Chapter 2

Structural breaks for models *without path dependence*

Chapter 2

Motivation

Motivation

- So far : Fixed parameters over time

  → Unlikely to hold for long time series

  → Many policy changes and turbulent financial periods
  - Should affect the dynamics of the series

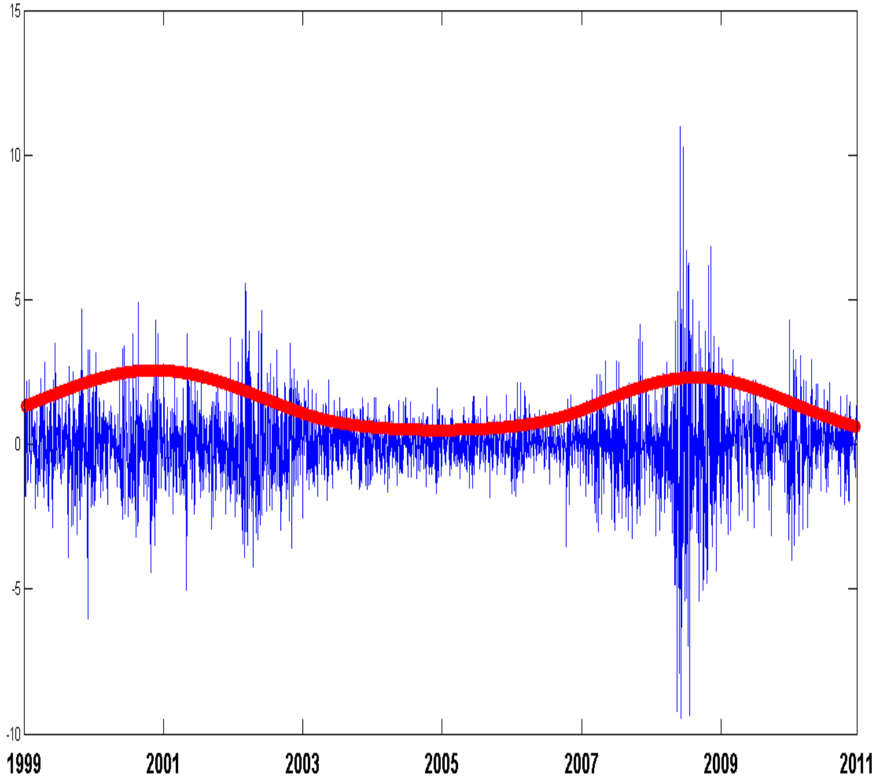
- Stylized fact of many time series
  - High persistence – almost integrated series

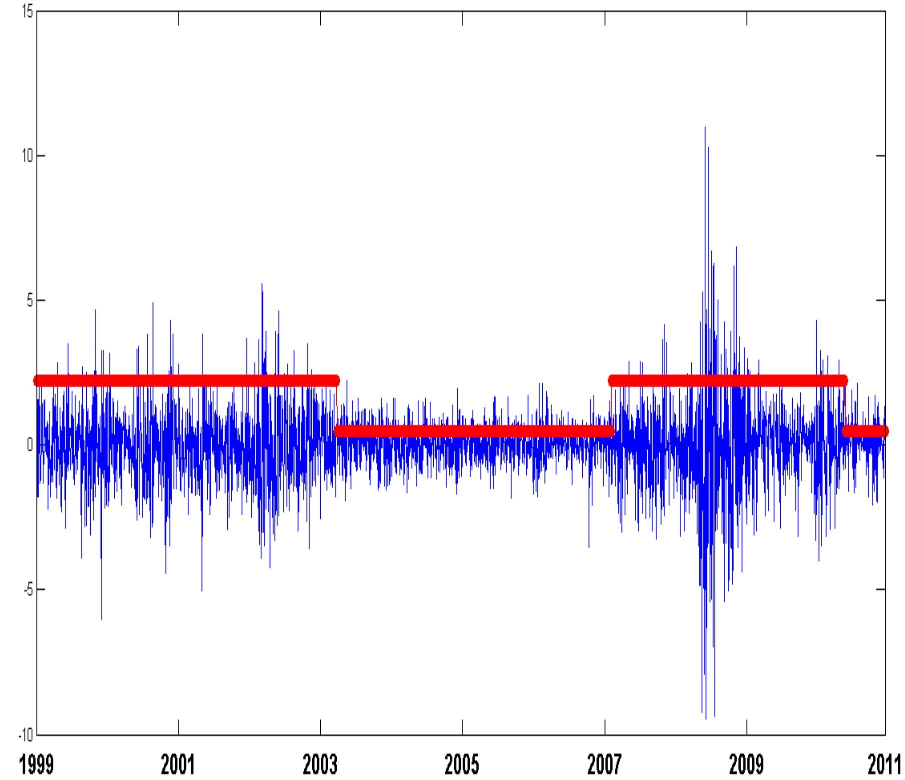
  → Unit root model → No predictability !
  - Long-run dynamic evolves over time

Structural breaks can cause these stylized facts

Long run dynamic : S&P 500

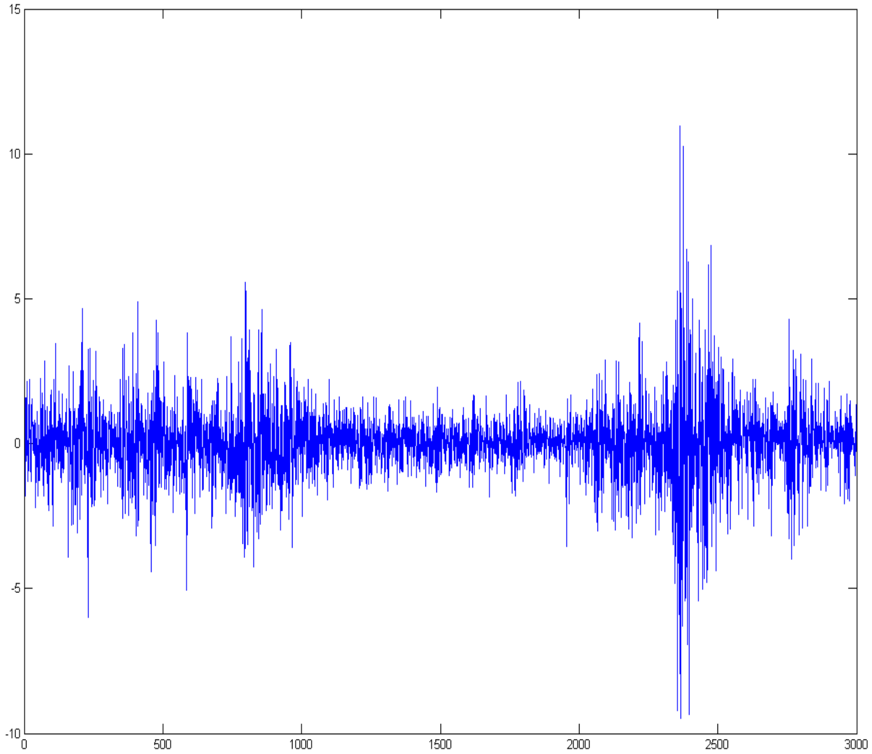
Spline GARCH (Engle, Rangel, 2013)	Markov-Switching GARCH (Bauwens,Dufays,Rombouts, 2013)

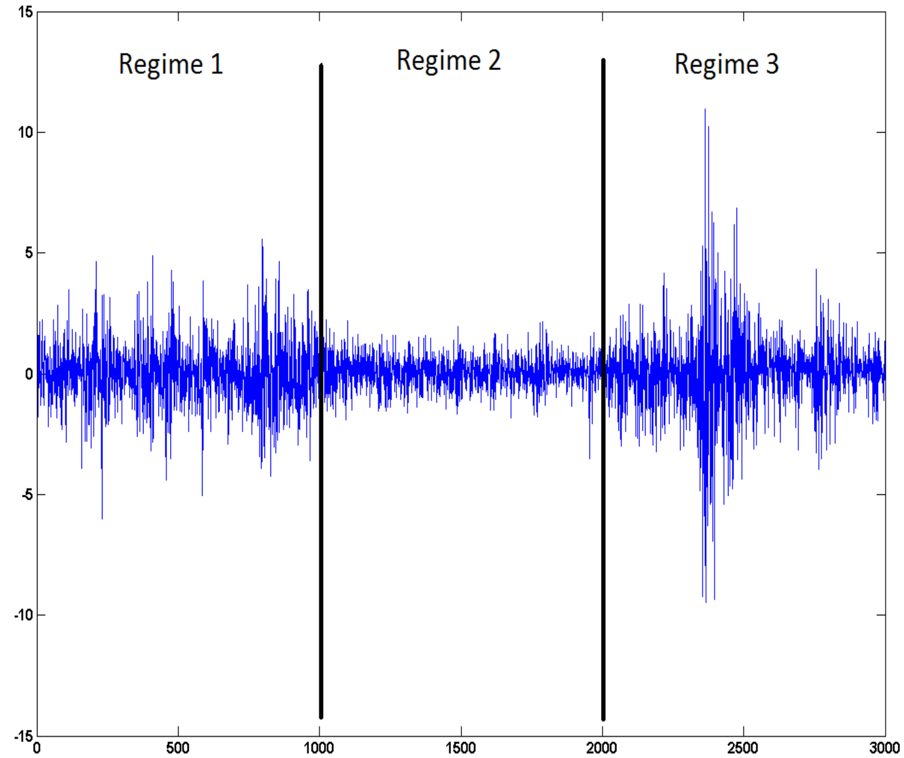




Long run volatility evolves over time

Example : S&P 500





$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1} + \beta\sigma_{t-1}^2$$

No parameter dynamic

$$\alpha + \beta = 0.99$$

Almost integrated

$$\sigma_t^2 = \omega_i + \alpha_i\epsilon_{t-1} + \beta_i\sigma_{t-1}^2$$

Less persistent

$$\begin{cases} \alpha_1 + \beta_1 = 0.95 \\ \alpha_2 + \beta_2 = 0.95 \\ \alpha_3 + \beta_3 = 0.99 \end{cases}$$

SB : Motivation

Why detecting structural breaks is relevant ?

- ***Historical analysis***

  → Better understanding of the time series dynamics

- ***Detection of instabilities***

  → Useful for systemic risk

- ***Forecasts***

  → Automatically select the optimal window size

  → Parameters adapted over time

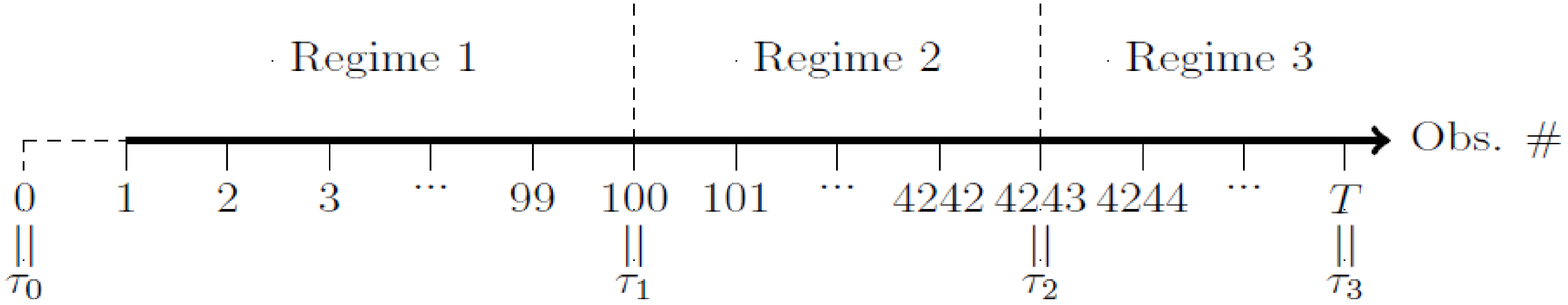
Change-point models

Change-point models

CP models :

- ***Non recurrent regimes***

  → Not stationary – In line with economic theories



- ***Forecasts***

  → Automatically select the optimal window size

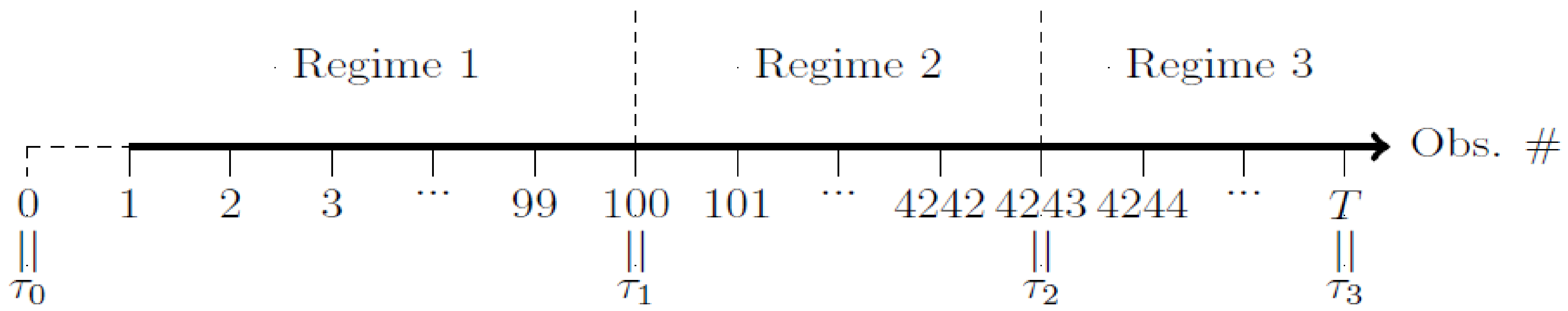
  → Predictions based on the last sub-sample

Change-point models

CP models : first attempt to model structural breaks

- *Chernoff and Zacks (1964) ; Carlin, Gelfand and Smith (1992) ; Stephens (1994)*

→ Modelling SB as discrete parameters to be estimated



Carlin, Gelfand and Smith (1992)

- Inference on one structural break (in mean or variance)
  - Estimation carried out by Gibbs sampler

- The model :

$$\begin{aligned} y_t|Y_{1:t-1} &\sim N(\theta_1'[1 \quad y_{t-1}]', \sigma_1^2) \text{ if } \tau_1 < t \\ &\sim N(\theta_2'[1 \quad y_{t-1}]', \sigma_2^2) \text{ otherwise} \end{aligned}$$

- The set of parameters : $\Sigma = \{\sigma_1^2, \sigma_2^2\}$ and $\tau_1 \in [1, T-1]$
  $\Theta = \{\theta_1, \theta_2\}$

- Prior distributions :

$$\begin{cases} \sigma_i^2 \sim IG(\alpha, \beta) \quad i \in [1,2] \\ \theta_i \sim N(\mu_0, \Sigma_0) \quad i \in [1,2] \\ \tau_1 \sim U(1, T-1) \end{cases}$$

Carlin, Gelfand and Smith (1992)

- Gibbs sampler :

$$\begin{aligned}
\theta_1|Y_{1:T},\tau_1,\theta_2,\sigma_1^2,\sigma_2^2 &\sim N(\bar{\mu}_1,\bar{\Sigma}_1)\\
\sigma_1^2|Y_{1:T},\tau_1,\theta_1,\theta_2,\sigma_2^2 &\sim IG(\alpha+\tau_1/2,\beta+\sum_{t=1}^{\tau_1}\epsilon_t)\\
\theta_2|Y_{1:T},\tau_1,\theta_1,\sigma_1^2,\sigma_2^2 &\sim N(\bar{\mu}_2,\bar{\Sigma}_2)\\
\sigma_2^2|Y_{1:T},\tau_1,\theta_1,\theta_2,\sigma_1^2 &\sim IG(\alpha+(T-\tau_1)/2,\beta+\sum_{t=\tau_1+1}^{T}\epsilon_t)\\
\tau_1|Y_{1:T},\theta_1,\theta_2,\sigma_1^2,\sigma_2^2 &\sim \text{Griddy-Gibbs}
\end{aligned}$$

with
$$\begin{cases}
\bar{\mu}_1=\bar{\Sigma}_1[\sigma_1^{-2}\sum_{t=1}^{\tau_1}x_ty_t+\Sigma_0^{-1}\mu_0]\\
\bar{\Sigma}_1=[\sigma_1^{-2}\sum_{t=1}^{\tau_1}(x_tx_t')+\Sigma_0^{-1}]^{-1}
\end{cases}$$

and
$$\begin{cases}
\bar{\mu}_2=\bar{\Sigma}_2[\sigma_2^{-2}\sum_{t=\tau_1+1}^{T}x_ty_t+\Sigma_0^{-1}\mu_0]\\
\bar{\Sigma}_2=[\sigma_2^{-2}\sum_{t=\tau_1+1}^{T}(x_tx_t')+\Sigma_0^{-1}]^{-1}
\end{cases}$$

Carlin, Gelfand and Smith (1992)

- Griddy-Gibbs :

$$\pi(\tau_1 = i | Y_{1:T}, \Theta, \Sigma) \propto f(Y_{1:T} | \Theta, \Sigma, \tau_1 = i) f(\tau_1 = i) \quad \forall i \in [1, T-1]$$

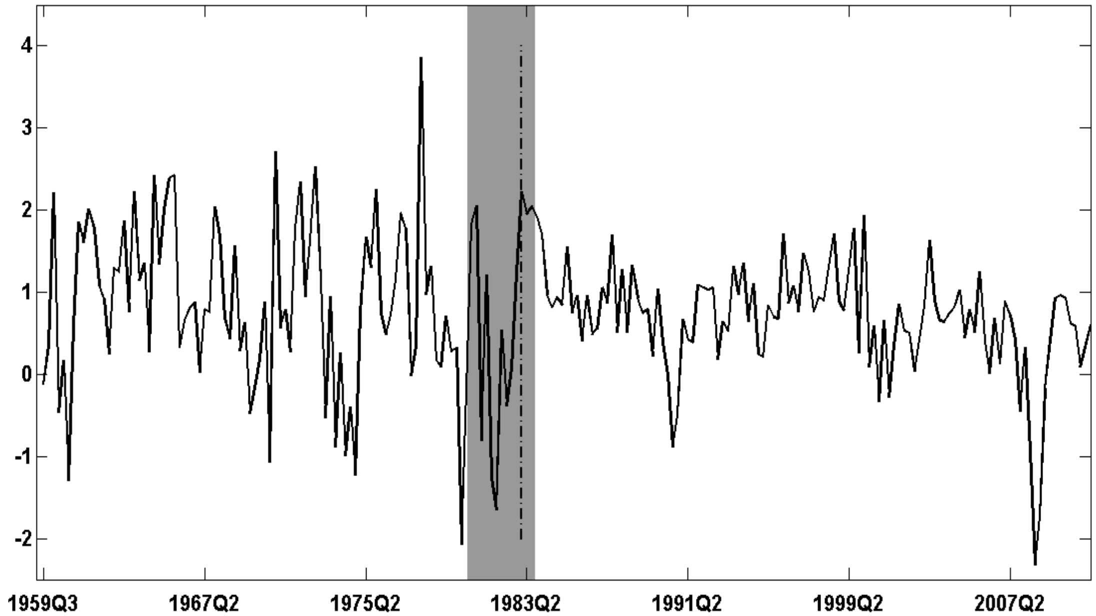
→ **Discrete conditional distribution**

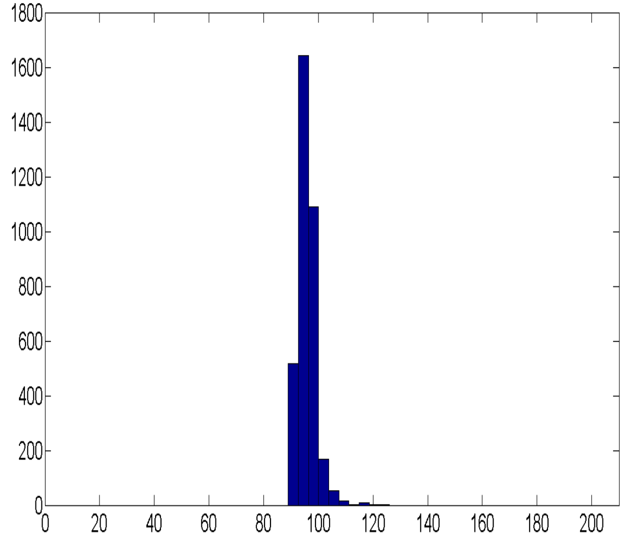
1) Compute the posterior density for each *i* and normalize.
2) Draw u ~ U(0,1)
3) Find τ such that

$$\sum_{t=1}^{\tau} \pi(\tau_1 = t | Y_{1:T}, \Theta, \Sigma) < u \leq \sum_{t=1}^{\tau+1} \pi(\tau_1 = t | Y_{1:T}, \Theta, \Sigma)$$

Example

- US GDP growth (1959 Q2-2011 Q3) : AR(1)





- Posterior means :

τ_1	σ_1^2	σ_2^2	θ_1	θ_2
1983 - Q1	1.1	0.34	$\begin{pmatrix}0.64\\0.22\end{pmatrix}$	$\begin{pmatrix}0.32\\0.53\end{pmatrix}$

Carlin, Gelfand and Smith (1992)

Advantages

- Generic method :
  - Works for many models (even with path dependence)
- Easily extendible to M-H inference

Drawbacks

- Limited to two regimes
- No criterion for selecting the number of regimes (one or two)
- Time-consuming if T large

Stephens (1994)

- Inference on multiple structural breaks (in mean or variance)
  - Extension of Carlin, Gelfand and Smith
  - Estimation carried out by Gibbs sampler
- Instead of one structural break : K breaks (K+1 regimes)
  - The parameter set is augmented by

→ K break date parameters : $\tau = \{\tau_0, \tau_1, ..., \tau_K, \tau_{K+1}\}$

→ Corresponding mean parameters : $\Theta = \{\theta_1', ..., \theta_{K+1}'\}'$

→ Corresponding var. parameters : $\Sigma = \{\sigma_1^2, ..., \sigma_{K+1}^2\}'$

Prior distributions on break parameters :

$\tau_0 = 0$ $\qquad$ $\tau_{K+1} = T$

$\tau_1 \sim U(1, T-(K+2))$ $\qquad$ $\tau_i | \tau_{i-1} \sim U(\tau_i + 1, T-(K+2-i))$

Stephens (1994)

- Gibbs sampler : if AR process as before

$$\theta_i | Y_{1:T}, \Theta_{-i}, \tau, \Sigma \sim N(\bar{\mu}_i, \bar{\Sigma}_i)$$

$$\sigma_i^2 | Y_{1:T}, \tau, \Theta, \Sigma_{-i} \sim IG(\alpha + (\tau_i - \tau_{i-1})/2, \beta + \sum_{t=\tau_{i-1}+1}^{\tau_i} \epsilon_t)$$

$$\tau_i | Y_{1:T}, \Theta, \Sigma, \tau_{-i} \sim \text{Griddy-Gibbs}$$

with

$$\begin{cases} \bar{\mu}_i = \bar{\Sigma}_i [\sigma_i^{-2} \sum_{t=\tau_{i-1}+1}^{\tau_i} x_t y_t + \Sigma_0^{-1} \mu_0] & \forall i \in [1, K+1] \\ \bar{\Sigma}_i = [\sigma_i^{-2} \sum_{t=\tau_{i-1}+1}^{\tau_i} (x_t x_t') + \Sigma_0^{-1}]^{-1} & \forall i \in [1, K+1] \end{cases}$$

Stephens (1994)

Advantages

- Generic method :
    - Works for many models (even with path dependence)
- Easily extendible to M-H inference

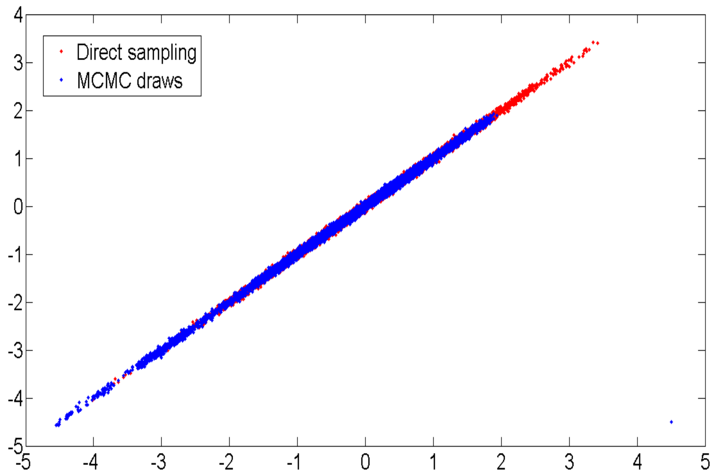
Drawbacks

- No criterion for selecting the number of regimes
- Time-consuming if T large
- Many MCMC iterations are required

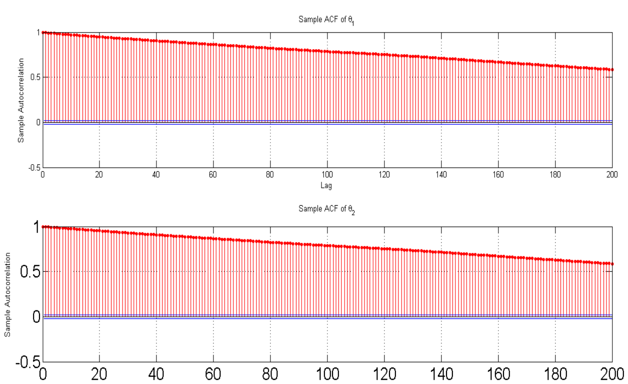
→ **May not converge in a finite amount of time !**

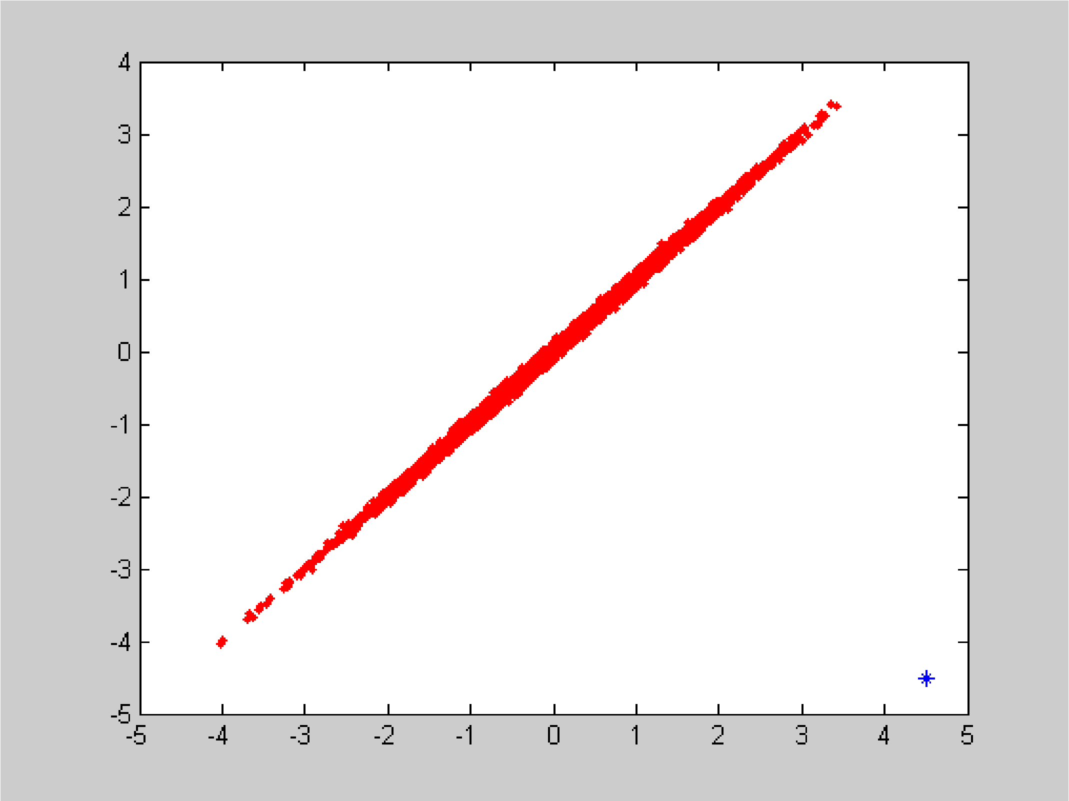
MCMC : Mixing problem

Invariant Dist. : $N(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0,999\\0,999 & 1\end{pmatrix})$



After 10000 draws





First 200 draws of the MCMC

→ Correlated parameters should be jointly sampled

Stephens (1994)

MCMC may not converge in a finite amount of time !

→ **True for Stephens' MCMC but for many others**

Other example :

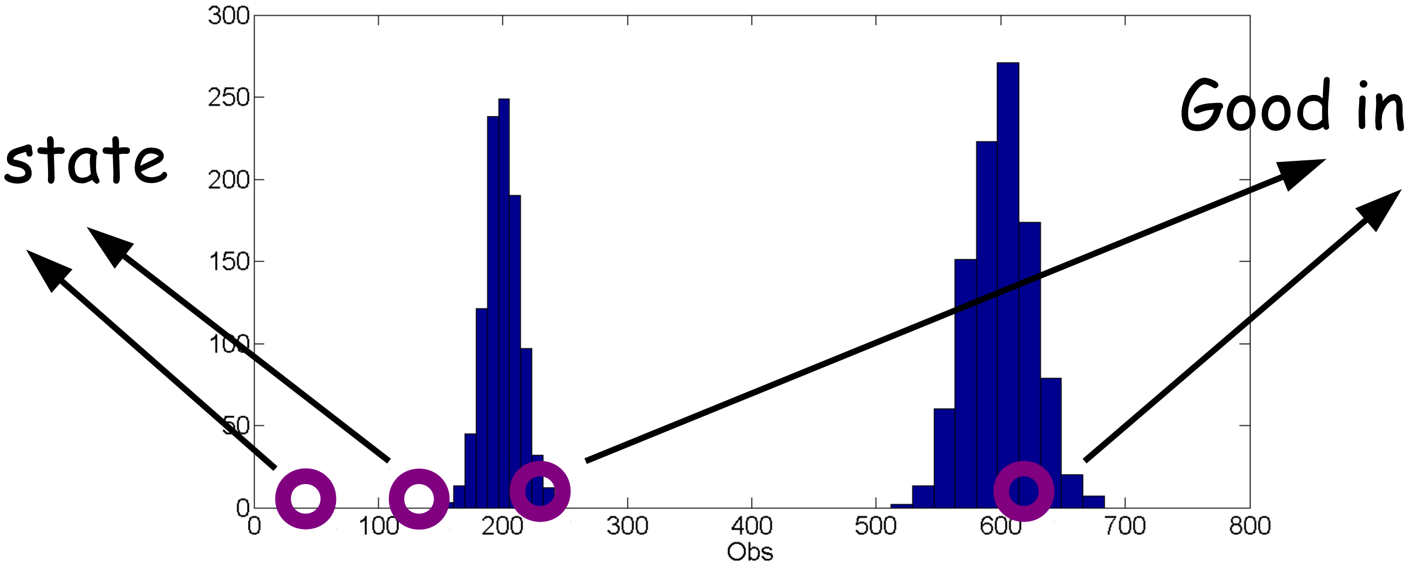
- Single-move of Bauwens,Preminger,Rombouts (2011)

Critical issues :

- Initial state of the MCMC

Bad initial state

Good initial state



Questions ?

MS and CP models

Chib (1996 - 1998)

Drawbacks (Stephens)

- No criterion for selecting the number of regimes (**ok in Chib**)
- Time-consuming if T large (**ok in Chib**)
- Many MCMC iterations are required (**ok in Chib**)

Moreover

- Algorithm for CP and MS models
- Estimation of the MLE

→ Perfect for the MCMC initial state

Chib (1996 - 1998)

- AR process of order 1

$$
\begin{aligned}
y_t | Y_{1:t-1}, s_t, \Theta, \Sigma &\sim N(\theta'_{s_t}[1 \quad y_{t-1}]', \sigma^2_{s_t}) \\
s_t | s_{t-1}, P &\sim p_{s_{t-1}, s_t}
\end{aligned}
$$

- The set of parameters :

$\Theta = \{\theta'_1, ..., \theta'_{K+1}\}'$ ⟶ Mean parameters

$\Sigma = \{\sigma^2_1, ..., \sigma^2_{K+1}\}'$ ⟶ Var. parameters

$S_{1:T} = \{s_1, s_2, ..., s_T\}$ ⟶ Discrete states

P ⟶ Transition matrix

Introduction of a latent state vector driven by a Markov chain

Chib (1996 - 1998)

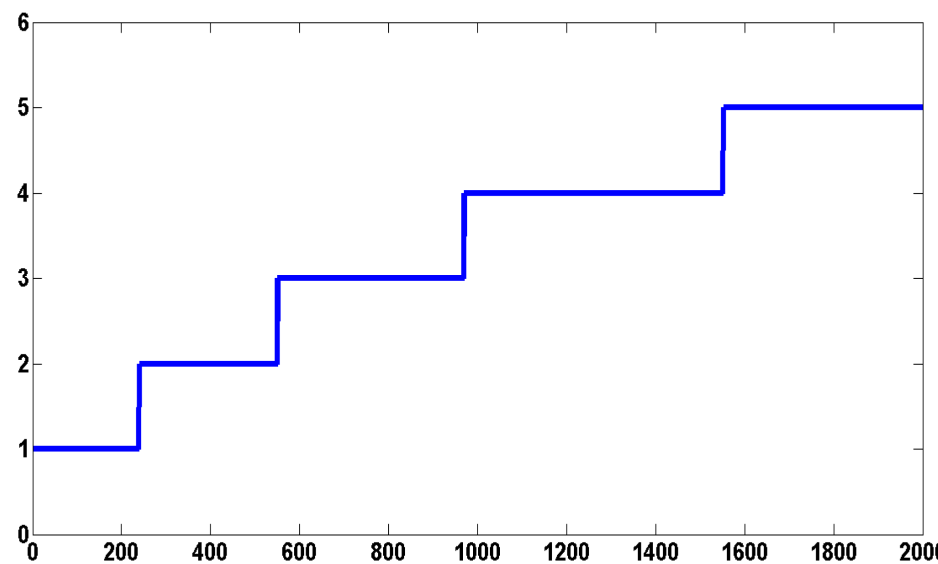
$S_{1:T}$ is driven by a Markov-chain with transition matrix P

Example :

$s_t = i \;\; \forall i \in [1, K+1]$ ⟶ at time t, active state : i

Change-point configuration

$$P = \begin{pmatrix} p_{1,1} & 1-p_{1,1} & 0 & \dots & 0 \\ 0 & p_{2,2} & 1-p_{2,2} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$



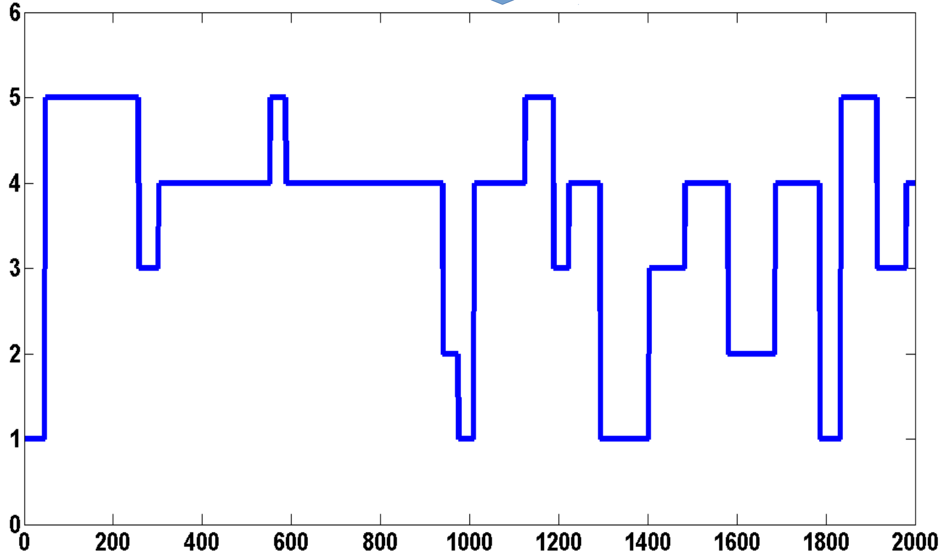
No recurrent state !

Chib (1996 - 1998)

$S_{1:T}$ is driven by a Markov-chain with transition matrix P

Markov-switching configuration

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,K+1} \\ p_{2,1} & p_{2,2} & p_{2,3} & \cdots & p_{2,K+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{K+1,1} & p_{K+1,2} & p_{K+1,3} & \cdots & p_{K+1,K+1} \end{pmatrix}$$



Recurrent states

Parsimonious model

Difficult to estimate

(label switching problem)

Chib (1996 - 1998)

Markov-switching configuration

Change-Point configuration

$$f(y_t|Y_{1:t-1}, S_{1:t-1}, \Theta, \Sigma) = \sum_{i=1}^{K+1} p_{s_{t-1},i} f(y_t|Y_{1:t-1}, \Theta, \Sigma, s_t = i)$$

CP and MS models are mixture models with time-varying probabilities

Chib (1996 - 1998)

- AR process of order 1

$$\begin{cases} y_t|Y_{1:t-1}, s_t & \sim & N(\theta'_{s_t}[1 \quad y_{t-1}]', \sigma^2_{s_t}) \\ s_t|s_{t-1}, P & \sim & p_{s_{t-1},s_t} \end{cases}$$

- Gibbs sampler :

$$\begin{aligned} \theta_i|Y_{1:T}, S_{1:T}, P, \Theta_{-i}, \Sigma &\sim N(\bar{\mu}_i, \bar{\Sigma}_i) \\ \sigma^2_i|Y_{1:T}, S_{1:T}, P, \Theta, \Sigma_{-i} &\sim IG(\alpha + (n_{i,i})/2, \beta + \sum_{t=1}^{T} \epsilon_t \delta_{s_t=i}) \\ P|Y_{1:T}, S_{1:T}, \Theta, \Sigma &\sim \prod_{i=1}^{K+1} \text{Dirichlet}(\eta + n_{i,1:K+1}) \\ S_{1:T}|Y_{1:T}, P, \Theta, \Sigma &\sim \text{Forward-Backward algorithm} \end{aligned}$$

Where $n_{i,j} = \sum_{t=1}^{T} \delta_{s_{t-1}=i, s_t=j}$ and $n_{i,1:j} = [n_{i,1} \quad n_{i,2} \quad \ldots \quad n_{i,j}]$

Chib samples the state vector in one block !

Forward-Backward algorithm

- Discrete version of the Kalman filter.

- Two steps :

  1) Compute the forward and the predictive probabilities
  2) Sample an entire state vector starting from s_T until s_1

- Observe that (without conditioning on $\{\Theta, \Sigma, P\}$ for clarity) :

$$\pi(S_{1:T}|Y_{1:T}) = \pi(s_T|Y_{1:T})\pi(s_{T-1}|Y_{1:T}, s_T) \ldots \pi(s_1|Y_{1:T}, S_{2:T})$$

- Sampling from $\pi(S_{1:T}|Y_{1:T})$ is equivalent to draw

s_T → s_{T-1} → … → s_1

$s_T \sim \pi(s_T|Y_{1:T})$, $s_{T-1} \sim \pi(s_{T-1}|Y_{1:T}, s_T)$, …, $s_1 \sim \pi(s_1|Y_{1:T}, S_{2:T})$

Forward-Backward algorithm

- The challenge is to compute : $\pi(s_t|Y_{1:T}, s_{t+1:T}) \quad \forall t \in [1, T-1]$

- **Assumption :** $f(y_t|Y_{1:t-1}, S_{1:t}) = f(y_t|Y_{1:t-1}, s_t)$

  ⇒ { $y_t|Y_{1:t-1}$ is independent of $S_{1:t-1}$ given s_t
  Model without path dependence

- AR process of order p : $y_t|Y_{1:t-1}, s_t \sim N(\theta'_{s_t} x_t, \sigma^2_{s_t})$

  with $x_t = [1 \quad y_{t-1} \quad \ldots \quad y_{t-p}]'$

then $f(y_t|Y_{1:t-1}, s_t) = (2\pi\sigma^2_{s_t})^{\frac{-1}{2}} e^{-(y_t - \theta'_{s_t} x_t)^2/(2\sigma^2_{s_t})}$

No path dependence for AR processes

- **ARMA and GARCH : path dependence problem**

Forward-Backward algorithm

- **Under assumption :** $f(y_t|Y_{1:t-1}, S_{1:t}) = f(y_t|Y_{1:t-1}, s_t)$
- The challenge is to compute : $\pi(s_t|Y_{1:T}, s_{t+1:T}) \quad \forall t \in [1, T-1]$

$$\begin{aligned}
\pi(s_t|Y_{1:T}, S_{t+1:T}) &\propto f(s_t|Y_{1:t}) f(S_{t+1:T}, Y_{t+1:T}|Y_{1:t}, s_t) \\
&\propto f(s_t|Y_{1:t}) f(S_{t+1:T}|Y_{1:t}, s_t) f(Y_{t+1:T}|Y_{1:t}, S_{t:T}) \\
&\propto f(s_t|Y_{1:t}) f(s_{t+1}|s_t)
\end{aligned}$$

↓ **Independent of** s_t

- Two terms :

1) $f(s_{t+1}|s_t) = p_{s_t, s_{t+1}}$ transition matrix of the MC

2) $f(s_t|Y_{1:t}) = \ ?$ probability of a state given the obs until t

New challenge : $f(s_t|Y_{1:t})$

Forward-Backward algorithm

- How to compute : $f(s_t|Y_{1:t})$

$$
\begin{aligned}
f(s_t|Y_{1:t}) &\propto f(y_t|Y_{1:t-1}, s_t) f(s_t|Y_{1:t-1}) \\
&\propto f(y_t|Y_{1:t-1}, s_t) \sum_{i=1}^{K+1} f(s_t = i|s_{t-1}) f(s_{t-1}|Y_{1:t-1}) \\
&\propto \underbrace{f(y_t|Y_{1:t-1}, s_t)}_{\text{computable}} \sum_{i=1}^{K+1} p_{s_{t-1}, s_t = i} \underbrace{f(s_{t-1}|Y_{1:t-1})}_{\text{Previous quantity !}}
\end{aligned}
$$

Starting point
$$
\left\{
\begin{array}{l}
f(s_1 = 1|y_1) \\
\quad \dots \\
f(s_1 = K+1|y_1)
\end{array}
\right.
\Rightarrow
\left\{
\begin{array}{l}
f(s_2 = 1|Y_{1:2}) \\
\quad \dots \\
f(s_2 = K+1|Y_{1:2})
\end{array}
\right.
\Rightarrow \text{until } T
$$

Forward-Backward algorithm

- To summarize :

1) Compute $f(s_t = i | Y_{1:t}) \ \ \forall i \in [1, K+1] \ \ and \ \ \forall t \in [1, T]$

Example :

$$f(s_1 = i | y_1) = \frac{f(y_1 | s_1 = i)}{\sum_{j=1}^{K+1} f(y_1 | s_1 = j)}$$

1st quantity

Prediction step

$$f(s_2 = i | y_1) = \sum_{j=1}^{K+1} p_{s_1 = j, s_2 = i} f(s_1 = j | y_1)$$

2nd quantity

Update step

$$f(s_2 = i | Y_{1:2}) = \frac{f(y_2 | y_1, s_2 = i) f(s_2 = i | y_1)}{\sum_{j=1}^{K+1} f(y_2 | y_1, s_2 = j) f(s_2 = j | y_1)}$$

Forward-Backward algorithm

- At the end of 1)

1) 'Forward' matrix $F \in \Re^{T,K+1}$

$$\begin{pmatrix} f(s_1 = 1|y_1) & ... & f(s_{K+1}|y_1) \\ f(s_2 = 1|Y_{1:2}) & ... & f(s_2 = K+1|Y_{1:2}) \\ ... & ... & ... \\ f(s_T = 1|Y_{1:T}) & ... & f(s_T = K+1|Y_{1:T}) \end{pmatrix}$$

2) Draw $s_T \sim f(s_T|Y_{1:T})$

3) Draw $s_t \sim f(s_t|Y_{1:T}, S_{t+1:T}) \;\; \forall\, t \in [1, T-1]$

with $f(s_t = i|Y_{1:T}, S_{t+1:T}) = \dfrac{f(s_t = i|Y_{1:t})f(s_{t+1}|s_t = i)}{\sum_{j=1}^{K+1} f(s_t = j|Y_{1:t})f(s_{t+1}|s_t = j)}$

Chib's Gibbs sampler

- Last item of the Gibbs : $P|Y_{1:T}, S_{1:T}, \Theta, \Sigma \quad \sim \quad \text{Dirichlet}(\alpha + n_{i,1:K+1})$

Prior on P :

$$f(P) = \prod_{i=1}^{K+1} f(p_{i,1:K+1})$$

$$p_{i,1:K+1} \sim \text{Dirichlet}(\eta_1, ..., \eta_{K+1}) \equiv \text{Dir}(\eta)$$

Posterior :

$$\pi(P|Y_{1:T}, S_{1:T}) \quad \propto \quad f(S_{1:T}|P) \prod_{i=1}^{K+1} f(p_{i,1:K+1})$$

$$\propto \quad \prod_{t=2}^{T} p_{s_{t-1},s_t} \prod_{i=1}^{K+1} [\prod_{i=j}^{K+1} p_{i,j}^{\eta_j - 1}]$$

$$\propto \quad \prod_{i=1}^{K+1} [\prod_{i=j}^{K+1} p_{i,j}^{n_{i,j}+\eta_j - 1}]$$

$$\sim \quad \prod_{i=1}^{K+1} \text{Dir}(\eta_1 + n_{i,1}, ..., \eta_{K+1} + n_{i,K+1})$$

with $n_{i,j}$
Number of times where the state moves from i to j

Model selection

- How to choose the number of breaks ?

→ **By Marginal likelihood**

- Choose a range of number of breaks : e.g. 0 to 5
  - Estimate each model from one break to five

No break	One break		Five breaks
$\pi(\Theta, \Sigma \mid Y_{1:T}, K=0)$	$\pi(\Theta, \Sigma, S_{1:T}, P \mid Y_{1:T}, K=1)$	...	$\pi(\Theta, \Sigma, S_{1:T}, P \mid Y_{1:T}, K=5)$

  - At each time, compute the marginal quantity

No break	One break		Five breaks
$f(Y_{1:T} \mid K=0)$	$f(Y_{1:T} \mid K=1)$	...	$f(Y_{1:T} \mid K=5)$

  - Find the number of breaks that maximizes the ML

Model selection

- How to compute the marginal likelihood ?

→ **By MCMC or by Importance sampling**

By MCMC : Bayes' rule

$$f(Y_{1:T}|K=k) = \frac{f(Y_{1:T}|\Theta^*,\Sigma^*,P^*)f(\Theta^*,\Sigma^*,P^*)}{\pi(\Theta^*,\Sigma^*,P^*|Y_{1:T},K=k)}$$

Posterior density

$$\pi(\Theta^*,\Sigma^*,P^*|Y_{1:T}) = \underbrace{\pi(P^*|Y_{1:T})}_{1}\underbrace{\pi(\Theta^*|Y_{1:T},P^*)}_{2}\underbrace{\pi(\Sigma^*|Y_{1:T},P^*,\Theta^*)}_{\text{computable}}$$

1) $\pi(P^*|Y_{1:T}) = \int \pi(P^*|Y_{1:T},S_{1:T})\pi(S_{1:T}|Y_{1:T})dS_{1:T}$

2) $\pi(\Theta^*|Y_{1:T},P^*) = \int \pi(\Theta^*|Y_{1:T},S_{1:T})\pi(S_{1:T}|Y_{1:T},P^*)dS_{1:T}$

→ **Auxiliary MCMC with fixed P***

Model selection

- How to compute the marginal likelihood ?

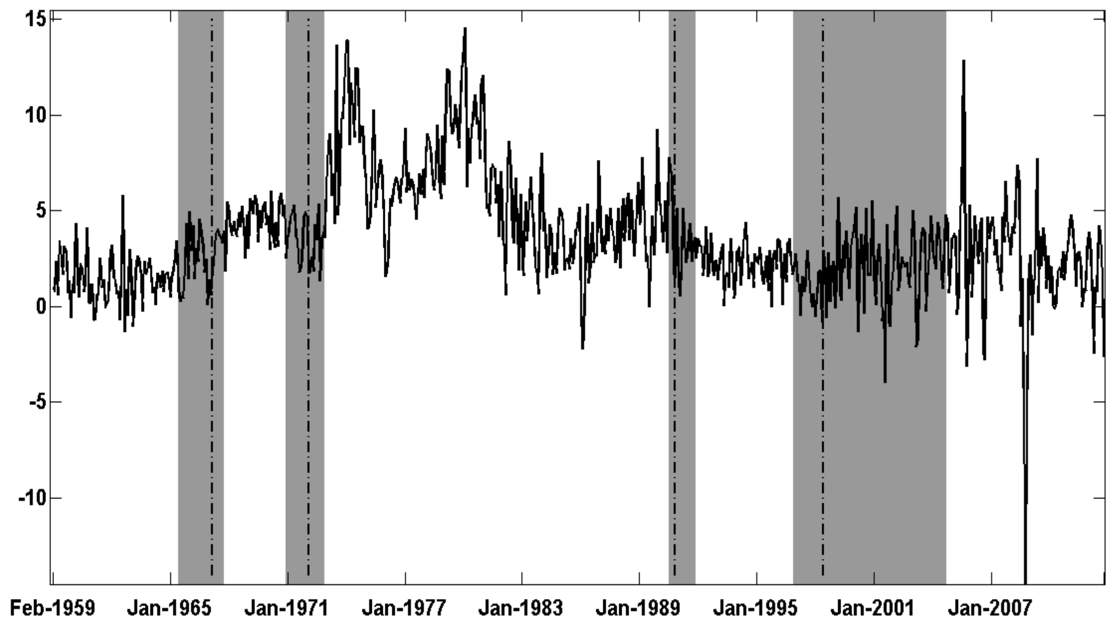
→ **By MCMC or by Bridge sampling**

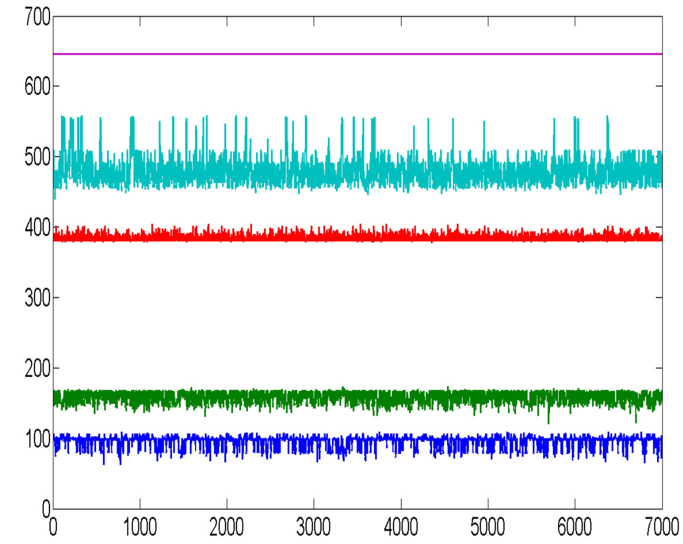
By Importance sampling

$$
\begin{aligned}
f(Y_{1:T}|K=k) &= \int f(Y_{1:T}|\Theta,\Sigma,P) f(\Theta,\Sigma,P) d\Theta d\Sigma dP \\
&= \int f(Y_{1:T}|\Theta,\Sigma,P) f(\Theta,\Sigma,P) \frac{q(\Theta,\Sigma,P)}{q(\Theta,\Sigma,P)} d\Theta d\Sigma dP \\
&\approx \frac{1}{G}\sum_{i=1}^{G} \frac{f(Y_{1:T}|\Theta^i,\Sigma^i,P^i) f(\Theta^i,\Sigma^i,P^i)}{q(\Theta^i,\Sigma^i,P^i)}
\end{aligned}
$$

Where $\{\Theta^i,\Sigma^i,P^i\} \sim q(-)$

US Monthly Inflation Rate





#regimes	1	2	3	4	5	6
MLL	-1424	-1409	-1391	-1386	-1384	-1388

Predictions of structural breaks

Pesaran, M. H.; Pettenuzzo, D. & Timmermann, A. *'Forecasting Time Series Subject to Multiple Structural Breaks'*, Review of Economic Studies, 2006, 73, 1057-1084

- Hierarchical distributions

Mean parameters

$$\begin{cases} \Theta | \mu_0, \Sigma_0 \sim \prod_{i=1}^{K+1} N(\mu_0, \Sigma_0) \\ \mu_0 \sim N(\underline{\mu}_\mu, \underline{\Sigma}_\mu) \\ \Sigma_0 \sim IW(\underline{v}, \underline{V}) \end{cases}$$

Var. parameters

$$\begin{cases} \Sigma | \alpha, \beta \sim \prod_{i=1}^{K+1} IG(\alpha, \beta) \\ \alpha \sim \exp(\lambda) \\ \beta \sim IG(e, f) \end{cases}$$

- The hierarchical (random) parameters gather information from the different regimes

→ **If new regime : draw parameters from the hier. dist.**

Predictions of structural breaks

- Hierarchical distributions

Mean parameters

$$\begin{cases} \mu_0 \sim N(\underline{\mu}_{\mu}, \underline{\Sigma}_{\mu}) \\ \Sigma_0 \sim IW(\underline{v}, \underline{V}) \end{cases}$$

$$\Theta | \mu_0, \Sigma_0 \sim \prod_{i=1}^{K+1} N(\mu_0, \Sigma_0)$$

- A new break happens after the end of the sample (regime K+2) :

→ New parameters have to be drawn !

θ_1, θ_2, ..., θ_{K+1} — 'Observations' — **Update** → $\left\{ \begin{matrix} \mu_0 \\ \Sigma_0 \end{matrix} \right\}$ → **New regimes** →

$$\theta_{K+2} \sim N(\mu_0, \Sigma_0)$$

$$\theta_{K+3} \sim N(\mu_0, \Sigma_0)$$

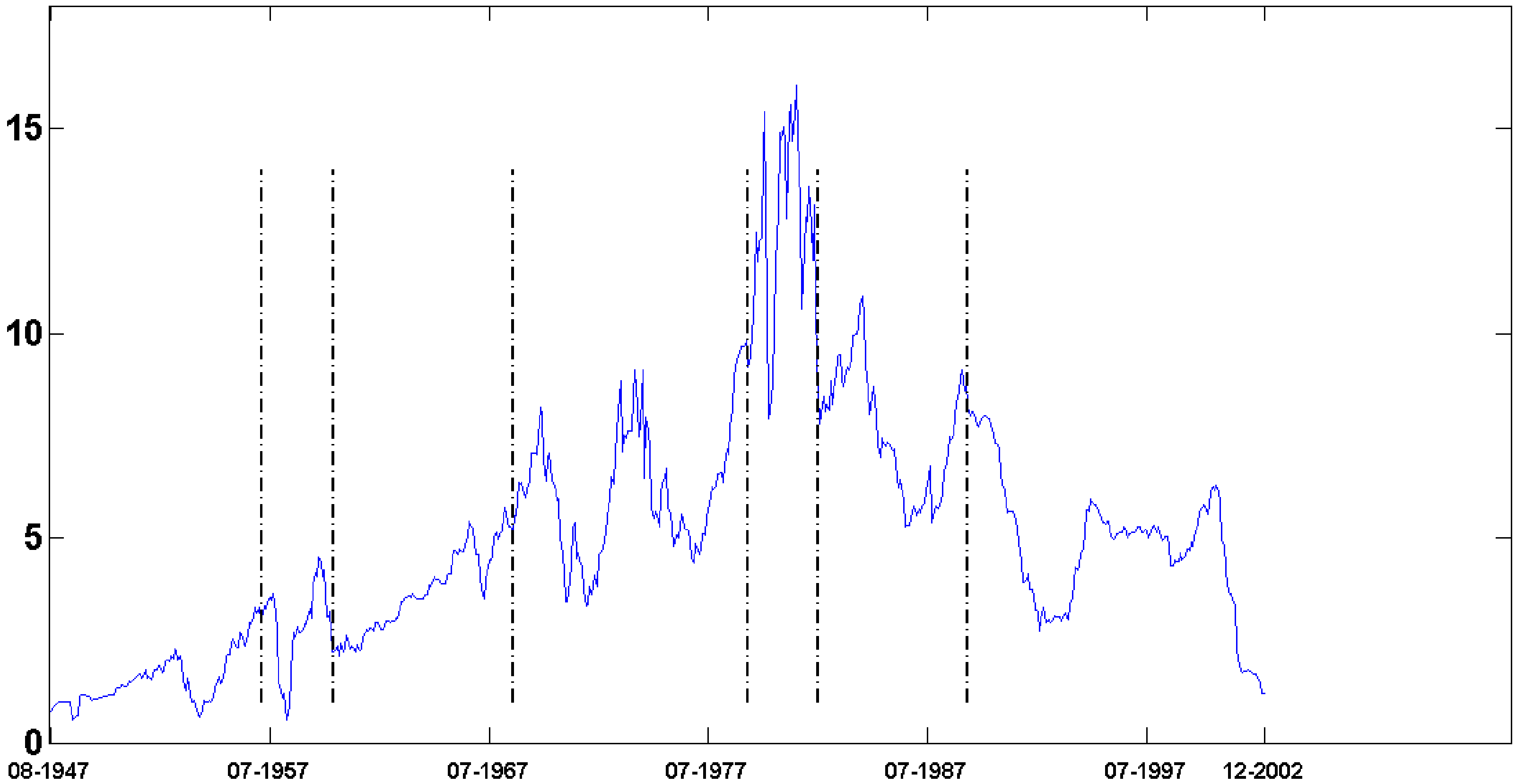
...

AR process : Gibbs sampler exists

Example :

US 3-Month Treasury Bill

CP-AR(1) model exhibiting 7 regimes (according to MLL)



Example :

US 3-Month Treasury Bill

CP-AR(1) model exhibiting 7 regimes (according to MLL)

TABLE 2

Posterior parameter estimates for the unconstrained AR(1) hierarchical HMC model with six break-points

	Regimes						
	1	2	3	4	5	6	7
Date	1947–1957	1957–1960	1960–1966	1966–1979	1979–1982	1982–1989	1989–2002
			Constant				
Mean	0·021	0·252	0·017	0·220	0·412	0·246	−0·004
S.E.	0·034	0·208	0·067	0·161	0·521	0·211	0·054
			AR(1) coefficient				
Mean	1·002	0·895	1·006	0·969	0·958	0·968	0·992
S.E.	0·020	0·071	0·020	0·026	0·045	0·027	0·011
			Variances				
Mean	0·023	0·256	0·015	0·260	2·558	0·161	0·048
S.E.	0·003	0·068	0·003	0·031	0·671	0·027	0·005
			Transition probability matrix				
Mean	0·988	0·960	0·979	0·991	0·961	0·981	1
S.E.	0·010	0·032	0·017	0·008	0·032	0·015	–
Mean duration	120	37	72	156	37	84	–

Notes: AR, autoregressive; S.E., standard error.

Label switching

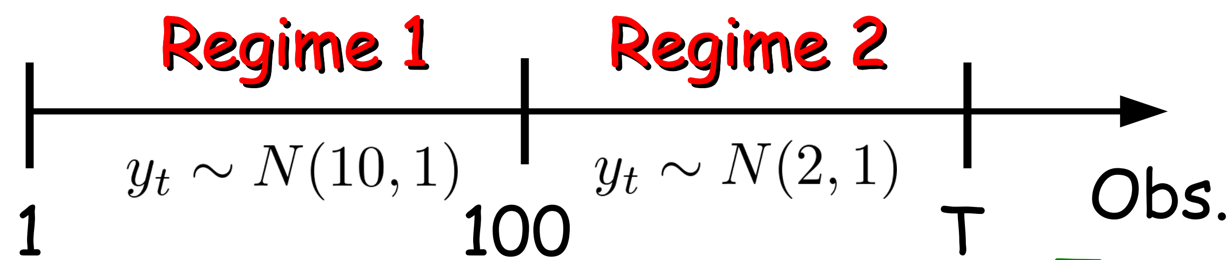
- Issue arises in MS specification (recurrent states)

Full transition matrix

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,K+1} \\ p_{2,1} & p_{2,2} & p_{2,3} & \cdots & p_{2,K+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{K+1,1} & p_{K+1,2} & p_{K+1,3} & \cdots & p_{K+1,K+1} \end{pmatrix}$$

- Posterior distribution : invariant to the label of the states

Example :



$\theta_1 = 10, \theta_2 = 2$ → $s_{1:100} = 1$ $s_{101:T} = 2$

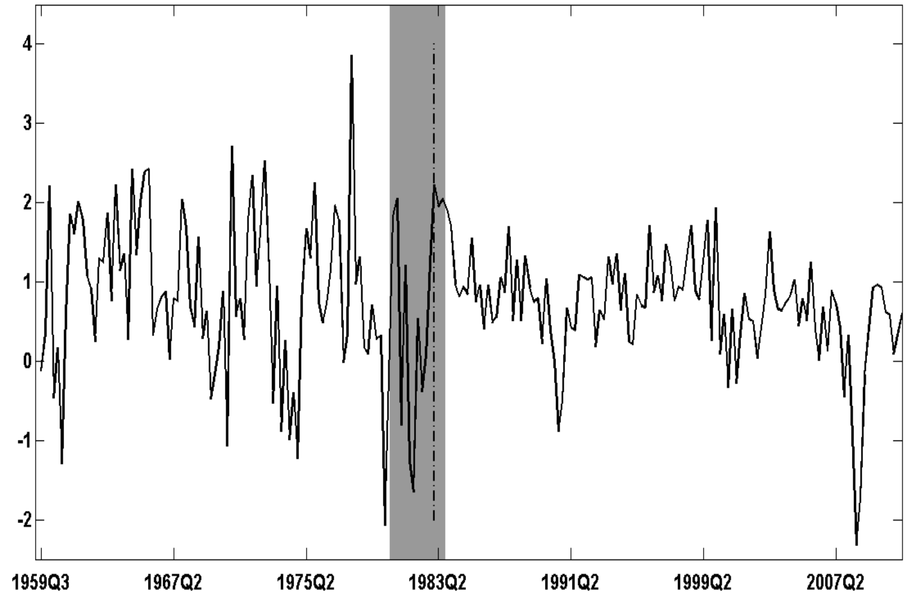
$\theta_1 = 2, \theta_2 = 10$ → $s_{1:100} = 2$ $s_{101:T} = 1$

} Same likelihood

Misleading if it happens during the MCMC algorithm

Example : US GDP Growth

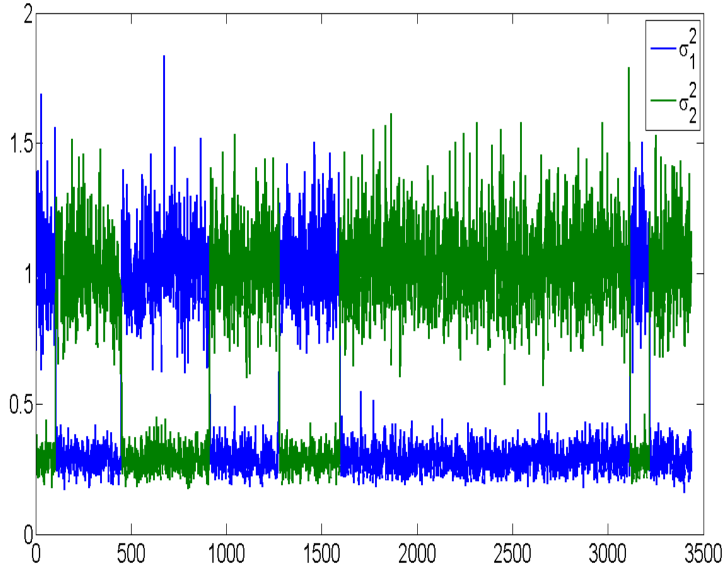
- Great moderation : Drop of the variance



- Could be estimated by an MS-AR model

Example : US GDP Growth

Label switching problem

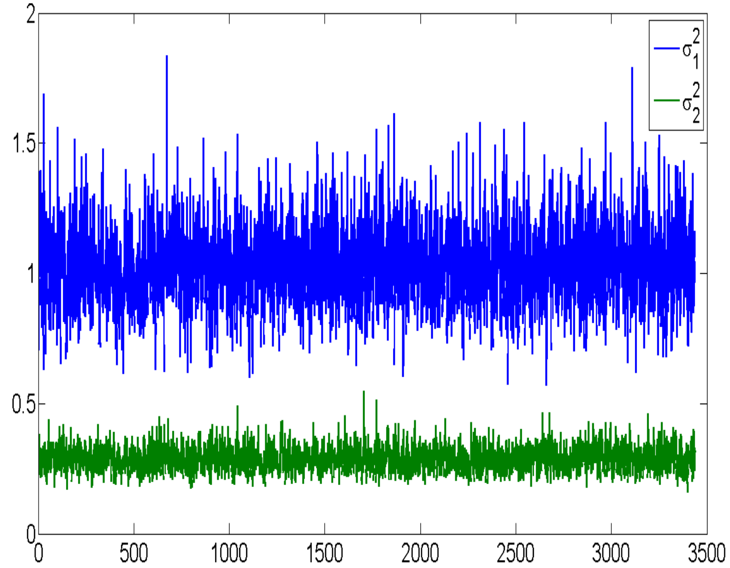


- Posterior means are not safe !

$$E(\sigma_1^2|Y_{1:T}) \neq \frac{1}{N}\sum_{i=1}^{N}\sigma_1^{2,i} = 0.5$$

$$E(\sigma_2^2|Y_{1:T}) \neq \frac{1}{N}\sum_{i=1}^{N}\sigma_2^{2,i} = 0.81$$

No label switching



- Posterior means are safe !

$$E(\sigma_1^2|Y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N}\sigma_1^{2,i} = 1.02$$

$$E(\sigma_2^2|Y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N}\sigma_2^{2,i} = 0.29$$

Label switching

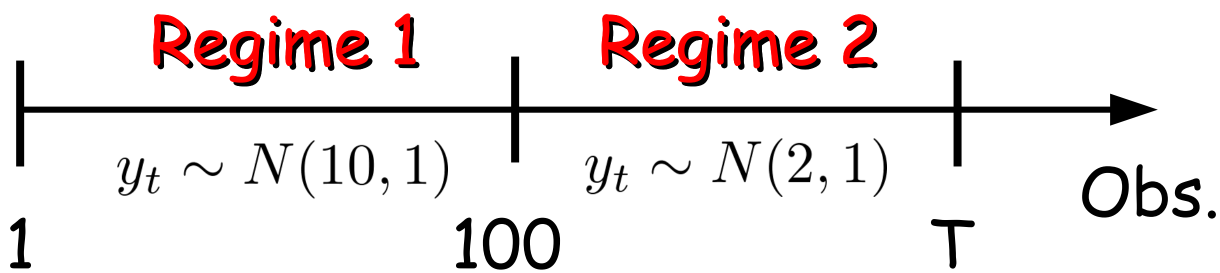
- Solutions

1) Posterior distribution : invariant to the label of the states

Only if the prior is also symetric to the labeling

→ Constrain the prior distribution

Example :



Prior distributions

$$\theta_1 \sim N(\mu_0, \Sigma_0)$$

$$\theta_2 | \theta_1 \sim N(\mu_0, \Sigma_0)\delta_{\theta_2 < \theta_1}$$

⇒ $\theta_1 = 10, \theta_2 = 2$

~~$\theta_1 = 2, \theta_2 = 10$~~ → **Impossible label**

Label switching

- Solutions

2) Sort out the MCMC draws after the algorithm **according to a loss function**

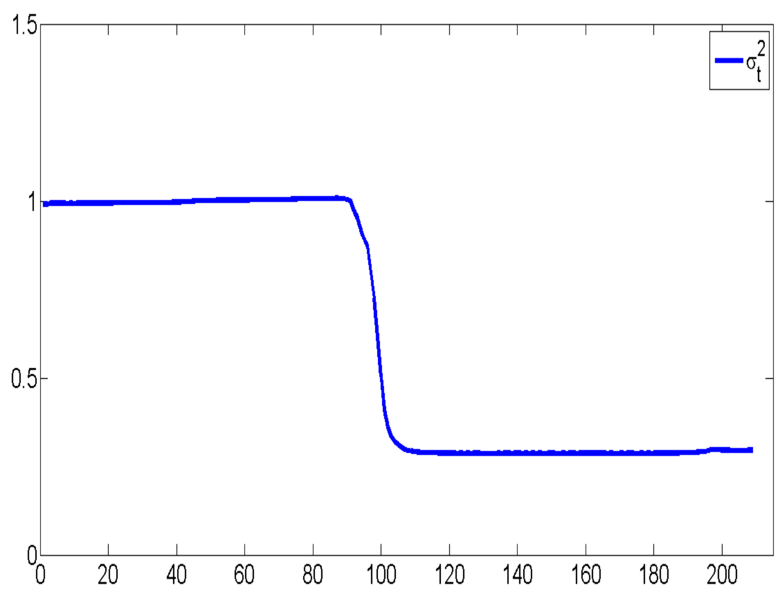
- 3) Only use summary statistics invariant to label switches

Example : US GDP growth

Posterior means over time

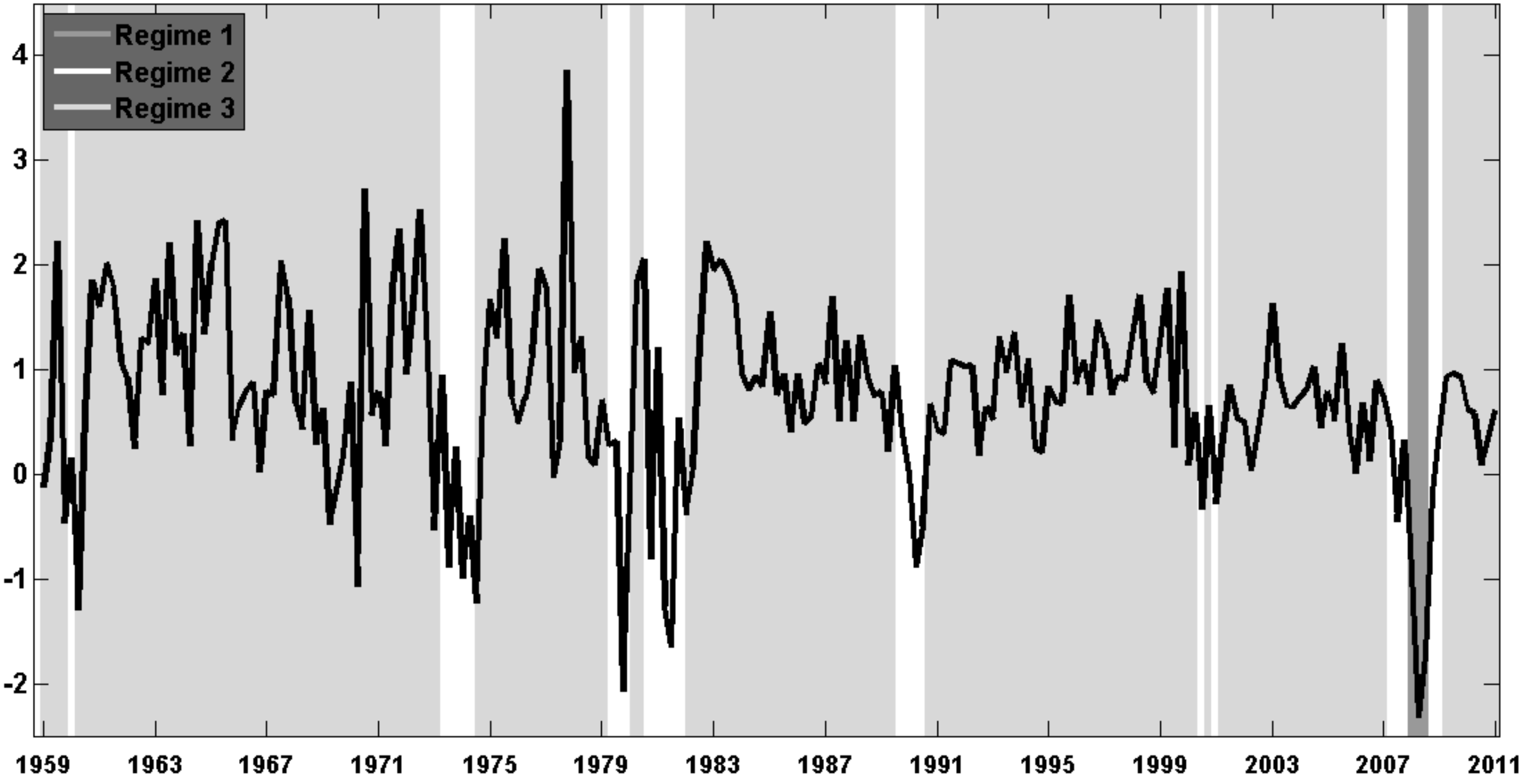
$$E(\sigma_t^2|Y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N}\sigma_{s_t}^{2,i}$$

Invariant to labeling



Example : MS

US GDP Growth rate – MS-ARMA with 3 regimes



Chib

Advantages

- Multiple breaks
- Recurrent or no recurrent states (Change-point/Markov-switching)
- MCMC with good mixing properties
- Allow to select an optimal number of regimes
- Forecast of structural breaks

→ **State of the art !**

Drawbacks

- Geometric distribution for the regime duration
- Many computation for selecting the number of regimes
- Not applicable to models with path dependence

Questions ?

References

References

- Carlin, Gelman and Smith (1992)

Carlin, B., A.E. Gelfand and A.F.M. Smith, 1992, *'Hierarchical Bayesian analysis of changepoint problems'*, Applied Statistics, 41, 389-405

- Stephens (1994)

Stephens, D. A. *'Bayesian Retrospective Multiple-Changepoint Identification'*, Applied Statistics, 1994, 1, 159-178

- Chib (1996)

Chib, S. *'Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models'*, Journal of Econometrics, 1996, 75, 79-97

References

- Chib (1998)

Chib, S. *'Estimation and comparison of multiple change-point models'*, Journal of Econometrics, 1998, 86, 221-241

- Pesaran, Pettenuzzo and Timmermann (2006)

Pesaran, M. H.; Pettenuzzo, D. & Timmermann, A. *'Forecasting Time Series Subject to Multiple Structural Breaks'*, Review of Economic Studies, 2006, 73, 1057-1084

Other CP and MS specs

- Koop and Potter (2007) – CP models

  CP models without geometric durations

  Inference of breaks without marginal likelihood

  Extremely demanding !

Koop, G. & Potter, S. *'Estimation and Forecasting with Multiple Breaks'*, Review of Economic Studies, 2007, 74, 763-789

- Giordani and Kohn (2008) – recurrent models

  Breaks modelled as mixtures (no prob. dynamic)

  Inference of breaks without marginal likelihood

  Parameters are subject to different breaks

Giordani, P. & Kohn, R. *'Efficient Bayesian inference for multiple change-point and mixture innovation models'* Journal of Business and Economic Statistics, 2008, 26, 66-77

Other CP and MS specs

- Maheu and Song (2013) – CP models

  Inference of breaks without marginal likelihood

  Only adapted to AR process with normal innovations

  Very fast !

Maheu, J. & Song, Y. *'A new structural break model, with an application to canadian inflation forecasting'*, International journal of forecasting, 2013, 30, 144-160

- Jochmann (2013) – CP and MS models

  Inference of breaks with Dirichlet processes

  Inference of CP and MS models at the same time

  Predictions that encompass different number of breaks

Jochmann, M. *'Modeling U.S. Inflation Dynamics : A Bayesian Nonparametric Approach'* Econometric Reviews, 2013, Forthcoming

Estimation by SMC

- Fearnhead and Liu (2007) – CP models
  Exact inference by SMC
  Very fast !

Fearnhead, P. & Liu, Z. *'On-line inference for multiple changepoint problems'*, Journal of Royal Statistical Society, Series B, 2007, 69 (4), 589-605

- Whiteley, Andrieu, Doucet (2013) – CP models
  Unknow number of breaks
  Exact inference by SMC
  Faster than O(T^2)

Whiteley, N.; Andrieu, C. & Doucet, A. *'Bayesian Computational Methods for Inference in Multiple Change-points Models'*, Discussion paper, University of Bristol, 2011

Other references

- Marginal likelihood by Importance sampling and Bridge sampling

Fruhwirth-Schnatter, S. *'Estimating Marginal Likelihoods for Mixture and Markov-switching Models Using Bridge Sampling Techniques'*, Econometrics Journal, 2004, 7, 143-167

- Label-switching

Geweke, J. *'Interpretation and Inference in Mixture Models: Simple MCMC works'* Computational Statistics and Data Analysis, 2007, 51, 3259-3550