Bayesian methods for structural break modellings

Arnaud Dufays
Centre de Recherche en Economie et Statistique (CREST)

Course Structure

- **Chapter 1 :**

  Bayesian concepts and inference

- **Chapter 2 :**

  Structural breaks for models *without path dependence*

- **Chapter 3 :**

  Structural breaks for models *with path dependence*

- **Chapter 4 :**

  Introduction to Bayesian econometrics using Matlab

Chapter 1

Bayesian inference : Principles

Bayesian inference

Differences from classic approach :

- Model parameters (random vs fixed)
- Finite sample vs asymptotic theory
- Statistical interpretation (subjective vs objective)

Historical consideration

From effects to causes

- Consider two non-independent events : A and B

- From basic axiom of probability :

$$P(A|B) = \frac{P(A,B)}{P(B)} \quad \textbf{and} \quad P(B|A) = \frac{P(A,B)}{P(A)}$$

Bayes' rule : $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

What is the probability of observing A if B has occurred ?

Generalization

- Multiple events :

Prior Distribution

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Posterior Distribution

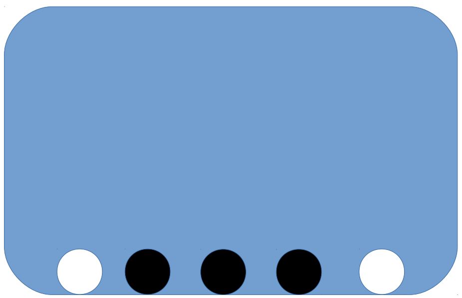
Normalizing constant / Marginal likelihood

Normalizing constant : $P(B) = \sum_{j=1}^{k} P(B|A_j)P(A_j)$

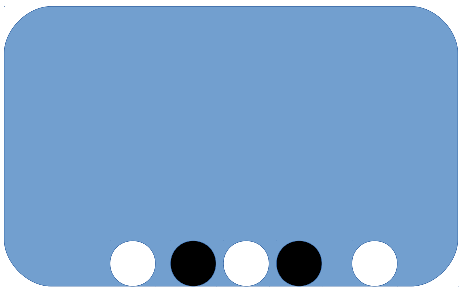
- Continuous case : Sum becomes an integral.

From effects to causes : Examples

Urn 1



Urn 2



One draw : A black ball

What is the probability that the ball comes from urn 1 ?

Prior distribution $P(U_1) = P(U_2) = 0,5$

Post distribution

$$P(U_1|B) = \frac{P(B|U_1)P(U_1)}{P(B|U_1)P(U_1) + P(B|U_2)P(U_2)}$$

$$P(U_1|B) = 0.6$$

From effects to causes : Examples

Unwin, S. D. , *'The probability of God : A simple Calculation that proves the ultimate truth'*, 2004

- A priori, God may or may not exist : P(G exists) = 0,5
- Observations : Miracles, Wars, Sins, ...

$$P(G \text{ exists}|B) \quad = \quad \frac{P(B|G \text{ exists})P(G \text{ exists})}{P(B|G \text{ exists})P(G \text{ exists}) + P(B|G \text{ not exist})P(G \text{ not exist})}$$

Conclusion... You should bet the Lord exists

Principles of Bayesian inference

- Any unobserved value is random since we shall do a probabilistic statement on it.

  → Distributions on model parameters (prior to observing data)

- Data supposedly generated by the model contain information about the parameters.

  → Likelihood principle

- Through Bayes theorem, the parameter distribution is updated.

  → Posterior distribution of the parameters

Statistical Example

- The model :

$$\begin{cases} y_t = \theta + \epsilon_t \\ \epsilon_t \sim \text{ i.i.d. } N(0,1) \end{cases}$$

- Observations :

$$Y_{1:T} = \{y_1, ..., y_T\}'$$

- Prior distribution :

$$\theta \sim N(\mu_0, n_0^{-1})$$

Example
$$\begin{cases} \mu_0 = 0 \\ n_0^{-1} = 100 \end{cases}$$

- Posterior distribution :

$$\theta | Y_{1:T} \sim N(\mu^*, \sigma^*)$$

$$\mu^* = \frac{\sum_{t=1}^{T} y_t + n_0 \mu_0}{T + n_0}$$

$$\sigma^* = (T + n_0)^{-1}$$

Prior beliefs

Based on another set of observations

- Reflects the uncertainty on the parameter

Calculus

Bayes' rule :

$$\pi(\theta|Y_{1:T}) = \frac{f(Y_{1:T}|\theta)f(\theta)}{f(Y_{1:T})}$$

$$\propto f(Y_{1:T}|\theta)f(\theta)$$

Likelihood function : $f(Y_{1:T}|\theta) = (2\pi)^{-\frac{T}{2}} e^{-0,5\sum_{t=1}^{T}(y_t-\theta)^2}$

Prior distribution :

$$f(\theta) = \frac{e^{\frac{-(\theta-\mu_0)^2}{2n_0}}}{\sqrt{2\pi n_0}}$$

The posterior must be a proper distribution :
insured by the normalizing constant

⟶ We can drop all the terms that do not depend on θ

and see if the posterior kernel comes from a known distribution

Calculus

Posterior kernel :

$$\begin{aligned} \pi(\theta|Y_{1:T}) &\propto e^{-0,5[n_0(\theta-\mu_0)^2+\sum_{t=1}^{T}(y_t-\theta)^2)]} \\ &\propto e^{-0,5[\theta^2(T+n_0)-2\theta(n_0\mu_0+\sum_{t=1}^{T} y_t)]} \\ &\propto e^{\frac{-(\theta-\mu^*)^2}{2\sigma^*}} \end{aligned}$$

The posterior kernel is a normal one with

$$\mu^* = \frac{\sum_{t=1}^{T} y_t + n_0\mu_0}{T+n_0}$$

$$\sigma^* = (T+n_0)^{-1}$$

Discussion

$$\text{If } T \to \infty,\ E(\theta|Y_{1:T}) \to \bar{y},\ V(\theta|Y_{1:T}) \to 0$$

Data increasingly dominate the prior information.

$$\text{If } \ n_0 = 0 \text{ and } \mu_0 = 0 \quad :$$

Same estimator as in the classical approach.

- Bayesian inference provides an entire distribution only based on the observed data.
- Delivers different statistical interpretations.

Summarizing the posterior

- Posterior means and standard deviations

$$E(\theta|Y_{1:T}) = \int_{-\infty}^{\infty} \theta \pi(\theta|Y_{1:T}) d\theta$$

$$V(\theta|Y_{1:T}) = \int_{-\infty}^{\infty} (\theta - E(\theta|Y_{1:T}))^2 \pi(\theta|Y_{1:T}) d\theta$$

- Credible intervals
- Posterior Covariance matrix
- Quantiles and graphics of the marginal distributions of the parameters

Criteria for statistical procedure

Classical

- Properties (Consistency, efficiency,...) from hypothetical repeated samples/ large sample.

Bayesian

- Only based on the observed sample used 'coherently' through the likelihood principle.

Caution :

- Some Bayesian state that Bayesian inference is 'exact' in finite sample

→ No meaning since what happens in repeated sample is not relevant for Bayesian inference.

→ Based on very restrictive assumptions.

Treatment of parameters

Classical

- Fixed parameters in reality

Bayesian

- Random parameters in reality
- Fixed parameters but subjective probabilistic statement

Eases the interpretation :

Classical

95 % Confidence interval covers the true value of the parameter in nineteen out of twenty trials on average.

Bayesian

95 % credible interval gives the region of the parameter space where the probability of covering θ is equal to 95.

Bayesian learning

The posterior distribution as the prior distribution ?

$$\pi(\theta|Y_{1:T}) = \frac{f(Y_{1:T}|\theta)f(\theta)}{f(Y_{1:T})}$$
$$= \frac{f(Y_{t+1:T}|\theta)\pi(\theta|Y_{1:t})}{f(Y_{t+1:T}|Y_{1:t})} \quad \forall \ t \in [1, T-1]$$

New information makes update our belief

Core idea of Sequential Monte Carlo

- Teglas, E. et al., '*Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference*'. 2011.

Complex Bayesian inference

What happens if the posterior distribution is not a known parametric distribution ?

Small dimension

- Deterministic integration
- Direct sampling
- Importance sampling

Medium dimension

- Sequential Monte Carlo (SMC)
- Annealed importance Sampling (SMC sampler)

High dimension

- Markov-chain Monte Carlo (MCMC)
- Approximate Bayesian Computation (ABC)
- Variational Bayes methods

Markov-chain Monte Carlo

Markov chain Monte Carlo

Posterior distribution :

1) Not a known parametric distribution.
2) Cannot sample directly from it.
3) Parameter dimension greater than 3.

MCMC principles :

1) Simulation of a Markov-chain exhibiting the posterior distribution of interest as invariant one.

MCMC Outputs :

1) (Correlated) draws of the posterior distribution.
2) LLN theorem allows to approximate any deterministic function of the posterior distribution.

Invariant distribution

Markov-chain :

- Characterizes by a number of reachable states (discrete or continuous), a transition probability matrix (P), a probability vector of being in the initial state.
- Chain that only depends on the current state for moving to the next

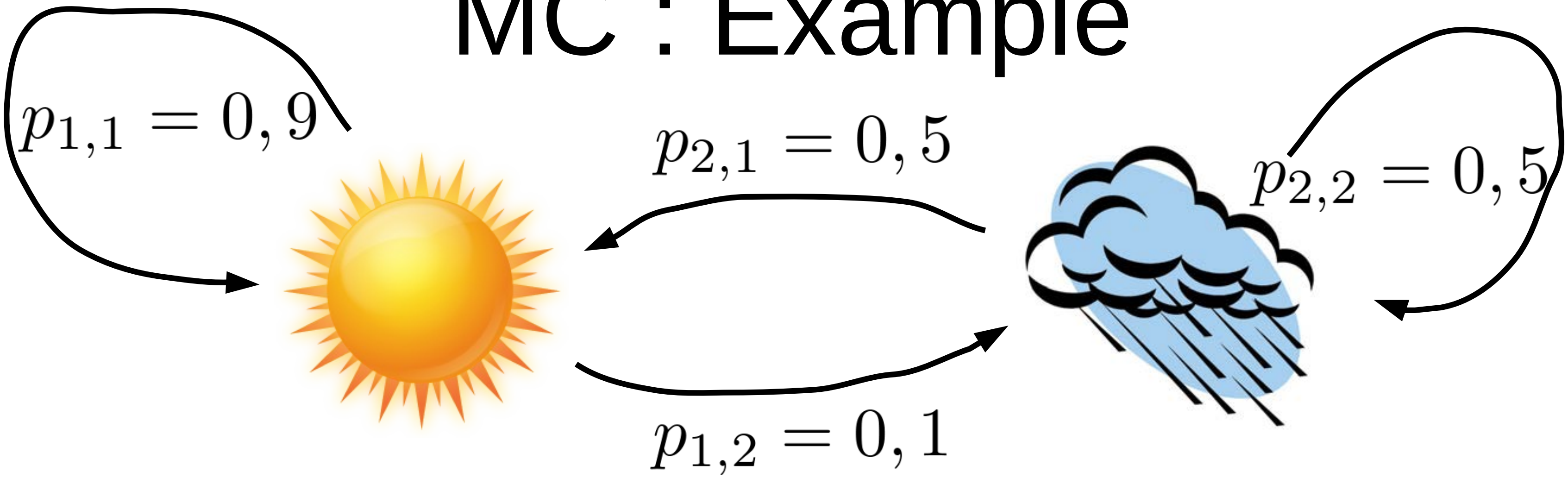
$$s_1 \rightarrow s_2 \rightarrow \ldots \rightarrow s_N$$

Moves according to the transition probability matrix P

Invariant distribution (q) : Independent from the initial state

$$q'P = q'$$ → Unchanged when transformed by P

MC : Example



Transition matrix :

$$P = \begin{pmatrix} 0,9 & 0,1 \\ 0,5 & 0,5 \end{pmatrix}$$

Invariant Distribution

$$q'P = q'$$

$$\downarrow$$

$$q' = [0,8333 \;\; 0,167]$$

Ergodicity

From any initial point, the MC converges to the invariant distribution.

$$P^{100} = \begin{pmatrix} 0,833 & 0,167 \\ 0,833 & 0,167 \end{pmatrix}$$

Based on 10000 simulated draws

$$\downarrow$$

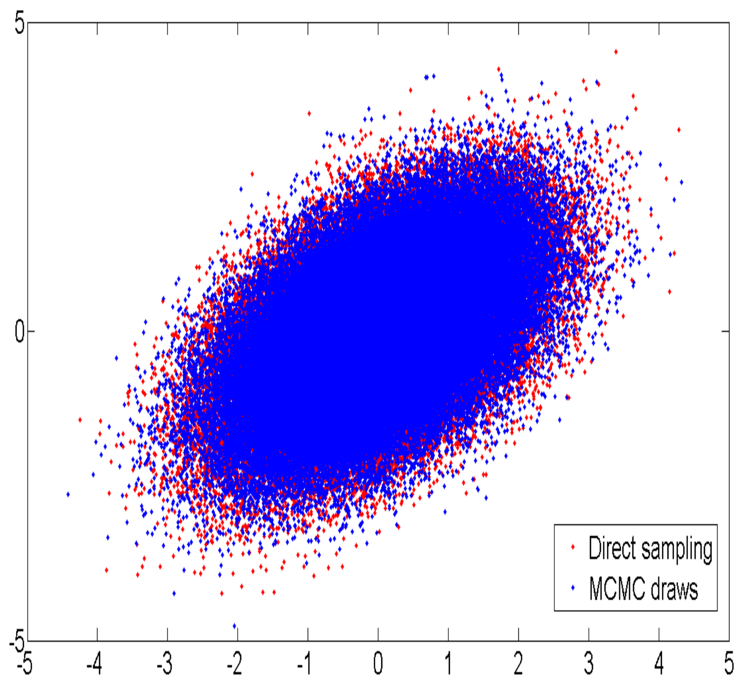
$$q' = [0,8332 \;\; 0,1668]$$

MCMC : Example

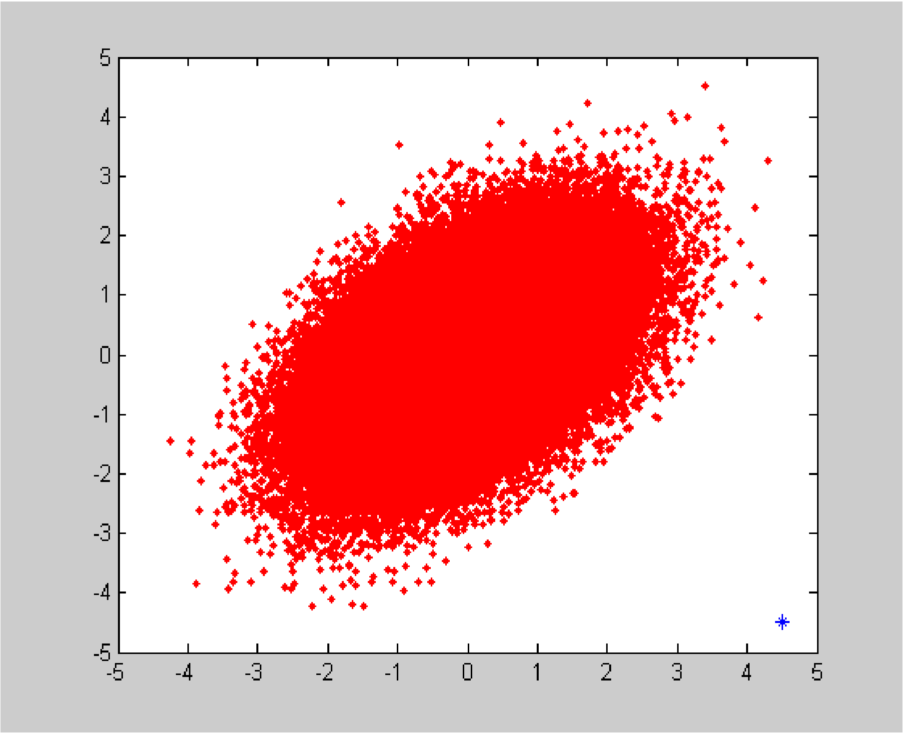
Inversion of the problem

Invariant Dist. : $N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0,5\\0,5 & 1\end{pmatrix}\right)$ + Ergodicity

Invariant Distribution



Ergodicity



MCMC

- Markov-chain exhibiting the posterior distribution as invariant one :

Sufficient condition : $$\pi(\theta|Y_{1:T})K(\theta^*|\theta) = \pi(\theta^*|Y_{1:T})K(\theta|\theta^*)$$

Reversibility condition – Detailed balance

- Markov-chain that is ergodic

→ From any initial point, the MC converges to the invariant distribution.

Sufficient condition :

Irreducible → Able to visit all sets A such that $\int_A \pi(\theta|Y_{1:T})d\theta > 0$ from any starting point

Aperiodic → Does not cycle through a finite number of sets

Positive Harris-recurrent

Beyond the scope of the course.

MCMC : pros and cons

- MCMC outputs : correlated draws of the posterior.

For any π-integrable real valued functions h,

$$\frac{1}{N}\sum_{i=1}^{N} h(\theta_i) \rightarrow \int h(\theta)\pi(\theta|Y_{1:T})d\theta \quad as \;\; N \rightarrow \infty, \;\; a.s.$$

- Posterior expectation : $h(\theta) = \theta$
- Posterior variance : $h(\theta) = (\theta - E(\theta|Y_{1:T}))(\theta - E(\theta|Y_{1:T}))'$

Rate of convergence :

Burn-in period

- Ergodicity implies convergence from any initial point but after how many iterations ?

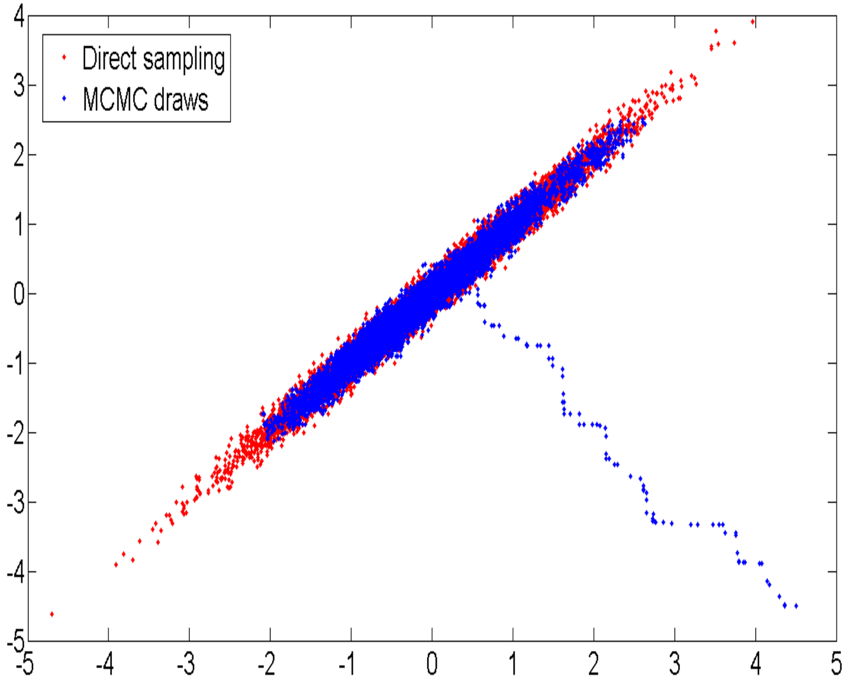
Criteria : Geweke, Gelman and Rubin, Cusum plot, ...

N ?

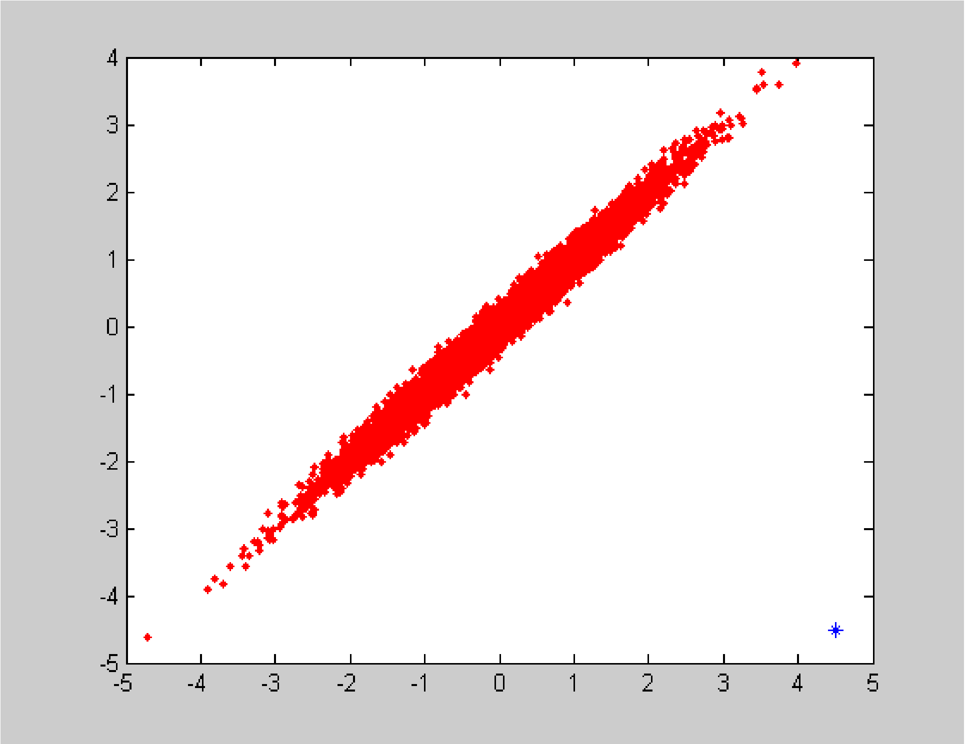
- How many iterations after convergence ? Criterion : Autocorrelation time

MCMC : Burn-in size

Invariant Dist. : $N(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & 0,99\\0,99 & 1\end{pmatrix})$



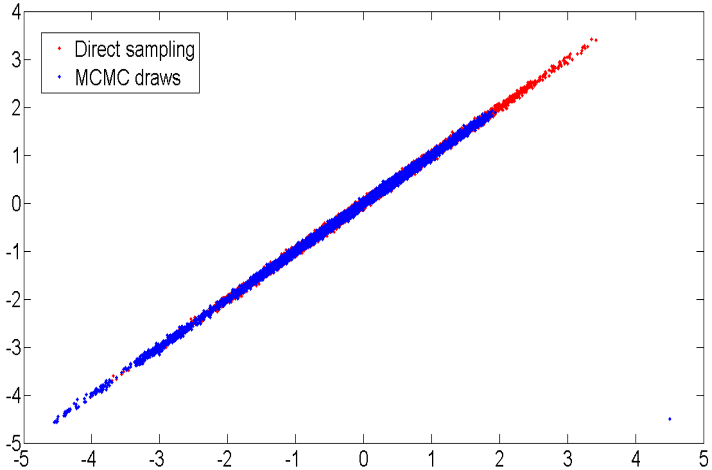
After 10000 draws



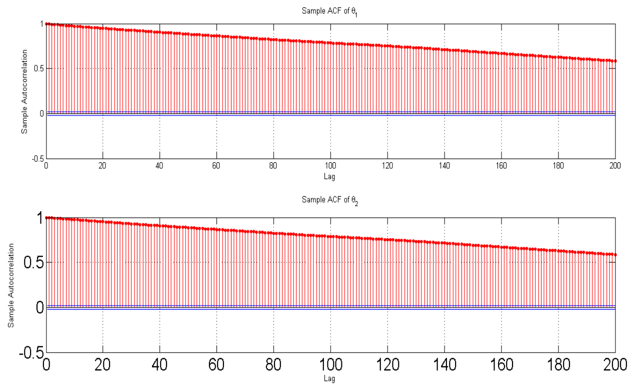
First 200 draws of the MCMC

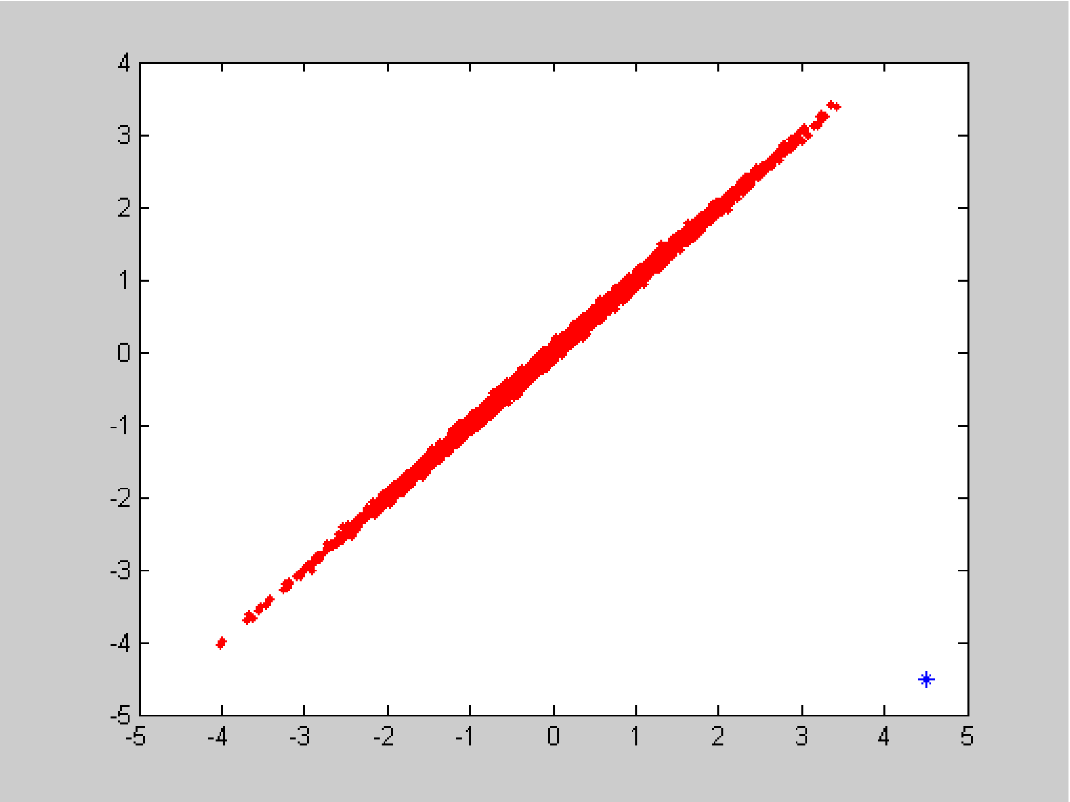
MCMC : Mixing problem

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



After 10000 draws





First 200 draws of the MCMC

MCMC : Gibbs sampler

- Iterations on *full conditional distributions*

Inference on $\theta = \{\theta_1', \theta_2', \theta_3'\}'$

Dimension of blocks can be higher than one

MCMC iterations

Initial state : $\{\theta_1^0, \theta_2^0, \theta_3^0\}$

1 $\theta_1^1 \sim \theta_1 | Y_{1:T}, \theta_2^0, \theta_3^0$ → $\theta_2^1 \sim \theta_2 | Y_{1:T}, \theta_1^1, \theta_3^0$ → $\theta_3^1 \sim \theta_3 | Y_{1:T}, \theta_1^1, \theta_2^1$

2 $\theta_1^2 \sim \theta_1 | Y_{1:T}, \theta_2^1, \theta_3^1$ → $\theta_2^2 \sim \theta_2 | Y_{1:T}, \theta_1^2, \theta_3^1$ → $\theta_3^2 \sim \theta_3 | Y_{1:T}, \theta_1^2, \theta_2^2$

... ...

Burn-in

Gibbs sampler : Example

- Auto-Regressive process (AR) : $Y_{1:T} = \{y_1, \ldots, y_T\}'$

$$
\begin{aligned}
y_t &= \theta_0 + \theta_1 y_{t-1} + \epsilon_t \\
&= \theta' x_t + \epsilon_t \\
\epsilon_t &\sim_{i.i.d.} N(0, \sigma^2)
\end{aligned}
$$

Prior distributions

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

$$\sigma^2 \sim IG(\alpha, \beta)$$

- Iterations on *full conditional distributions*

$$
\begin{aligned}
\pi(\theta | Y_{1:T}, \sigma^2) &\propto e^{\overbrace{-0{,}5(\theta-\mu_0)'\Sigma_0^{-1}(\theta-\mu_0)}^{\text{Prior component}}\overbrace{-0{,}5\sum_{t=1}^{T}\frac{(y_t-\theta' x_t)^2}{\sigma^2}}^{\text{Likelihood}}} \\
&\sim N(\bar{\mu}, \bar{\Sigma})
\end{aligned}
$$

$$\bar{\mu} = \bar{\Sigma}[\sigma^{-2}\sum_{t=1}^{T} x_t y_t + \Sigma_0^{-1}\mu_0] \qquad \bar{\Sigma} = [\sigma^{-2}\sum_{t=1}^{T}(x_t x_t') + \Sigma_0^{-1}]^{-1}$$

Gibbs sampler : Example

- Auto-Regressive process (AR) :

$$\begin{aligned} y_t &= \theta' x_t + \epsilon_t \\ \epsilon_t &\sim_{i.i.d.} N(0, \sigma^2) \end{aligned}$$

Prior distributions

$$\theta \sim N(\mu_0, \Sigma_0)$$
$$\sigma^2 \sim IG(\alpha, \beta)$$

- Iterations on *full conditional distributions*

$$\pi(\theta | Y_{1:T}, \sigma^2) \sim N(\bar{\mu}, \bar{\Sigma})$$

$$\pi(\sigma^2 | Y_{1:T}, \theta) \propto \underbrace{(\sigma^2)^{-\alpha-1} e^{-\beta\sigma^{-2}}}_{\text{Prior component}} \underbrace{(\sigma^2)^{-T/2} e^{-\sigma^{-2}[\sum_{t=1}^{T} \epsilon_t^2 / 2]}}_{\text{Likelihood}}$$

$$\sim IG(\alpha + T/2, \beta + \sum_{t=1}^{T} \epsilon_t^2 / 2)$$

- If Z~IG(a,b) then 1/Z~G(a,1/b)

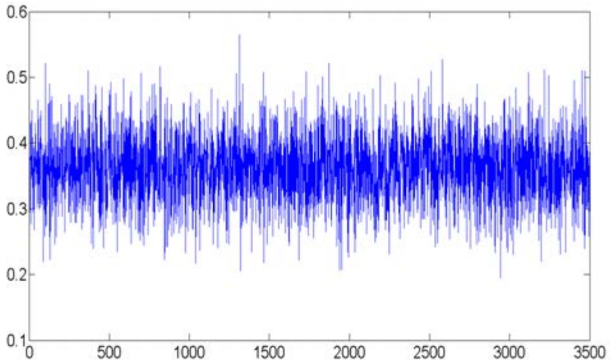
Gibbs sampler : Example

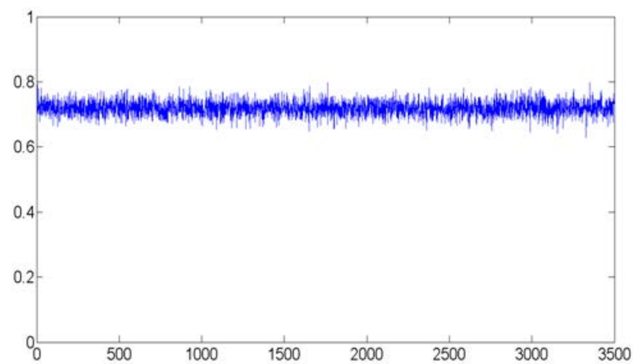
- AR(1) : 1000 Observations

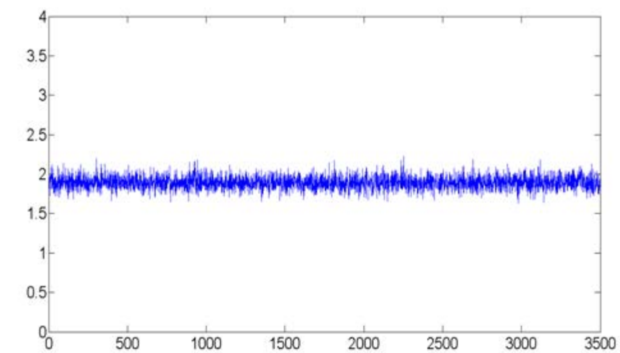
$$y_t = 0,4 + 0,7 y_{t-1} + \epsilon_t$$
$$\epsilon_t \sim_{i.i.d.} N(0,2)$$

Prior distributions

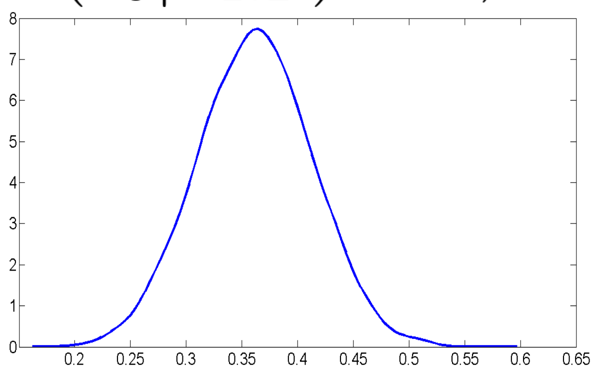
$$\theta \sim N([0 \quad 0]', 100 I_2)$$
$$\sigma^2 \sim IG(2,2)$$

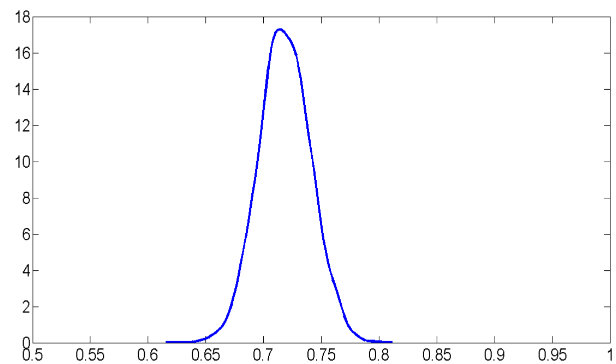


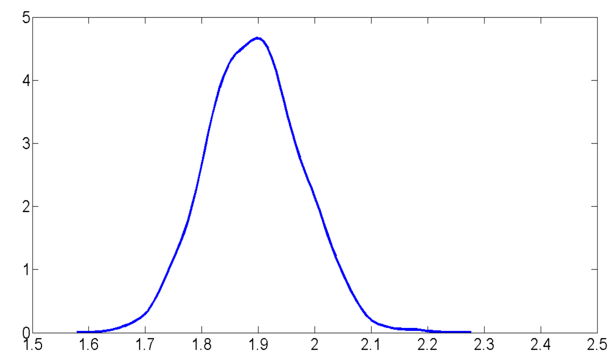




$$E(\theta_0 | Y_{1:T}) \approx 0,36 \qquad E(\theta_1 | Y_{1:T}) \approx 0,72 \qquad E(\sigma^2 | Y_{1:T}) \approx 1,89$$





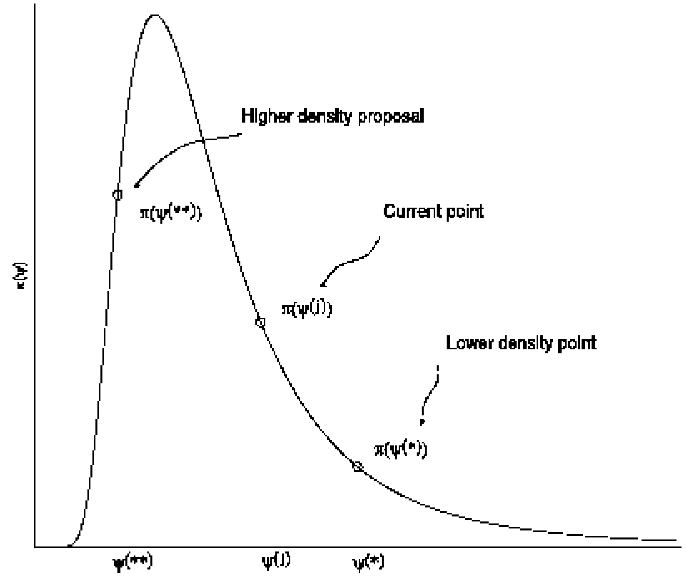


Metropolis-Hastings

- Conditional posterior distributions : too restrictive

- Metropolis-Hastings :

1- Draw a proposal parameter from any chosen distribution
2- Accept or reject the draw according to a probability which ensures that the invariant distribution of the MC is the posterior distribution of interest.



Metropolis-Hastings

- Let q be the proposal distribution (e.g. Normal)
- How to determine the probability function ?

Sufficient condition : $\pi(\theta|Y_{1:T})K(\theta^*|\theta) = \pi(\theta^*|Y_{1:T})K(\theta|\theta^*)$

Let assume that $\pi(\theta|Y_{1:T})q(\theta^*|\theta) > \pi(\theta^*|Y_{1:T})q(\theta|\theta^*)$

Move from $\theta \to \theta^*$ too often

Move from $\theta^* \to \theta$ too rarely

$$\pi(\theta|Y_{1:T})q(\theta^*|\theta)\alpha(\theta,\theta^*) = \pi(\theta^*|Y_{1:T})q(\theta|\theta^*)$$

$$\Updownarrow$$

$$\alpha(\theta^*,\theta) = \min[\frac{\pi(\theta^*|Y_{1:T})q(\theta|\theta^*)}{\pi(\theta|Y_{1:T})q(\theta^*|\theta)}, 1]$$

Since it is a probability

Metropolis-Hastings

- Initialize the MCMC
- At each MCMC iteration :
  - Generate a candidate from the proposal distribution

$$\theta^* \sim q(-|\theta)$$

  - Accept or reject the draw according to the probability

$$\phi(\theta^*, \theta) = \min[\frac{\pi(\theta^*|Y_{1:T})q(\theta|\theta^*)}{\pi(\theta|Y_{1:T})q(\theta^*|\theta)}, 1]$$

No need of the normalizing constant !

$$\phi(\theta^*, \theta) = \min[\frac{f(|Y_{1:T}|\theta^*)f(\theta^*)q(\theta|\theta^*)}{f(Y_{1:T}|\theta)f(\theta)q(\theta^*|\theta)}, 1]$$

M-H : Comments

- If q is symmetric : Metropolis algorithm

$$q(\theta|\theta^*) = q(\theta^*|\theta) \quad \text{then} \quad \alpha(\theta^*,\theta) = \min[\frac{f(|Y_{1:T}|\theta^*)f(\theta^*)}{f(Y_{1:T}|\theta)f(\theta)},1]$$

- Random Walk Metropolis : $q(\theta^*|\theta) \sim N(\theta,\bar{\Sigma})$

- Independent M-H : $q(\theta^*|\theta) \equiv q(\theta^*) \sim N(\bar{\theta},\bar{\Sigma})$

Caution : proposal dist : thicker tail than post dist.

- Gibbs sampler : $q(\theta^*|\theta) = \pi(\theta^*|Y_{1:T})$

$$\phi(\theta^*,\theta) = \min[\frac{\pi(\theta^*|Y_{1:T})\pi(\theta|Y_{1:T})}{\pi(\theta|Y_{1:T})\pi(\theta^*|Y_{1:T})},1] = 1 \quad \forall\ \theta^*$$

M-H can also be applied to multiple blocks

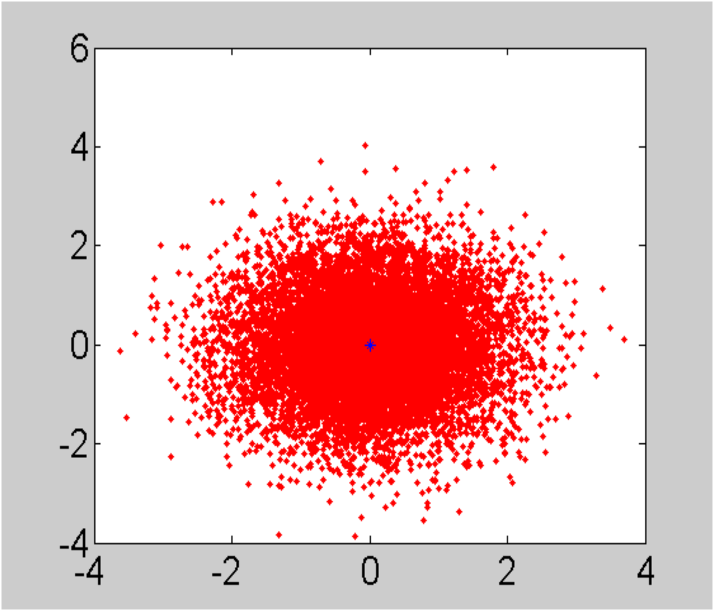
M-H : Comments

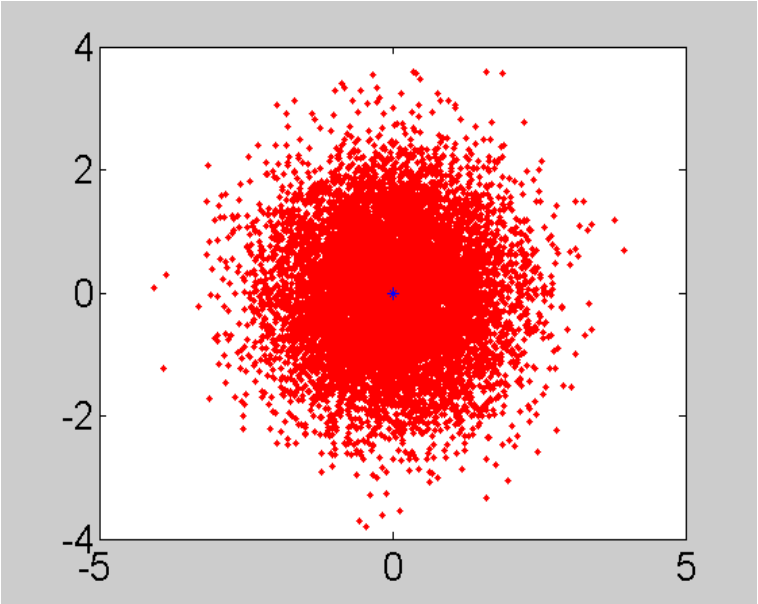
- Most frequent : Random Walk Metropolis - $q(\theta^*|\theta) \sim N(\theta, \bar{\Sigma})$
- How to choose the variance parameter ?

Hot topic in the literature

If $\bar{\Sigma}$ too small $\longrightarrow$ Too many similar parameter values : Slow exploration of the support

If $\bar{\Sigma}$ too big $\longrightarrow$ Too many rejected values : Slow exploration of the support





M-H : Comments

- Most frequent : Random Walk Metropolis - $q(\theta^*|\theta) \sim N(\theta, \bar{\Sigma})$
- How to choose the variance parameter ?

Hot topic in the literature

If Posterior distribution = Multivariate Normal distribution

Optimal acceptance rate

Dimension 1	Dimension 5	...	Large Dimension
Accept. Rate = 44 %	Accept. Rate = 28 %		Accept. Rate = 23,4 %

Reference : Roberts, G. O. & Rosenthal, J. S. *'Optimal scaling for various Metropolis-Hastings algorithms'*, Statistical Science, 2001, 16, 351-367

M-H : Comments

- Most frequent : Random Walk Metropolis - $q(\theta^*|\theta) \sim N(\theta, \bar{\Sigma})$
- How to choose the variance parameter ?

Common practice : Find the Optimal acceptance rate

- By trials and errors → Very demanding, Model dependent
- By adaptive RW metropolis

Reference : Atchadé, Y. & Rosenthal, J. *'On adaptive Markov chain Monte Carlo algorithms'*, Bernoulli, 2005, 11(5), 815-828

- At MCMC iteration *i*

$$\bar{\Sigma}_i = \bar{\Sigma}_{i-1} + (\phi_r^{i-1} - \phi_{target})/(i^c) \quad \text{if} \quad \Sigma_{Low} < \bar{\Sigma}_i < \Sigma_{High}$$

With ϕ_r^{i-1} the acceptance rate at iteration *i-1*

ϕ_{target} the targeted acceptance rate

$$0.5 < c < 1$$

MH sampler : Example

- GARCH process : $Y_{1:T} = \{y_1, ..., y_T\}'$

$$
\begin{aligned}
y_t &= \epsilon_t \\
\sigma_t^2 &= \omega + \alpha\epsilon_{t-1} + \beta\sigma_{t-1}^2 \\
\epsilon_t | Y_{1:t-1} &\sim_{i.i.d.} N(0, \sigma_t^2)
\end{aligned}
$$

Prior distributions

$$\omega \sim U(0, 10) \;|\; \alpha \sim U(0, 1)$$

$$\beta | \alpha \sim U(0, 1 - \alpha)$$

- Adaptive RW Metropolis with block of one dimension

DGP of the simulated time series

$$T = 3000 \qquad \omega = 0.1 \qquad \alpha = 0.15 \qquad \beta = 0.8$$

Choice of the MCMC parameters

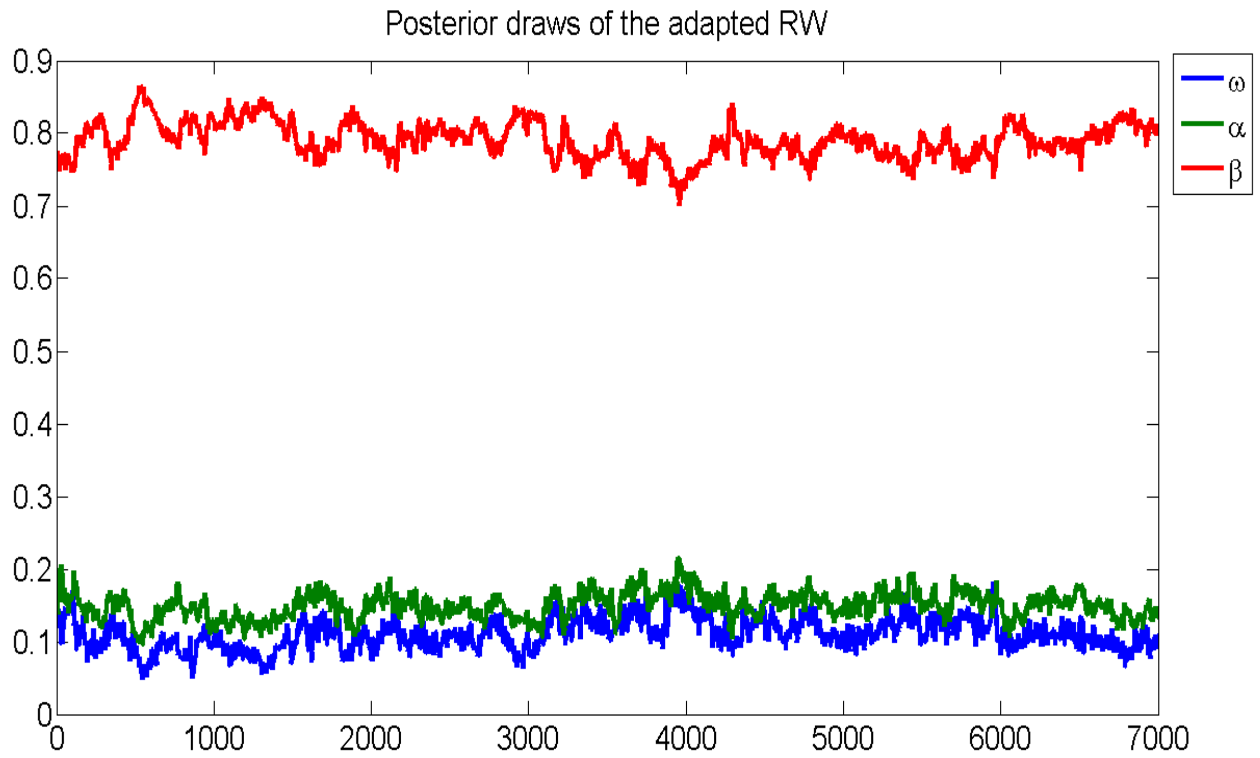
$$\phi_{target} = 0.44 \qquad c = 0.6 \qquad \Sigma_{Low} = 1e - 5 \qquad \Sigma_{High} = 10$$

$$N = 10000 \qquad \text{Burn-in} = 3000$$

MH sampler : Example

- GARCH process :

$$T = 3000 \qquad \omega = 0.1 \qquad \alpha = 0.15 \qquad \beta = 0.8$$



- Acc. Rate :

$$\phi_\omega = \phi_\alpha = \phi_\beta = 0,44$$

- Post means :

$$E(\omega|Y_{1:T}) \approx 0.11$$

$$E(\alpha|Y_{1:T}) \approx 0.15$$

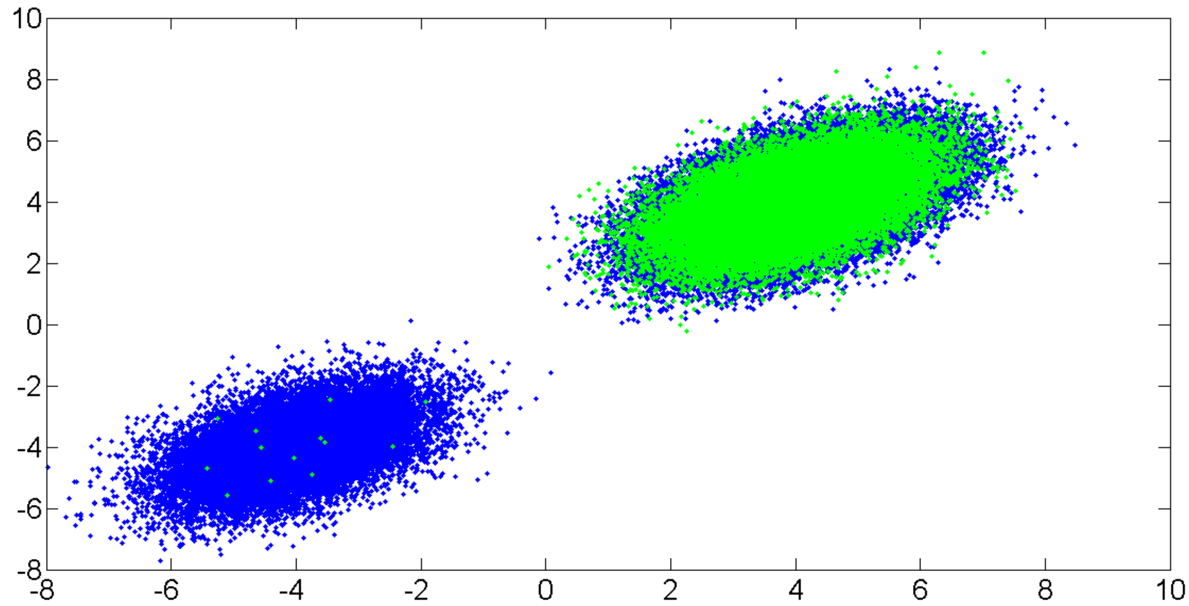
$$E(\beta|Y_{1:T}) \approx 0.79$$

MCMC : limitations

- Difficult to assess if the MC has converged to the post. Dist.
- Difficult to assess how much MCMC draws are required.
- Difficult to infer on multi-modal posterior distributions :

RW Metropolis : Jump from the current MCMC parameter

→ Unlikely to jump from one mode to another.



Questions ?

Model selection

Criteria

- Two popular Criteria :

Marginal likelihood

Intuitive criterion derived from Bayes' rule

Deviance Information Criterion (DIC)

Perfect when Marginal likelihood is out of reach

Spiegelhalter, D.; Best, D.; Carlin, B. & van der Linde, A. *'Bayesian measures of model complexity and fit'*, Journal of Royal Statistical Society, Series B, 2002, 64, 583-639

→ Focus on Marginal likelihood

Marginal likelihood

- Marginal likelihood = Normalizing constant :

How to Choose between two models : M_1 or M_2 ?

Bayes' theorem

$$\pi(M_1|Y_{1:T}) = \frac{f(Y_{1:T}|M_1)f(M_1)}{f(Y_{1:T}|M_1)f(M_1) + f(Y_{1:T}|M_2)f(M_2)}$$

$$\text{Choose } M_1 \text{ if } \pi(M_1|Y_{1:T}) > \pi(M_2|Y_{1:T})$$

If no subjective idea over the two models : $f(M_1) = f(M_2) = 0.5$

$$\pi(M_1|Y_{1:T}) > \pi(M_2|Y_{1:T}) \Leftrightarrow f(Y_{1:T}|M_1) > f(Y_{1:T}|M_2)$$

Marginal likelihood

- Multiple Models :

Bayes' theorem

$$\pi(M_j|Y_{1:T}) = \frac{f(Y_{1:T}|M_j)f(M_j)}{\sum_{i=1}^{k} f(Y_{1:T}|M_i)f(M_i)}$$

$$\text{Choose } M_j \text{ that maximizes } \pi(M_j|Y_{1:T}) \ \ \forall\, j \in [1,k]$$

If no subjective idea over the different models : $f(M_j) = \frac{1}{k}$

$$\text{Choose } M_j \text{ that maximizes } f(Y_{1:T}|M_j) \ \ \forall\, j \in [1,k]$$

Bayesian Model Averaging (BMA)

- Multiple Models :

Bayes' theorem

$$\pi(M_j|Y_{1:T}) = \frac{f(Y_{1:T}|M_j)f(M_j)}{\sum_{i=1}^{k} f(Y_{1:T}|M_i)f(M_i)}$$

Instead of choosing one model, keep them all

→ Take into account the model uncertainty

Predictive density : $\pi(y_{T+1}|Y_{1:T}) = \sum_{j=1}^{k} \pi(y_{T+1}|Y_{1:T}, M_j)\pi(M_j|Y_{1:T})$

Point forecast : $E(y_{T+1}|Y_{1:T}) = \sum_{j=1}^{k} [\int y_{T+1}\pi(y_{T+1}|Y_{1:T}, M_j)dy_{T+1}]\pi(M_j|Y_{1:T})$

Marginal likelihood

- Quantity of interest : $f(Y_{1:T}|M_j)$
- Local Formula :

$$f(Y_{1:T}|M_j) = \frac{\overbrace{f(Y_{1:T}|\Theta^*)}^{\text{Likelihood}}\overbrace{f(\Theta^*)}^{\text{Prior}}}{\underbrace{\pi(\Theta^*|Y_{1:T})}_{\text{Posterior}}} \quad \forall\ \Theta^* \text{such that } \pi(\Theta^*|Y_{1:T}) > 0$$

- Ockham's razor :

Likelihood	increases as long as the model complexity grows
Prior and Posterior	penalize the model complexity

Marginal likelihood

- Quantity of interest : $f(Y_{1:T}|M_j)$

$$f(Y_{1:T}|M_j) = \int_{\Omega} f(Y_{1:T}|M_j,\Theta) f(\Theta|M_j) d\Theta$$

If complex model with many parameters :

→ Highly dimensional integration : difficult to compute.

Dimension < 3	Middle Dimension	High Dimension
• Numerical Integration • Importance sampling	• Bridge sampling	• Path sampling • SMC sampler • MCMC • Variational Bayes

Marginal likelihood by MCMC

- Quantity of interest : $f(Y_{1:T}|M_j)$

$$f(Y_{1:T}|M_j) = \int_{\Omega} f(Y_{1:T}|M_j,\Theta) f(\Theta|M_j) d\Theta$$

- Local Formula :

$$f(Y_{1:T}|M_j) = \frac{\overbrace{f(Y_{1:T}|\Theta^*)}^{\text{Likelihood}}\overbrace{f(\Theta^*)}^{\text{Prior}}}{\underbrace{\pi(\Theta^*|Y_{1:T})}_{\text{Posterior}}} \quad \forall\ \Theta^* \text{such that } \pi(\Theta^*|Y_{1:T}) > 0$$

<u>The posterior density is the most tricky part.</u>

Marginal likelihood by MCMC

- Local Formula :

$$f(Y_{1:T}|M_j) = \frac{\overbrace{f(Y_{1:T}|\Theta^*)}^{\text{Likelihood}}\overbrace{f(\Theta^*)}^{\text{Prior}}}{\underbrace{\pi(\Theta^*|Y_{1:T})}_{\text{Posterior}}} \quad \forall\ \Theta^* \text{such that } \pi(\Theta^*|Y_{1:T}) > 0$$

- Marginal likelihood from Gibbs sampler :

> Chib, S. *'Marginal Likelihood from the Gibbs Output'*, Journal of the American Statistical Association, 1995, 90, 1313-1321

- Marginal likelihood from MH sampler :

> Chib, S. & Jeliazkov, I. *'Marginal Likelihood from the Metropolis-Hastings Output'*, Journal of the American Statistical Association, 2001, 96, 270-281

Marginal likelihood : Example

- ML for AR processes : $y_t|Y_{1:t-1} \quad \sim \quad N(\theta' x_t, \sigma^2)$

MCMC iterations

$$\begin{cases} \pi(\theta|Y_{1:T}, \sigma^2) & \sim \quad N(\bar{\mu}, \bar{\Sigma}) \\ \pi(\sigma^2|Y_{1:T}, \theta) & \sim \quad IG(\alpha + T/2, \beta + \sum_{t=1}^{T} \epsilon_t^2/2) \end{cases}$$

Choose a high density point : $\{\theta^*, \sigma^{2*}\}$

1) **Likelihood :** $f(Y_{1:T}|\theta^*, \sigma^{2*}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^{2*}}} e^{-0,5(y_t - \theta'^* x_t)^2/\sigma^{2*}}$

2) **Prior :** $f(\theta^*) \sim N(\mu_0, \Sigma_0) \qquad f(\sigma^{2*}) \sim IG(\alpha, \beta)$

3) **Posterior :** $\pi(\theta^*, \sigma^{2*}|Y_{1:T}) = \pi(\sigma^{2*}|Y_{1:T})\pi(\theta^*|Y_{1:T}, \sigma^{2*})$

NB : $\pi(\sigma^{2*}|Y_{1:T}) = \int \pi(\sigma^{2*}|Y_{1:T}, \theta)\pi(\theta|Y_{1:T})d\theta \approx \frac{1}{N}\sum_{i=1}^{N} \pi(\sigma^{2*}|Y_{1:T}, \theta^i)$

Marginal likelihood : Example

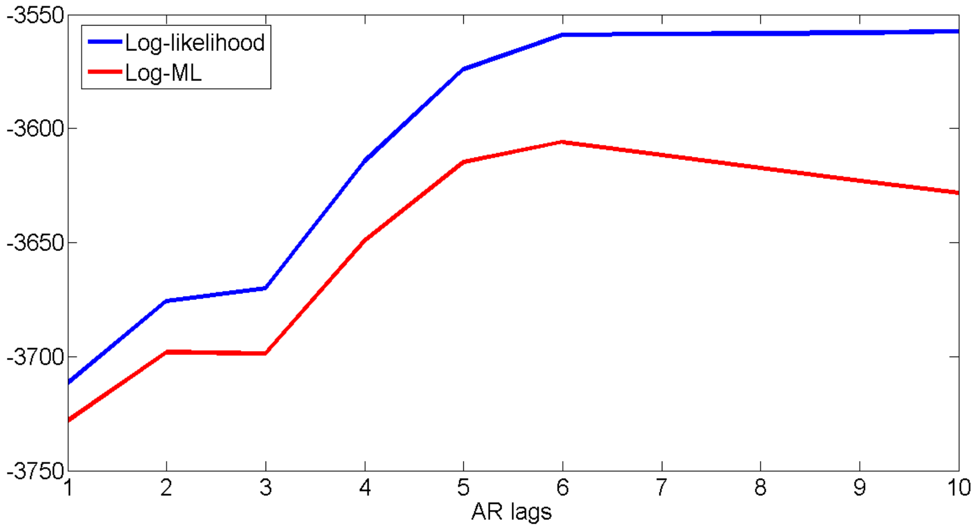
- ML for AR processes : $y_t | Y_{1:t-1} \sim N(\theta' x_t, \sigma^2)$

DGP of the simulated time series

$$T = 2000 \qquad \text{AR lags} = 6$$

- Log Marginal Likelihood :

1	2	3	4	5	6	7	8	9	10
-3728	-3698	-3697	-3649	-3614	-3606	-3611	-3617	-3623	-3628



Questions ?

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

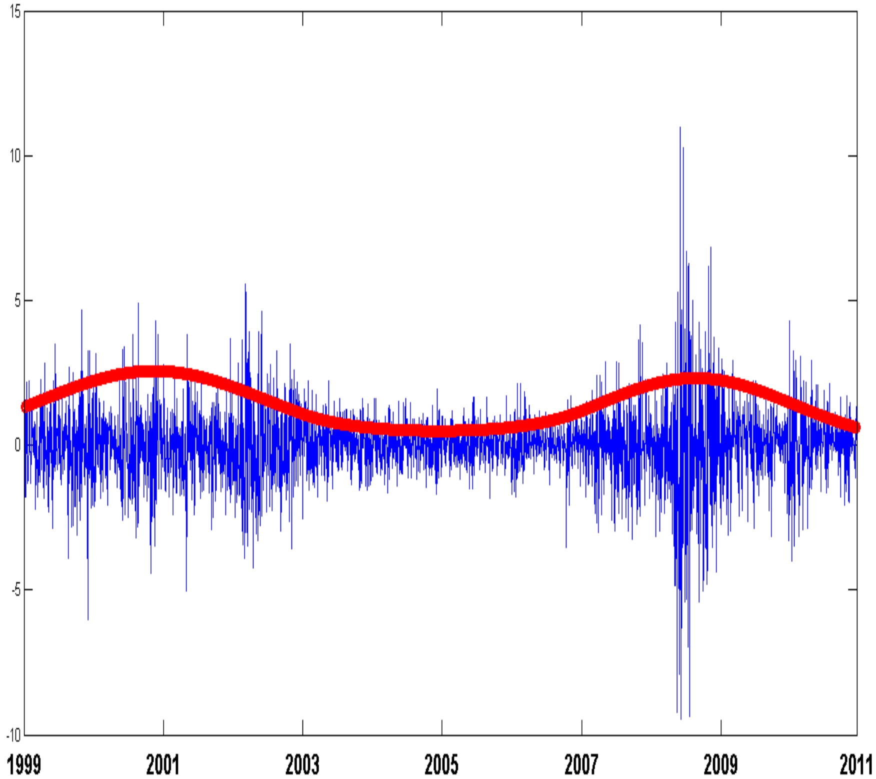
- Stylized fact of many time series
  - High persistence – almost integrated series

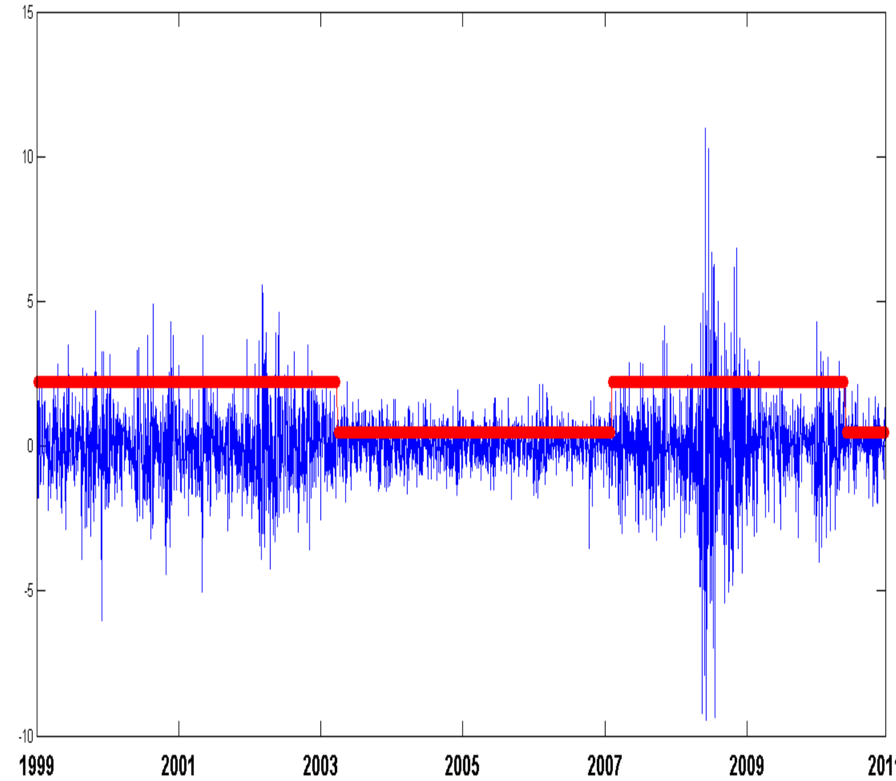
  → Unit root model → No predictability !
  - Long-run dynamic evolves over time

Structural breaks can cause these stylized facts

Long run dynamic : S&P 500

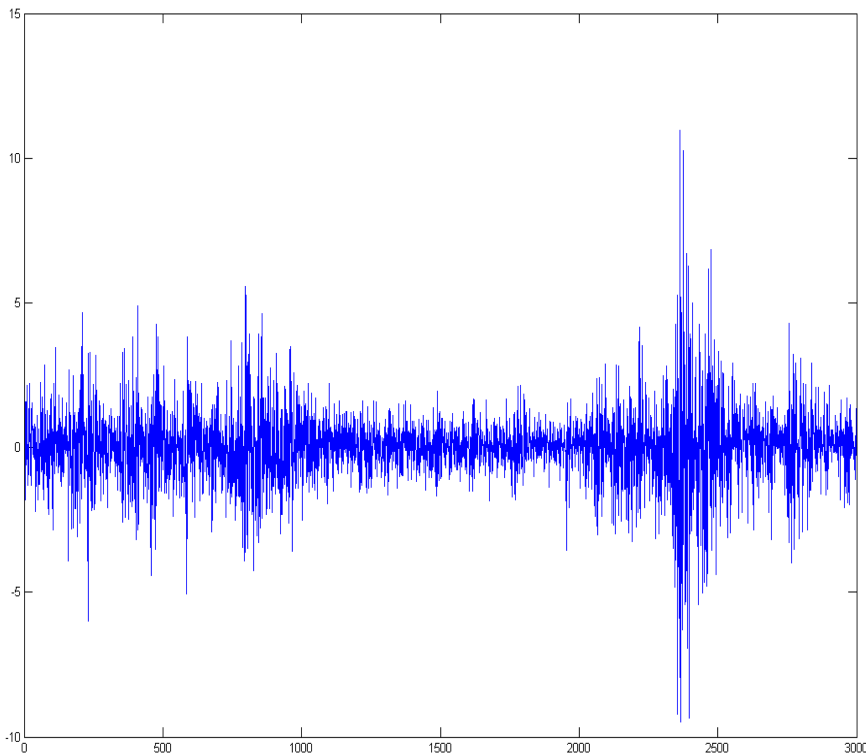
Spline GARCH (Engle, Rangel, 2013)	Markov-Switching GARCH (Bauwens,Dufays,Rombouts, 2013)





Long run volatility evolves over time

Example : S&P 500

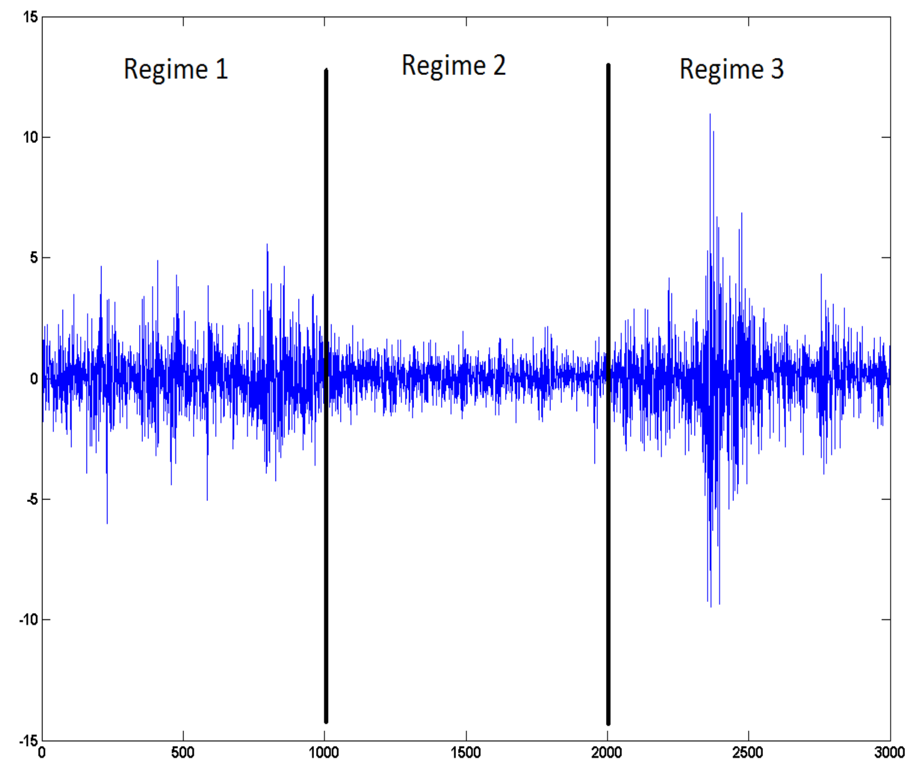


$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1} + \beta\sigma_{t-1}^2$$

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

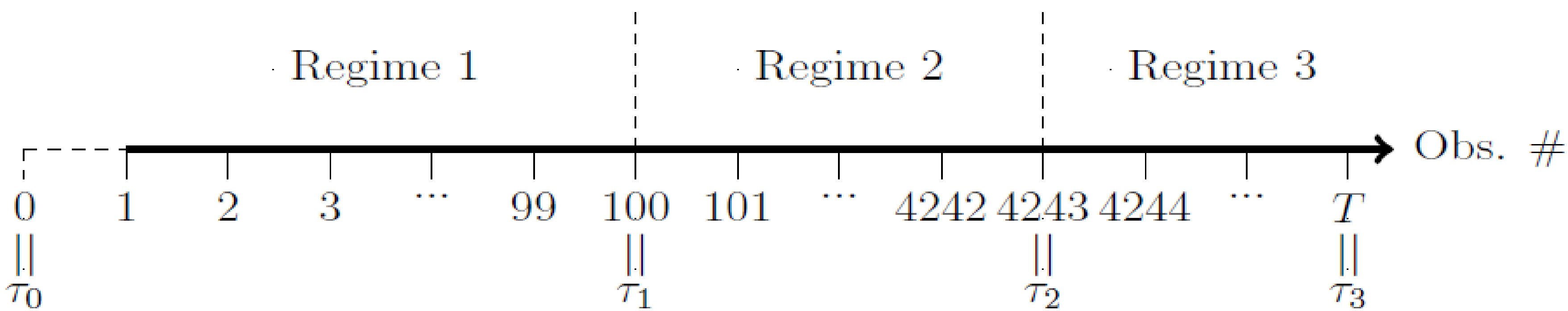
Change-point models

Change-point models

CP models :

- ***Non recurrent regimes***

→ Not stationary – In line with economic theories



- ***Forecasts***

→ Automatically select the optimal window size

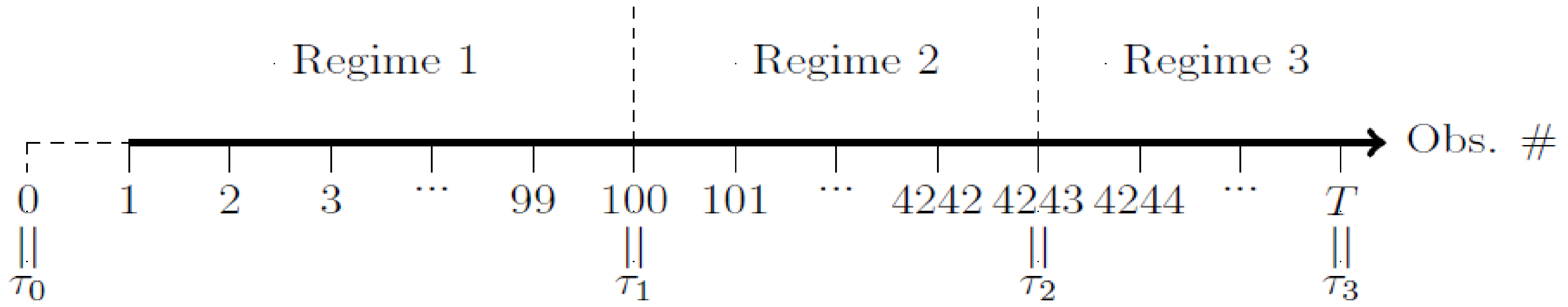
→ Predictions based on the last sub-sample

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 :

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

- 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 :

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

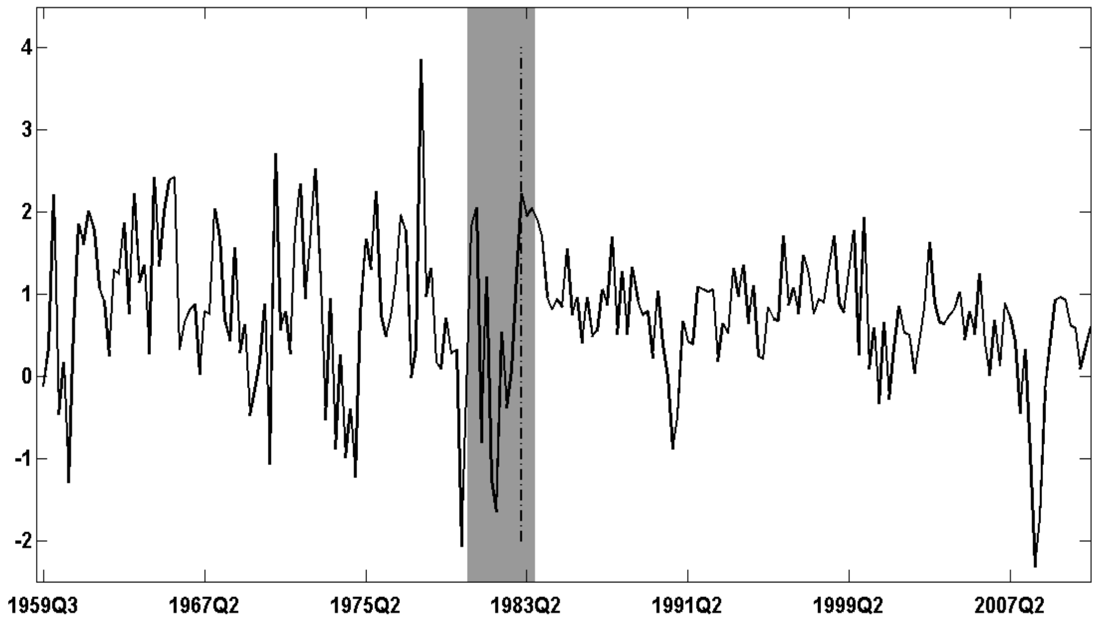
→ Discrete conditional distribution

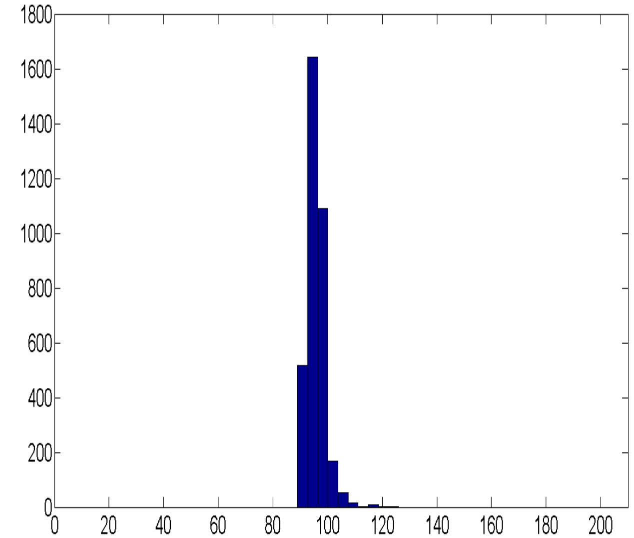
1) Compute the posterior density for each *i* and normalize.
2) Draw u ~ U(0,1)
3) Find τ such that

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

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 :

$\tau_0 = 0$

$\tau_{K+1} = T$

$\tau_1 \sim U(1, T - (K+2))$

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

Stephens (1994)

- Gibbs sampler : if AR process as before

$$\begin{aligned}
\theta_i | Y_{1:T}, \Theta_{-i}, \tau, \Sigma &\sim N(\bar{\mu}_i, \bar{\Sigma}_i) \\
\sigma_i^2 | Y_{1:T}, \tau, \Theta, \Sigma_{-i} &\sim IG(\alpha + (\tau_i - \tau_{i-1})/2, \beta + \sum_{t=\tau_{i-1}+1}^{\tau_i} \epsilon_t) \\
\tau_i | Y_{1:T}, \Theta, \Sigma, \tau_{-i} &\sim \text{Griddy-Gibbs}
\end{aligned}$$

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

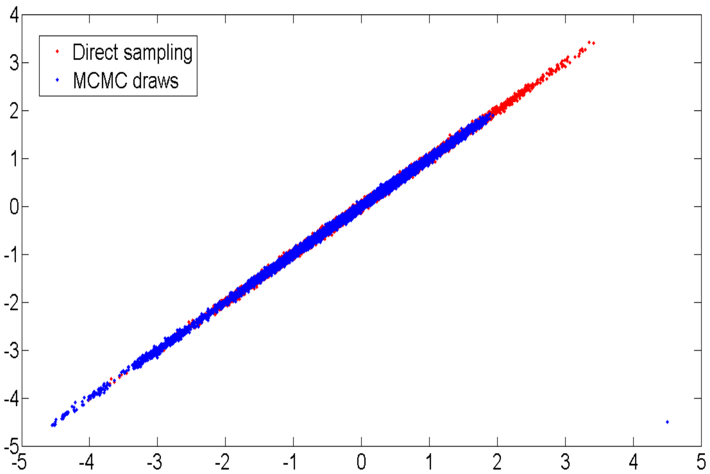
Drawbacks

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

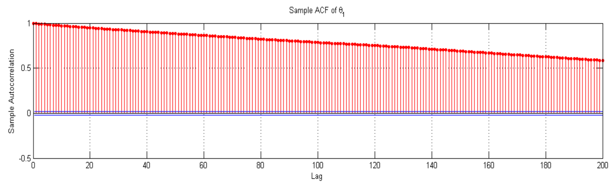
→ **May not converge in a finite amount of time !**

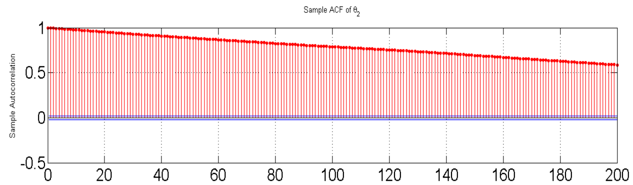
MCMC : Mixing problem

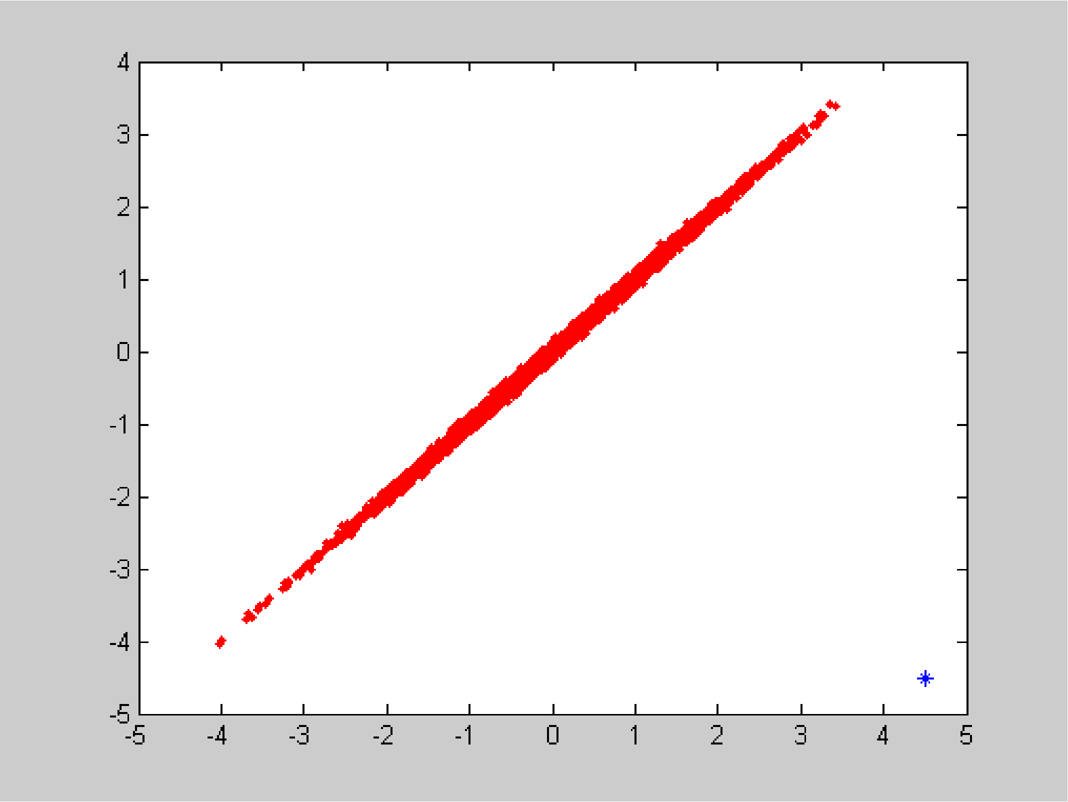
Invariant Dist. : $N(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0,999\\0,999 & 1\end{pmatrix})$



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 :

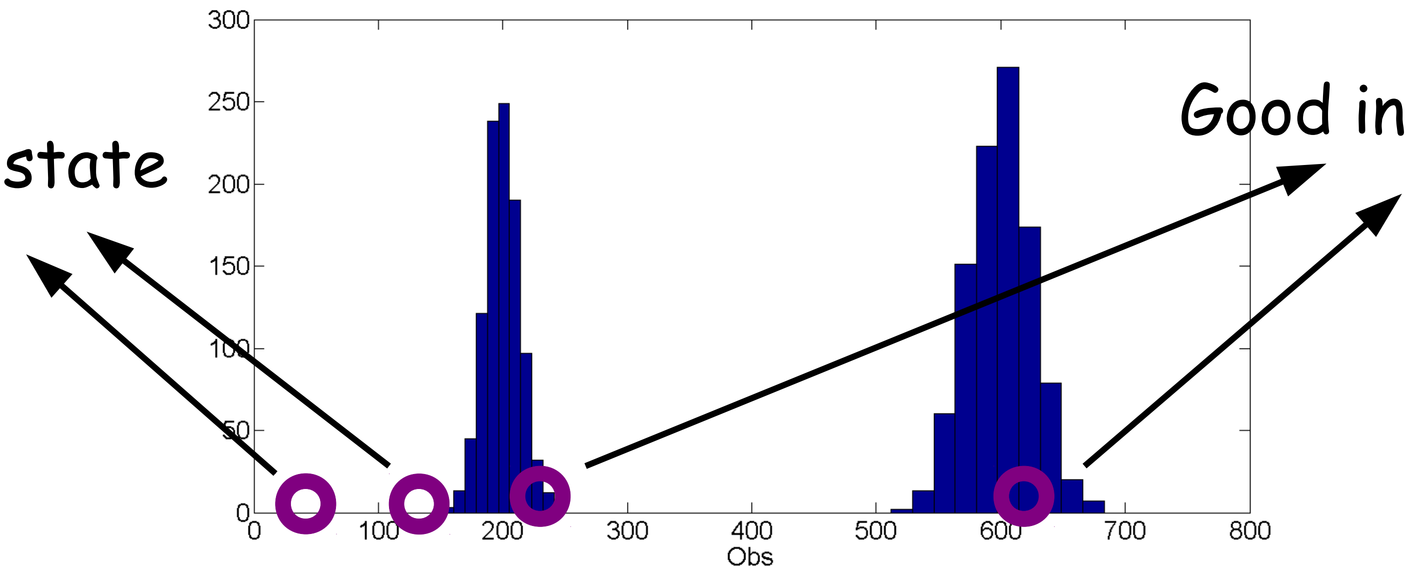
- Single-move of Bauwens,Preminger,Rombouts (2011)

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)

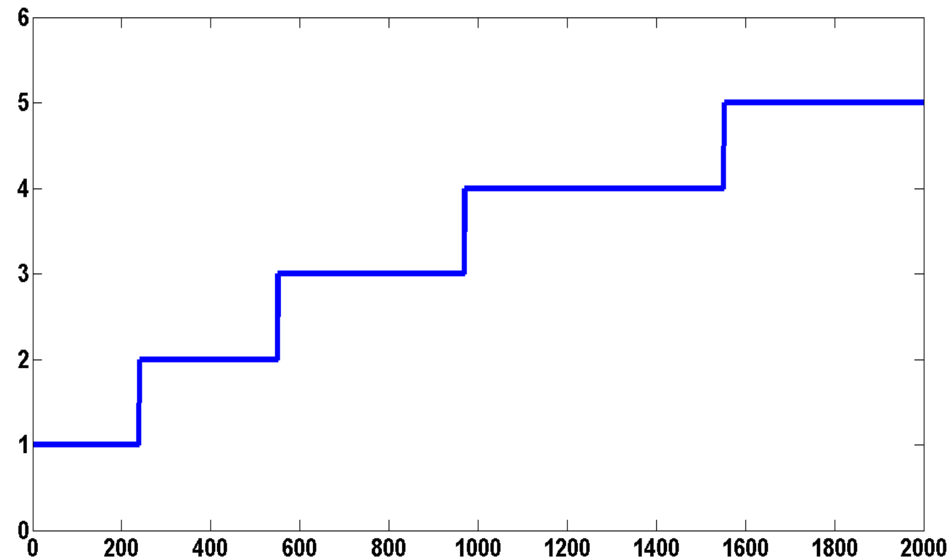
$S_{1:T}$ is driven by a Markov-chain with transition matrix P

Example :

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

Change-point configuration

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



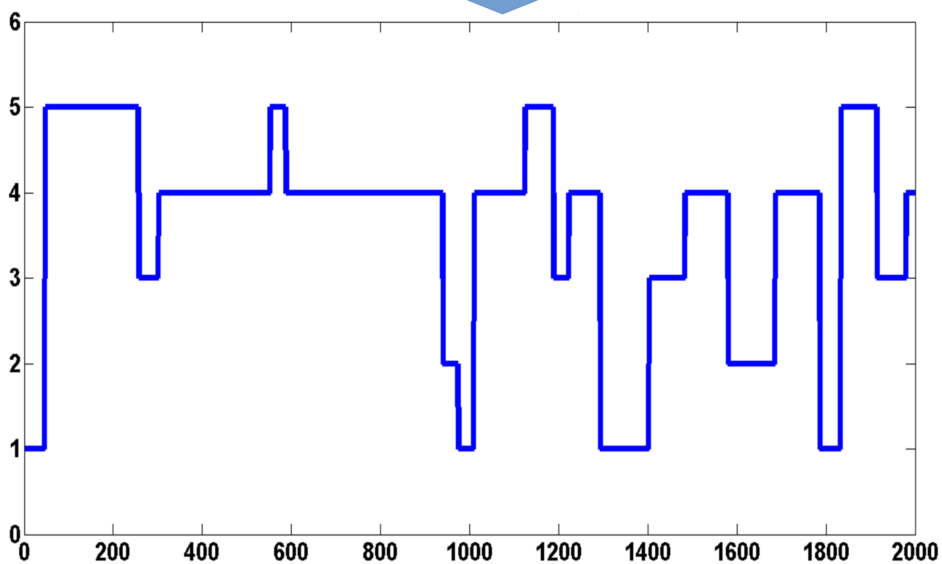
No recurrent state !

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 :

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

$$\sigma^2_i | Y_{1:T}, S_{1:T}, P, \Theta, \Sigma_{-i} \sim IG(\alpha + (n_{i,i})/2, \beta + \sum_{t=1}^{T} \epsilon_t \delta_{s_t = i})$$

$$P | Y_{1:T}, S_{1:T}, \Theta, \Sigma \sim \prod_{i=1}^{K+1} \text{Dirichlet}(\eta + n_{i,1:K+1})$$

$$S_{1:T} | Y_{1:T}, P, \Theta, \Sigma \sim \text{Forward-Backward algorithm}$$

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

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

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

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 : $\pi(s_t|Y_{1:T}, s_{t+1:T}) \;\; \forall t \in [1, T-1]$

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

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 :

$$\begin{cases} f(P) = \prod_{i=1}^{K+1} f(p_{i,1:K+1}) \\ p_{i,1:K+1} \sim \text{Dirichlet}(\eta_1, ..., \eta_{K+1}) \equiv \text{Dir}(\eta) \end{cases}$$

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 ?

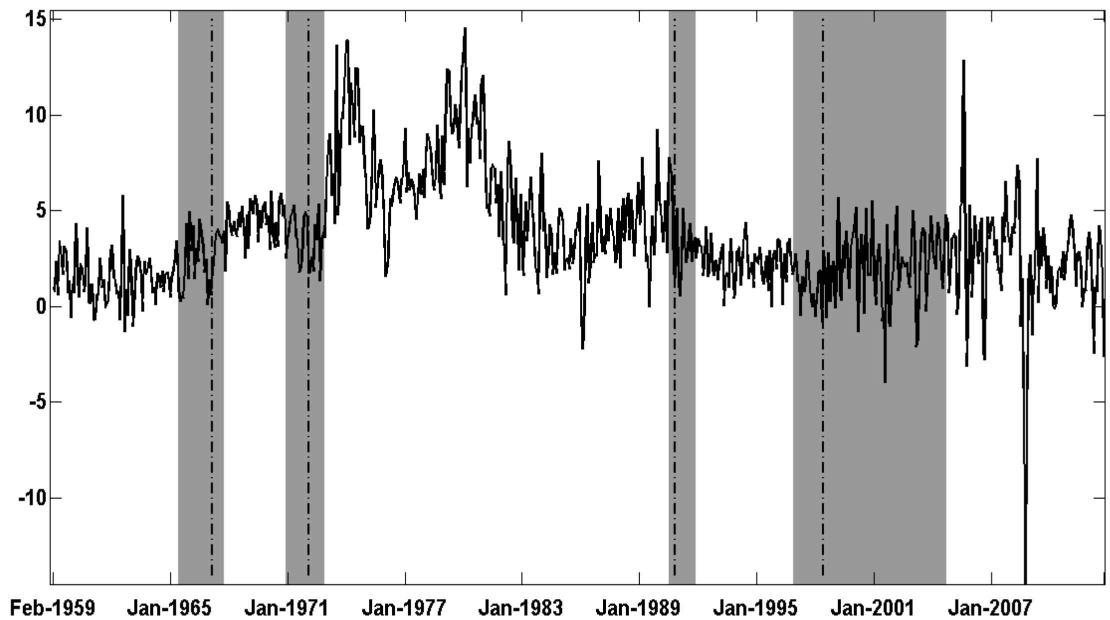
$\longrightarrow$ **By MCMC or by Bridge sampling**

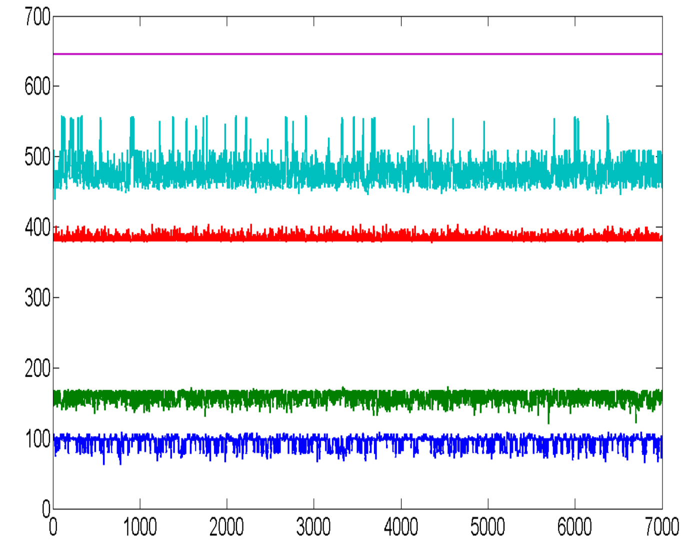
By Importance sampling

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

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

$\longrightarrow$ **If new regime : draw parameters from the hier. dist.**

Predictions of structural breaks

- Hierarchical distributions

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

Mean parameters

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

- 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 → $\left\{ \mu_0, \Sigma_0 \right\}$ → New regimes

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

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

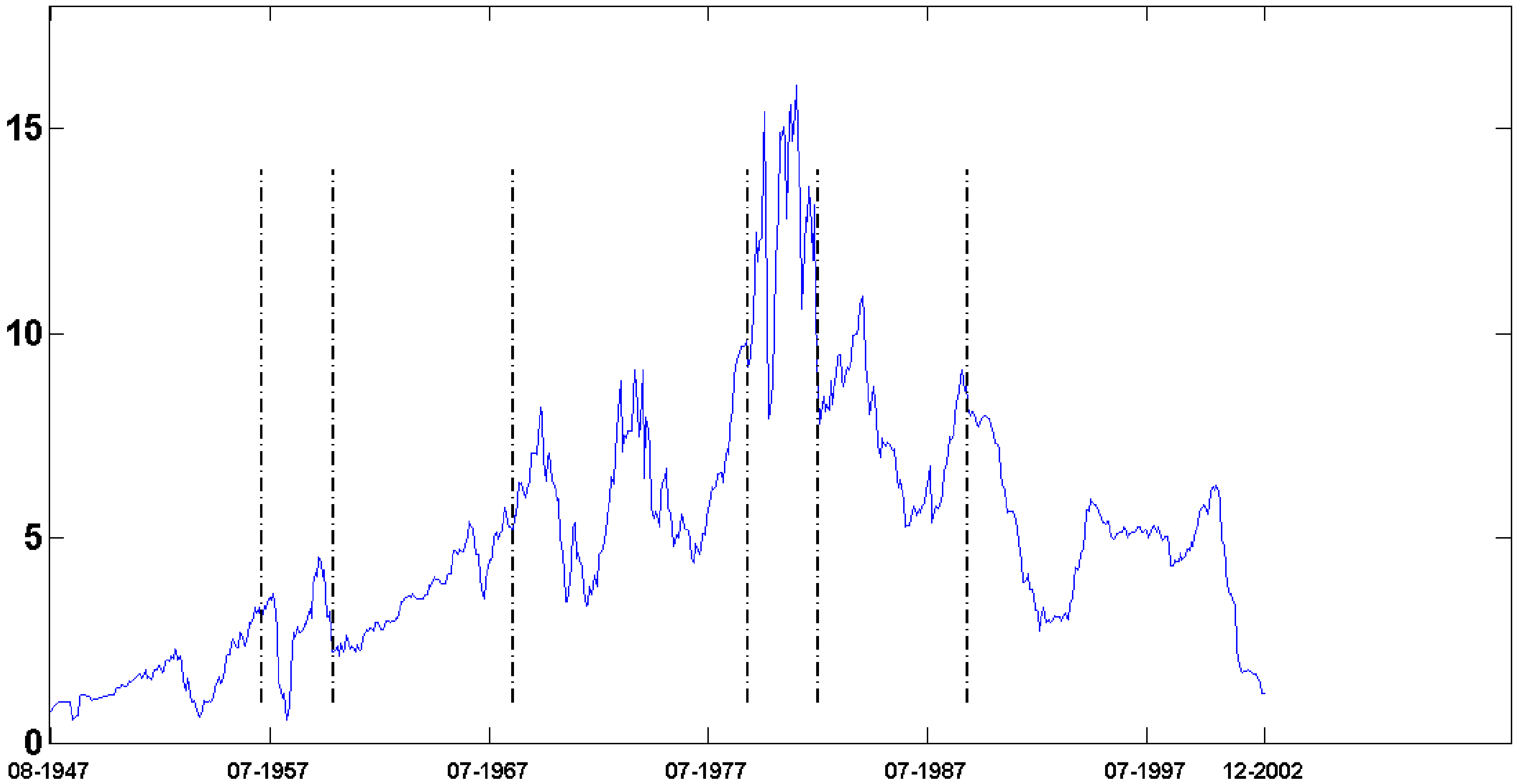
...

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

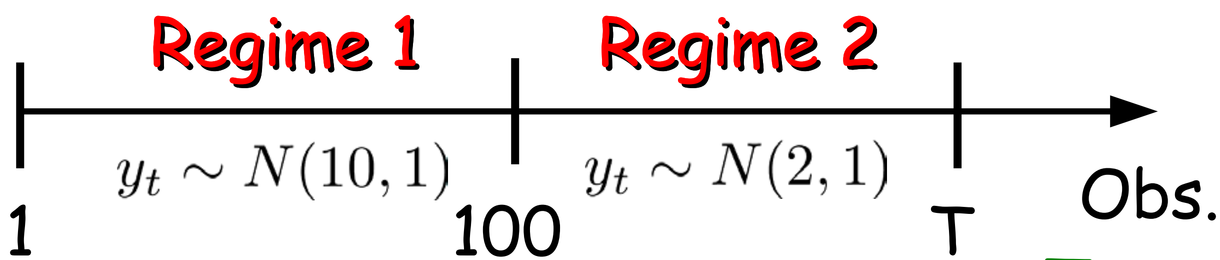
- Issue arises in MS specification (recurrent states)

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 :



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

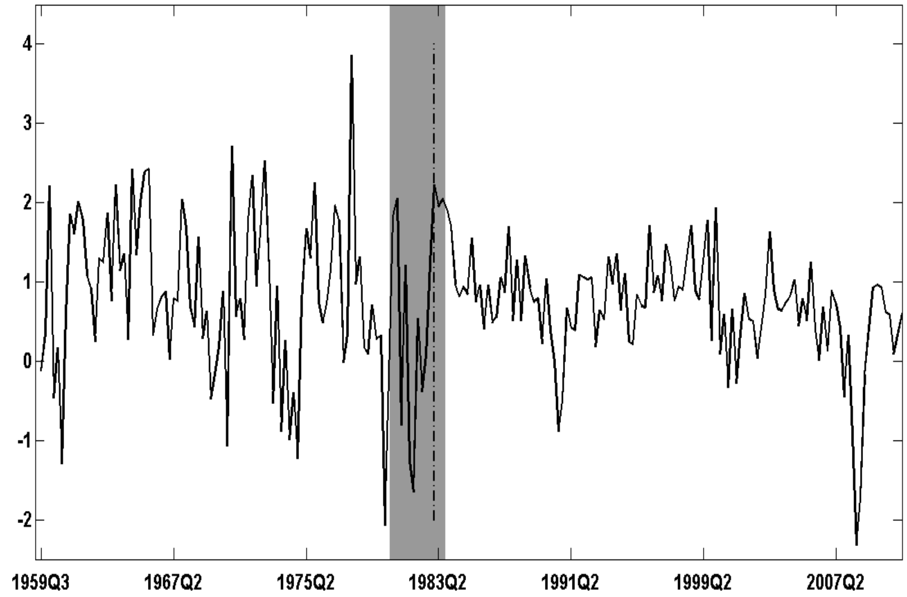
$\theta_1 = 2, \theta_2 = 10$ → $s_{1:100} = 2$ $\quad s_{101:T} = 1$

} Same likelihood

Misleading if it happens during the MCMC algorithm

Example : US GDP Growth

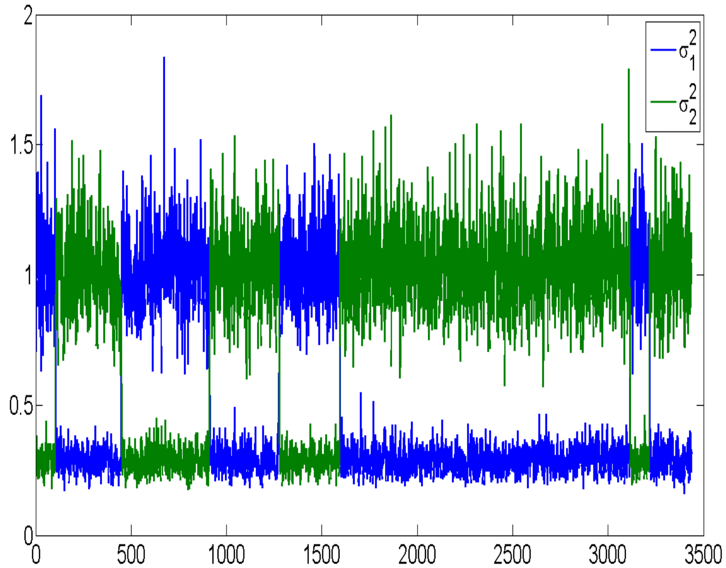
- Great moderation : Drop of the variance



- Could be estimated by an MS-AR model

Example : US GDP Growth

Label switching problem

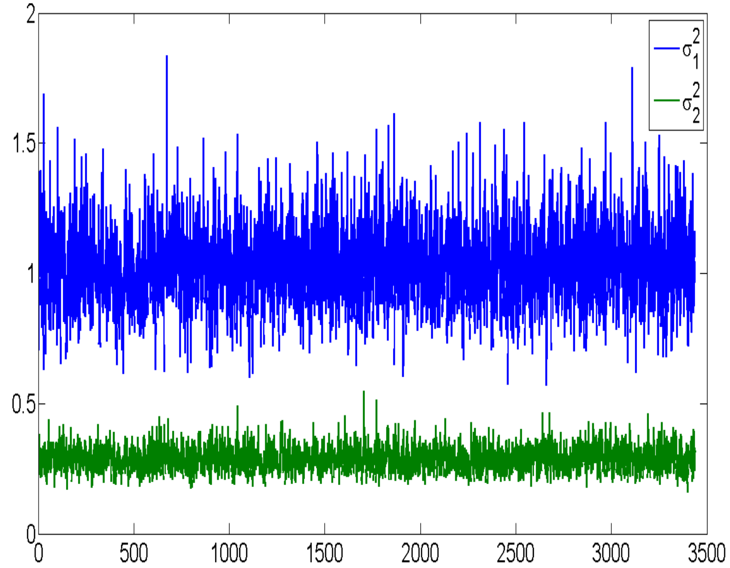


- Posterior means are not safe !

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

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

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

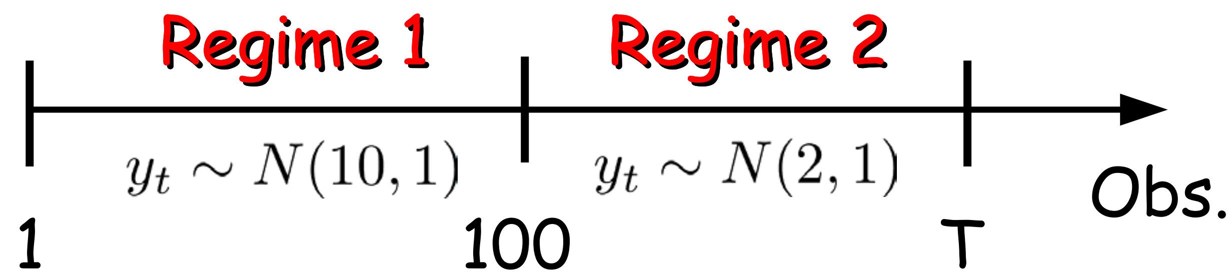
- Solutions

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**

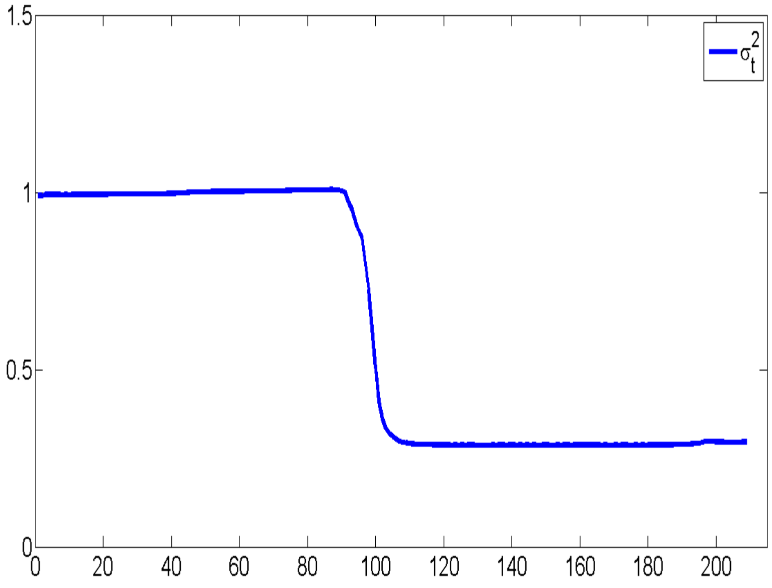
- 3) Only use summary statistics invariant to label switches

Example : US GDP growth

Posterior means over time

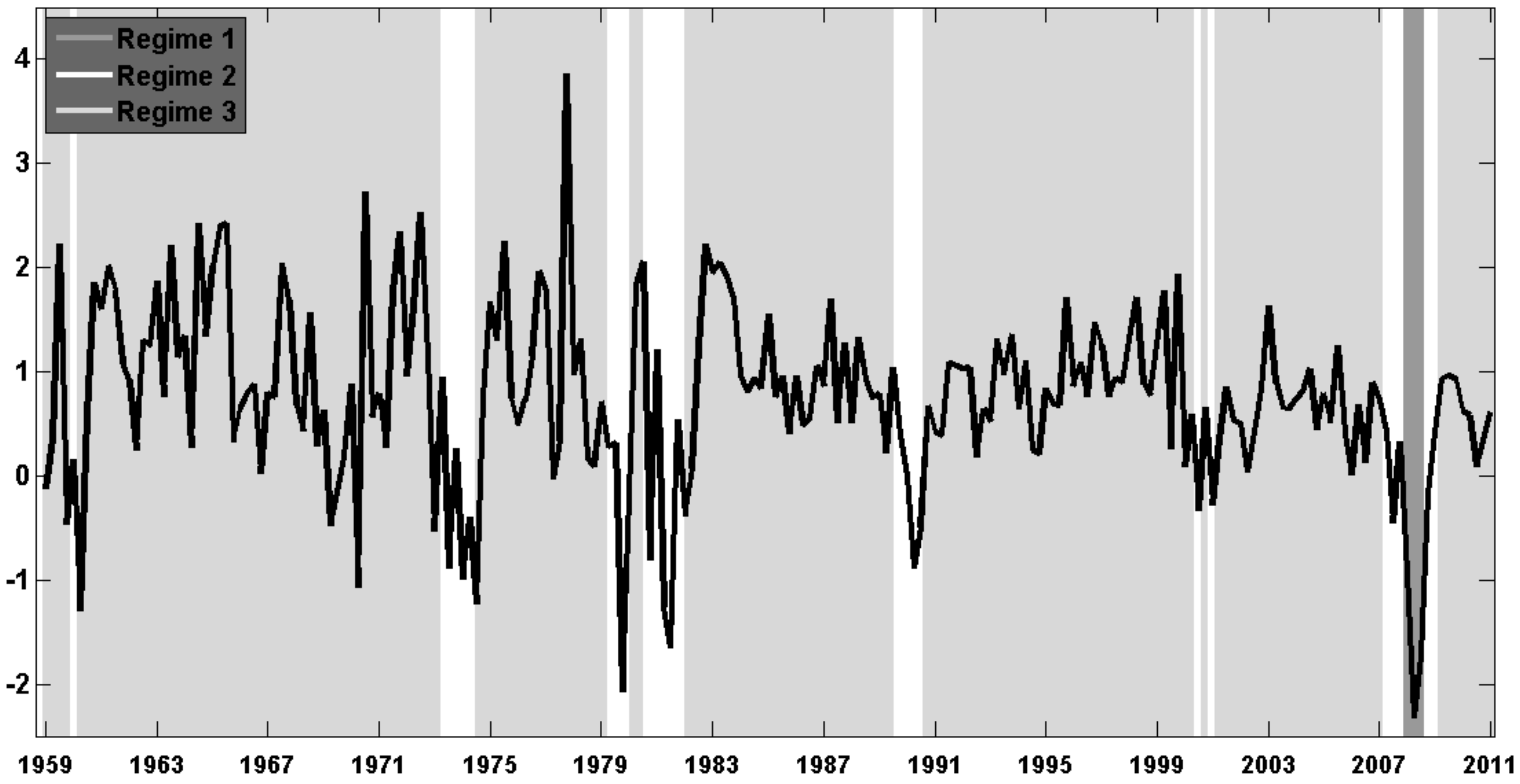
$$E(\sigma_t^2|Y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N}\sigma_{s_t}^{2,i}$$

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

Chapter 3

Structural breaks for models *with path dependence*

Chapter 3

Path dependence

Chib's specification

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**

Chib's specification

Why not applicable ?

- Simplification in the Forward-backward algorithm :

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

- If assumption does not hold :

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

Chib's algorithm not available for

State-space model with structural breaks in parameters

Example : ARMA, GARCH

Path dependent models

CP- and MS-ARMA models

$$\begin{aligned} y_t &= \mu_{s_t} + \theta_{s_t} y_{t-1} + \phi_{s_t} \epsilon_{t-1} + \epsilon_t \\ s_t | s_{t-1}, P &\sim p_{s_{t-1}, s_t} \end{aligned}$$

CP- and MS-GARCH models

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2 \\ s_t | s_{t-1}, P &\sim p_{s_{t-1}, s_t} \end{aligned}$$

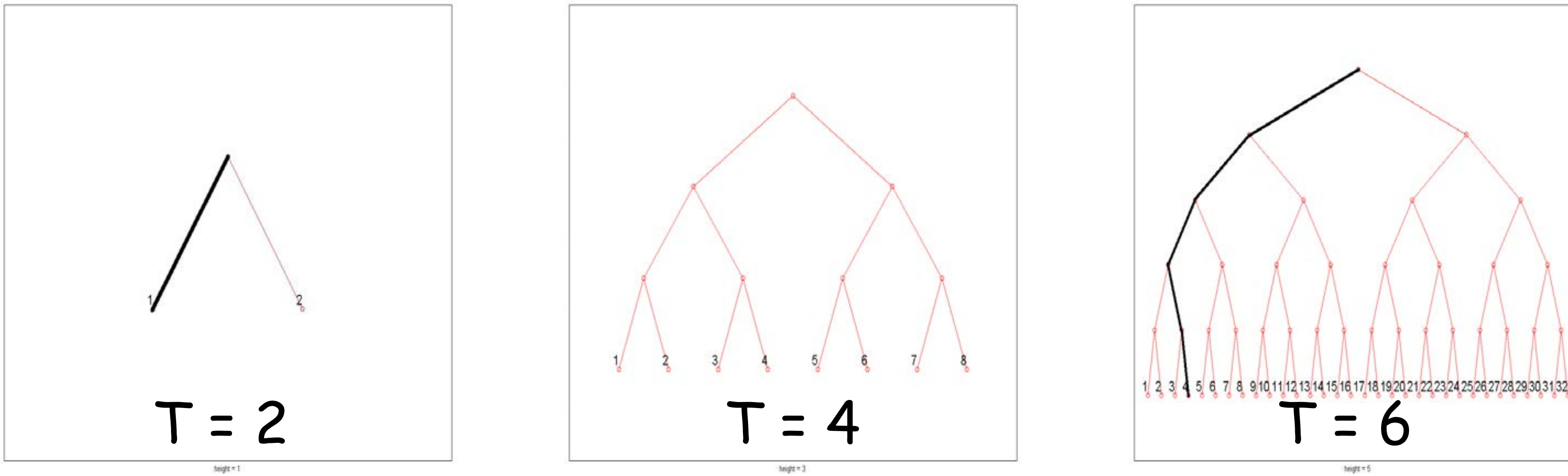
Change-point

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

Markov-switching

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

Path dependence problem



Function of $S_{1:t-1}$

ARMA $y_t = \mu_{s_t} + \theta_{s_t} y_{t-1} + \phi_{s_t} \epsilon_{t-1}(S_{1:t-1}) + \epsilon_t$

GARCH $\sigma_t^2(S_{1:t}) = \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2(S_{1:t-1})$

Likelihood at time t depends on the whole path that has been followed so far

T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Poss. paths	1	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192	16384

Path dependence problem

Solutions ?

1) Use of approximate models without path dependence

- Gray (1996), Dueker (1997), Klaassen (2002)
- Haas, Mittnik, Poella (2004)

$$
\begin{aligned}
y_t &= \sigma_{t,s_t}\epsilon_t \\
\sigma^2_{t,1} &= \omega_1 + \alpha_1 y^2_{t-1} + \beta_1 \sigma^2_{t-1,1} \\
\sigma^2_{t,2} &= \omega_2 + \alpha_2 y^2_{t-1} + \beta_2 \sigma^2_{t-1,2} \\
&\cdots \\
\sigma^2_{t,K+1} &= \omega_{K+1} + \alpha_{K+1} y^2_{t-1} + \beta_{K+1} \sigma^2_{t-1,K+1}
\end{aligned}
$$

Path dependence problem

Solutions ?

2) Stephens (1994) : Inference on multiple breaks

Drawbacks

- Time-consuming if T large
- Many MCMC iterations are required

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

3) Bauwens, Preminger, Rombouts (2011) :

- Single-move MCMC

Single-move MCMC

CP- and MS-GARCH models

$$
\begin{aligned}
y_t &= \sigma_t \epsilon_t \\
\sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2 \\
s_t | s_{t-1}, P &\sim p_{s_{t-1}, s_t}
\end{aligned}
$$

Change-point

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

Markov-switching

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

Single-move MCMC

Metropolis-Hastings sampler : $\Theta = \{\omega_1, \alpha_1, \beta_1, ..., \omega_{K+1}, \alpha_{K+1}, \beta_{K+1}\}$

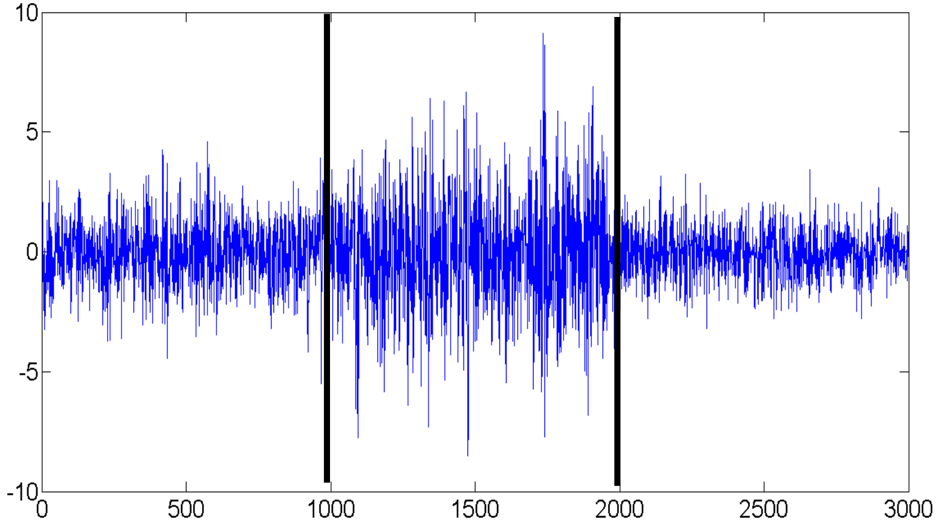
$$\begin{aligned} \theta_i | Y_{1:T}, S_{1:T}, P, \Theta_{-i} &\sim \text{Griddy-Gibbs} \\ P | Y_{1:T}, S_{1:T}, \Theta &\sim \text{Dirichlet}(\eta + n_{i,1:K+1}) \\ s_t | Y_{1:T}, P, \Theta, S_{1:t-1}, S_{t+1:T} &\sim \text{single-move} \end{aligned}$$

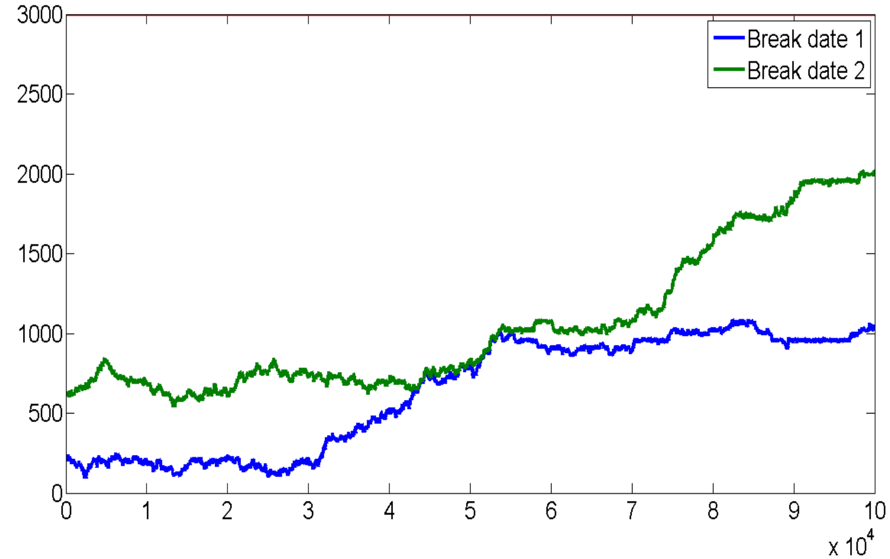
One state updated at a time !

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

Likelihood | Transition matrix

Example





Simulated series : $T = 3000$

Initial state : $[200 \quad 600]$

$$\mathrm{Corr}(\tau_1^i, \tau_1^{i-200}) = 0.996$$

Convergence after 100.000 MCMC iterations !

Single-move

Advantages

- Generic method :
    - Works for many CP and MS models

Drawbacks

- No criterion for selecting the number of regimes
- Very Time-consuming if T large (especially for MS)
- Many MCMC iterations are required :

{ <u>Very difficult to assess convergence</u>

<u>May not converge in a finite amount of time !</u>

Questions ?

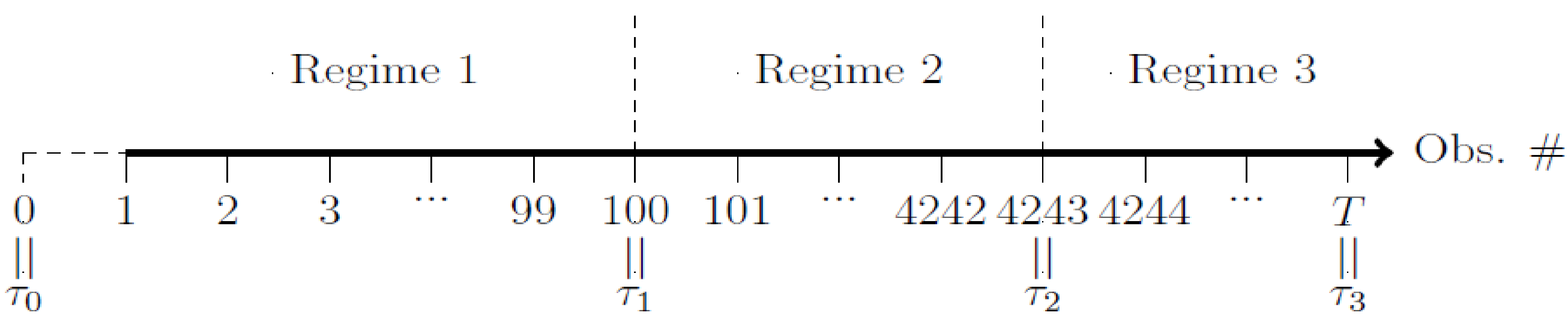
Change-point models

D-DREAM algorithm

CP-GARCH models :

$$
\begin{aligned}
y_t &= \mu_k + \epsilon_t, \\
\epsilon_t &= \sigma_t \eta_t, \ \ with\ \eta_t \sim i.i.d.\, N(0,1), \\
\sigma_t^2 &= \omega_k + \alpha_k \epsilon_{t-1}^2 + \beta_k \sigma_{t-1}^2,
\end{aligned}
$$

if $t \in]\tau_{k-1}, \tau_k]$, with $k = \{1, 2, ..., K+1\}$ denoting the regime.



Come back to the Stephens' specification !

D-DREAM algorithm

Problem with Stephens' inference :

- Break dates sample one at a time (single-move)

  → MCMC mixing issue

- Very demanding if T is large

Discrete-DREAM MCMC :

- Metropolis algorithm
  - Jointly sample the break dates
  - Very fast (faster than Forward-Backward)

D-DREAM algorithm

- Two sets of parameters to be estimated :

Continuous $\begin{cases} \Theta = (\theta_1, \theta_2, ..., \theta_{K+1})' \\ \theta_k = (\mu_k, \omega_k, \alpha_k, \beta_k) \end{cases}$

Discrete $\Gamma = (\tau_1, \tau_2, ..., \tau_K)'$

- MCMC scheme :

$\pi(\Theta | Y_{1:T}, \Gamma)$ **Iterations** $\pi(\Gamma | Y_{1:T}, \Theta)$

Not a standard dist.	**Not a standard dist.**
Metropolis	**Metropolis**
Proposal : DREAM	**Proposal : D-DREAM**

D-DREAM algorithm

DiffeRential Adaptative Evolution Metropolis
(Vrugt et al. 2009)

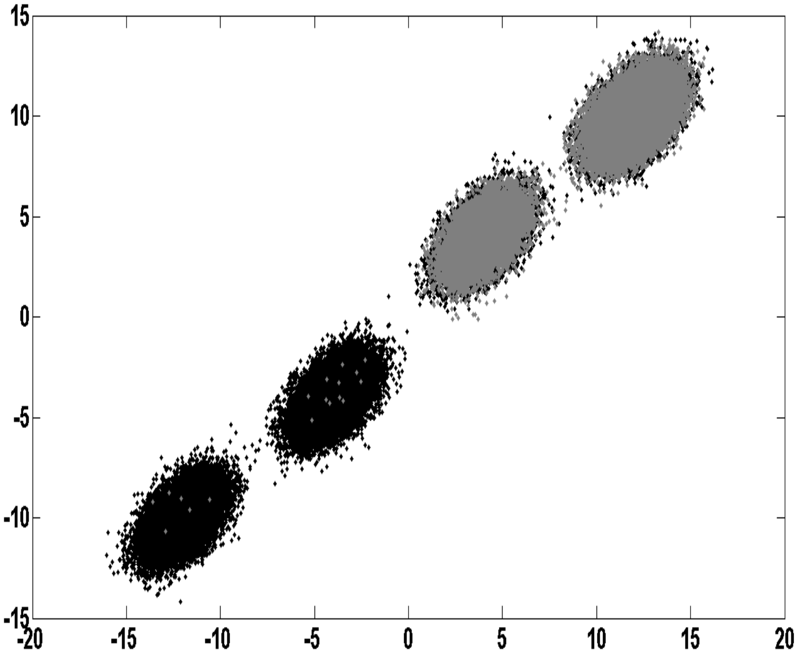
- DREAM automatically determines the **size** of the jump.
- DREAM automatically determines the **direction** of the jump
- DREAM is well suited for **multi-modal** post. dist.
- DREAM is well suited for **high dimensional** sampling
- DREAM is **symmetric** : only a Metropolis ratio

Nevertheless only applicable to continuous parameters

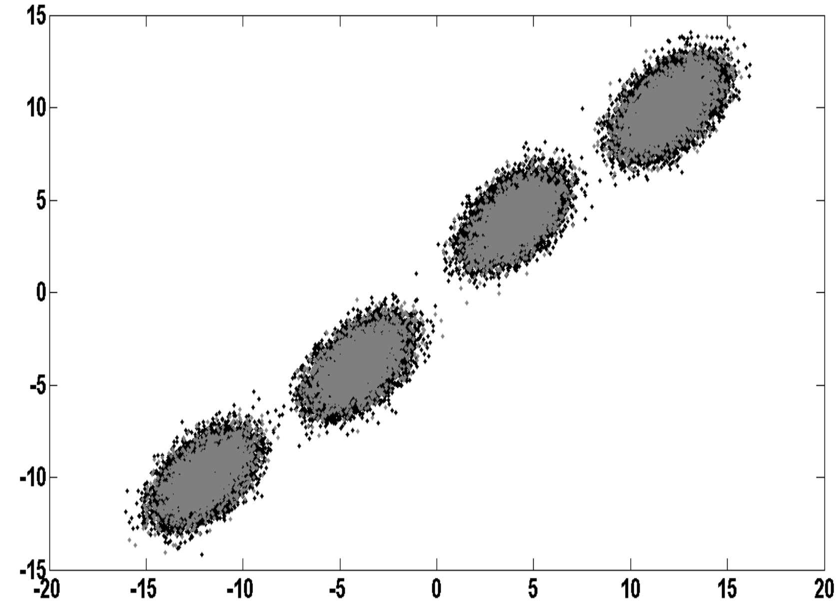
Extension for discrete parameter : Discrete-DREAM

DREAM : Example

Adaptive RW



DREAM



DREAM algorithm

M parallel MCMC chains :

$$\Theta_1^i \qquad \Theta_2^i \qquad \ldots \qquad \Theta_M^i$$

Proposal distribution :

$$\delta \in [1, \lfloor M/2 \rfloor] \qquad \zeta \sim N(0, 10^{-8})$$

$$\tilde{\Theta}_j^i = \Theta_j^i + \gamma(\delta, d)\left(\sum_{g=1}^{\delta} \Theta_{r_1(g)}^i - \sum_{h=1}^{\delta} \Theta_{r_2(h)}^i\right) + \zeta$$

$$\gamma(\delta, d) = 2.38/\sqrt{2\delta d} \qquad r_1(g) \neq r_2(h) \neq j$$

Symmetric proposal dist :

- Accept/reject the draw according to the probability

$$\min\left[\frac{f(Y_{1:T}|\tilde{\Theta}_j^i, \Gamma_j^i) f(\tilde{\Theta}_j^i)}{f(Y_{1:T}|\Theta_j^i, \Gamma_j^i) f(\Theta_j^i)}, 1\right]$$

D-DREAM algorithm

M parallel MCMC chains : $\{\Theta_j^i, \Gamma_j^i\}_{i=1,j=1}^{N,M}$

Continuous

$$\Theta = (\theta_1, \theta_2, ..., \theta_{K+1})'$$
$$\theta_k = (\mu_k, \omega_k, \alpha_k, \beta_k)$$

↓

Proposal distribution :

$$\tilde{\Theta}_j^i = \Theta_j^i + \gamma(\delta, d)(\sum_{g=1}^{\delta} \Theta_{r_1(g)}^i - \sum_{h=1}^{\delta} \Theta_{r_2(h)}^i) + \zeta$$

↓

Accept with probability

$$\min[\frac{f(Y_{1:T}|\tilde{\Theta}_j^i, \Gamma_j^i) f(\tilde{\Theta}_j^i)}{f(Y_{1:T}|\Theta_j^i, \Gamma_j^i) f(\Theta_j^i)}, 1]$$

Discrete

$$\Gamma = (\tau_1, \tau_2, ..., \tau_K)'$$

↓

Proposal distribution :

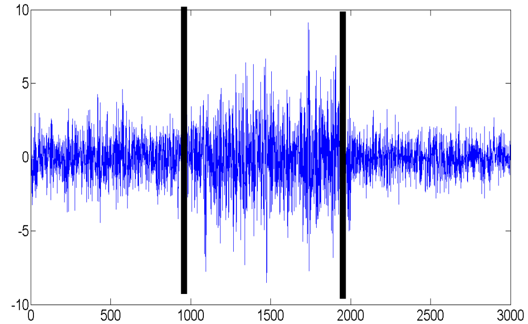
$$\tilde{\Gamma}_j^i = \Gamma_j^i + \mathrm{round}[\gamma(\delta, d)(\sum_{g=1}^{\delta} \Gamma_{r_1(g)}^i - \sum_{h=1}^{\delta} \Gamma_{r_2(h)}^i) + \zeta]$$

↓

Accept with probability

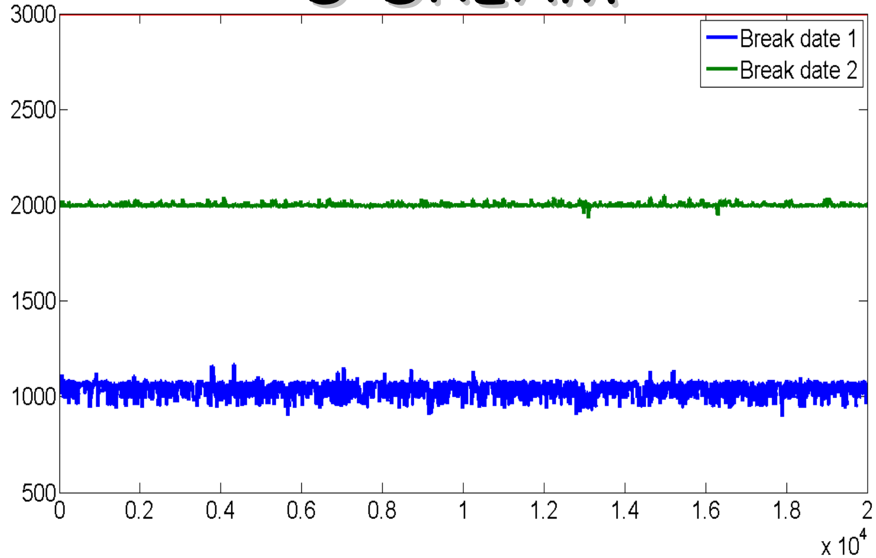
$$\min[\frac{f(Y_{1:T}|\Theta_j^i, \tilde{\Gamma}_j^i) f(\tilde{\Gamma}_j^i)}{f(Y_{1:T}|\Theta_j^i, \Gamma_j^i) f(\Gamma_j^i)}, 1]$$

Example

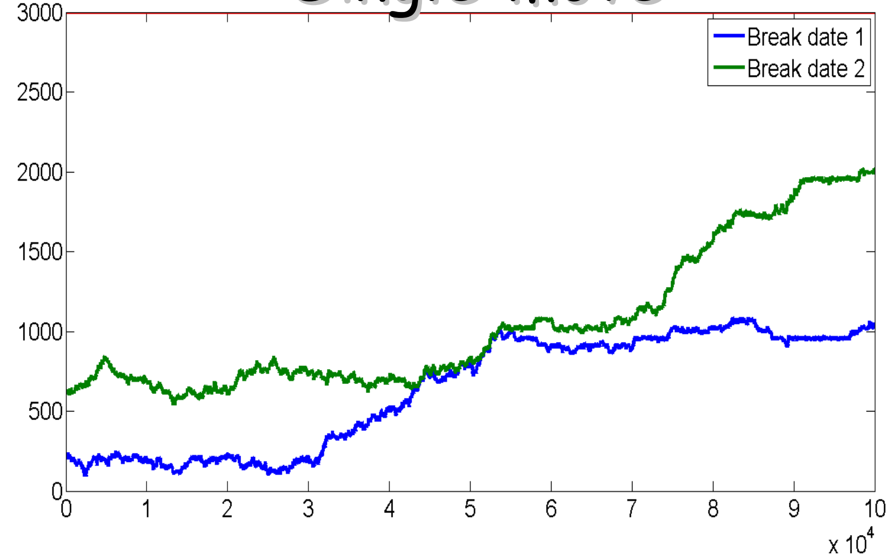


$T = 3000$

D-DREAM



Single-move



Initial states around $[200 \quad 600]$

$$\mathrm{Corr}(\tau_1^i, \tau_1^{i-200}) = -0.005$$

$$\mathrm{Corr}(\tau_1^i, \tau_1^{i-10}) = 0.54$$

Convergence after 3.000 MCMC iterations !

Initial state : $[200 \quad 600]$

$$\mathrm{Corr}(\tau_1^i, \tau_1^{i-200}) = 0.996$$

Convergence after 100.000 MCMC iterations !

D-DREAM (2014)

Advantages

- Generic method for CP models
- Inference on multiple breaks by marginal likelihood
- Very fast compared to existing algorithms

Drawbacks

- Model selection based on many estimations
- Only applicable to CP models and specific class of recurrent states

CP and MS models

Particle MCMC

CP- and MS-GARCH models

$$
\begin{aligned}
y_t &= \sigma_t \epsilon_t \\
\sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2 \\
\epsilon_t &\sim \text{i.i.d.} N(0,1) \\
s_t | s_{t-1}, P &\sim p_{s_{t-1}, s_t}
\end{aligned}
$$

Change-point

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

Markov-switching

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

Particle MCMC

Sets of parameters :

Continuous
$$\begin{cases} \Theta = (\theta_1, \theta_2, ..., \theta_{K+1})' \\ \theta_k = (\omega_k, \alpha_k, \beta_k) \\ P \end{cases}$$

State var. $S_{1:T} = \{s_1, ..., s_T\}$

MCMC scheme :

1) $\Theta | Y_{1:T}, P, S_{1:T} \sim \text{Metropolis-Hastings}$

2) $P | Y_{1:T}, \Theta, S_{1:T} \sim \prod_{i=1}^{K+1} \text{Dir}(\eta_1 + n_{i,1}, ..., \eta_1 + n_{i,K+1})$

3) $S_{1:T} | Y_{1:T}, P, \Theta \sim \text{Particle-Gibbs}$

Sampling a full state vector is unfeasible due to the path dependence issue

Particle MCMC

3) $S_{1:T}|Y_{1:T}, P, \Theta \sim \text{Particle-Gibbs}$

Idea : Approximate the distribution with a SMC algorithm

→ Does not keep invariant the posterior distribution

<u>*Andrieu, Doucet and Holenstein (2010)*</u>

- Show how to incorporate the SMC into an MCMC
- Allow for Metropolis and Gibbs algorithms
- Introduce the concept of conditional SMC

→ With a conditional SMC, the MCMC exhibits the posterior distribution as invariant one.

Particle MCMC

3) $S_{1:T}|Y_{1:T}, P, \Theta \sim \text{Particle-Gibbs}$

$$f(S_{1:t}|Y_{1:t}) \propto f(y_t|Y_{1:t-1}, S_{1:t})f(s_t|S_{1:t-1})\underbrace{f(S_{1:t-1}|Y_{1:t-1})}_{\text{Previous value}}$$

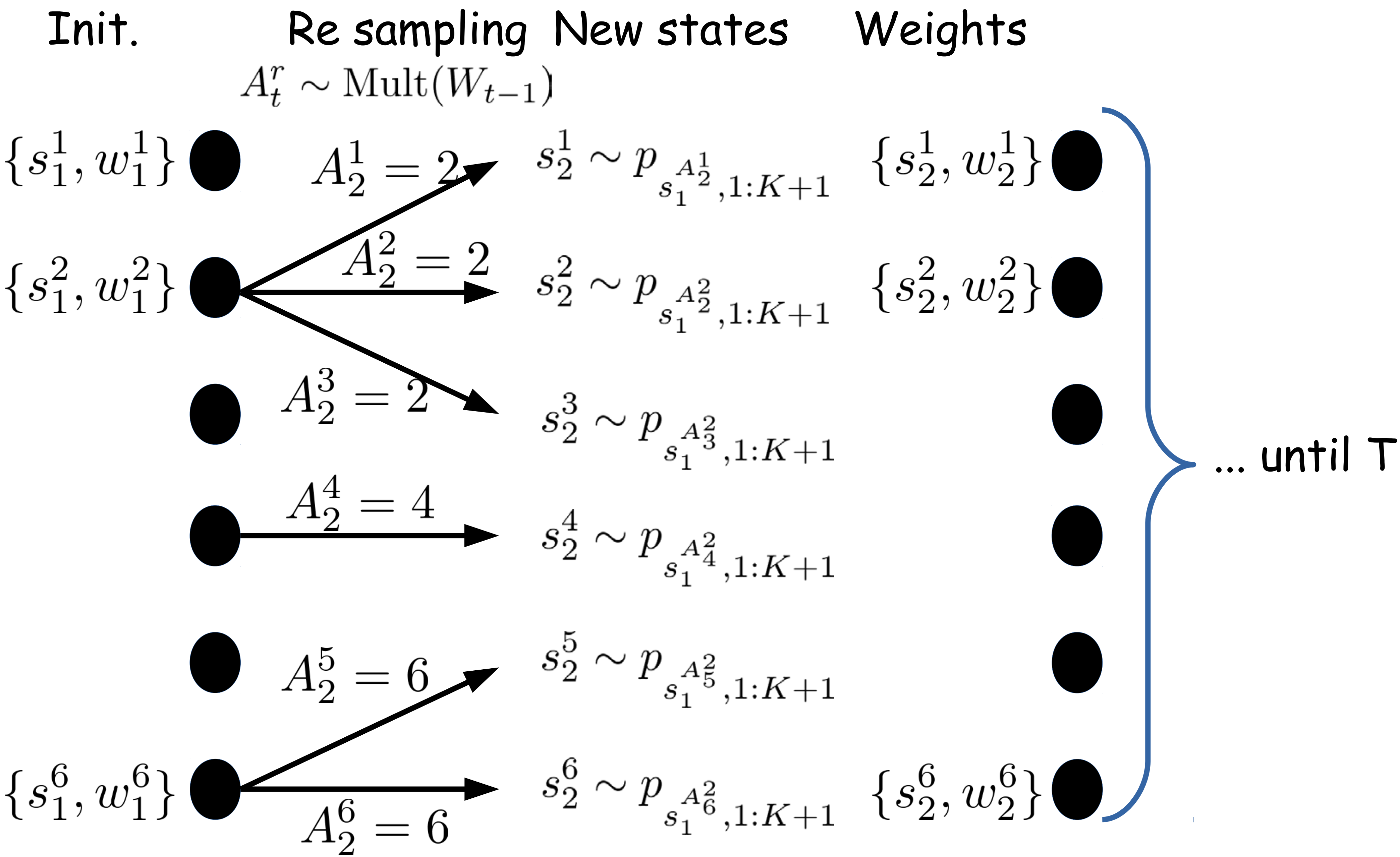
SMC :

1) Initialisation of the particles and weights: $\begin{cases} \{s_1^r\}_{r=1}^R \\ w_r \propto f(y_1|s_1^r) \end{cases}$

Iterations
$\forall t \in [2, T]$

- Re-sample the particles
  $\forall r \in [1, R]; A_t^r \sim \text{Mult}(W_{t-1})$
- Generate new states
  $\forall r \in [1, R]; s_t^r \sim \text{Mult}(p_{s_{t-1}^{A_t^r}, s_t})$
- Compute new weights
  $\forall r \in [1, R]; w_t^r \propto f(y_t|S_{1:t-1}^{A_{1:t-1}^r}, s_t^r)$ and $W_t^r = w_t^r / (\sum_{i=1}^R w_t^i)$

SMC


Init.
Re sampling
New states
Weights
$A_t^r \sim \mathrm{Mult}(W_{t-1})$
$\{s_1^1, w_1^1\}$
$\{s_1^2, w_1^2\}$
$\{s_1^6, w_1^6\}$
$A_2^1 = 2$
$A_2^2 = 2$
$A_2^3 = 2$
$A_2^4 = 4$
$A_2^5 = 6$
$A_2^6 = 6$
$s_2^1 \sim p_{s_1^{A_2^1}, 1:K+1}$
$s_2^2 \sim p_{s_1^{A_2^2}, 1:K+1}$
$s_2^3 \sim p_{s_1^{A_3^2}, 1:K+1}$
$s_2^4 \sim p_{s_1^{A_4^2}, 1:K+1}$
$s_2^5 \sim p_{s_1^{A_5^2}, 1:K+1}$
$s_2^6 \sim p_{s_1^{A_6^2}, 1:K+1}$
$\{s_2^1, w_2^1\}$
$\{s_2^2, w_2^2\}$
$\{s_2^6, w_2^6\}$
... until T

Particle Gibbs

- **Conditional SMC :** SMC where the previous MCMC state vector is ensured to survive during the entire SMC sequence.

3) $S_{1:T}|Y_{1:T}, P, \Theta \sim \text{Particle-Gibbs}$

- Launch a conditional SMC
- Sample a state vector as follows :

1) $r \sim \text{Mult}(W_T^1, ..., W_T^R)$ and set $s_T = s_T^r$

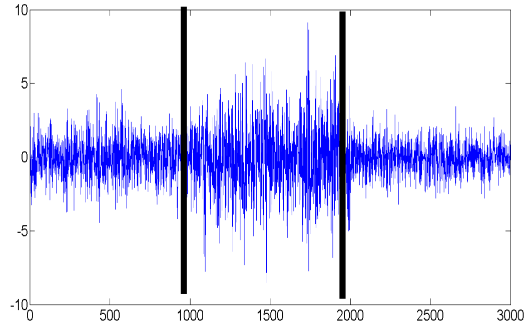
2) From T-1 until 1, retrieve the path of the state : $s_t = s_t^{A_{t+1}^r}$

- Improvements :

  1) Incorporation of the APF in the conditional SMC

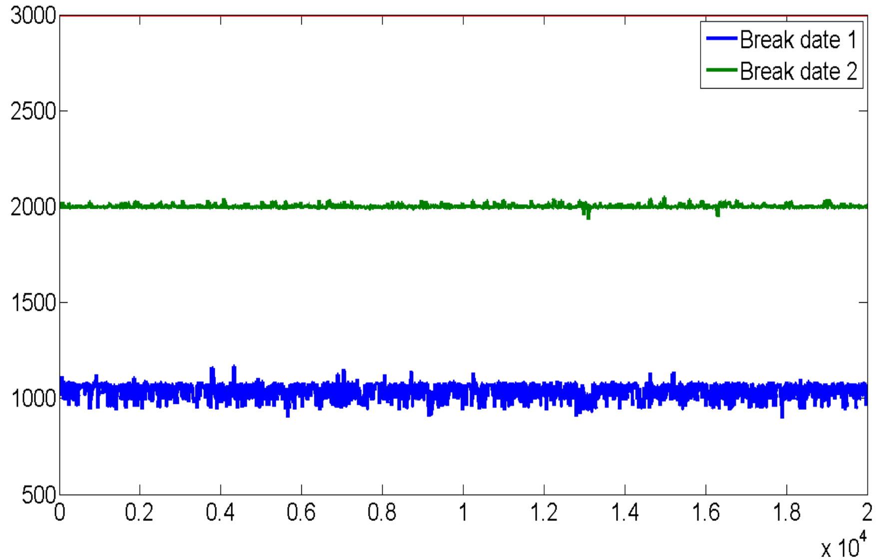
  2) Backward sampling as Godsill, Doucet and West (2004)

Example



$$T = 3000$$

D-DREAM

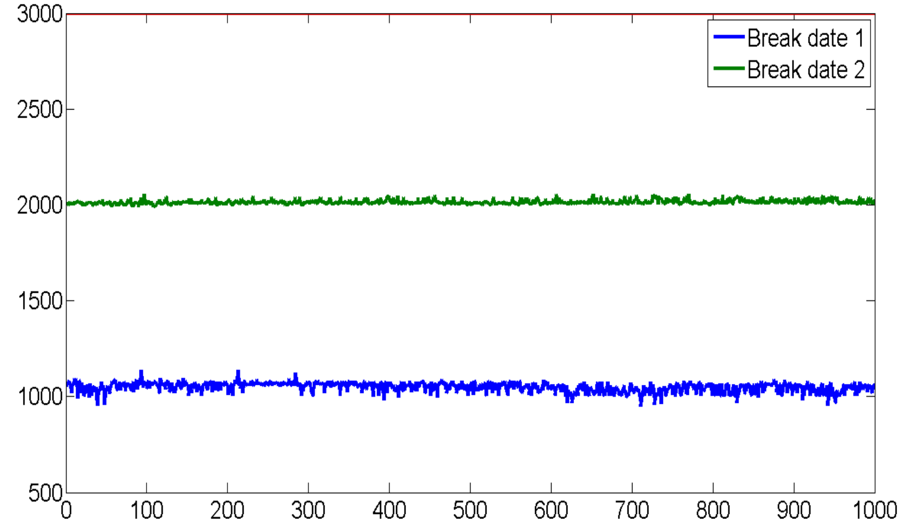


Initial states around $[200 \quad 600]$

$$\mathrm{Corr}(\tau_1^i, \tau_1^{i-200}) = -0.005$$

$$\mathrm{Corr}(\tau_1^i, \tau_1^{i-10}) = 0.54$$

PMCMC



Initial state : $[200 \quad 600]$

$$\mathrm{Corr}(\tau_1^i, \tau_1^{i-200}) = 0.05$$

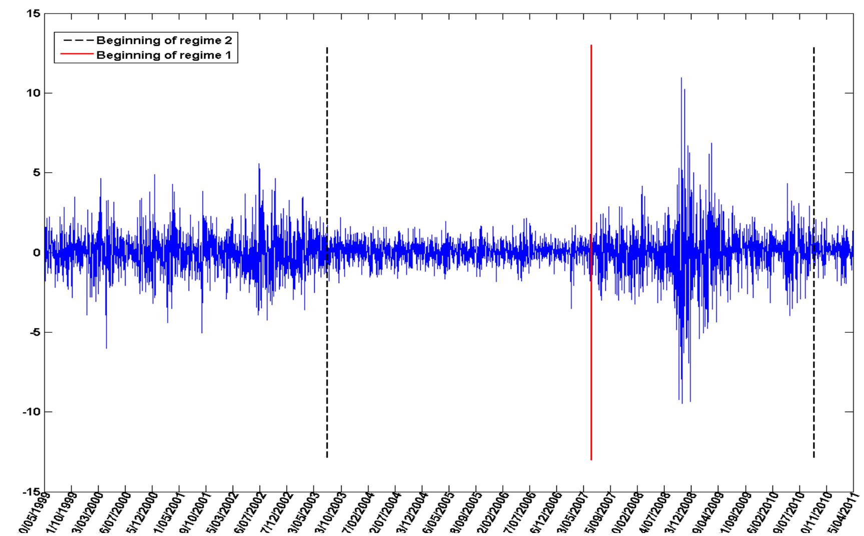
$$\mathrm{Corr}(\tau_1^i, \tau_1^{i-10}) = 0.21$$

PMCMC

S&P 500 daily percentage returns from May 20,1999 to April 25, 2011

Table: Marginal log-likelihood values for S&P 500 data

Regimes	1	2	3	4
Change-Point				
BS	-4505.33	-4505.83	-4503.05	-4519.23
Chib	-4504.95	-4505.93	-4502.97	-4516.16
Markov-switching				
BS	-4505.31	-4497.99	-4502.74	
Chib	-4505.08	-4496.04	-4497.73	



Regime	GARCH σ^2	GARCH α	GARCH β	CP-GARCH σ^2	CP-GARCH α	CP-GARCH β	MS-GARCH σ^2	MS-GARCH α	MS-GARCH β
1	1.55	0.075	0.915	1.95	0.0849	0.868	2.32	0.089	0.891
2				0.45	0.023	0.931	0.46	0.031	0.901
3				2.75	0.098	0.890			

PMCMC

Various financial time series

Series	Spline-GARCH		GARCH	CP-GARCH		MS-GARCH		
	knots	log-BF	MLL	K+1	log-BF	K+1	log-BF	nswitch
S&P 500	3	5.21	-4505.33	3	2.28	2	7.34	3
DJIA	3	2.99	-4333.43	1	0	2	4.7	3
NASDAQ	3	3.20	-5429.84	1	0	2	1.94	7
NYSE	3	2.40	-4380.62	1	0	2	3.91	13
BAC	4	16.62	-6127.39	3	50.12	3	79.49	11
BA	4	9.10	-6174.57	2	8.9	2	11.48	6
JPM	3	8.82	-6400.27	3	5.17	3	7.22	9
MRK	5	48.78	-6209.73	5	215.39	3	335.23	56
PG	4	16.34	-4842.02	3	24.23	2	33.6	9
Metals	2	6.66	-5267.44	2	11.33	2	14.68	5
Yen/Dollar	1	-3.34	-2982.33	1	0	2	3.05	7

PMCMC (2013)

Advantages

- Generic method for CP and MS models
- Inference on multiple breaks by marginal likelihood
- Very good mixing properties

Drawbacks

- Model selection based on many estimations
- Very computationally demanding
- Difficult to calibrate the number of particles
- Difficult to implement

IHMM-GARCH

CP- and MS-GARCH models

$$
\begin{aligned}
y_t &= \sigma_t \epsilon_t \\
\sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2 \\
\epsilon_t &\sim \text{i.i.d.} N(0,1) \\
s_t | s_{t-1}, P &\sim p_{s_{t-1}, s_t}
\end{aligned}
$$

Change-point

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

Markov-switching

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

IHMM-GARCH

Sets of parameters :

Continuous
$$\begin{cases} \Theta = (\theta_1, \theta_2, ..., \theta_{K+1})' \\ \theta_k = (\omega_k, \alpha_k, \beta_k) \\ P \end{cases}$$

State var. $S_{1:T} = \{s_1, ..., s_T\}$

MCMC scheme :

1) $\Theta | Y_{1:T}, P, S_{1:T} \sim \text{Metropolis-Hastings}$

2) $P | Y_{1:T}, \Theta, S_{1:T} \sim \prod_{i=1}^{K+1} \text{Dir}(\eta_1 + n_{i,1}, ..., \eta_1 + n_{i,K+1})$

3) $S_{1:T} | Y_{1:T}, P, \Theta \sim \text{M-H based on approximate models}$

IHMM-GARCH

3) $S_{1:T}|Y_{1:T}, P, \Theta \sim \text{M-H based on approximate models}$

Sampling a full state vector is infeasible
due to the path dependence issue

↓

Sampling a full state vector from an approximate model
Klaassen or Haas, Mittnik and Paolela

↓

Accept/reject according to the Metropolis-hastings ratio

IHMM-GARCH

Moreover, Hierarchical dirichlet processes are used

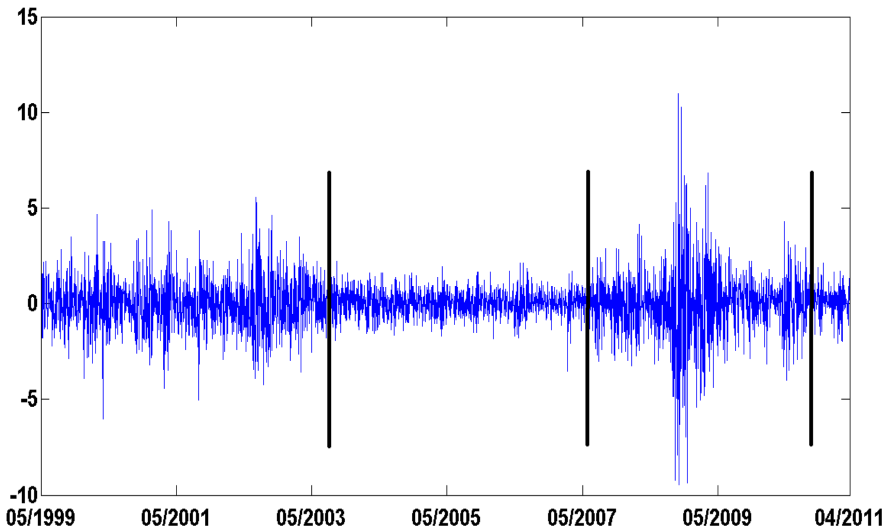
- To infer the number of regime in one estimation
- To include both CP and MS specification in one model

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

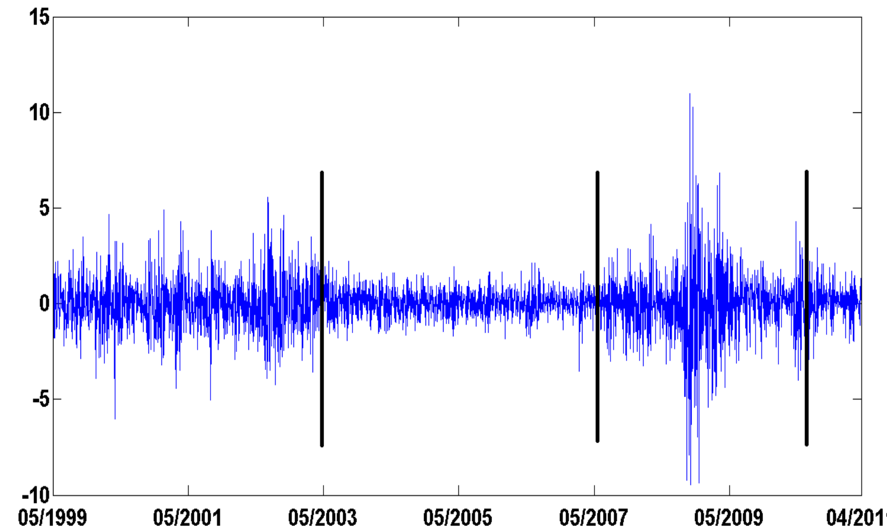
IHMM-GARCH

S&P 500 daily percentage returns
from May 20,1999 to April 25, 2011

PMCMC



IHMM-GARCH



	Regime 1	Regime 2	Regime 3	Regime 4	Regime 5	Regime 6	Regime 7
Prob.	0	**0.6046**	0.2075	0.1455	0.0224	0.0196	0.0004

Table: Posterior probabilities of the number of regimes for the *S*&*P*500 daily index

IHMM-GARCH (2014)

Advantages

- Generic method for CP and MS models
- Self-determination of the number of breaks
- Self-determination of the specification (CP and/or MS)
- Predictions of breaks
- Very good mixing properties
- Fast MCMC estimation

Drawbacks

- Difficult to implement

References

References

- Bauwens, Preminger, Rombouts (2010)

Bauwens, L.; Preminger, A. & Rombouts, J. *'Theory and Inference for a Markov-switching GARCH Model'*, Econometrics Journal, 2010, 13, 218-244

- Bauwens, Dufays, De Backer (2011)

Bauwens, L.; Dufays, A. & De Backer, B. *'Estimating and forecasting structural breaks in financial time series'*, CORE discussion paper, 2011/55, 2011

- Bauwens, Dufays, Rombouts (2013)

Bauwens, L.; Dufays, A. & Rombouts, J. 'Marginal Likelihood for Markov Switching and Change-point GARCH Models', Journal of Econometrics, 2013, 178 (3), 508-522

References

- Dufays (2012)

- Dufays, A. *'Infinite-State Markov-switching for Dynamic Volatility and Correlation Models'*, CORE discussion paper, 2012/43, 2012

SMC algorithm :

- He, Maheu (2010)

He, Z. & Maheu, J. M. *'Real time detection of structural breaks in GARCH models'*, Computational Statistics & Data Analysis, 2010, 54, 2628-2640

Without path dependence

- Gray (1996)

Gray, S. *'Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process'*, Journal of Financial Economics, 1996, 42, 27-62

- Dueker (1997)

Dueker, M. *'Markov Switching in GARCH Processes in Mean Reverting Stock Market Volatility'*, Journal of Business and Economics Statistics, 1997, 15, 26-34

- Klaassen (2002)

Klaassen, F. *'Improving GARCH volatility forecasts with regime-switching GARCH'*, Empirical Economics, 2002, 27, 363-394

- Haas, Mittnik and Paolella (2004)

Haas, M.; Mittnik, S. & Paolella, M. *'A New Approach to Markov-Switching GARCH Models'*, Journal of Financial Econometrics, 2004, 2, 493-530

Sparse Change-point model

Arnaud Dufays
(Centre de Recherche en Economie et Statistique)

May 19, 2014

Motivation

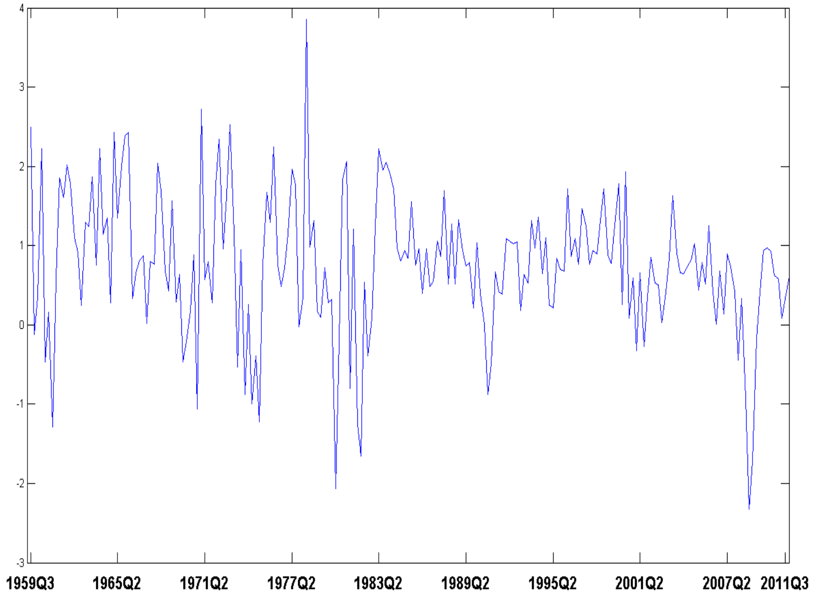
Limitations of standard CP models :

1. Classical inferences available for AR and ARCH models
   $\rightarrow$ Difficult to estimate path dependence models (CP-ARMA,CP-GARCH).
2. Optimal number of regimes computed by Marginal likelihood
   $\rightarrow$ Many useless estimations and uncontrolled penalty.
3. Each new regime increases the number of model parameters
   $\rightarrow$ Over-parametrization.
4. Forecasts based on the last regime
   $\rightarrow$ Uncertainties on parameters and inaccurate predictions.

Example

ARMA model

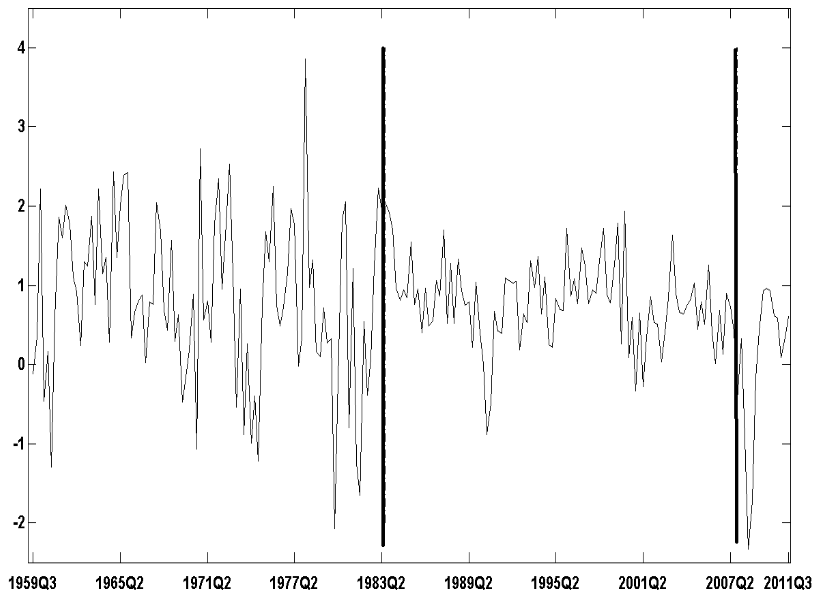
Standard



$$y_t = c + \beta y_{t-1} + \phi \epsilon_{t-1} + \epsilon_t$$

No dynamic for c, β, ϕ

Change-Point



$$y_t = c_i + \beta_i y_{t-1} + \phi_i \epsilon_{t-1} + \epsilon_t$$

A latent variable governs the dynamic of breaks

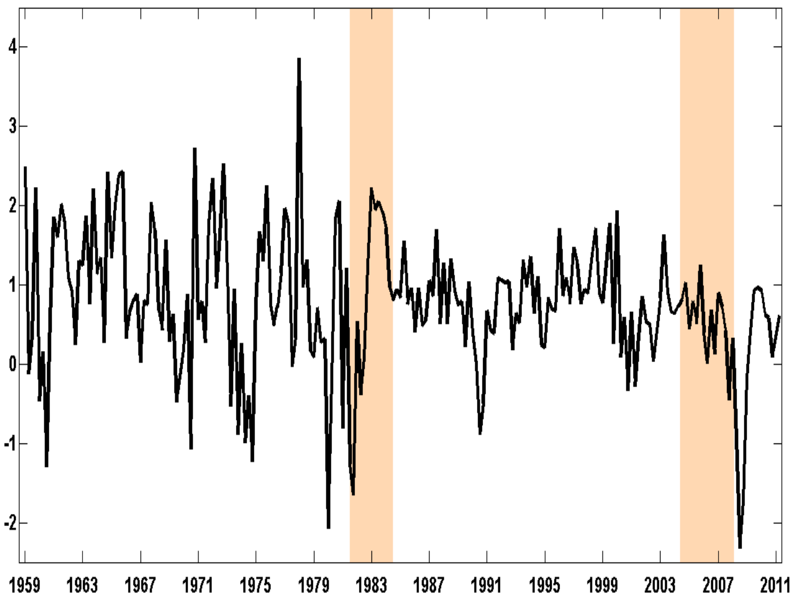
Contribution

CP models using shrinkage priors :

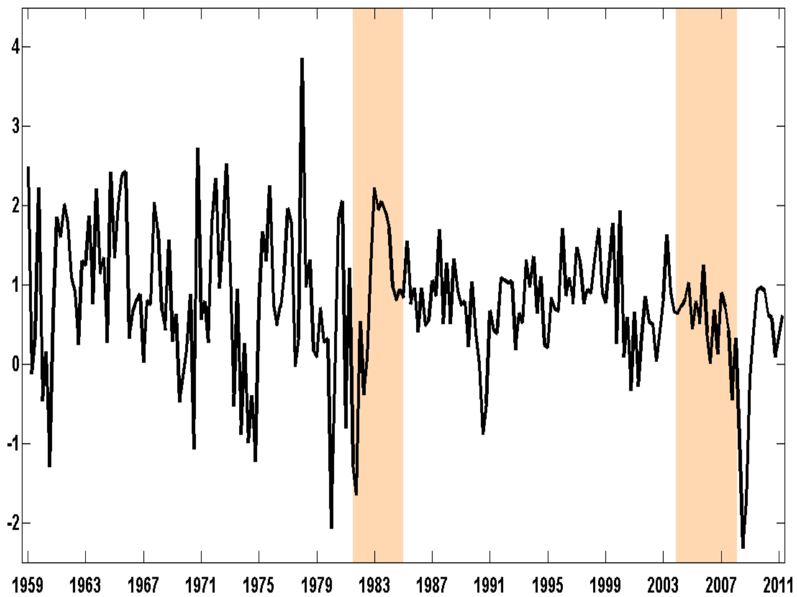
1. Adapted to estimate path dependence models (CP-ARMA,CP-GARCH).
2. Optimal number of regimes obtained in one estimation with user-specified penalty.
3. Controls the over-parametrization.
   $\rightarrow$ Only a few parameter evolves from one regime to another.
4. Very good forecast performances.

Example

CP-ARMA



CP-ARMA with Shrinkage priors



	CP-ARMA			Our CP-ARMA		
σ^2	1.11	0.25	0.53	1.11	0.27	0.27
	(0.14)	(0.04)	(0.20)	(0.15)	(0.04)	(0.04)
MA term	-0.08	-0.37	0.48	-0.34	-0.34	0.5
	(0.15)	(0.13)	(0.29)	(0.09)	(0.09)	(0.19)

Outline

The CP-Model and its estimation

Shrinkage methods

Empirical applications

Conclusion

The CP-ARMA model

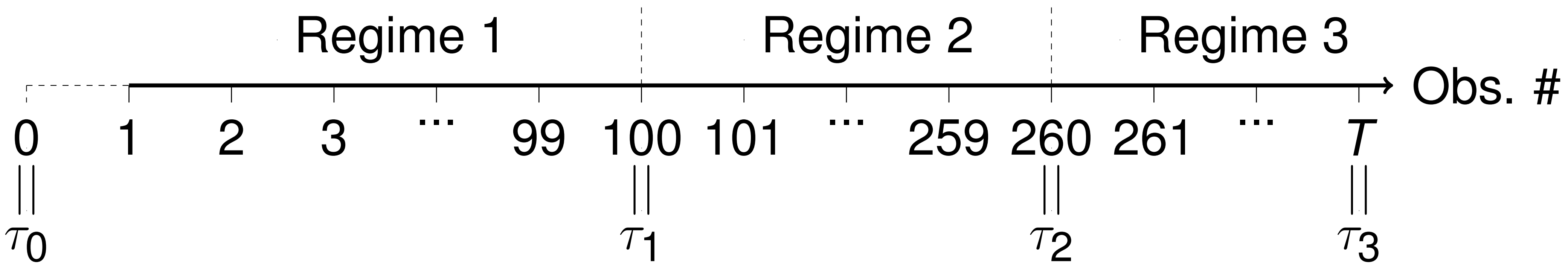
Let $Y_T = \{y_1, ..., y_T\}$ be a time series of T observations.

The CP-ARMA model defined by :

$$
\begin{aligned}
y_t &= c_1 + \beta_1 y_{t-1} + \phi_1 \epsilon_{t-1} + \epsilon_t \text{ with } \epsilon_t \sim N(0, \sigma_1^2) \text{ for } t \leq \gamma_1 \\
y_t &= c_2 + \beta_2 y_{t-1} + \phi_2 \epsilon_{t-1} + \epsilon_t \text{ with } \epsilon_t \sim N(0, \sigma_2^2) \text{ for } t \leq \gamma_2 \\
\cdots & \cdots \cdots \\
y_t &= c_{K+1} + \beta_{K+1} y_{t-1} + \phi_{K+1} \epsilon_{t-1} + \epsilon_t \text{ for } \gamma_K < t \leq T
\end{aligned} \tag{1}
$$

Focus on the break dates $\Gamma = (\tau_1, ..., \tau_K)'$ instead of a state vector $S_T = \{s_1, ..., s_T\}$

Let $\Theta = (c_1, \beta_1, \phi_1, \sigma_1^2, ..., c_{K+1}, \beta_{K+1}, \phi_{K+1}, \sigma_{K+1}^2)'$.



D-DREAM algorithm : specification

The MCMC scheme is

1. $\pi(\Theta|\Gamma, Y_T)$
2. $\pi(\Gamma|\Theta, Y_T)$

We use a Metropolis algorithm :

DiffeRential Adaptative Evolution Metropolis (Vrugt et al. (2009))

- DREAM automatically determines the **size** of the jump.
- DREAM automatically determines the **direction** of the jump.
- DREAM is well suited for **multi-modal** posterior distributions and for **high dimensional** parameters.
- DREAM proposal is **symmetric**.

However only suited for continuous parameters.
D-DREAM for Discrete parameters (Bauwens et al. (2011))

Outline

The CP-Model and its estimation

Shrinkage methods

Empirical applications

Conclusion

Shrinkage priors

How to determine which model parameter(s) evolves from one regime to another ?

Testing all the possibilities and computing the Marginal likelihood

1. Too many posterior distributions to be estimated.

   $\rightarrow$ For 4 regimes and 4 parameters by regime : 256 models.
   $\rightarrow$ In the empirical example : 8 regimes (4096 models).

2. No proof that the Marginal likelihood will choose the right spec.

3. Break date parameters for each parameters : Over-parametrization.

Shrinkage priors

How to determine which model parameter(s) evolves from one regime to another ?

Keep the specification of a standard CP-ARMA but shrink the irrelevant parameters and break dates toward zero.

1. Only one estimation is required.
2. Break is identified only if it improves the likelihood.
3. Long regimes : estimators tend to their distribution without shrinkage.

Transformation of the model

For applying a shrinkage prior, the CP-ARMA model becomes :

$$y_t = (c_1 + \sum_{i=2}^{k} \Delta c_i) + (\beta_1 + \sum_{i=2}^{k} \Delta \beta_i) y_{t-1} + (\phi_1 + \sum_{i=2}^{k} \Delta \phi_i) \epsilon_{t-1}$$

$$+ \epsilon_t \text{ with } \epsilon_t \sim N(0, \sigma_1^2 + \sum_{i=2}^{k} \Delta \sigma_i^2) \text{ for } t \in]\gamma_{k-1}, \gamma_k]$$

with $\Delta c_i = c_i - c_{i-1}$.

The first regime : $\{c_1, \beta_1, \phi_1\} \sim N(0, 10 I_3)$ and $\sigma_1^2 \sim U[0, 30]$.

Shrinkage on the other parameters :

For Example : $\Delta c_i | \tau \sim N(0, \tau)$ and $\tau \sim Q$.

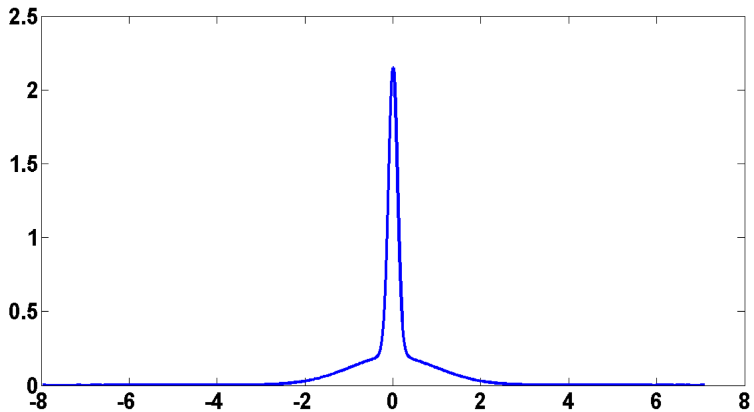
Shrinkage priors

Shrinkage is about the distribution Q : $\Delta c_i | \tau \sim N(0, \tau)$ and $\tau \sim Q$.

- The absolutely continuous spike-and-slab prior (Ishwaran and Rao (2000)) :

  $\tau = \sigma^2 \kappa$ with $\sigma^{-2} \sim G(\frac{v}{2}, \frac{v}{2})$ and $\kappa | \omega = \omega \delta_{\kappa=0.00001} + (1-\omega)\delta_{\kappa=1}$.

The marginal distribution of $\Delta c_i | \omega$ is a mixture of two student distributions :



Spike and Slab marginal distribution

Outline

The CP-Model and its estimation

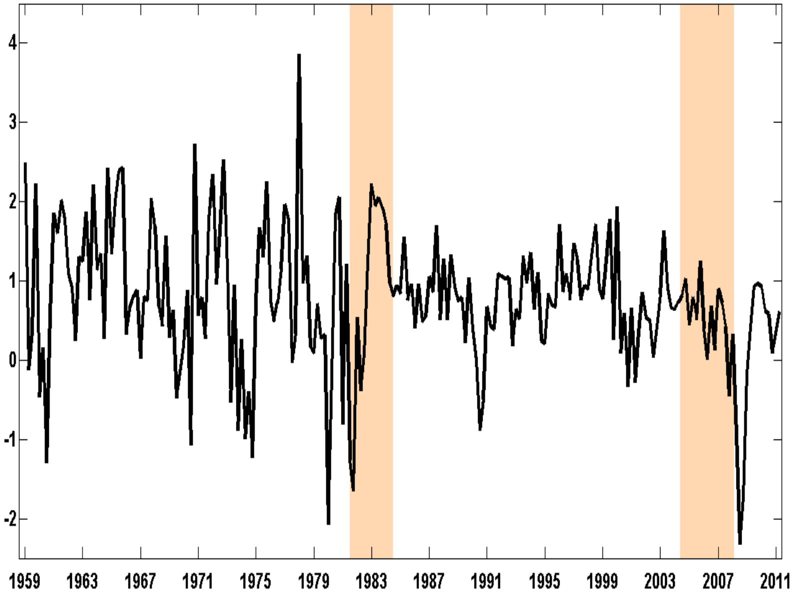
Shrinkage methods

Empirical applications

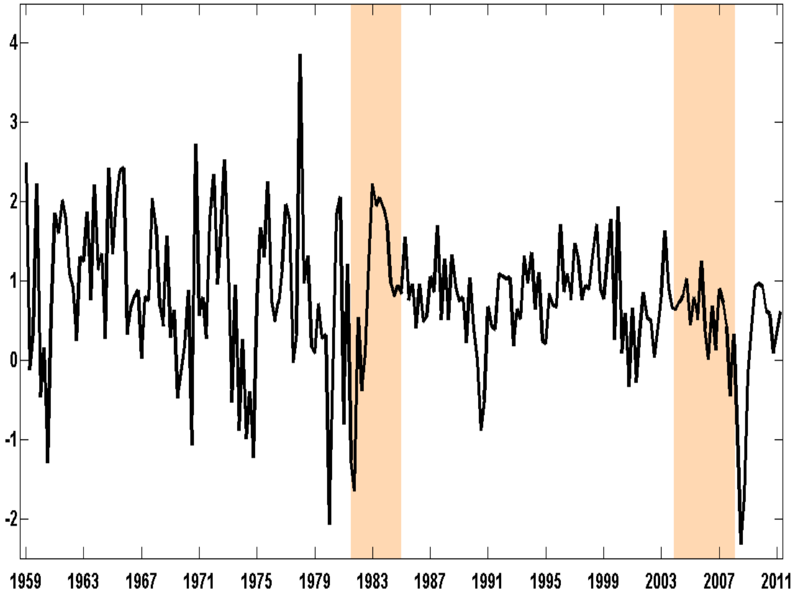
Conclusion

US GDP growth rate 1959-2011

CP-ARMA



CP-ARMA with Shrinkage priors



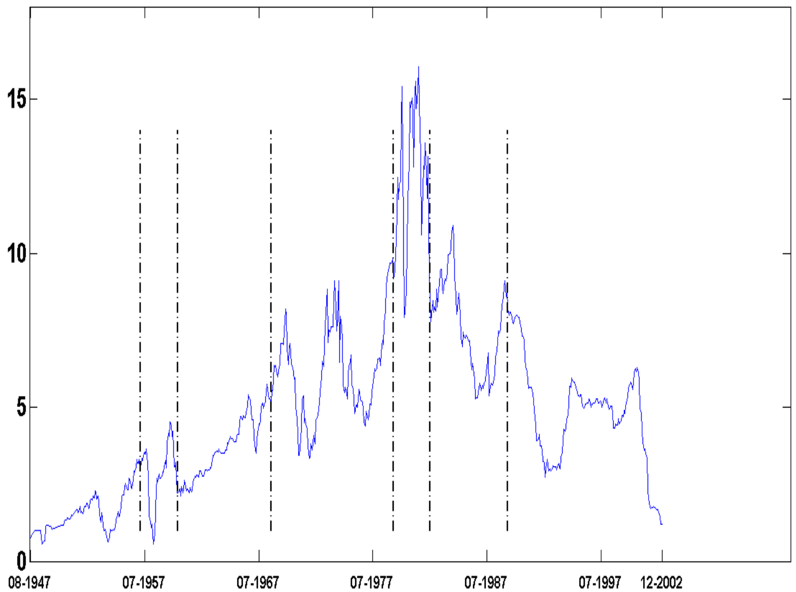
CP-ARMA with Spike and Slab

Regime	2	3
c	0.03	0.04
β	0.03	0.03
ϕ	0.03	**1.00**
σ^2	**1.00**	0.06

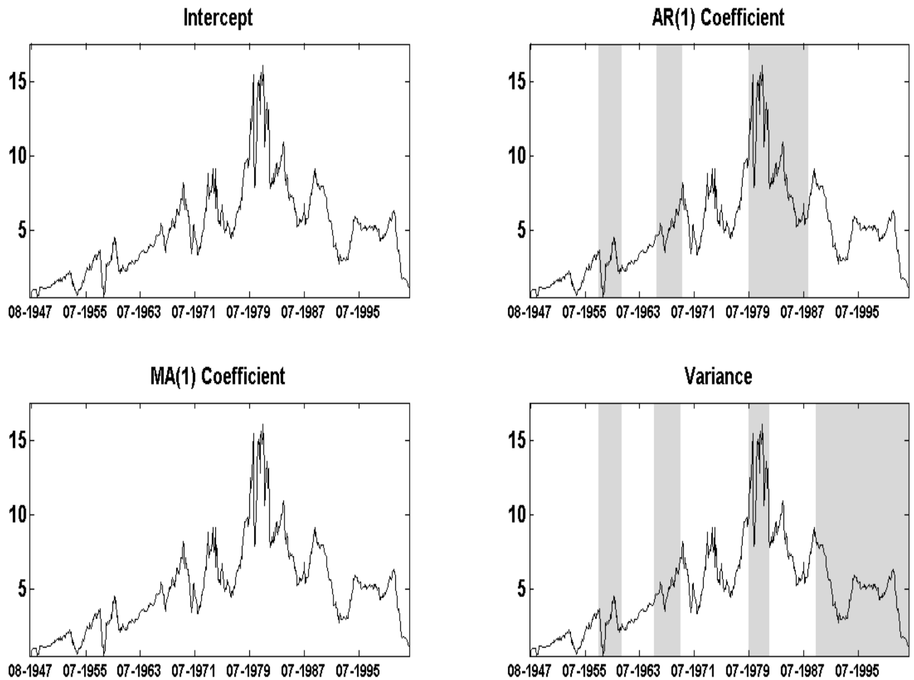
Only the variance and the MA parameter change over time

Monthly 3-Month US T-bill rate 1947-2002

Pesaran et al.



CP-ARMA with Shrinkage priors



Outline

The CP-Model and its estimation

Shrinkage methods

Empirical applications

Conclusion

Conclusion

$\rightarrow$ Algorithm : Inference for ARMA models with structural breaks.

- Detects the parameters that change from one regime to another.
- Shrinks all the irrelevant parameters toward zero.
- Shrinks all the irrelevant regime.
  $\rightarrow$ No need of the marginal likelihood.

$\rightarrow$ Empirical enhancements :

- Could improve the interpretation of the presence of structural breaks.
- Very good prediction performances.

Chapter 4

Numerical session on matlab

Chapter 4

Introduction

Introduction

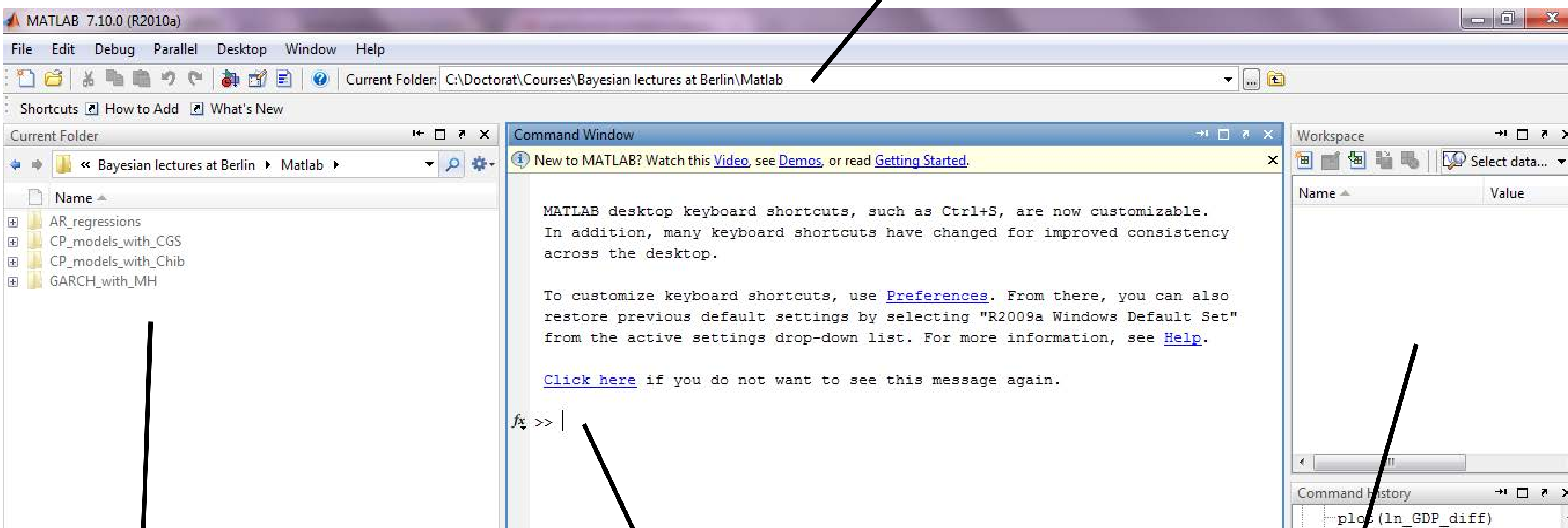
Quick start

- Four directories :
  - AR_regressions – ***MCMC for simulating posteriors of AR models***
  - GARCH_with_MH – ***MH MCMC for simulating posteriors of a GARCH(1,1) model***
  - CP_models_with_CGS – ***MCMC for drawing posteriors of AR models with 2 regimes using the Carlin, Gelman and Smith approach***
  - CP_models_with_Chib – ***MCMC for posterior distributions of AR models with k regimes based on Chib's algorithm***

Introduction

How to use the programs ?

Current folder : Matlab directory



Navigator

Command window

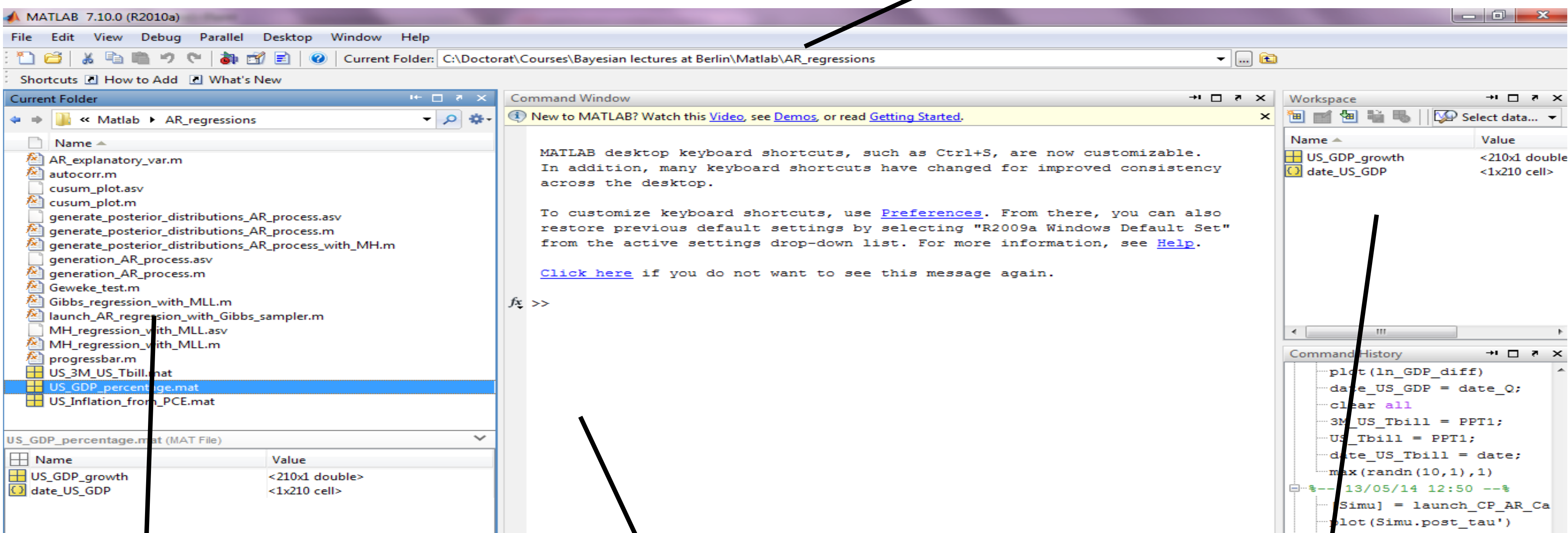
Workspace

Introduction

Go to the AR_regressions directory

Click on US_GDP_percentage

Current folder : Matlab directory



Navigator

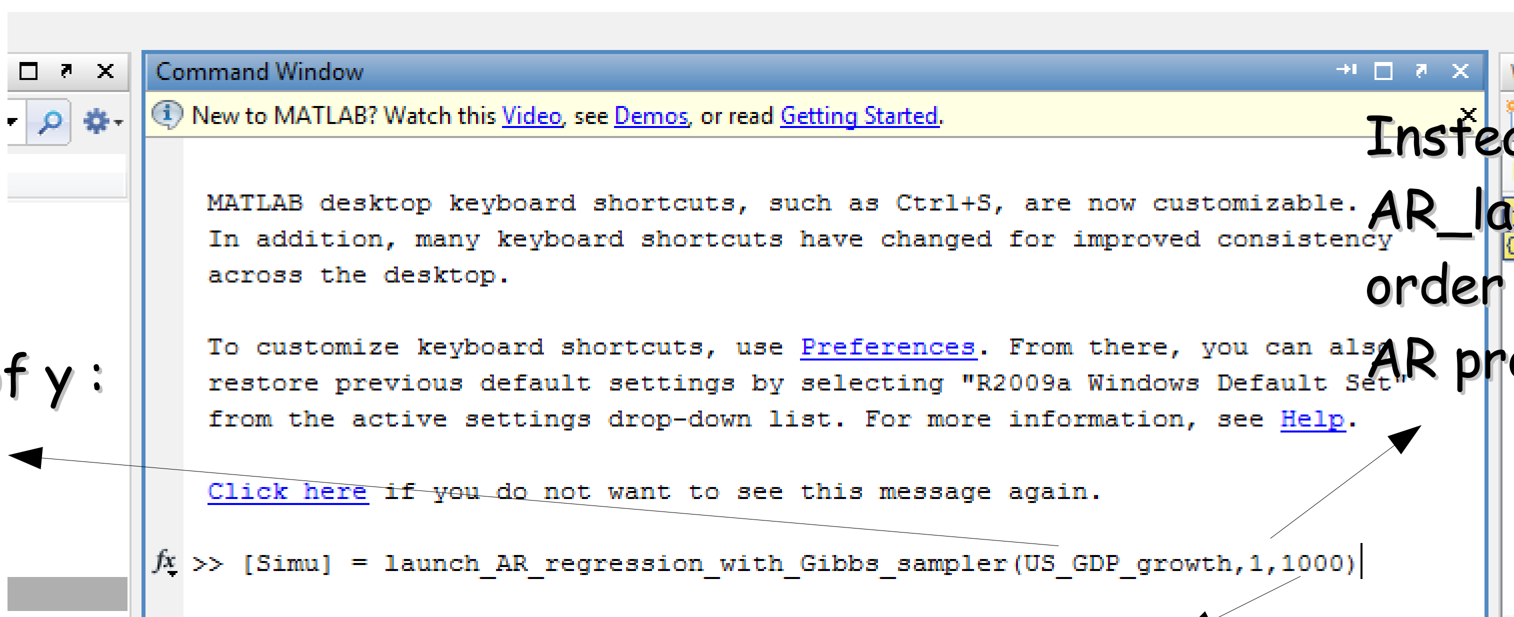
Command window

Workspace with data

Introduction

How to run a program ?

- Open the program :
  launch_AR_regression_with_Gibbs_sampler
  - Comments to run the function are in green
- Copy and paste the first line without 'function' in the command window



Instead of y : the data

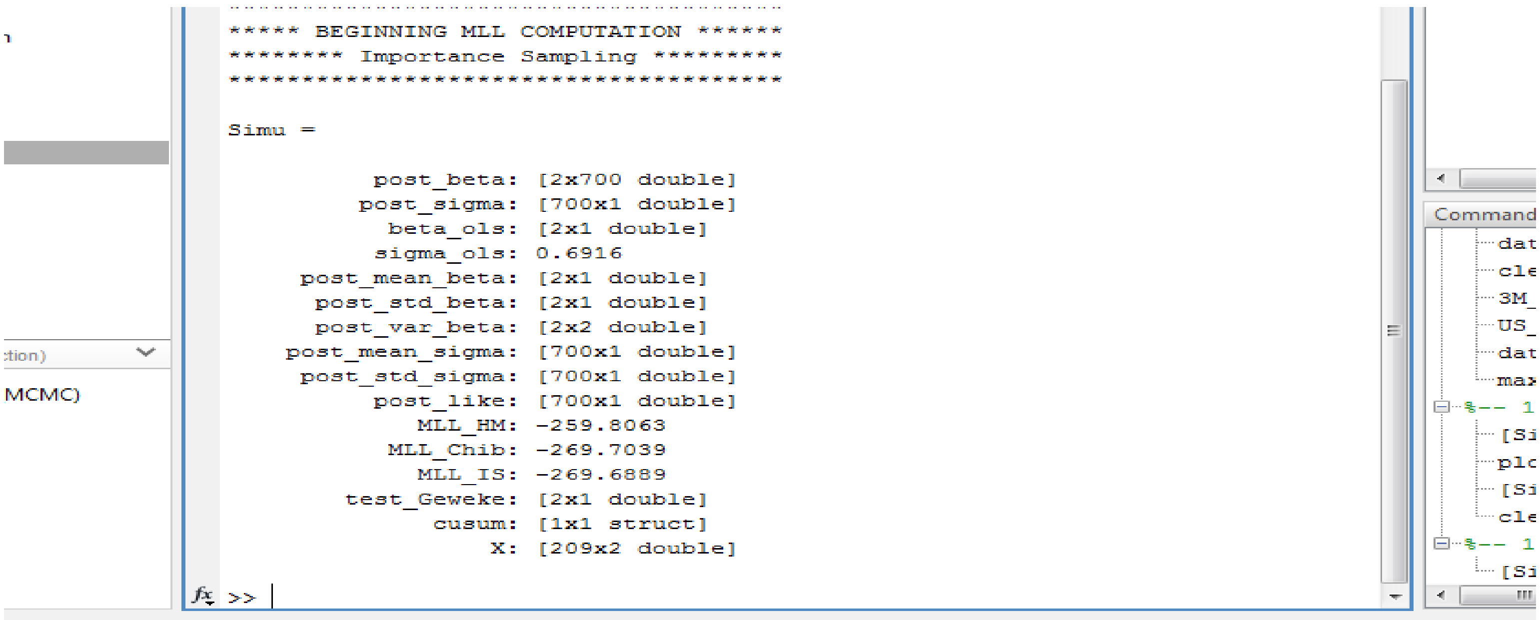
Instead of AR_lags : the order of the AR process

nb_MCMC : 1000 iterations

Introduction

What is a structure ?

- Simu is a matlab structure containing many tables



- To access the matrix 'post_beta' :

→ Type in command window : Simu.post_beta

Simple MCMC

Gibbs sampler

- **The model**

$$\begin{cases} y_t & = & \beta' x_t + \epsilon_t \\ \epsilon_t & \sim & \text{i.i.d.} N(0, \sigma^2) \end{cases}$$

- **The prior distributions**

$$\beta \sim N(\beta_0, \Sigma_0) \qquad \sigma^2 \sim IG(IG_a, IG_b)$$

- **Conditional posterior distributions :**

$$\pi(\theta | Y_{1:T}, \sigma^2) \quad \sim \quad N(\bar{\mu}, \bar{\Sigma}) \qquad \begin{cases} \bar{\Sigma} = [\sigma^{-2} \sum_{t=1}^{T} (x_t x_t') + \Sigma_0^{-1}]^{-1} \\ \bar{\mu} = \bar{\Sigma}[\sigma^{-2} \sum_{t=1}^{T} x_t y_t + \Sigma_0^{-1} \beta_0] \end{cases}$$

$$\pi(\sigma^2 | Y_{1:T}, \theta) \quad \sim \quad IG(IG_a + T/2, IG_b + \sum_{t=1}^{T} \epsilon_t^2 / 2)$$

Gibbs sampler

- Initial values :

MLE estimates

$$\beta = (\sum_{t=1}^{T} x_t x_t')^{-1} (\sum_{t=1}^{T} x_t y_t)$$

$$\sigma^2 = \sum_{t=1}^{T} (y_t - \beta' x_t)^2 / T$$

- Gibbs sampler :

$$\beta \sim \beta | Y_{1:T}, \sigma^2 \qquad \sigma^2 \sim \sigma^2 | Y_{1:T}, \beta$$

Discard the first draws as burn-in period

Use the rest as a sample of the posterior distribution :

$$\beta, \sigma^2 | Y_{1:T}$$

Gibbs sampler

- **Open the function :** ***Gibbs_regression_with_MLL***
  - Gibbs sampler of a simple regression
  - Provide the marginal log-likelihood (MLL) estimated from three different approaches (Chib, Importance sampling and Harmonic mean)

→ **Harmonic mean : very bad estimate**

- **The prior distributions**

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%% Set the prior values
%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Prior :
%%% beta ~ N(beta_0,Sigma_0)
%%% sigma^2 ~ IG(a,b)
%%%%%%%%%%%%%%%%%%%%%%%%%%%
IG_b = 1;
IG_a = 1;
var_uninformative = 100; %% We fix the variance of each beta equal to
beta_0 = zeros(dimension,1);
Sigma_0 = diag(ones(dimension,1)*(var_uninformative));
inv_Sigma_0 = diag(ones(dimension,1)*(1/var_uninformative));
```

Hyper-parameters of the variance

Hyper-parameters of the mean parameters

Gibbs sampler

```
%%%%%%%%%%%%%%%
%%%% MCMC starting values
%%%%%%%%%%%%%%%
beta = beta_ols;
sigma = sigma_ols;
iter = 1;
for i=1:nb_MCMC
    progressbar(i/nb_MCMC);
    %%%%%%%%%%%
    %%%% Saving posterior samples
    %%%%%%%%%%%
    if(i>burn_in)
        beta_post(:,iter) = beta;
        sigma_post(iter) = sigma;
        like_post(iter) = (-T/2)*log(2*pi*sigma) - (y-X*beta)'*(y-X*beta)/(2*sigma);
        iter = iter +1;
    end
    %%%%%%%%%%%%%
    %%% Update of sigma|beta
    %%%%%%%%%%%%%
    IG_post_b_MCMC = IG_b + 0.5*sum((y-X*beta).^2);
    inv_sigma = gamrnd(IG_post_a_MCMC,1/IG_post_b_MCMC);
    sigma = 1/inv_sigma;
    %%%%%%%%%%%%%
    %%% Update of beta|Sigma
    %%%%%%%%%%%%%
    Sigma_cond = inv(inv_sigma*Mat_XX + inv_Sigma_0);
    mu_cond = Sigma_cond*(inv_sigma*Vec_Xy + inv_Sigma_0*beta_0);
    beta = mvnrnd(mu_cond,Sigma_cond)';
end
%%%%%%%%%%%%%
```

Starting values

Saving posterior draws

First block - the variance

First block - the mean parameters

Gibbs sampler

```
%%%%%%%%%%%%%%%
%%%% MCMC starting values
%%%%%%%%%%%%%%%
beta = beta_ols;
sigma = sigma_ols;
iter = 1;
for i=1:nb_MCMC
    progressbar(i/nb_MCMC);
    %%%%%%%%%%%
    %%%% Saving posterior samples
    %%%%%%%%%%%
    if(i>burn_in)
        beta_post(:,iter) = beta;
        sigma_post(iter) = sigma;
        like_post(iter) = (-T/2)*log(2*pi*sigma) - (y-X*beta)'*(y-X*beta)/(2*sigma);
        iter = iter +1;
    end
    %%%%%%%%%%%%%
    %%% Update of sigma|beta
    %%%%%%%%%%%%%
    IG_post_b_MCMC = IG_b + 0.5*sum((y-X*beta).^2);
    inv_sigma = gamrnd(IG_post_a_MCMC,1/IG_post_b_MCMC);
    sigma = 1/inv_sigma;
    %%%%%%%%%%%%%
    %%% Update of beta|Sigma
    %%%%%%%%%%%%%
    Sigma_cond = inv(inv_sigma*Mat_XX + inv_Sigma_0);
    mu_cond = Sigma_cond*(inv_sigma*Vec_Xy + inv_Sigma_0*beta_0);
    beta = mvnrnd(mu_cond,Sigma_cond)';
end
%%%%%%%%%%%%%
```

Starting values

Saving posterior draws

First block - the variance

First block - the mean parameters

Gibbs sampler

- **Open the function :** ***launch_AR_regression_with_Gibbs_sampler***

In the command window, type :

[Simu] =
launch_AR_regression_with_Gibbs_sampler(US_GDP_growt
h,1,3000,1)

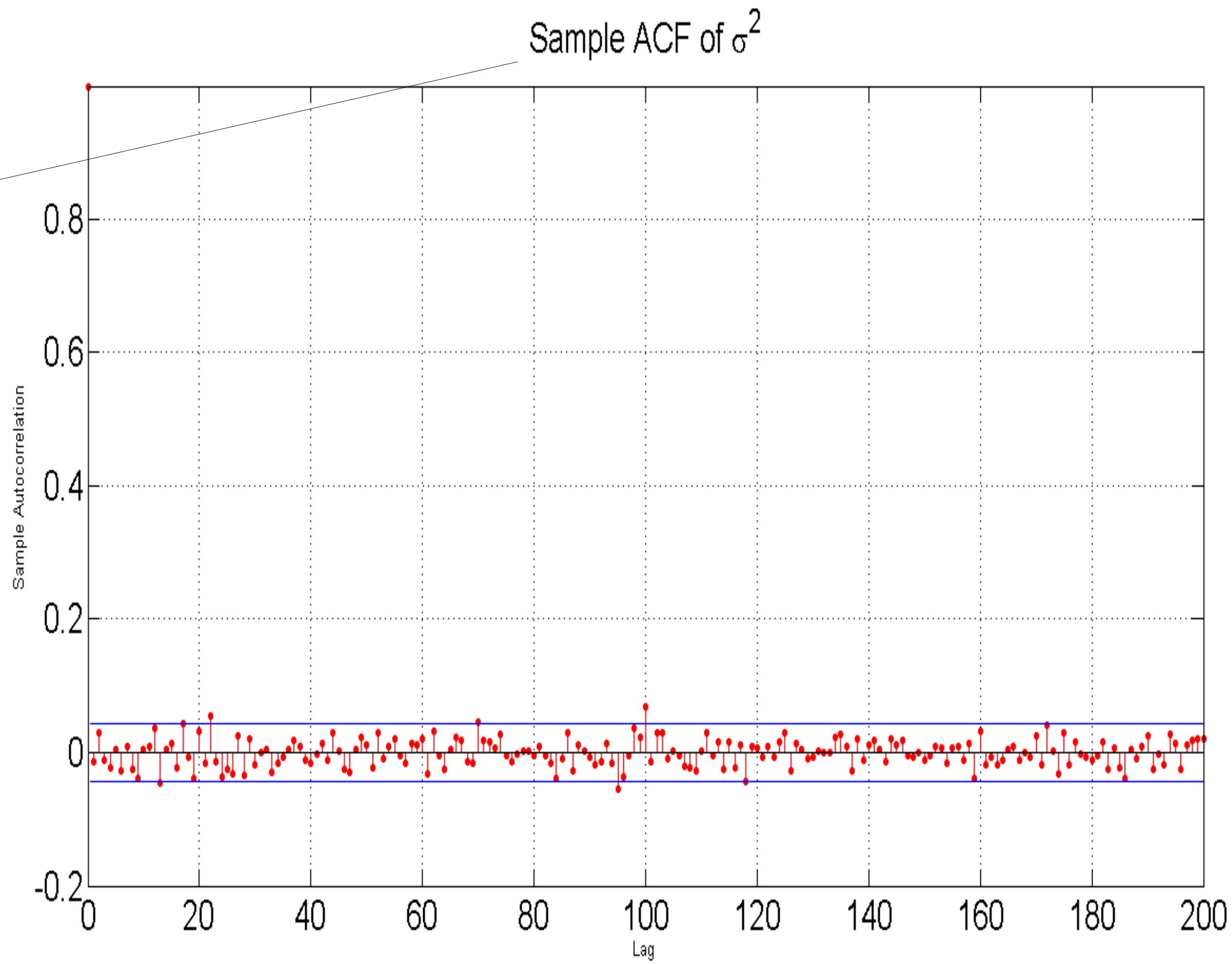
- The data : US_GDP_growth
- AR order : 1
- Nb MCMC iterations : 3000
- Convergence Graphics : on

Gibbs sampler

- **Graphics :**

Empirical correlations between MCMC draws

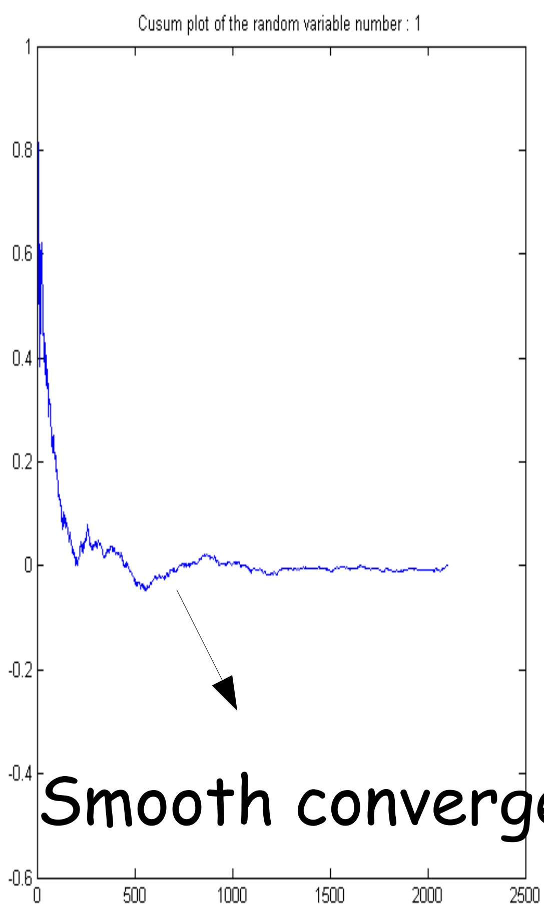
Variance

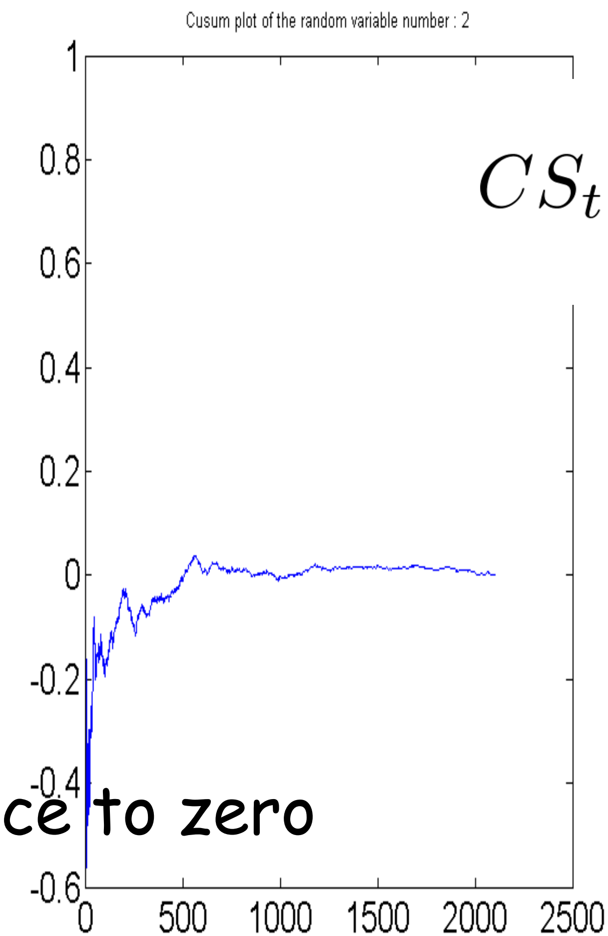


Gibbs sampler

- **Graphics :**

Cumsum plots of the two mean parameters





Smooth convergence to zero

$$CS_t = (\frac{1}{i}\sum_{j=1}^{i}\theta^j - m_\theta)/s_\theta$$

Empirical std of the parameter

Empirical mean of the parameter

$$CS_N = 0$$

Gibbs sampler

- **Geweke's test of MCMC convergence :**

Comparison of two empirical means well separated

In the command window :

Simu.test_Geweke

→ Test for each mean parameters and return zero if the hypothesis is not rejected at 95%

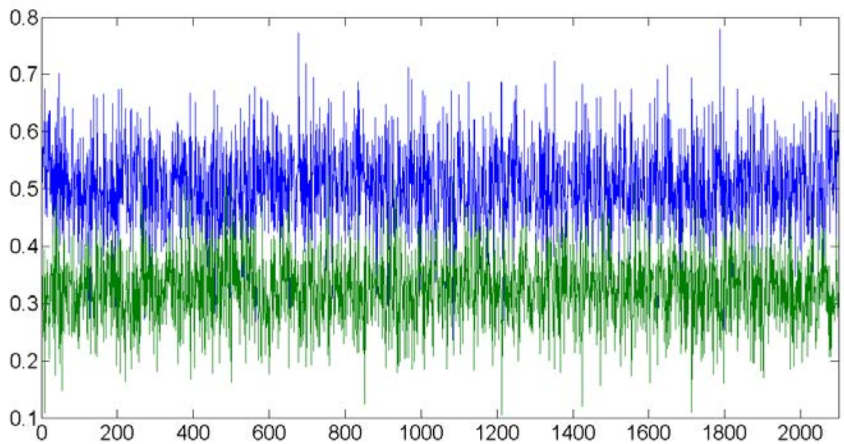
```
>> set(gca,'fontSize',20);
>> Simu.test_Geweke

ans =

     0
     0

>>
```

Convergence

Gibbs sampler

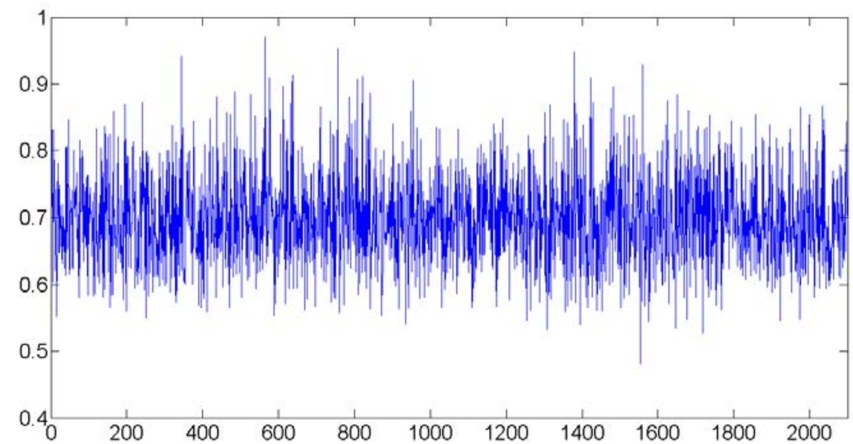
- **Results :**

plot(Simu.post_beta')



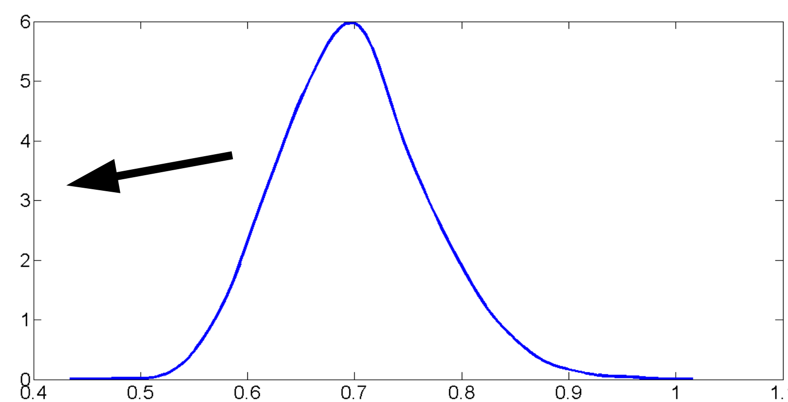
mean(Simu.post_beta')= 0,50 ; 0,32
std(Simu.post_beta') = 0,08 ; 0,06

plot(Simu.post_sigma')



mean(Simu.post_sigma')= 0,70
std(Simu.post_sigma') = 0,07

Marginal distribution
of the variance



Gibbs sampler

- Testing for the order of the AR process :
  - Launch several times the program with different # of lags
  - Compare the MLL obtained from the different simulations

	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
MLL HM	-274,16	-259,89	-256,36	-257,16	-254,75
MLL Chib	-279,12	-269,7	-269,71	-273,72	-275,46
MLL IS	-279,12	-269,7	-269,71	-273,72	-275,46

- Harmonic mean : not reliable
- Same Estimates from the local and the global formula

MH sampler

- **Open the function :** ***MH_regression_with_MLL***

  - MH sampler of a simple regression
  - Provide the marginal log-likelihood (MLL) estimated from three different approaches (Chib, Importance sampling and Harmonic mean)

- **Running an estimation of AR model with MH sampler :** ***launch_AR_regression_with_Gibbs_sampler***

Gibbs sampler

- **Adaptation of the proposal distribution**

(Atchadé and Rosenthal (2005))

$$\bar{\Sigma}_i = \bar{\Sigma}_{i-1} + (\phi_r^{i-1} - \phi_{target})/(i^c) \;\; \text{if} \;\; \Sigma_{Low} < \bar{\Sigma}_i < \Sigma_{High}$$

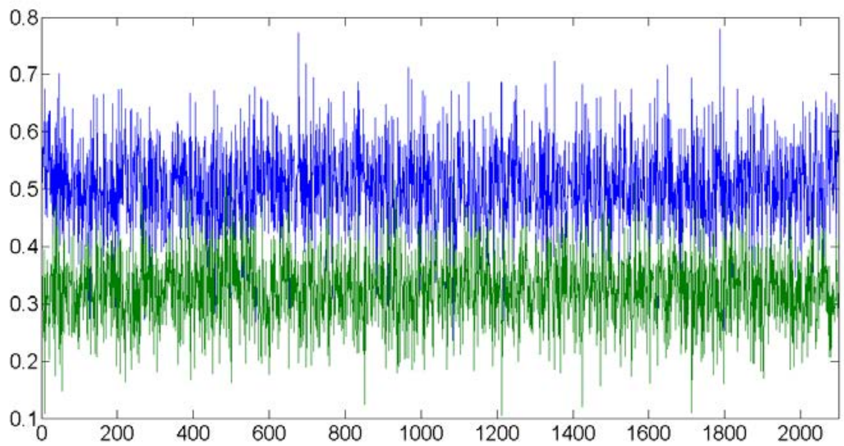
```
    for q=1:dimension+1
        adapt_rate(q) = max(min_max_var(1),adapt_rate(q) + (accept_rate(q)/i - 0.45)/(i^eta));
        if(adapt_rate(q)>min_max_var(2))
            adapt_rate(q) = min_max_var(2);
        end
    end
```

ϕ_{target}

Comparison Samplers

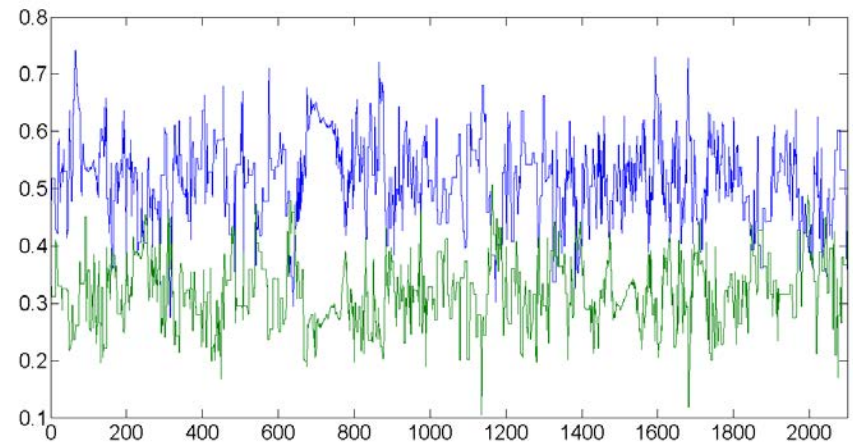
- AR(1)

Gibbs sampler



Mean= 0,50 ; 0,32
Std = 0,08 ; 0,06

Metropolis-Hastings



Mean = 0,51 ; 0,32
Std = 0,07 ; 0,06

Acceptance rate :
45 % ; 44 % ; 45 %

MH sampler

- Testing for the order of the AR process :
  - Launch several times the program with different # of lags
  - Compare the MLL obtained from the different simulations

	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
MLL Chib (MH)	-279,12	-269,7	-269,71	-273,72	-275,46
MLL Chib (Gibbs)	-279,12	-269,7	-269,71	-273,72	-275,46

Same Estimates from Gibbs or MH samplers

Marginal likelihood depends on the prior!

- Test the order of the AR process with different priors

GARCH estimation by MH MCMC

- Change the directory and go to *GARCH_with_MH*
- Import financial data by clicking on *SP500_percentage_returns.mat*
- Main matlab program *MCMC_GARCH_RW*

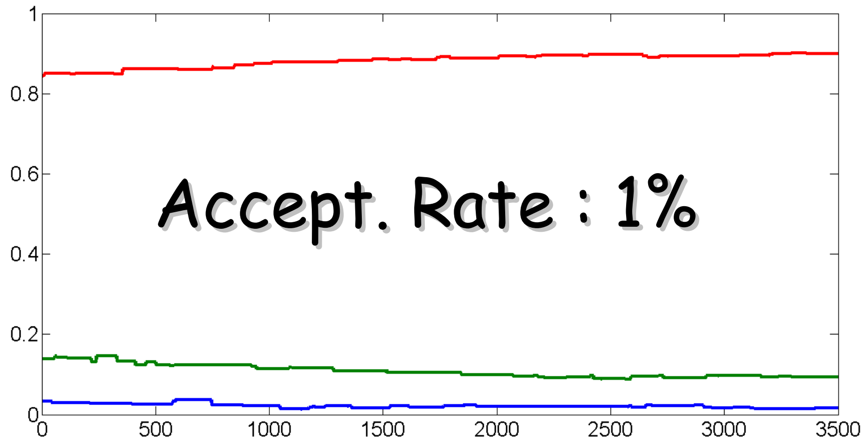
  Estimate a GARCH(1,1) model with
    - non adaptive RW
    - adaptive RW

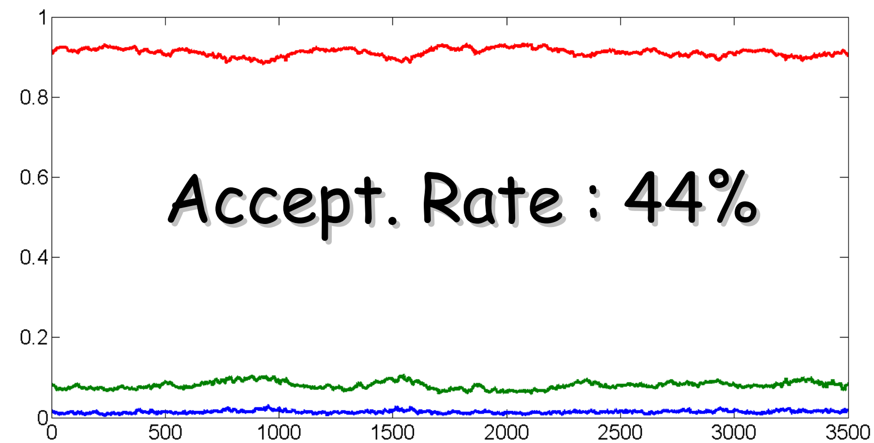
Inputs :

- Y : financial time series **(here SP500)**
- nb_MCMC : number of MCMC iterations
- RW_step : Variance of the proposal distribution
- Graph : Convergence graphics

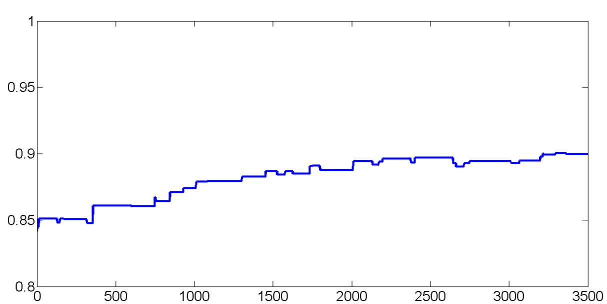
GARCH estimation by MH MCMC

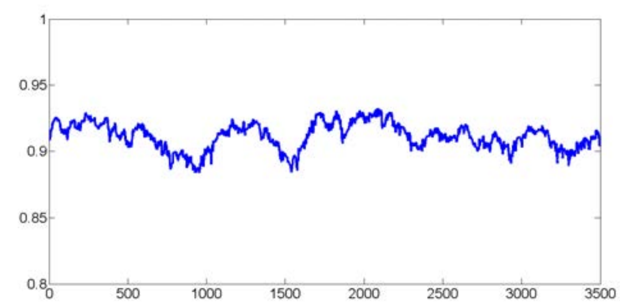
- Choose a RW variance and run the program
- [Simu] = MCMC_GARCH_RW(SP500,5000,0.1)





- Focus on parameter β in $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$





Change-Point AR models

CP-AR models with CGS

- Change the directory and go to *CP_models_with_CGS*
- Import data by clicking on US_GDP_percentage.*mat*
- Main matlab program
  Gibbs_regression_Carlin_Gelman_Smith
  - Estimates a CP model with 2 regimes using CGS's algorithm

Inputs :

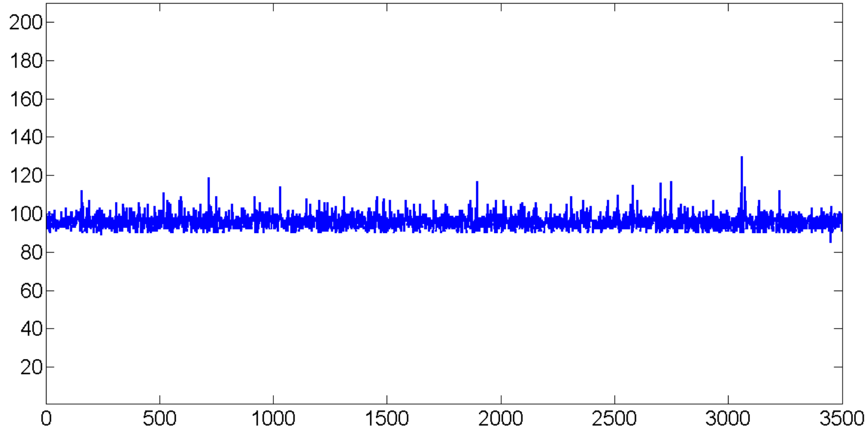
  - Y : a time series (here US_GDP_growth)
  - X : explanatory variables
  - nb_MCMC : number of MCMC iterations

- Program for estimating a CP-AR(q) model
  Launch_CP_AR_Carlin_Gelman_Smith

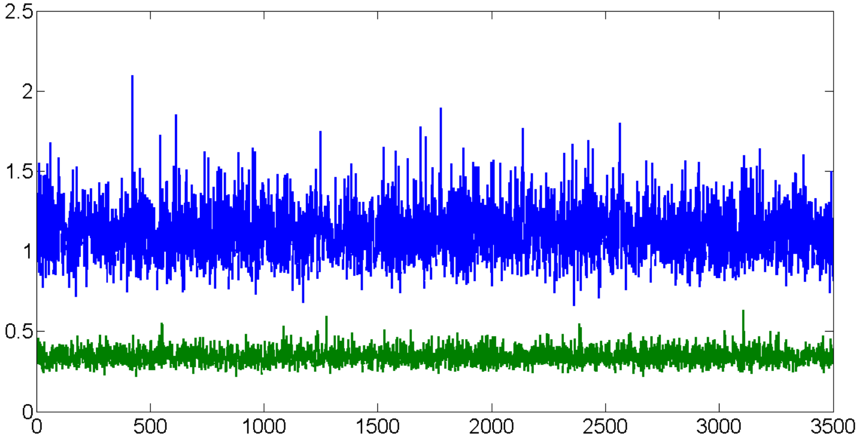
CP-AR models with CGS

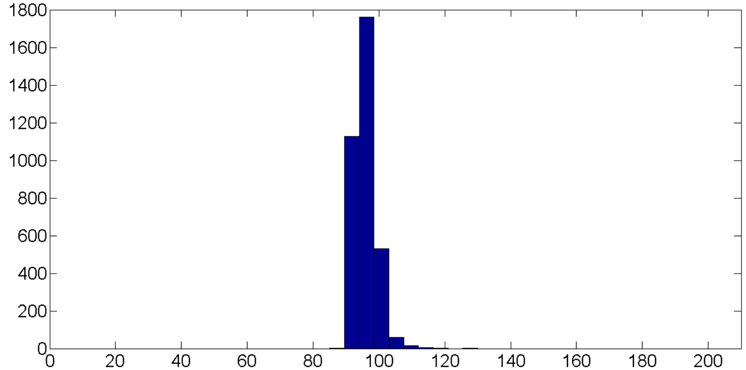
- Run an estimation of a CP-AR(1) model
  [Simu] = launch_CP_AR_Carlin_Gelman_Smith(US_GDP_growth,1,5000)

plot(Simu.post_tau')



plot(Simu.post_sigma')





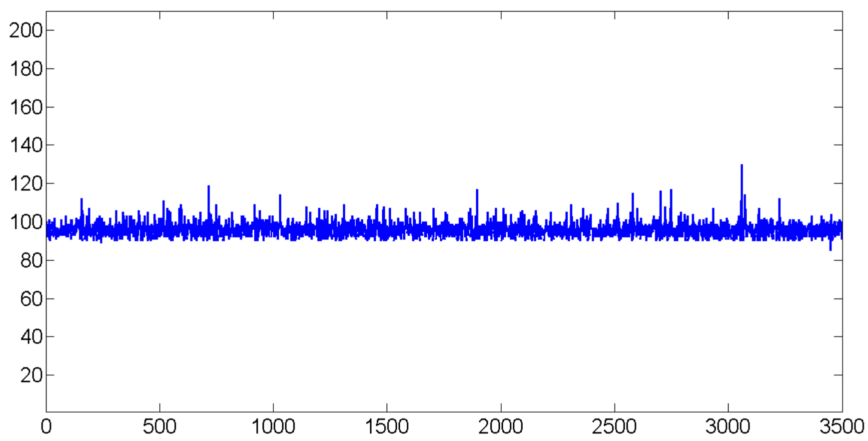
- Great moderation
- Not a symmetric distribution

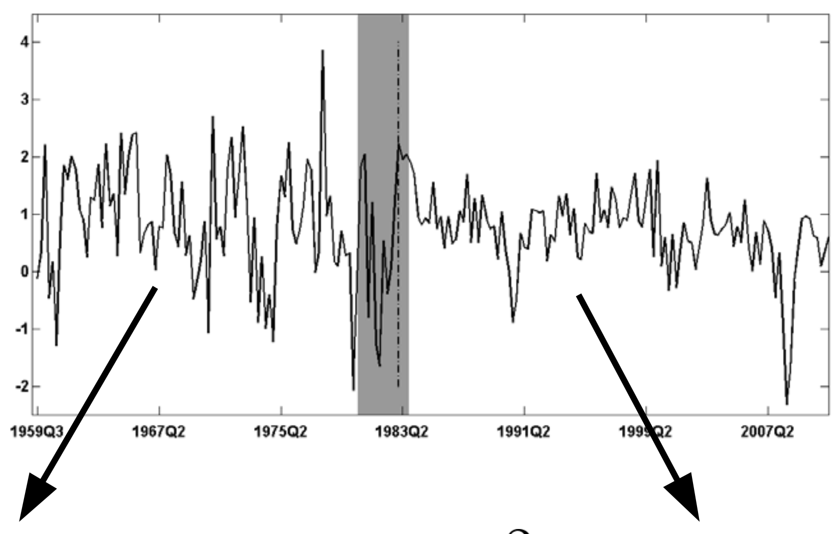
CP-AR models with CGS

- Run an estimation of a CP-AR(1) model

  [Simu] = launch_CP_AR_Carlin_Gelman_Smith(US_GDP_growth,1,5000)

plot(Simu.post_tau')





$$E(\sigma_1^2|Y_{1:T}) \approx 1.11 \qquad E(\sigma_2^2|Y_{1:T}) \approx 0.34$$

CP-AR models with Chib

- **Change the directory and go to *CP_models_with_Chib***
- **Import data by clicking on US_GDP_percentage.*mat***
- **Main matlab program *Gibbs_regression_chib***
  - Estimates a CP model with k regimes using Chib's algorithm

Inputs :

- Y : a time series **(here US_GDP_growth)**
- X : explanatory variables
- Regime : number of regimes
- nb_MCMC : number of MCMC iterations
- MLL_computation : =1 if MLL must be estimated

CP-AR models with Chib

- **Main matlab program** ***Gibbs_regression_chib***
  - Estimates a CP model with k regimes using Chib's algorithm
- **Matlab program** ***launch_CP_model_estimations***
  - Estimates CP-AR models from one up to k regimes

Inputs :

- Y : a time series **(here US_GDP_growth)**
- AR_lags : The order of the AR process
- upper_bound_regime : Max. considered number of regimes
- nb_MCMC : number of MCMC iterations

CP-AR models with Chib

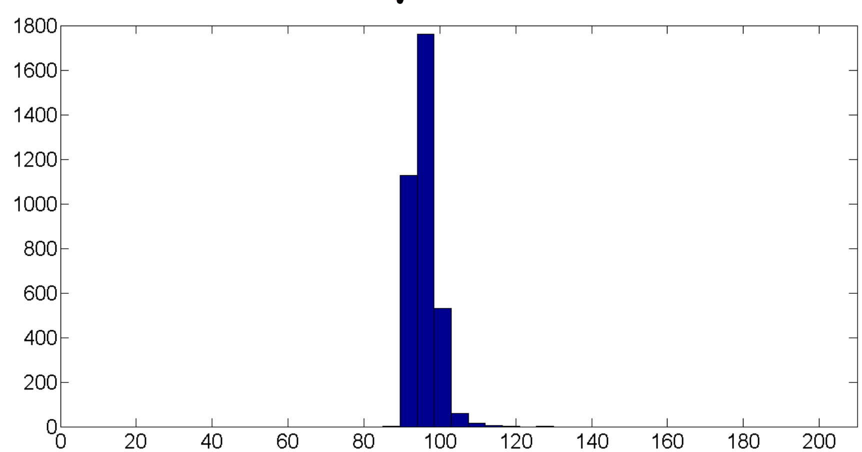
- Run an estimation of CP-AR(1) models from one to 3 regimes

[Simu MLL] =
launch_CP_model_estimations(US_GDP_growth,1,3,10000)

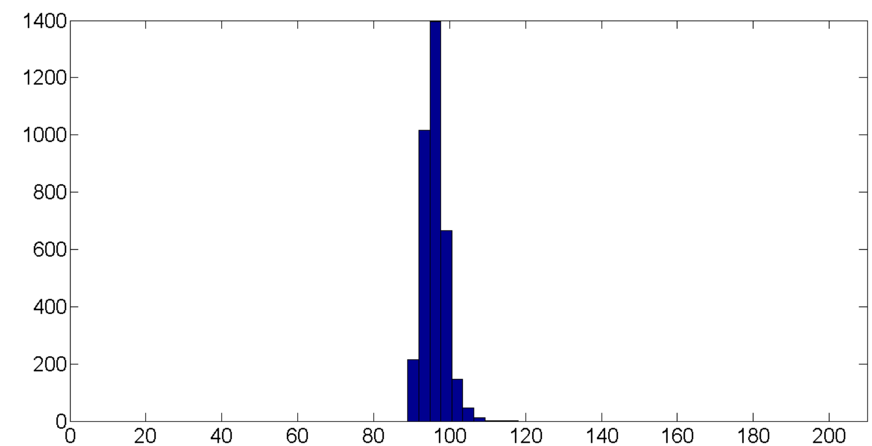
nb_MCMC

- Two regimes

Griddy-Gibbs



Chib



Gibbs sampler

- **The model**

$$\begin{cases} y_t &= \beta'_{s_t} x_t + \epsilon_t \\ \epsilon_t &\sim \text{i.i.d.} N(0, \sigma^2_{s_t}) \end{cases}$$

- **The prior distributions**

$\beta_i \sim N(\beta_0, \Sigma_0) \ \ \forall \ i \in [1, K+1]$ and $\sigma_i^2 \sim IG(IG_a, IG_b) \forall i \in [1, K+1]$

Transition prob. From *i* to *i* : $p_{ii} \sim \text{Beta}(\alpha_p, \beta_p)$

Trans. State :
$$\begin{cases} P(s_t = s_{t-1} | s_{t-1}, P) = p_{s_{t-1}, s_{t-1}} \\ P(s_t = s_{t-1} + 1 | s_{t-1}, P) = 1 - p_{s_{t-1}, s_{t-1}} \forall s_{t-1} \in [1, K] \end{cases}$$

Gibbs sampler

- **The model**

$$\begin{cases} y_t &= \beta'_{s_t} x_t + \epsilon_t \\ \epsilon_t &\sim \text{i.i.d.} N(0, \sigma^2_{s_t}) \end{cases}$$

Priors in the program : Gibbs_regression_chib

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%% Set the Hyper-parameters
%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Prior :
%%% beta_i ~ N(beta_0,var_uninformative)
%%% sigma^-2 ~ G(a,b)
%%% p_ii ~ Beta(alpha,beta)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
var_uninformative = 10; %% We fix the variance of each beta equal to var_u
beta_0 = zeros(dimension,1);
inv_Sigma_0 = diag(ones(dimension,1)*(1/var_uninformative));
det_inv_Sigma_0 = det(inv_Sigma_0);

%%% hyper-parameters for the variances
IG_b = 1;
IG_a = 1;

%%% hyper-parameters for each probability of staying in the same regime
alpha_p = 10;
beta_p = 1;
```

$$\beta_i \sim N(\beta_0, \Sigma_0)$$

$$\sigma_i^2 \sim IG(IG_a, IG_b)$$

$$p_{ii} \sim \text{Beta}(\alpha_p, \beta_p)$$

Forward-Backward

- **Gibbs step :** $S_{1:T}|Y_{1:T},\beta_1,...,\beta_{K+1},\sigma_1^2,...,\sigma_{K+1}^2,P$

Program for sampling a state vector : Forward_Backward

Two steps :

1) Compute the forward prob. $f(s_t|Y_{1:t})\forall t\in[1,T]$
2) Sample a state using the decomposition

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

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%% Backward algorithm : sampling a state vector
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
sn = zeros(T,1);
sn(T) = regime;
for t=T-1:-1:1
    etat_fut = sn(t+1);
    back = forward(t,:)'.*P(:,etat_fut);
    back = back/sum(back);
    sn(t) =  multinomialrnd(back);
end
```

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

Marginal likelihood

- **Local formula :**

Likelihood (by F-B) — Priors

$$f(Y_{1:T}) = \frac{f(Y_{1:T}|\beta_1^*,\dots,\beta_{K+1}^*,\sigma_1^{2,*},\dots,\sigma_{K+1}^{2,*},P^*)\,f(\beta_1^*,\dots,\beta_{K+1}^*,\sigma_1^{2,*},\dots,\sigma_{K+1}^{2,*},P^*)}{\pi(\beta_1^*,\dots,\beta_{K+1}^*,\sigma_1^{2,*},\dots,\sigma_{K+1}^{2,*},P^*|Y_{1:T})}$$

- **Third term :**

$$\pi(\beta_1^*,\dots,\beta_{K+1}^*,\sigma_1^{2,*},\dots,\sigma_{K+1}^{2,*},P^*|Y_{1:T}) = \underbrace{\pi(P^*|Y_{1:T})}_{1}\underbrace{\pi(\sigma_1^{2,*},\dots,\sigma_{K+1}^{2,*}|P^*,Y_{1:T})}_{2}\underbrace{\pi(\beta_1^*,\dots,\beta_{K+1}^*|\sigma_1^{2,*},\dots,\sigma_{K+1}^{2,*},P^*,Y_{1:T})}_{3}$$

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

$$\pi(P^*|Y_{1:T}) \approx \frac{1}{N}\sum_{i=1}^{N}\pi(P^*|Y_{1:T},S_{1:T}^i)$$

Marginal likelihood

2) $$\pi(\sigma_1^{2,*}, ..., \sigma_{K+1}^{2,*}|P^*, Y_{1:T}) = \int \pi(\sigma_1^{2,*}, ..., \sigma_{K+1}^{2,*}|P^*, \beta_1, ..., \beta_{K+1}, S_{1:T}, Y_{1:T}) \pi(S_{1:T}, \beta_1, ..., \beta_{K+1}|P^*, Y_{1:T}) dS_{1:T} d\beta_1 ... d\beta_{K+1}$$
$$\approx \frac{1}{G_1} \sum_{i=1}^{G_1} \pi(\sigma_1^{2,*}, ..., \sigma_{K+1}^{2,*}|P^*, \beta_1^i, ..., \beta_{K+1}^i, S_{1:T}^i, Y_{1:T})$$

- Run an auxiliary MCMC with fixed P*

3) $$\pi(\beta_1^*, ..., \beta_{K+1}^*|P^*, \sigma_1^{2,*}, ..., \sigma_{K+1}^{2,*}, Y_{1:T}) = \int \pi(\beta_1^*, ..., \beta_{K+1}^*|S_{1:T}, \sigma_1^{2,*}, ..., \sigma_{K+1}^{2,*}, Y_{1:T}) \pi(S_{1:T}|P^*, \sigma_1^{2,*}, ..., \sigma_{K+1}^{2,*}, Y_{1:T}) dS_{1:T}$$
$$\approx \frac{1}{G_2} \sum_{i=1}^{G_2} \pi(\beta_1^*, ..., \beta_{K+1}^*|S_{1:T}^i, \sigma_1^{2,*}, ..., \sigma_{K+1}^{2,*}, Y_{1:T})$$

- Run second auxiliary MCMC with fixed $\mathrm{P}^*, \sigma_1^{2,*}, ..., \sigma_{K+1}^{2,*}$

Model selection

- Run an estimation of CP-AR(1) models from one to 3 regimes
  [Simu MLL] = launch_CP_model_estimations(US_GDP_growth,1,3,10000)

Uninformative prior :

$\beta_i \sim N(\underline{0}, 1000 I_d)$ $\sigma_i^{-2} \sim G(0.01, 100)$ $p_{ii} \sim \mathrm{Beta}(0.5, 0.5)$

#Regime	1	2	3
MLL	-275.57	-269.25	-275.10

Informative prior :

$\beta_i \sim N(\underline{0}, I_d)$ $\sigma_i^{-2} \sim G(1, 1)$ $p_{ii} \sim \mathrm{Beta}(0.5, 0.5)$

#Regime	1	2	3
MLL	-265.28	-249.96	-243.52

Model selection

- Choose your prior according to 'your break sensitivity'
- Or use another criterion such as the predictive likelihood

→ Less impacted by the prior distributions

$$
\begin{aligned}
f(Y_{t+1:T}|Y_{1:t}) &= \frac{f(Y_{1:T})}{f(Y_{1:t})} \\
&= \frac{f(Y_{1:T}|\theta^*)f(\theta^*)}{\pi(\theta^*|Y_{1:T})} \frac{\pi(\theta^*|Y_{1:t})}{f(Y_{1:t}|\theta^*)f(\theta^*)} \\
&= \frac{f(Y_{1:T}|\theta^*)}{f(Y_{1:t}|\theta^*)} \frac{\pi(\theta^*|Y_{1:t})}{\pi(\theta^*|Y_{1:T})}
\end{aligned}
$$

No prior distributions! Impact through the posterior