimization algorithms that mimic natural evolution.

- Start with an initial set of *chromosomes*, or population, of ***parameter values*** that possible solutions to the optimization problem.
- Parent chromosomes are randomly selected (proportional to the rank of their ***MDL***), 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 produces future generations that give a *smaller **MDL** value.*

AR Simulation Example

Simulated data from Fearnhead (2005):

True model has 9 changepoints (autoregressive)

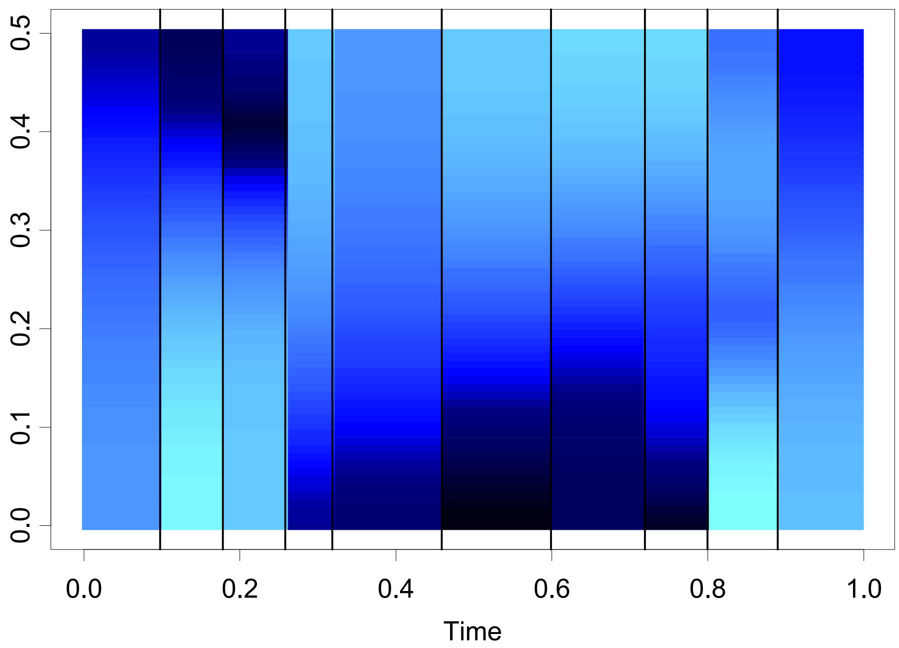


black=true

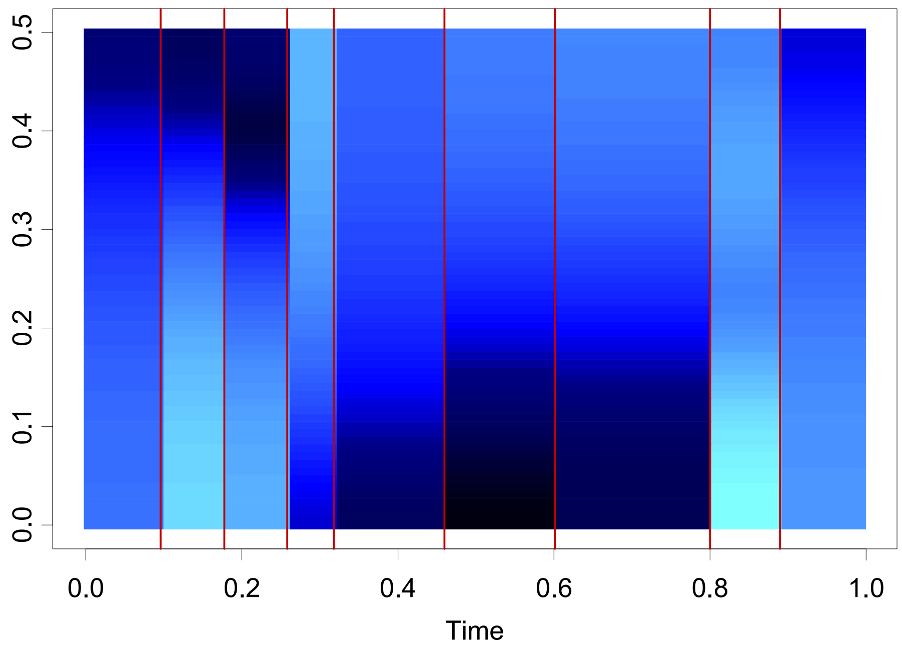
red=APARM

Fearnhead example (cont)

True Model



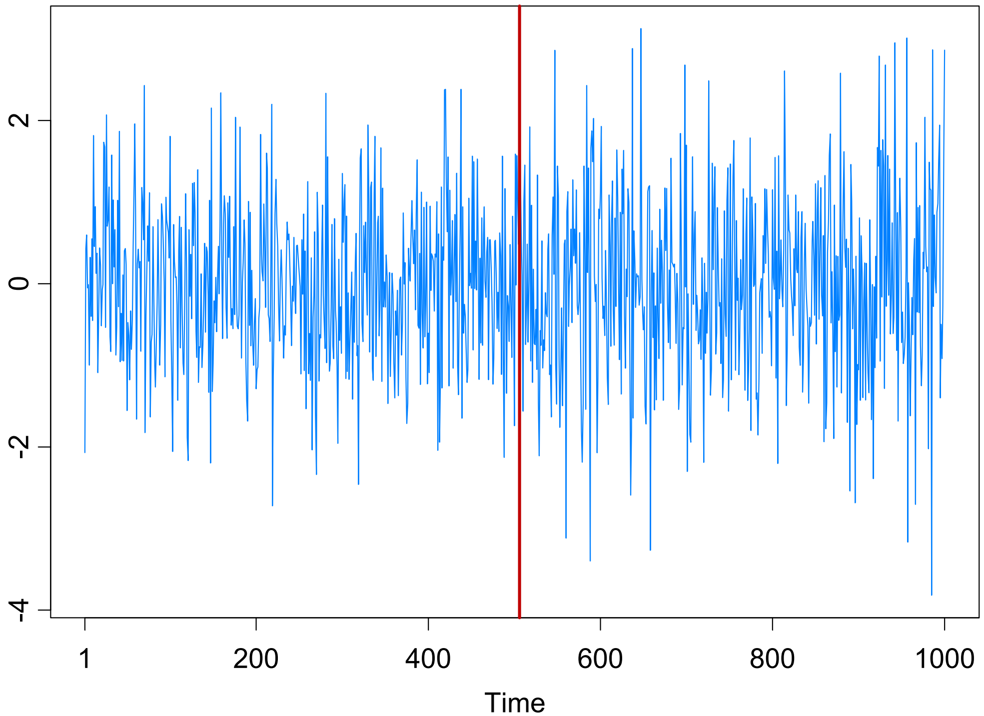
Fitted APARM Model



Application to GARCH

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

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



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

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

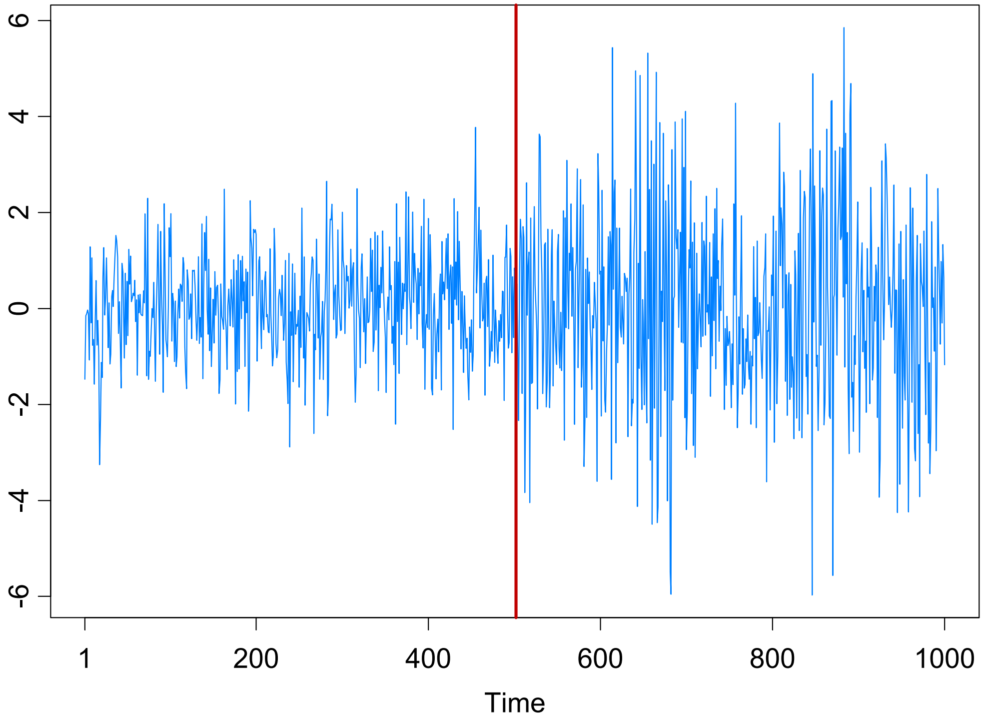
CP estimate = 506

AG = Andreou and Ghysels (2002)

Application to GARCH (cont)

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

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



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

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

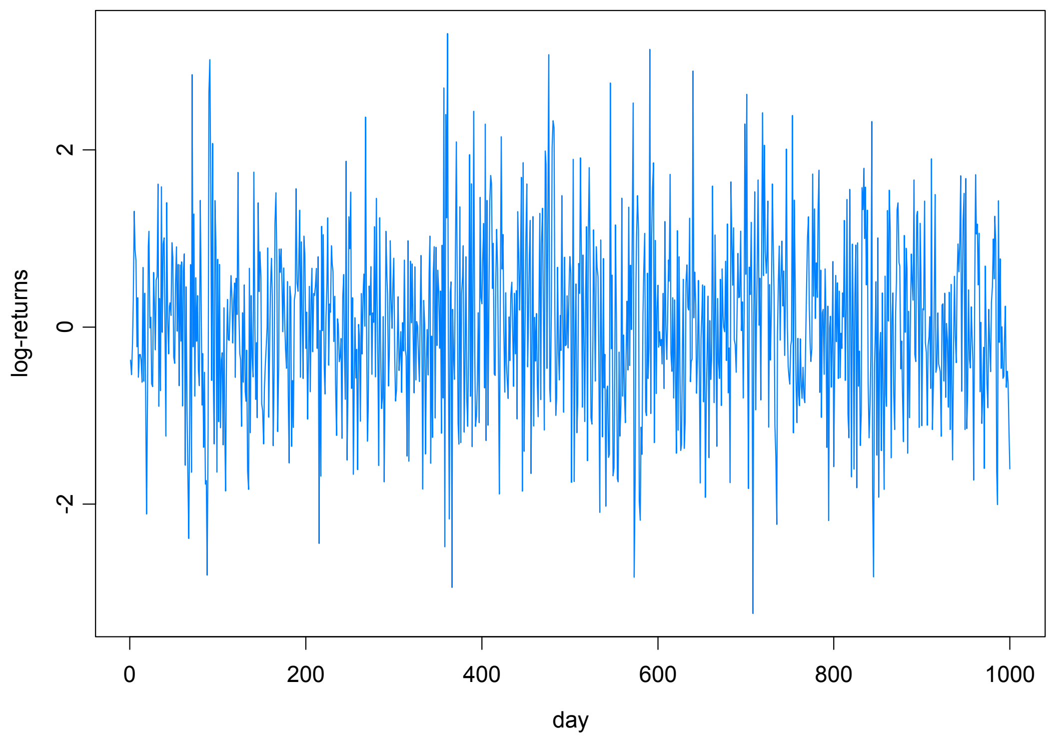
CP estimate = 502

AG = Andreou and Ghysels (2002)

Application to GARCH (cont)

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

$$\sigma_t^2 = 0.1 + 0.1 Y_{t-1}^2 + 0.8 \sigma_{t-1}^2$$



No break.

# of CPs	AutoSeg %	AG %
0	95.6	88.0
1	4.5	7.0
≥ 2	0.0	5.0

Application to GARCH (cont)

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

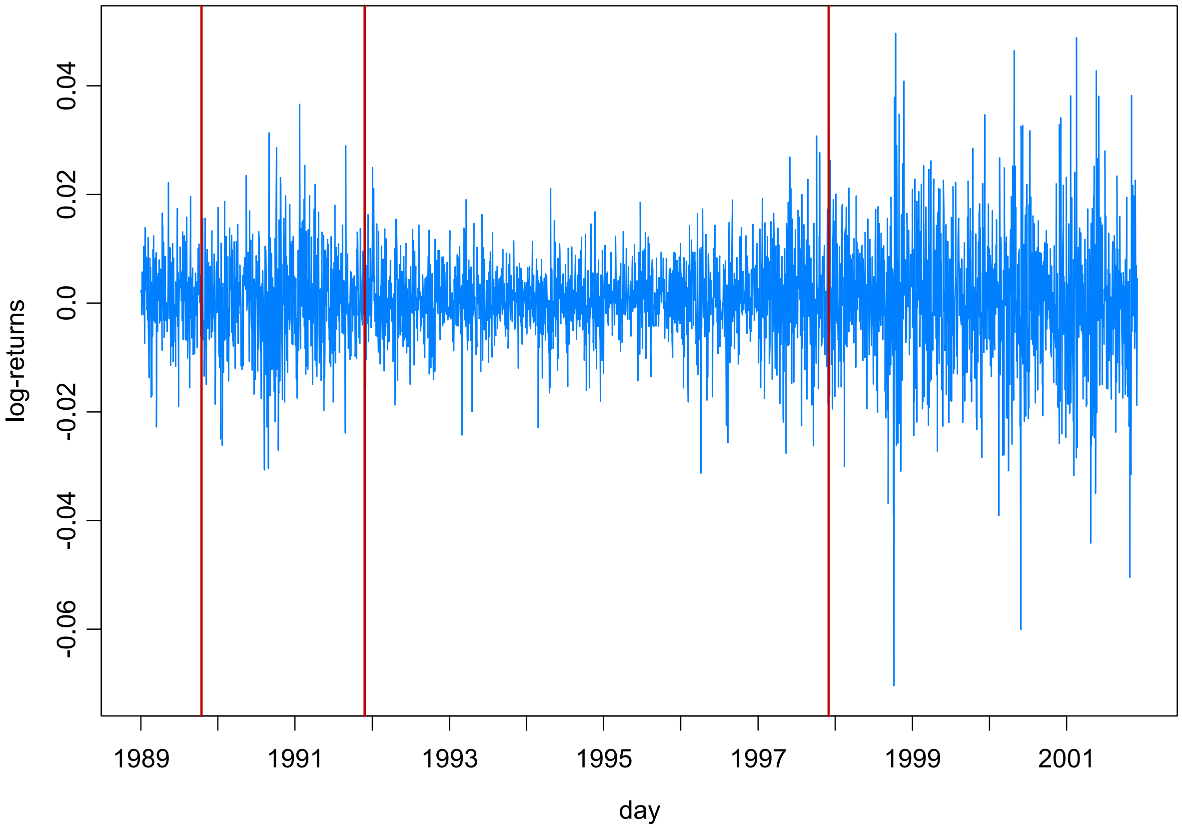
$$\sigma_t^2 = \begin{cases} .05 + .4Y_{t-1}^2 + .3\sigma_{t-1}^2, & \text{if } 1 \le t < \tau_1, \\ 1.00 + .3Y_{t-1}^2 + .2\sigma_{t-1}^2, & \text{if } \tau_1 \le t < 1000. \end{cases}$$

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

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

Log-returns for S&P 500, 4 Jan 1989 to 19 Oct 2001 (N=3230)

Andreou and Ghysels (2002) examined impact of Asisan and Russian financial crises (July 1997—Dec 1998) on S&P 500



AutoSeg:
Oct 13, 1989
Nov 15,1991
Oct 27, 1997

A-G based on $|Y_t|$:
Dec 27, 1989
Jan 1, 1996
Jul 28, 1998

A-G based on Y_t^2:
Oct 14, 1997

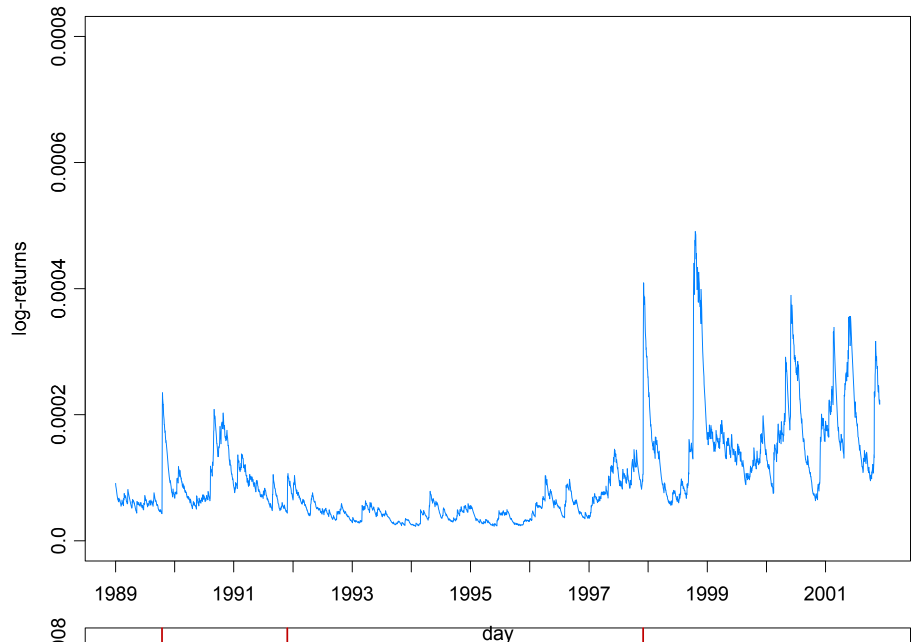
Log-returns for S&P, 4 Jan 1989 to 19 Oct 2001 (N=3230)

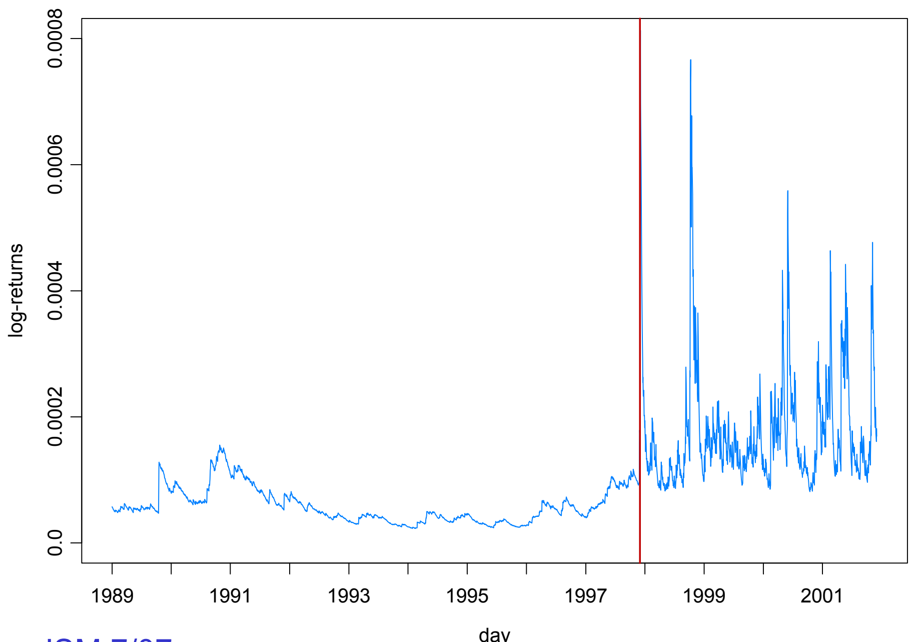
Estimates of volatility assuming:

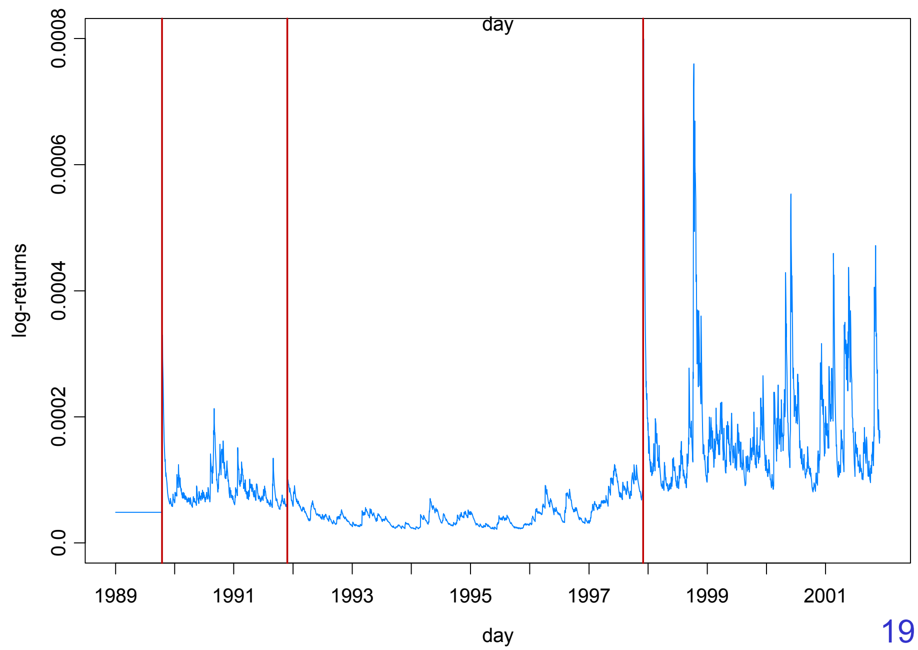
a) No breaks

b) 1 break

c) 3 breaks







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 $\{\varepsilon_t\} \sim$ 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 Parameter Driven SS Models—(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

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

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

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

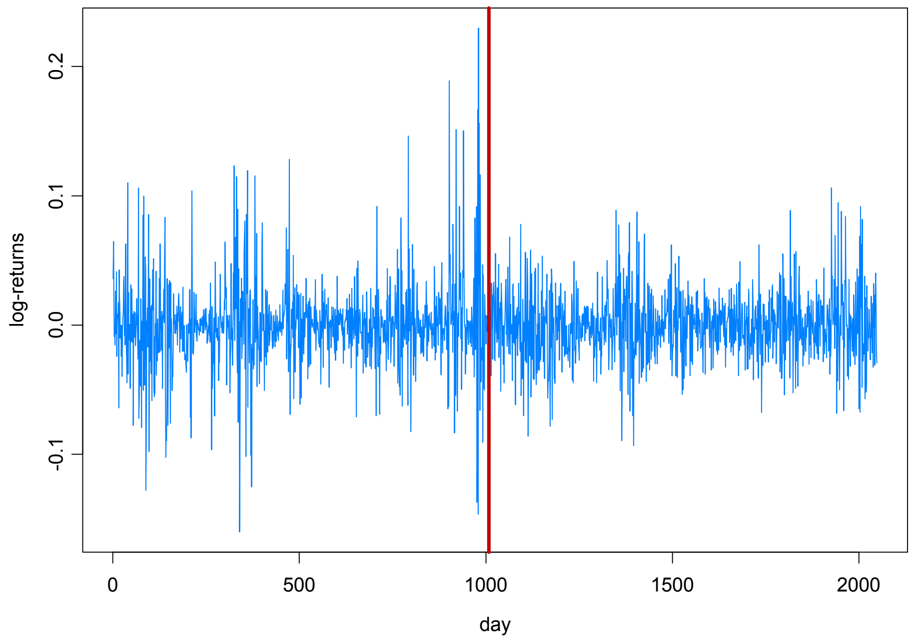
Application to a Stochastic Volatility Model

SV model:

$$Y_t = \sigma_t \varepsilon_t = e^{\alpha_t/2}\varepsilon_t, \qquad \{\varepsilon_t\} \sim \text{IID}(0,1)$$

$$\alpha_t = \gamma + \phi\alpha_{t-1} + \eta_t, \qquad \{\eta_t\} \sim \text{IIDN}(0,\sigma^2)$$

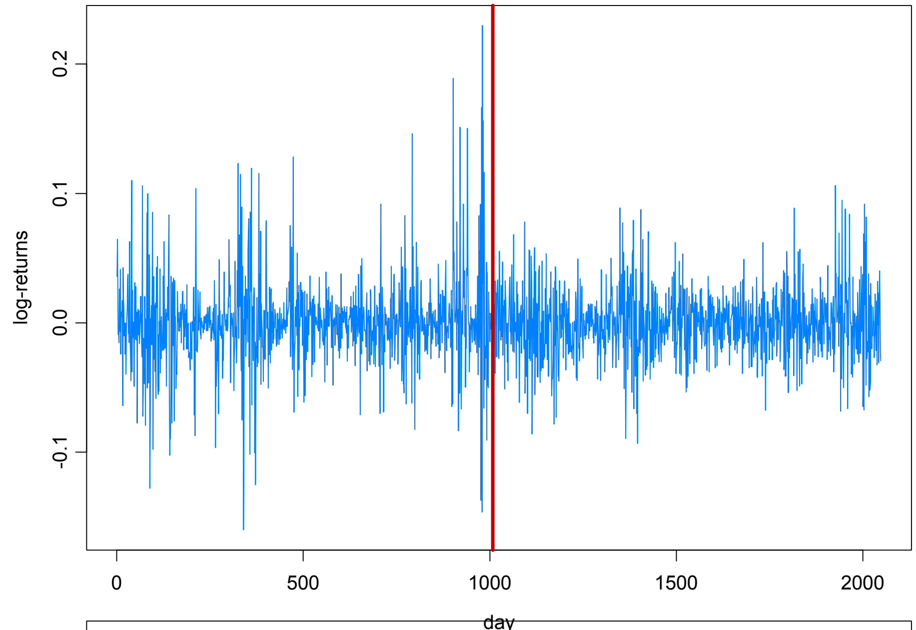
$$\begin{cases} \gamma = -.81067,\ \phi = .90,\ \sigma^2 = .4556, & \text{if } 1 \le t < 1024, \\ \gamma = -.39737,\ \phi = .95,\ \sigma^2 = .0676, & \text{if } 1024 \le t < 2048. \end{cases} \qquad \begin{aligned} \sigma_y^2 &= 0.0010 \\ \sigma_y^2 &= 0.0005 \end{aligned}$$



AutoSeg estimate = 1008

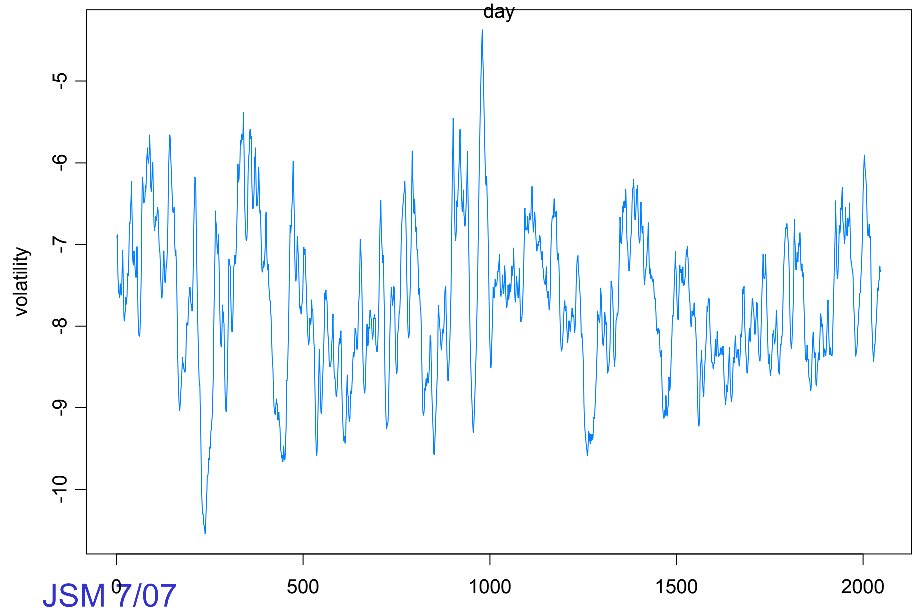
# of CPs	AutoSeg %
0	1.2
1	98.8
≥ 2	0.0

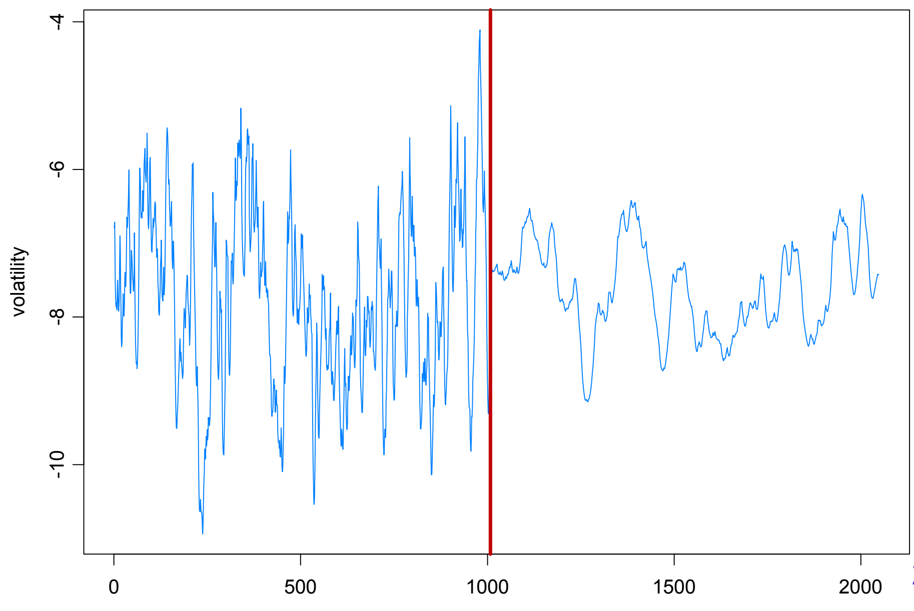
Application to a Stochastic Volatility Model (cont)



Two figures below are the posterior modes of the state process α_t.

1. Assuming no break.
2. Assuming one break.





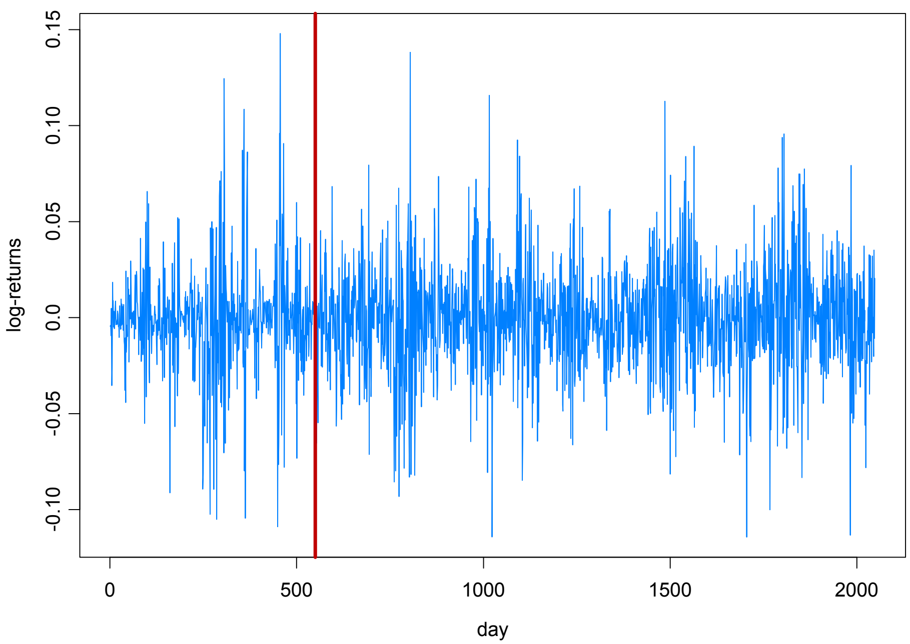
Second Stochastic Volatility Model Example

SV model:

$$Y_t = \sigma_t \varepsilon_t = e^{\alpha_t/2}\varepsilon_t, \qquad \{\varepsilon_t\} \sim \mathrm{IID}(0,1)$$

$$\alpha_t = \gamma + \phi\alpha_{t-1} + \eta_t, \quad \{\eta_t\} \sim \mathrm{IIDN}(0,\sigma^2)$$

$$\begin{cases} \gamma = -.81067,\ \phi = .90,\ \sigma^2 = .4556, & \text{if } 1 \le t < 513, \qquad \sigma_y^2 = 0.0010 \\ \gamma = -.37387,\ \phi = .95,\ \sigma^2 = .0676, & \text{if } 513 \le t < 2048. \qquad \sigma_y^2 = 0.0008 \end{cases}$$

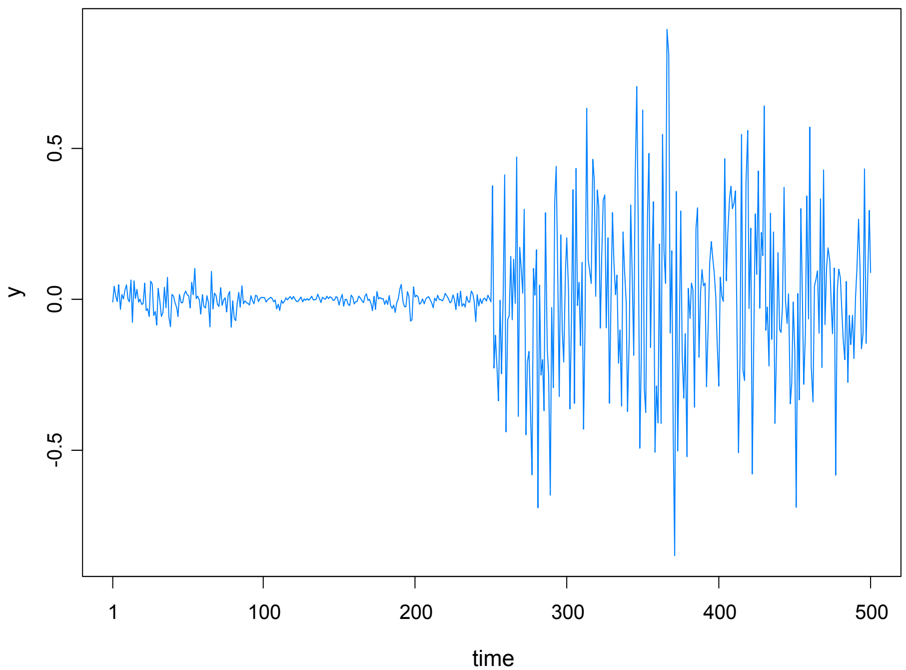


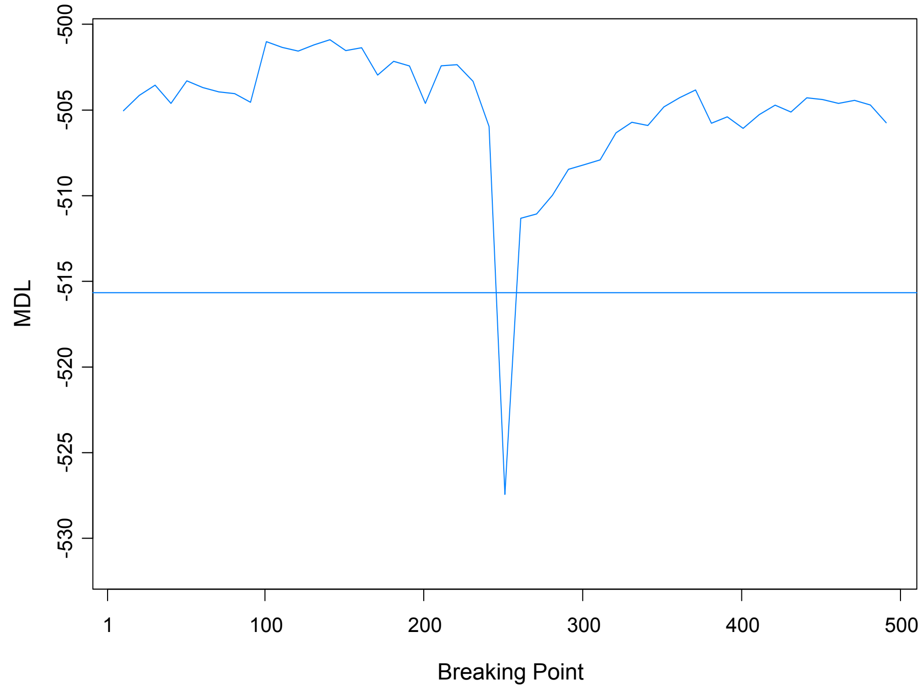
AutoSeg estimate = 550

# of CPs	AutoSeg %
0	18.2
1	82.8
≥ 2	0.0

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:

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

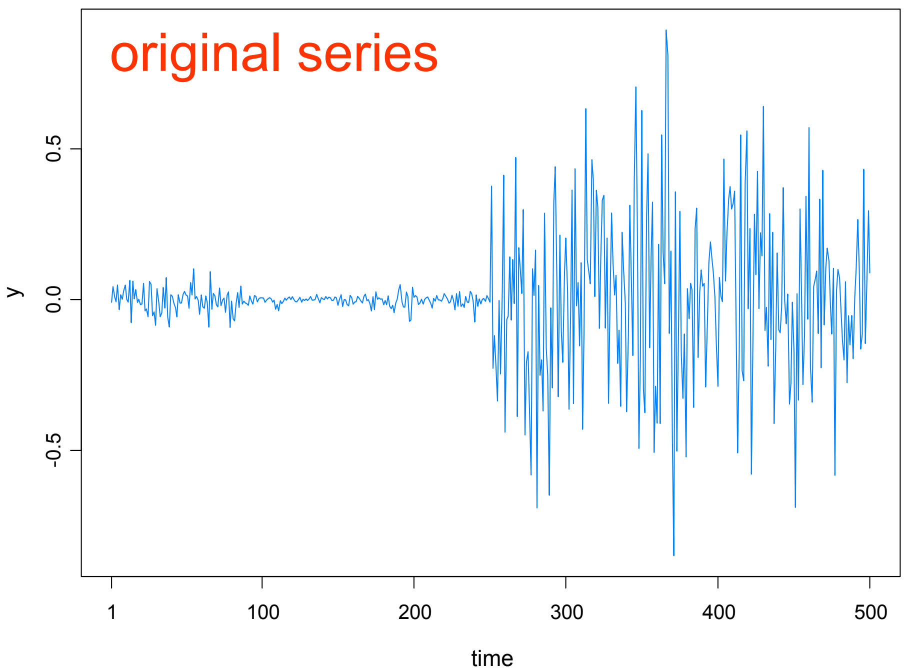
SV Process Example-(cont)

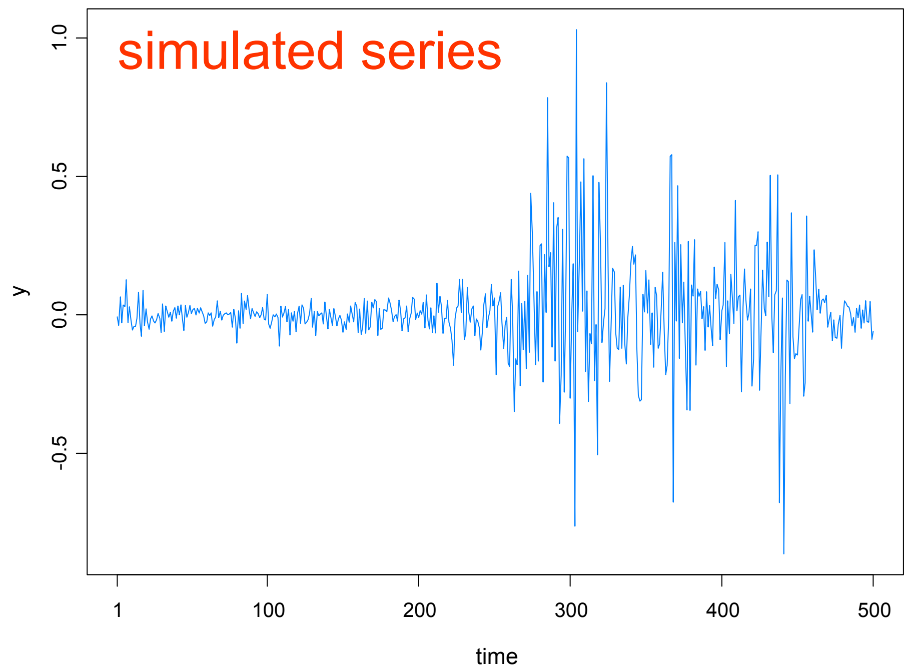
True model:

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

Fitted model based on no structural break:

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

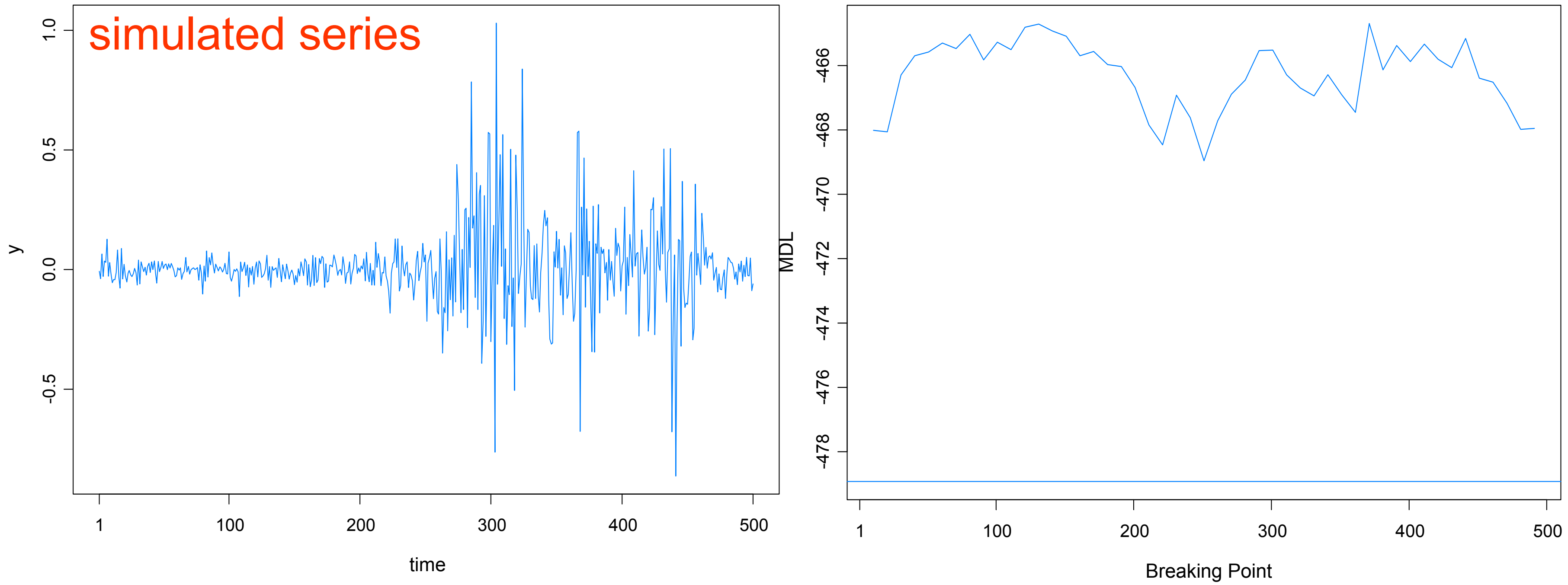




SV Process Example-(cont)

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)



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.