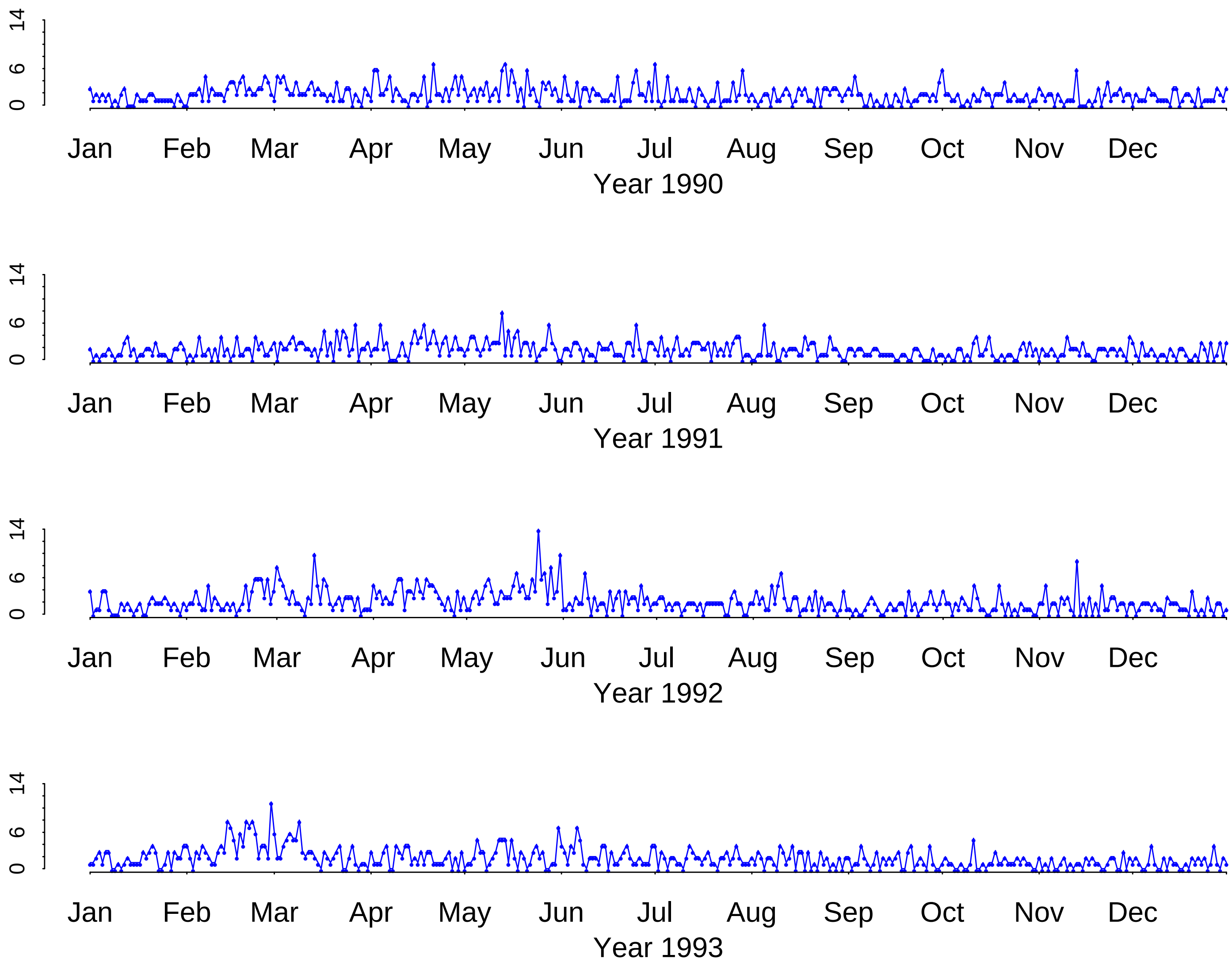
Estimation for a Class of State-Space Models:

Richard A. Davis and Gabriel Rodriguez-Yam
Colorado State University

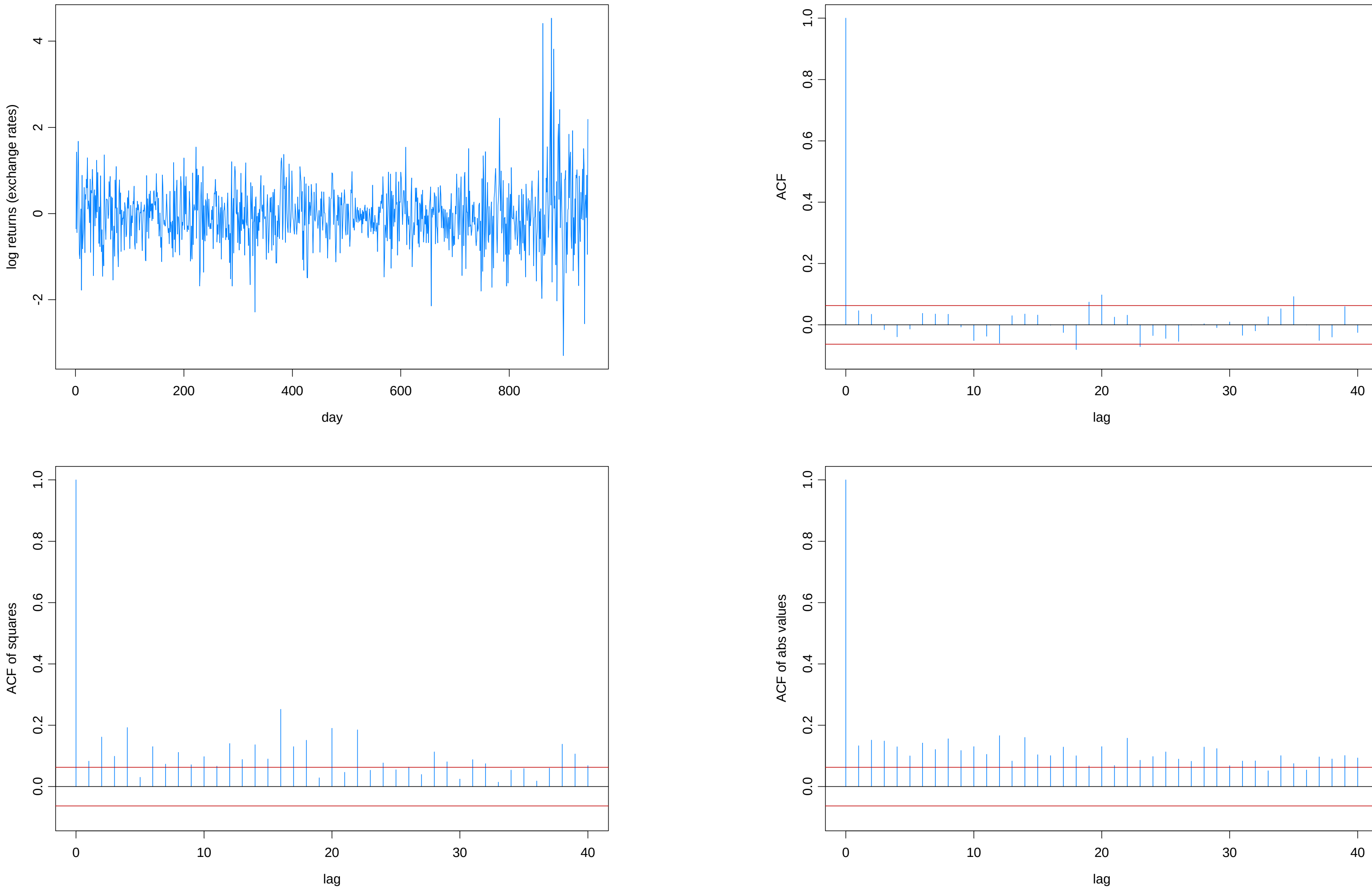
(http://www.stat.colostate.edu/~rdavis/lectures)

Joint work with:
William Dunsmuir, University of New South Wales
Ying Wang, Dept of Public Health, W. Virginia

Example: Daily Asthma Presentations (1990:1993)



Example: Pound-Dollar Exchange Rates
(Oct 1, 1981 – Jun 28, 1985; Koopman website)



- Motivating Examples
  - Time series of counts
  - Stochastic volatility
- Generalized state-space models
- Model setup and estimation
  - Exponential family
    - 2 examples
  - Estimation
    - Importance sampling
    - Approximation to the likelihood
- Simulation and Application
  - Time series of counts
  - Stochastic volatility
- How good is the posterior approximation?
  - Posterior mode vs posterior mean
- Application to estimating structural breaks
  - Poisson model
  - Stochastic volatility model

Generalized State-Space Models

Observations: $y^{(t)} = (y_1, \ldots, y_t)$

States: $\alpha^{(t)} = (\alpha_1, \ldots, \alpha_t)$

Observation equation:

$$p(y_t \mid \alpha_t) := p(y_t \mid \alpha_t, \alpha^{(t-1)}, y^{(t-1)})$$

State equation:

-observation driven

$$p(\alpha_{t+1} \mid y^{(t)}) := p(\alpha_{t+1} \mid \alpha_t, \alpha^{(t-1)}, y^{(t)})$$

-parameter driven

$$p(\alpha_{t+1} \mid \alpha_t) := p(\alpha_{t+1} \mid \alpha_t, \alpha^{(t-1)}, y^{(t)})$$

Exponential Family Setup for Parameter-Driven Model

Time series data: $Y_1, \ldots, Y_n$

Regression (explanatory) variable: x_t

Observation equation:

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

State equation: $\{\alpha_t\}$ follows an autoregressive process satisfying the recursions

$$\alpha_t = \gamma + \phi_1 \alpha_{t-1} + \phi_2 \alpha_{t-2} + \ldots + \phi_p \alpha_{t-p} + \varepsilon_t ,$$

where $\{\varepsilon_t\} \sim$ IID $N(0,\sigma^2)$.

Note: $\alpha_t = 0$ corresponds to standard generalized linear model.

Original primary objective: Inference about β.

Examples of parameter driven models

Poisson model for time series of counts

Observation equation:

$$p(y_t \mid \alpha_t) = \frac{e^{(\beta^T x_t + \alpha_t) y_t} e^{-e^{(\beta^T x_t + \alpha_t)}}}{y_t!}, \quad y_t = 0, 1, \ldots,$$

State equation: State variables follow a Gaussian AR(1) process

$$\alpha_t = \phi \alpha_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{IID N}(0, \sigma^2)$$

The resulting transition density of the state variables is

$$p(\alpha_{t+1} \mid \alpha_t) = n(\alpha_{t+1};\ \phi\, \alpha_t, \sigma^2)$$

Remark: The case $\sigma^2 = 0$ corresponds to a log-linear model with Poisson noise.

Examples of parameter driven models-cont

A stochastic volatility model for financial data (Taylor `86):

Model:

$$Y_t = \sigma_t Z_t, \ \{Z_t\} \sim \text{IID N}(0,1)$$

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

where $\alpha_t = 2 \log \sigma_t$.

The resulting observation and state transition densities are

$$p(y_t | \alpha_t) = n(y_t; 0, \exp(2\alpha_t))$$

$$p(\alpha_{t+1} | \alpha_t) = n(\alpha_{t+1}; \gamma + \phi\alpha_t, \sigma^2)$$

Properties:

- Martingale difference sequence.
- Stationary.
- Strongly mixing at a geometric rate.

Estimation Methods for Parameter Driven Models

- Estimating equations (Zeger `88): Let $\hat{\beta}$ be the solution to the equation

$$\frac{\partial\mu}{\partial\beta}\Gamma_n(y_n-\mu)=0,$$

where $\mu = \exp(X\beta)$ and $\Gamma_n = \mathrm{var}(Y_n)$.

- Monte Carlo EM (Chan and Ledolter `95)
- GLM (ignores the presence of the latent process, i.e., $\alpha_t = 0$.)

- Importance sampling (Durbin & Koopman `01, Kuk `99, Kuk & Chen `97):
- Approximate likelihood (Davis, Dunsmuir & Wang '98)

Estimation Methods — Importance Sampling (Durbin and Koopman)

Model:

$$Y_t \mid \alpha_t\, ,\, x_t \sim Pois(\exp(x_t^T \beta + \alpha_t\,))$$

$$\alpha_t = \phi\, \alpha_{t-1} + \varepsilon_t\, , \quad \{\varepsilon_t\} \sim \text{IID } N(0, \sigma^2)$$

Relative Likelihood: Let $\psi=(\beta, \phi, \sigma^2)$ and suppose $g(y_n, \alpha_n; \psi_0)$ is an approximating joint density for $Y_n = (Y_1, \ldots, Y_n)'$ and $\alpha_n = (\alpha_1, \ldots, \alpha_n)'$.

$$L(\psi) = \int p(y_n \mid \alpha_n) p(\alpha_n) d\alpha_n$$

$$= \int \frac{p(y_n \mid \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} g(y_n, \alpha_n; \psi_0) d\alpha_n$$

$$= \int \frac{p(y_n \mid \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} g(\alpha_n \mid y_n; \psi_0) g(y_n; \psi_0) d\alpha_n$$

$$\frac{L(\psi)}{L_g(\psi_0)} = \int \frac{p(y_n \mid \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} g(\alpha_n \mid y_n; \psi_0) d\alpha_n$$

Importance Sampling (cont)

$$\frac{L(\psi)}{L_g(\psi_0)} = \int \frac{p(y_n \mid \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} g(\alpha_n \mid y_n; \psi_0) d\alpha_n$$

$$= E_g\left[\frac{p(y_n \mid \alpha_n) p(\alpha_n)}{g(y_n, \alpha_n; \psi_0)} \mid y_n; \psi_0) \right]$$

$$\sim \frac{1}{N} \sum_{j=1}^{N} \frac{p(y_n \mid \alpha_n^{(j)}) p(\alpha_n^{(j)})}{g(y_n, \alpha_n^{(j)}; \psi_0)},$$

where $\{\alpha_n^{(j)}; j = 1, \ldots, N\} \sim \text{iid } g(\alpha_n \mid y_n; \psi_0)$.

Notes:

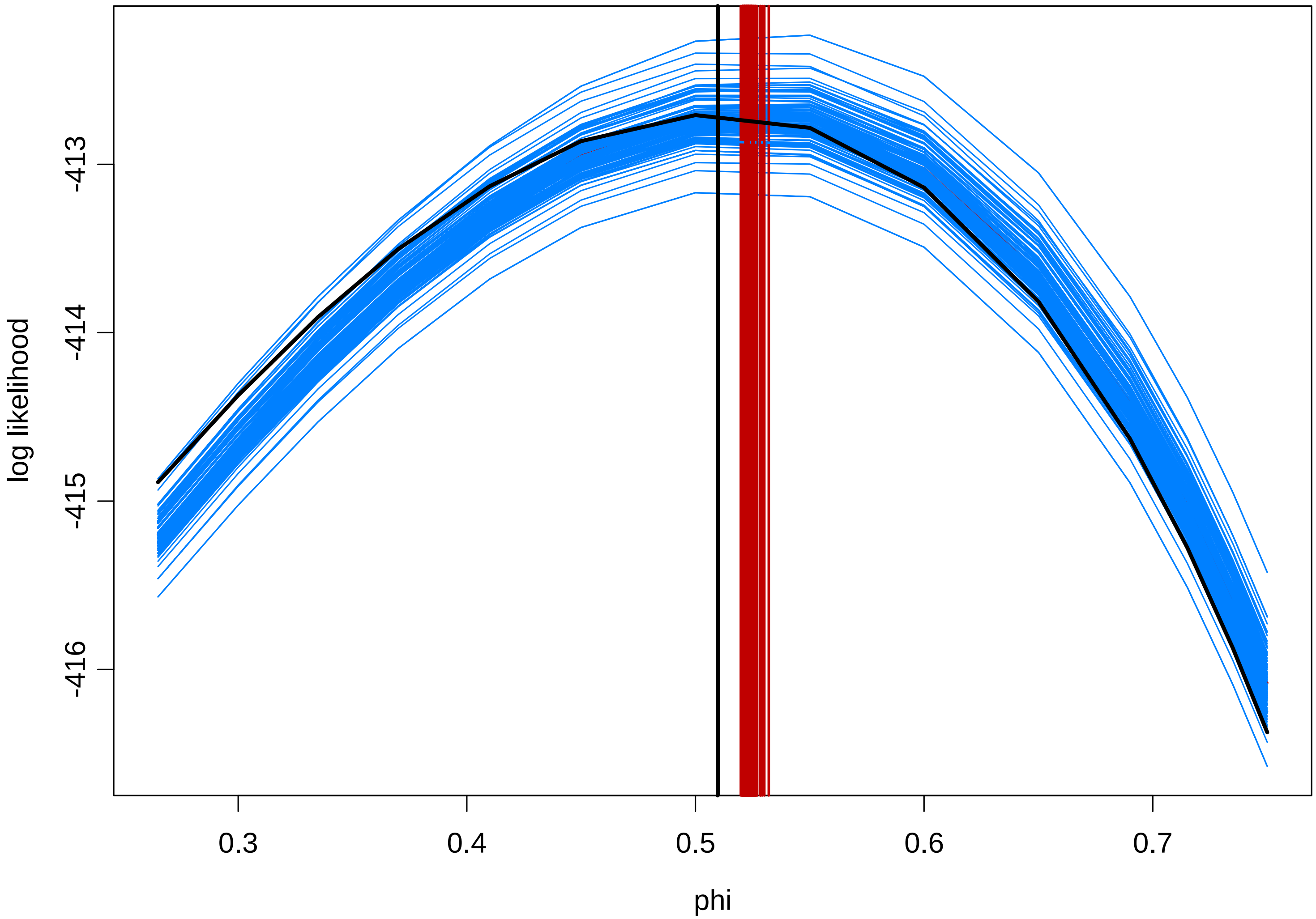
- This is a "one-sample" approximation to the relative likelihood. That is, for one realization of the α's, we have, in principle, an approximation to the whole likelihood function.

- Approximation is only good in a neighborhood of ψ_0. Geyer suggests maximizing ratio wrt ψ and iterate replacing ψ_0 with $\hat{\psi}$.

Importance Sampling — example

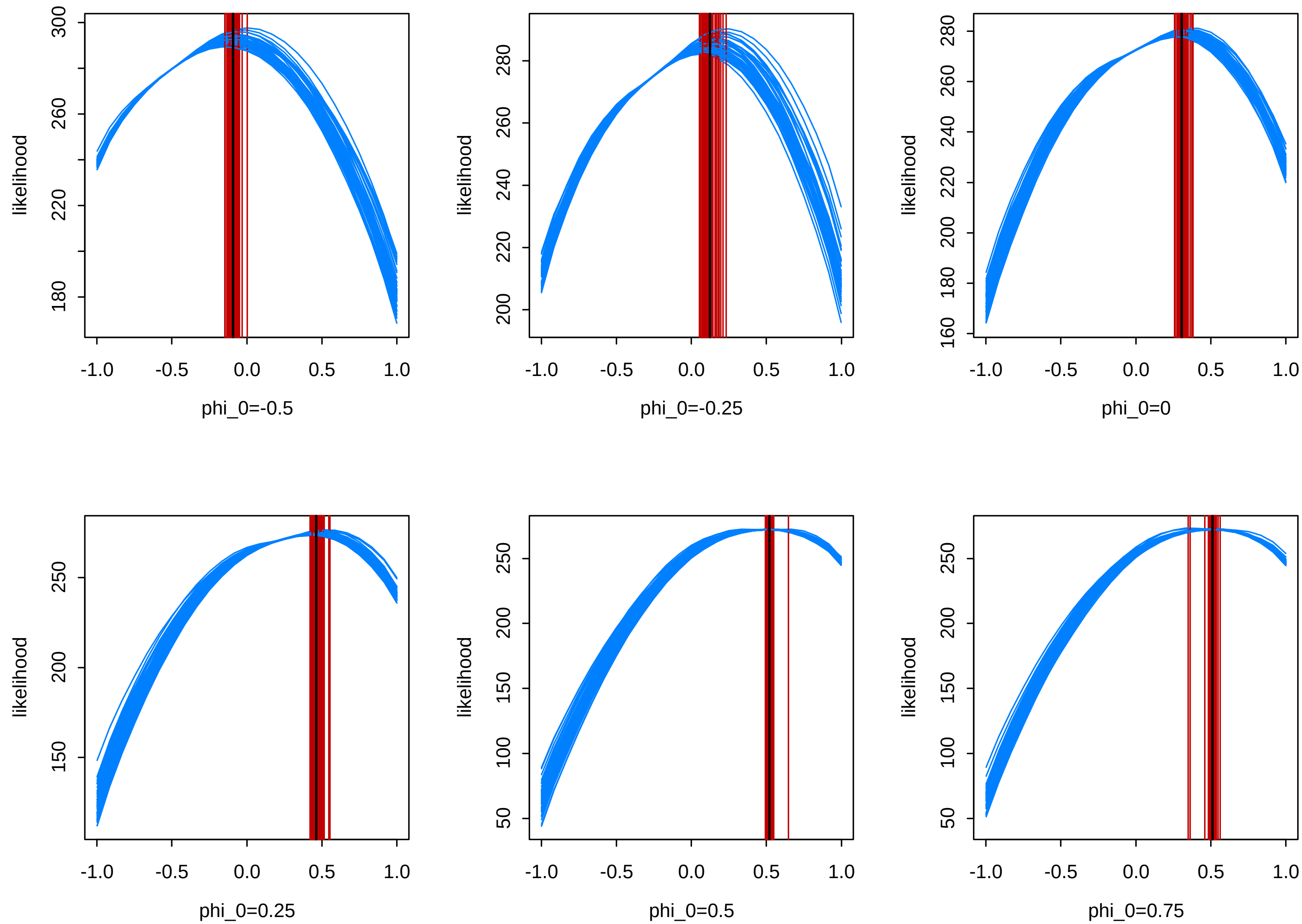
Simulation example: $Y_t \mid \alpha_t \sim Pois(\exp(.7 + \alpha_t))$,

$\alpha_t = .5\,\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\}$~IID N(0, .3), $n = 200$, $N = 1000$



Importance Sampling — example

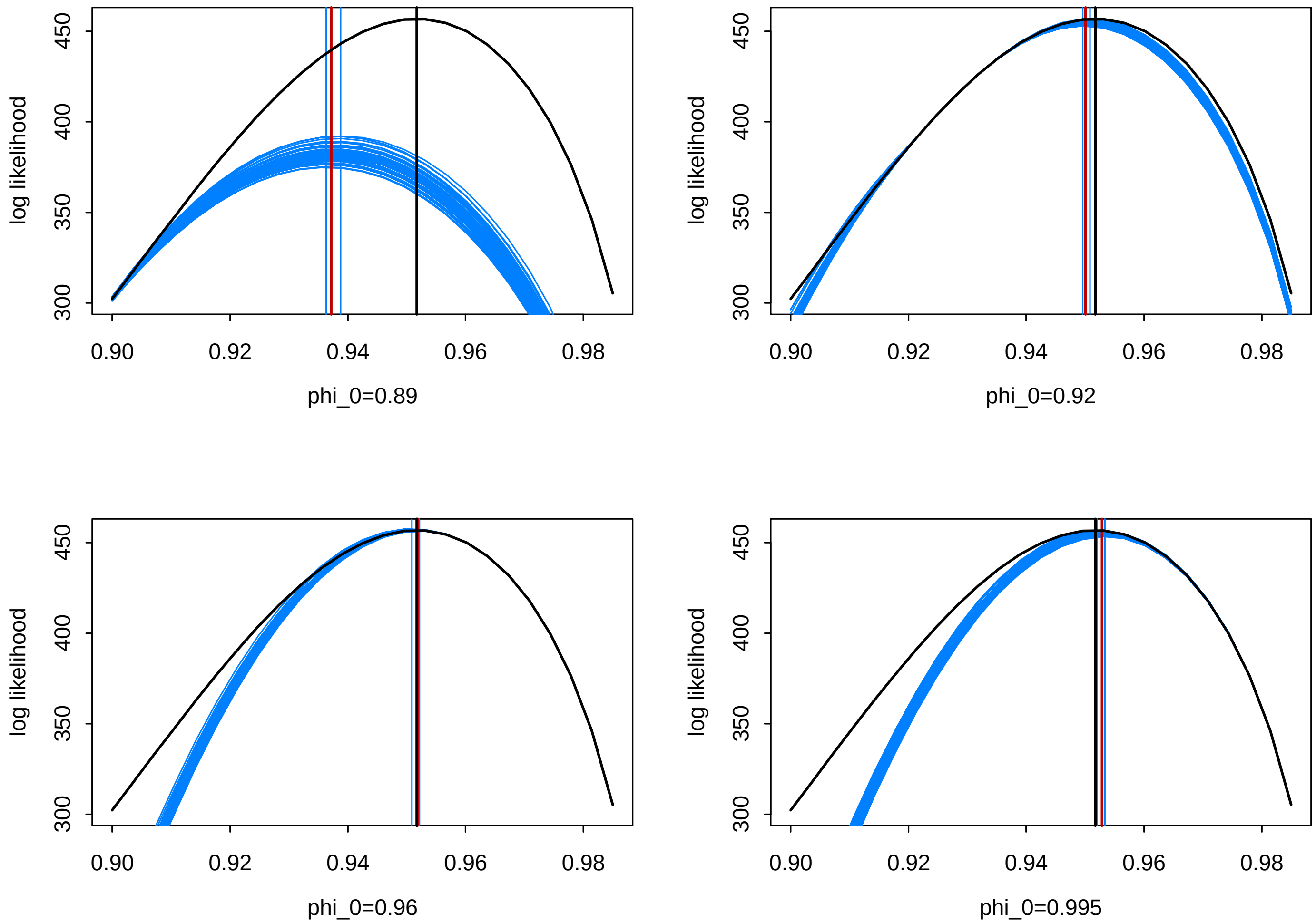
Simulation example: $\beta = .7$, $\phi = .5$, $\sigma^2 = .3$, $n = 200$, $N = 1000$, 50 realizations plotted



Importance Sampling — example

SV process: $Y_t \mid \alpha_t \sim N(0, \exp\{2\alpha_t\})$,

$$\alpha_t = -.368 + .95\,\alpha_{t-1} + \varepsilon_t\,, \quad \{\varepsilon_t\} \sim \text{IID } N(0, .0676),\ n = 200,\ N = 1000$$



Estimation Methods — Approximation to the likelihood

General setup:

$$p(y_n, \alpha_n) \propto p(y_n \mid \alpha_n) \det(G_n)^{1/2} \exp\{-(\alpha_n - \mu)^T G_n (\alpha_n - \mu)/2\}$$

where

$$G_n^{-1} = E(\alpha_n - \mu)^T (\alpha_n - \mu)$$

Likelihood:

$$L(\psi) = \int p(y_n \mid \alpha_n) p(\alpha_n) d\alpha_n$$

Consider a Gaussian approximation $p_a(\alpha_n \mid y_n) = \phi(\alpha_n ; \mu_0 , \Sigma_0)$ to the posterior

$$p(\alpha_n \mid y_n) \propto p(\alpha_n \mid y_n)\, p(\alpha_n)$$

Setting equal the respective posterior modes α_a^* and α^* of $p_a(\alpha_n \mid y_n)$ and $p(\alpha_n \mid y_n)$, we have $\mu_0 = \alpha^*$, where α^* is the solution of the equation

$$\frac{\partial}{\partial \alpha_n} \log p(y_n \mid \alpha_n, \psi) - G_n(\alpha_n - \mu) = 0$$

Estimation Methods — Approximation to the likelihood (cont)

Matching Fisher information matrices:

$$\Sigma_0 = \left(-\frac{\partial^2}{\partial\alpha\partial\alpha^T} \log p(\mathrm{y}_n \mid \alpha_n, \psi)\Big|_{\alpha_n=\alpha^*} + G_n \right)^{-1}$$

Approximating posterior:

$$p_a(\alpha_n \mid \mathrm{y}_n, \psi) = \phi(\alpha_n, \alpha^*, \left(-\frac{\partial^2}{\partial\alpha\partial\alpha^T} \log p(\mathrm{y}_n \mid \alpha_n, \psi)\Big|_{\alpha_n=\alpha^*} + G_n \right)^{-1}$$

Notes:

1. This approximating posterior is identical to the importance sampling density used by Durbin and Koopman.
2. In traditional Bayesian setting, posterior is approximately p_a for large n (see Bernardo and Smith, 1994).
3. Obtain same result if one applies a Taylor series expansion to the joint likelihood and ignore terms of order > 2.

Estimation Methods — Approximation to the likelihood (cont)

Approximate likelihood: Note that

$$p(\alpha_n \mid y_n) = \frac{p(y_n \mid \alpha_n)\, p(\alpha_n)}{L(\psi; y_n)},$$

which by solving for L in the expression,

$$p_a(\alpha_n^* \mid y_n, \psi) = p(\alpha_n^* \mid y_n, \psi),$$

we obtain

$$L_a(\psi; y_n) = p(y_n \mid \alpha^*, \psi)\, p(\alpha^*, \psi) / p_a(\alpha^* \mid y_n, \psi)$$

$$= \frac{|G_n|^{1/2}\, p(y_n \mid \alpha^*, \psi) \exp\{-(\alpha^* - \mu)^T G_n (\alpha^* - \mu)/2\}}{\det\left(-\frac{\partial^2}{\partial\alpha\partial\alpha^T} \log p(y_n \mid \alpha_n, \psi)\Big|_{\alpha^*} + G_n \right)^{1/2}}$$

Estimation Methods — Approximation to the likelihood (cont)

Case of exponential family:

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

where

$$K = \mathrm{diag}\{\frac{\partial^2}{\partial \alpha_t^2} b_t(\alpha_t)\Big|_{\alpha_t^*}\},$$

and α^* is the solution to the equation

$$\mathrm{y}_n - \frac{\partial}{\partial \alpha_n} \mathrm{b}(\alpha_\mathrm{n}) - \mathrm{G_n}(\alpha_\mathrm{n} - \mu) = 0.$$

Using a Taylor expansion, the latter equation can be solved iteratively.

Estimation Methods — Approximation to the likelihood

Implementation:

1. Let $\alpha^* = \alpha^*(\psi)$ be the converged value of $\alpha^{(j)}(\psi)$, where

$$\alpha^{(j+1)}(\psi) = (\ddot{b}^j + G_n)^{-1} \tilde{y}_n^j(\psi),$$

and

$$\tilde{y}_n^j = y_n - \dot{b}^j + \ddot{b}^j \alpha^{(j)} + G_n \mu.$$

2. Maximize $L_a(\psi; y_n)$ with respect to ψ.

Simulation Results

Model: $Y_t \mid \alpha_t \sim Pois(\exp(.7 + \alpha_t))$, $\alpha_t = .5\,\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim$ IID N(0, .3), $n = 200$

Estimation methods:

- Importance sampling (N=1000, ψ_0 updated a maximum of 10 times)

	beta	phi	sigma2
mean	0.6982	0.4718	0.3008
std	**0.1059**	**0.1476**	**0.0899**

- Approximation to likelihood

	beta	phi	sigma2
mean	0.7036	0.4579	0.2962
std	**0.0951**	**0.1365**	**0.0784**

Simulation Results

Stochastic volatility model:

$Y_t = \sigma_t Z_t$, $\{Z_t\}$~IID N(0,1)

$\alpha_t = \gamma + \phi\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\}$~IID N(0,σ^2), where $\alpha_t = 2 \log \sigma_t$; n=1000, NR=500

CV=10

	True	AL	RMSE	IS	RMSE
γ	–.411	–.491	.210	–.490	.216
ϕ	0.950	0.940	.025	0.940	.026
σ	0.484	**0.478**	.065	**0.481**	.073

CV=1

	True	AL	RMSE	IS	RMSE
γ	–.368	–.499	.341	–.485	.324
ϕ	0.950	0.932	.046	0.934	.043
σ	0.260	**0.270**	.068	**0.268**	.068

Is the posterior distribution close to normal?

Compare posterior mean with posterior mode: Can compute the posterior mean using *SIR* (sampling importance-resampling) or particle filtering.

Posterior mode: The mode of $p(\alpha_n \mid y_n)$ is α^* found at the last iteration of AL.

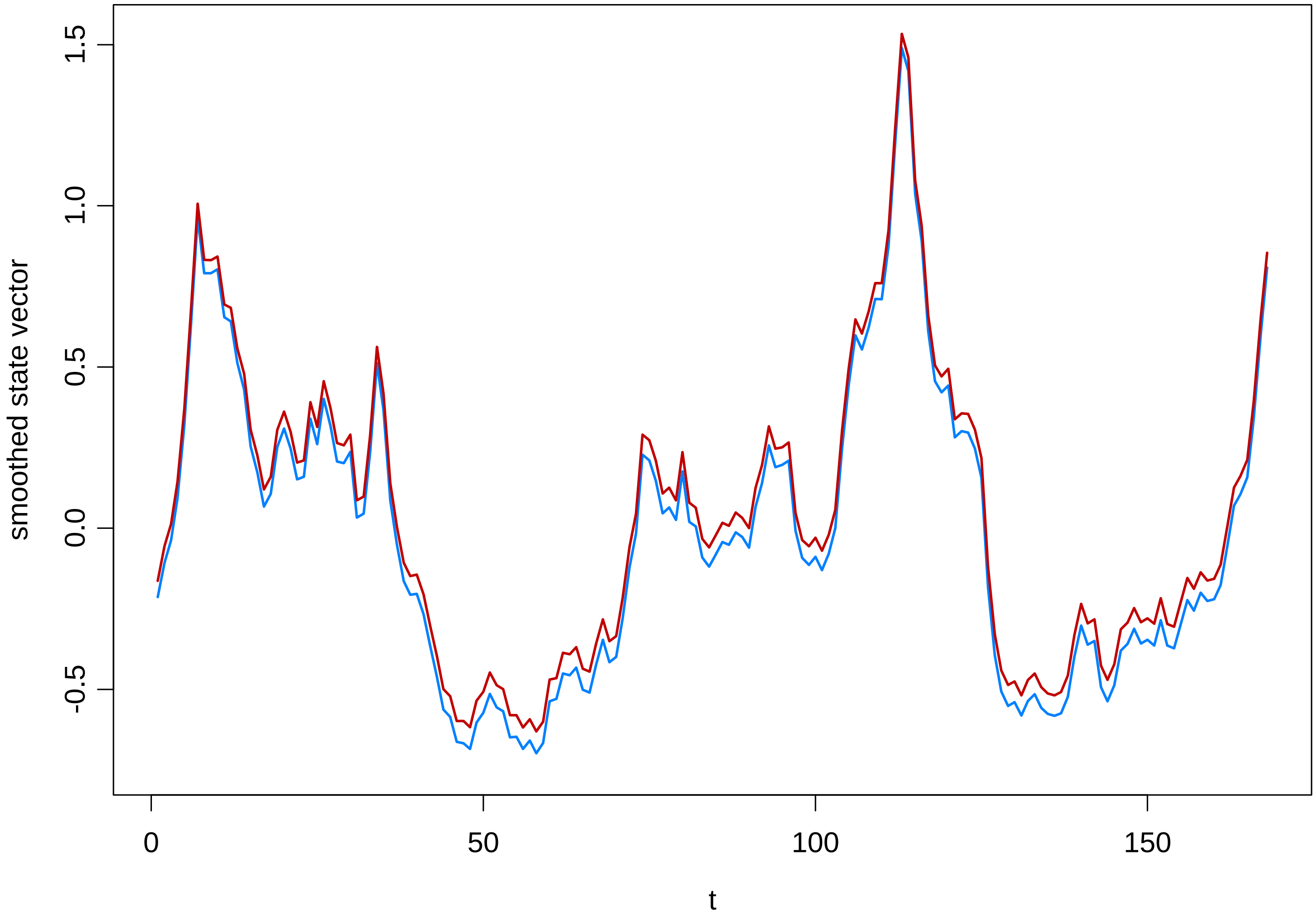
Posterior mean: The mean of $p(\alpha_n \mid y_n)$ can be found using SIR.

Let $\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(N)}$ be independent draws from the multivariate distr $p_a(\alpha_n \mid y_n)$. For N large, an approximate iid sample from $p(\alpha_n \mid y_n)$ can be obtained by drawing a random sample from $\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(N)}$ with probabilities

$$p_i = \frac{w_i}{\sum_{i=1}^{N} w_i}, \quad w_i = \frac{p(\alpha^{(i)} \mid y_n)}{p_a(\alpha^{(i)} \mid y_n)} \propto \frac{L(\psi; y_n, \alpha^{(i)})}{p_a(\alpha^{(i)} \mid y_n)}, \quad i = 1, \ldots, N.$$

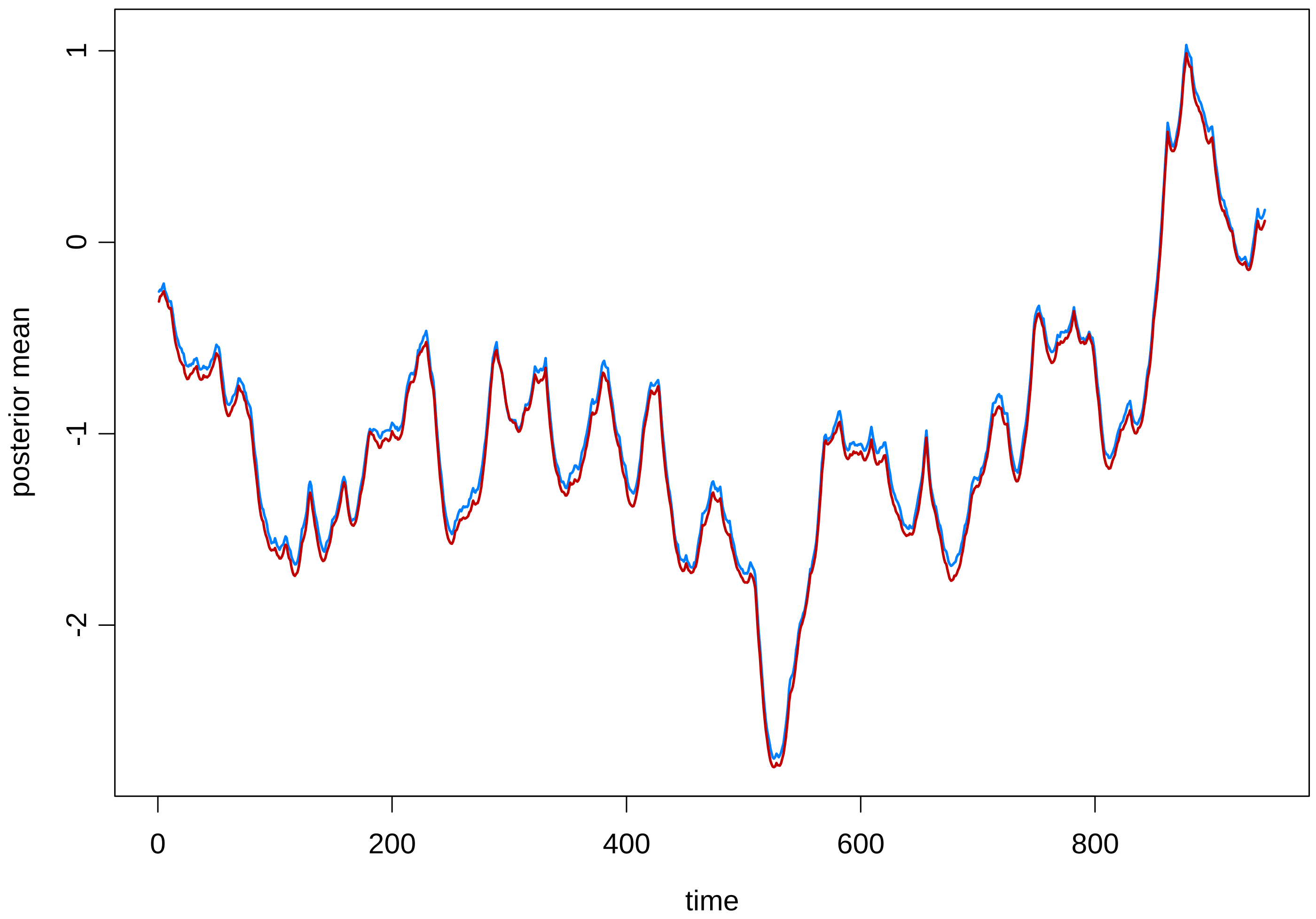
Posterior mean vs posterior mode?

Polio data: blue = mean, red = mode



Posterior mean vs posterior mode?

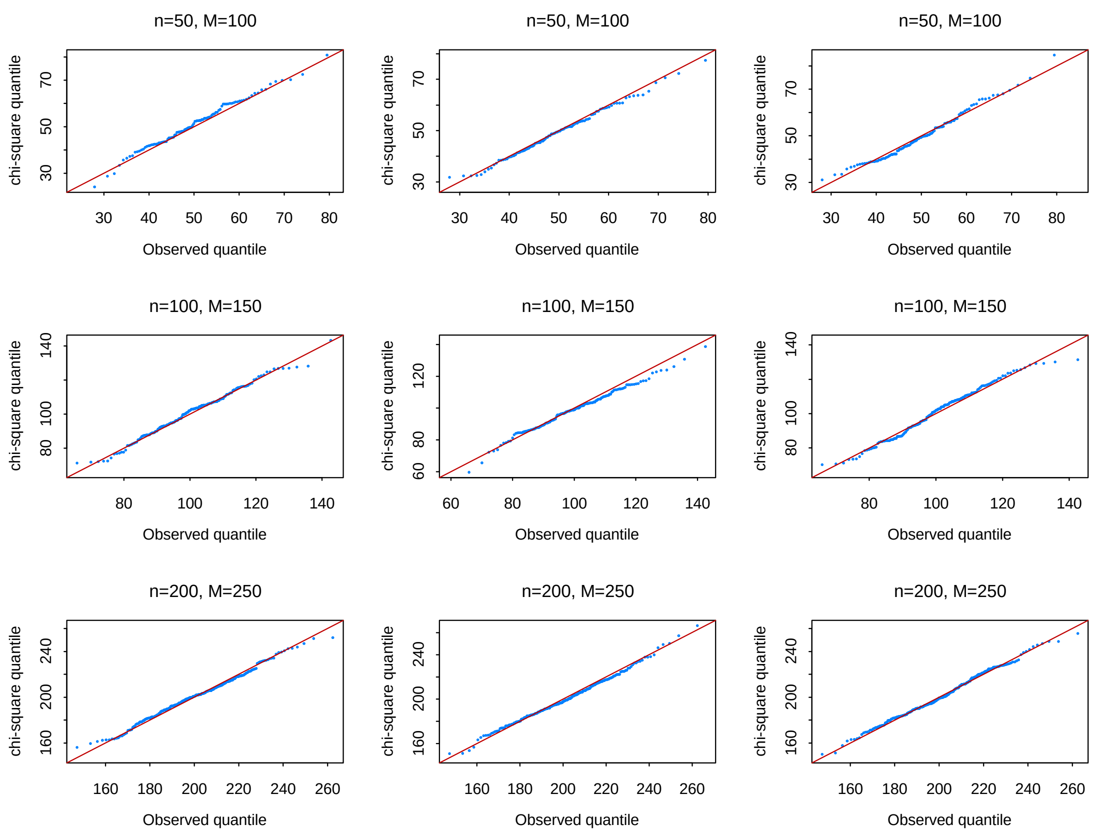
Pound/US exchange rate data: blue = mean, red = mode



Is the posterior distribution close to normal?

Suppose $\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(M)}$ are independent draws from the multivariate distr $p(\alpha_n \mid y_n)$, generated using SIR. Then

$$d_j^2 = (\alpha^{(j)} - \alpha^*)^T (K + G_n)(\alpha^{(j)} - \alpha^*) \overset{iid}{\sim} \chi_n^2$$



Correlations are *not significant*.

Application to Structural Breaks (Davis, Lee, Rodriguez-Yam)

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 Structural Breaks—(cont)

Estimation: For $(m, \tau_1, \ldots, \tau_m)$ fixed, calculate the approximate likelihood evaluated at the "MLE", i.e.,

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

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.

Goal: Optimize an *objective function* over $(m, \tau_1, \ldots, \tau_m)$.

Implementation problems:

- choice of objective function?
- how to optimize over $(m, \tau_1, \ldots, \tau_m)$?

Solutions(?):

- use minimum description length (MDL) as an objective function
- use genetic algorithm for optimization

Application to Structural Breaks—(cont)

Minimum Description Length (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 store the data).

Code Length("data") = CL("fitted model") + CL("data | fitted model")

~ CL("parameters") + CL("residuals")

$$MDL(m, \tau_1, \ldots, \tau_m) = \underbrace{\log(m) + m\log(n) + 1.5\sum_{j=1}^{m}\log(\tau_j - \tau_{j-1})}_{CL(\text{"Parameters"})} - \underbrace{\sum_{j=1}^{m}\log(L_a(\hat{\psi}_j; y_{\tau_{j-1}:\tau_j}))}_{CL(\text{"residuals"})}$$

Generalization: AR(p) segments can have unknown order.

$$MDL(m, (\tau_1, p_1), \ldots, (\tau_m, p_m)) = \log(m) + m\log(n) + 0.5\sum_{j=1}^{m}(p_j + 2)\log(\tau_j - \tau_{j-1}) - \sum_{j=1}^{m}\log(L_a(\hat{\psi}_j; y_{\tau_{j-1}:\tau_j}))$$

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

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

Where

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

For example,

c = (2, -1, -1, -1, -1, 0, -1, -1, -1, -1, 0, -1, -1, -1, -1, 3, -1, -1, -1, -1)

t: 1 6 11 16

would correspond to a process as follows:

AR(2), t=1:5; AR(0), t=6:10; AR(0), t=11:15; AR(3), t=16:20

Implementation of Genetic Algorithm—(cont)

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

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 $M(c_i)$.
  - k^{th} gene of child is $\delta_k = \delta_{i,k}$ w.p. 1/2 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. α_p

Implementation of Genetic Algorithm—(cont)

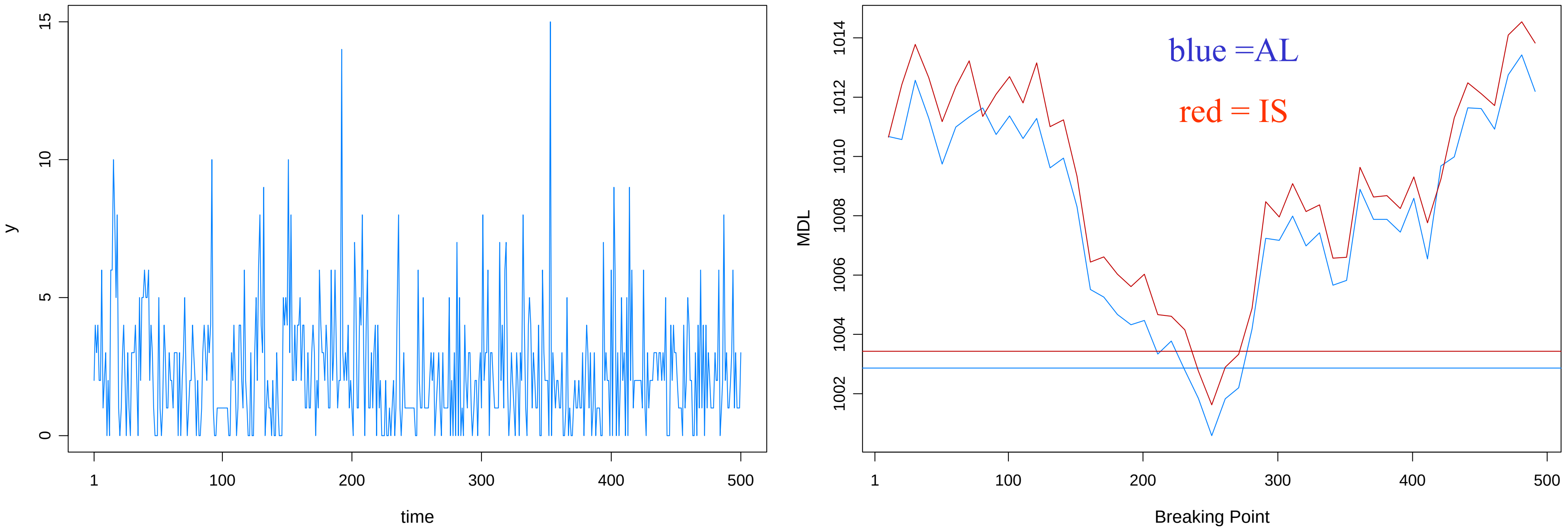
Execution of GA: Run GA for numerous generations or until *convergence*.

Various Strategies:

- include the *top* ten chromosomes from past generations in future generations.
- use multiple *islands*, in which populations run independently, and then allow *migration* after a fixed number of generations.

Count Data Example

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

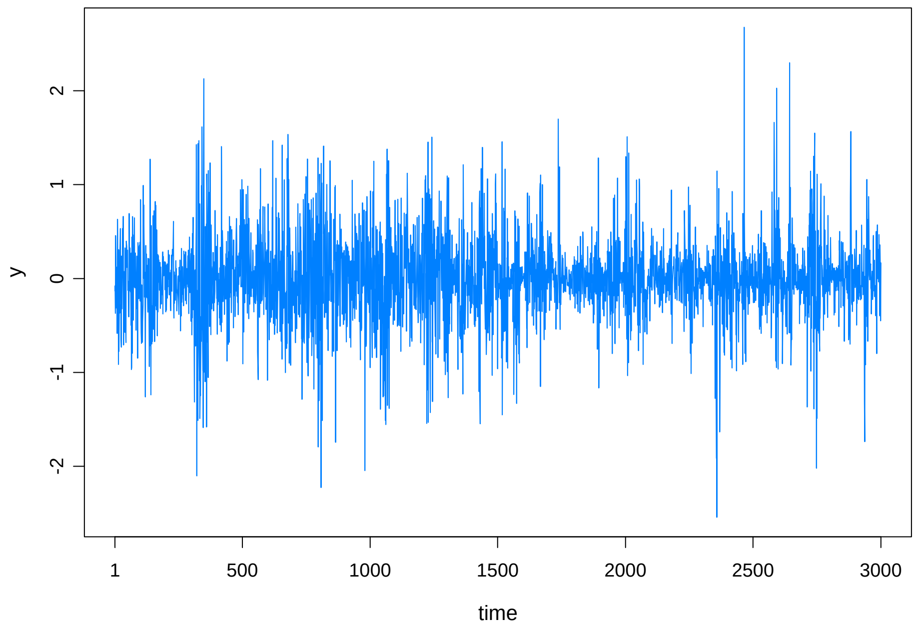


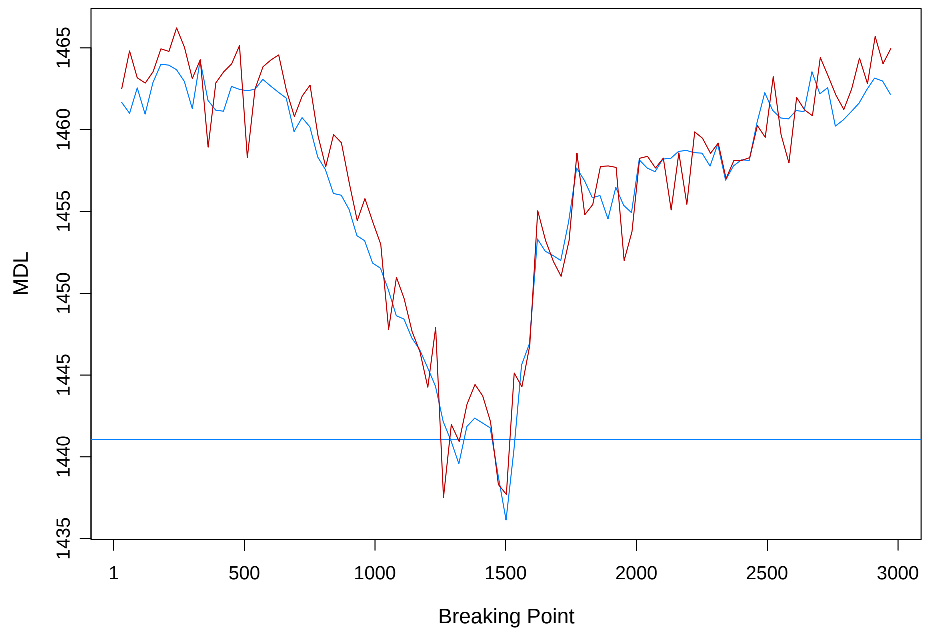
True model:

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

SV Process Example

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



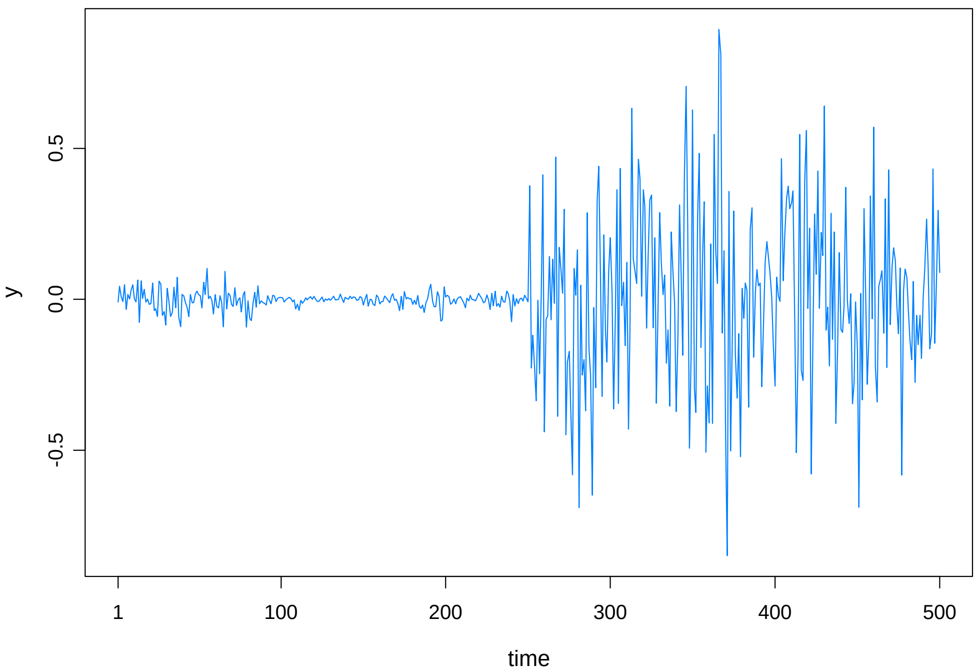


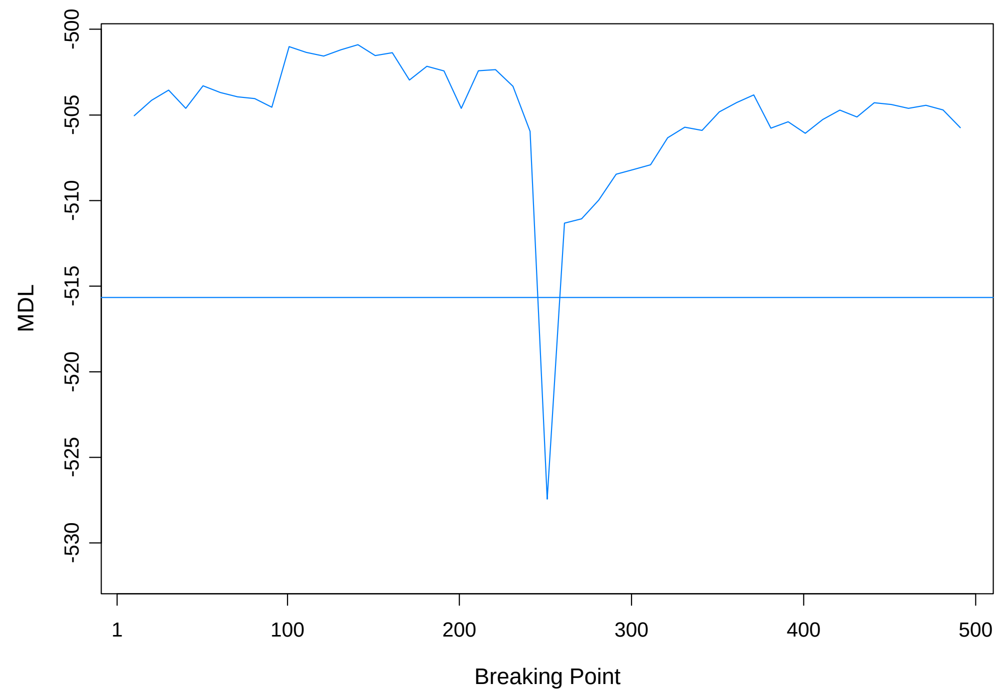
True model:

- $Y_t \mid \alpha_t \sim N(0, \exp\{2\alpha_t\})$, $\alpha_t = -.05 + .975\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim \text{IID } N(0, .05)$, $t < 1500$
- $Y_t \mid \alpha_t \sim N(0, \exp\{2\alpha_t\})$, $\alpha_t = -.25 + .900\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim \text{IID } N(0, .25)$, $t > 1500$.
- GA estimate 1502, time 1049secs

SV Process Example

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





True model:

- $Y_t \mid \alpha_t \sim N(0, \exp\{2\alpha_t\})$, $\alpha_t = -.175 + .977\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim$ IID $N(0, .1810)$, $t < 251$
- $Y_t \mid \alpha_t \sim N(0, \exp\{2\alpha_t\})$, $\alpha_t = -.010 + .996\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\} \sim$ IID $N(0, .0089)$, $t > 250$.
- GA estimate 251, time 269s

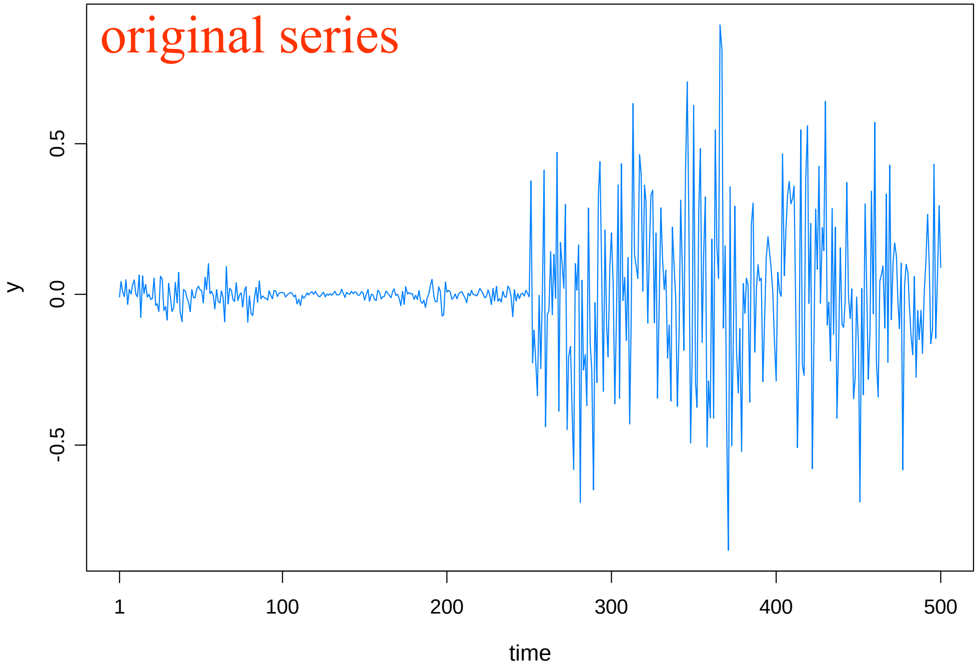
SV Process Example-(cont)

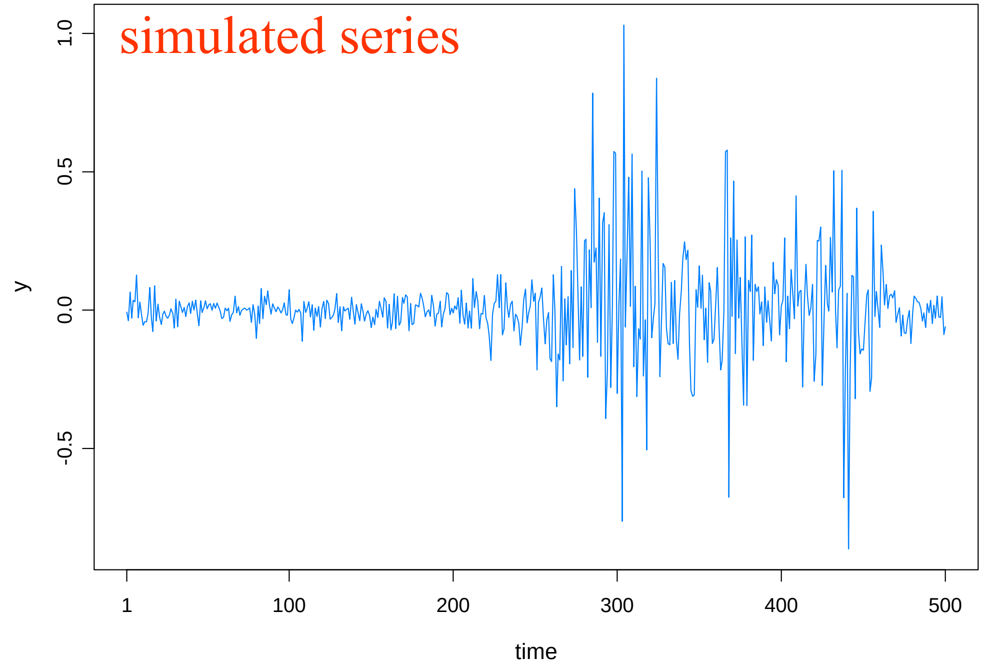
True model:

- $Y_t \mid \alpha_t \sim N(0, \exp\{2\alpha_t\})$, $\alpha_t = -.175 + .977\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\}\sim$IID N(0, .1810), $t < 251$
- $Y_t \mid \alpha_t \sim N(0, \exp\{2\alpha_t\})$, $\alpha_t = -.010 + .996\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\}\sim$IID N(0, .0089), $t > 250$.

Fitted model based on no structural break:

- $Y_t \mid \alpha_t \sim N(0, \exp\{2\alpha_t\})$, $\alpha_t = -.0645 + .9889\alpha_{t-1} + \varepsilon_t$, $\{\varepsilon_t\}\sim$IID N(0, .0935)

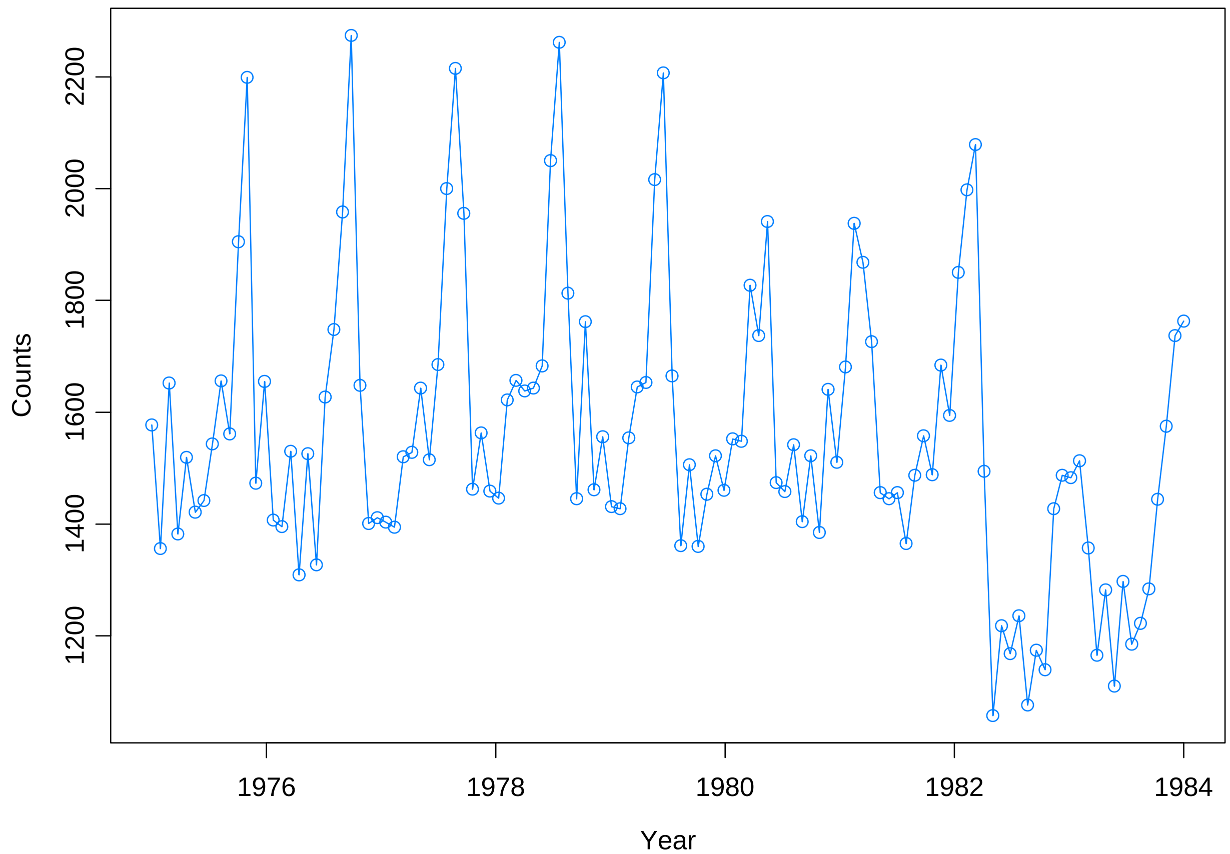




Linear Process Example (Monthly Deaths & Serious Injuries, UK)

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

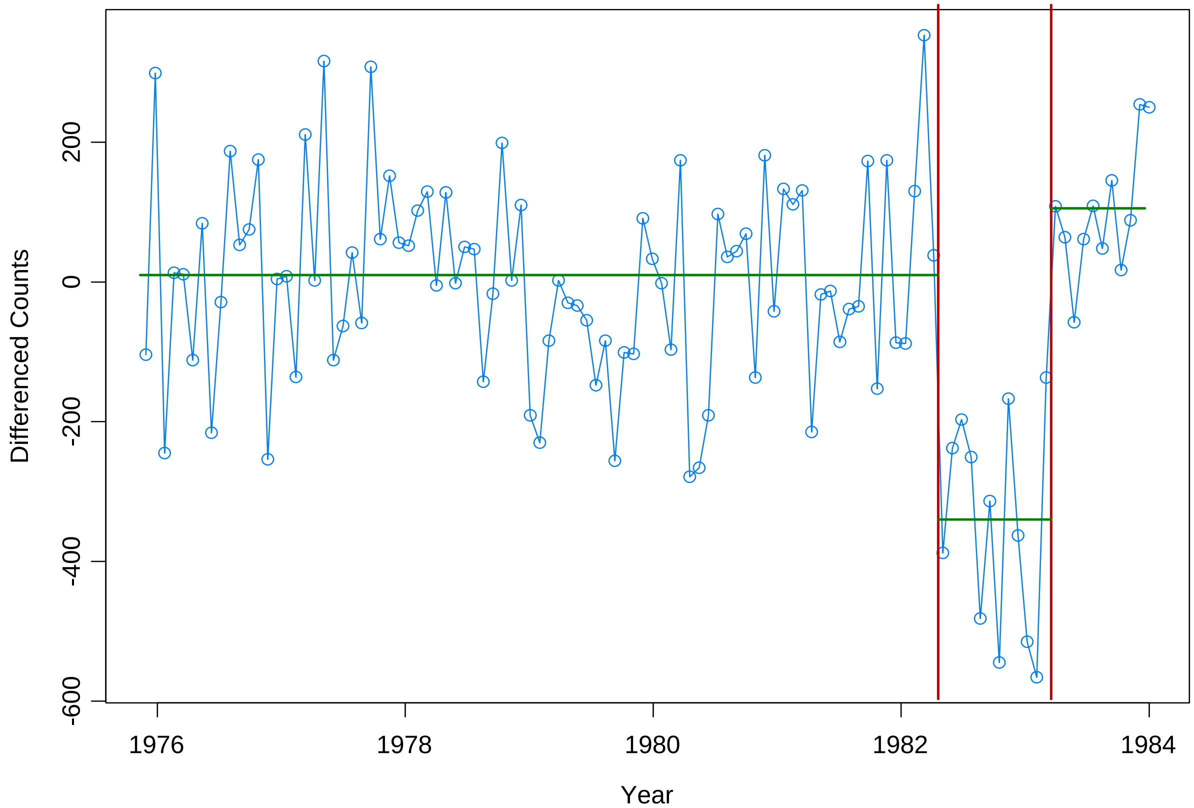
Remark: Seat belt legislation introduced in Feb `83 (t = 99).



Linear Process 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)

Summary Remarks

1. Importance sampling offers a nice clean method for estimation in parameter driven models.

2. Relative likelihood approach is a one-sample based procedure, but may have convergence problems.

3. Approximation to the likelihood is a non-simulation based procedure which may have great potential especially with large sample sizes and/or large number of explanatory variables.

5. Approximate likelihood approach is amenable to bootstrapping procedures for bias correction.

6. Posterior mode matches posterior mean reasonably well.

7. Approximate likelihood approach may be useful to the problem of structural break detection.