Chapter 2

Structural breaks for models *without path dependence*
Chapter 2

- Motivation (p. 3)
- Change-point models (p. 8)
- Markov-switching and Change-point models (p. 22)
  - Forward-backward algorithm
  - Label switching
- References (p. 51)
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

\[
\begin{align*}
\alpha_1 + \beta_1 &= 0.95 \\
\alpha_2 + \beta_2 &= 0.95 \\
\alpha_3 + \beta_3 &= 0.99
\end{align*}
\]
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:
  \[ y_t | Y_{1:t-1} \sim N(\theta'_1 [1 \ y_{t-1}]', \sigma_1^2) \text{ if } \tau_1 < t \]
  \[ \sim N(\theta'_2 [1 \ y_{t-1}]', \sigma_2^2) \text{ otherwise} \]

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

- Prior distributions:
  \[ \sigma_i^2 \sim IG(\alpha, \beta) \quad i \in [1, 2] \]
  \[ \theta_i \sim N(\mu_0, \Sigma_0) \quad i \in [1, 2] \]
  \[ \tau_1 \sim U(1, T - 1) \]
Carlin, Gelfand and Smith (1992)

- **Gibbs sampler** :

\[
\theta_1|Y_{1:T}, \tau_1, \theta_2, \sigma_1^2, \sigma_2^2 \sim N(\bar{\mu}_1, \bar{\Sigma}_1)
\]

\[
\sigma_1^2|Y_{1:T}, \tau_1, \theta_1, \theta_2, \sigma_2^2 \sim IG(\alpha + \tau_1/2, \beta + \sum_{t=1}^{\tau_1} \epsilon_t)
\]

\[
\theta_2|Y_{1:T}, \tau_1, \theta_1, \sigma_1^2, \sigma_2^2 \sim N(\bar{\mu}_2, \bar{\Sigma}_2)
\]

\[
\sigma_2^2|Y_{1:T}, \tau_1, \theta_1, \theta_2, \sigma_1^2 \sim IG(\alpha + (T - \tau_1)/2, \beta + \sum_{t=\tau_1+1}^{T} \epsilon_t)
\]

\[
\tau_1|Y_{1:T}, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2 \sim \text{Griddy-Gibbs}
\]

**with**

\[
\begin{aligned}
\bar{\mu}_1 &= \bar{\Sigma}_1[\sigma_1^{-2} \sum_{t=1}^{\tau_1} x_t y_t + \Sigma_0^{-1} \mu_0] \\
\bar{\Sigma}_1 &= [\sigma_1^{-2} \sum_{t=1}^{\tau_1} (x_t x_t') + \Sigma_0^{-1}]^{-1}
\end{aligned}
\]

**and**

\[
\begin{aligned}
\bar{\mu}_2 &= \bar{\Sigma}_2[\sigma_2^{-2} \sum_{t=\tau_1+1}^{T} x_t y_t + \Sigma_0^{-1} \mu_0] \\
\bar{\Sigma}_2 &= [\sigma_2^{-2} \sum_{t=\tau_1+1}^{T} (x_t x_t') + \Sigma_0^{-1}]^{-1}
\end{aligned}
\]
Carlin, Gelfand and Smith (1992)

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

  \[ \rightarrow \textbf{Discrete conditional distribution} \]

1) Compute the posterior density for each \( i \) and normalize.
2) Draw \( u \sim U(0,1) \)
3) Find \( \tau \) 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:

  \[ \begin{array}{c|c|c|c|c}
  \tau_1 & \sigma_1^2 & \sigma_2^2 & \theta_1 & \theta_2 \\
  1983 - Q1 & 1.1 & 0.34 & (0.64) & (0.32) \end{array} \]
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

  \[ \tau = \{ \tau_0, \tau_1, ..., \tau_K, \tau_{K+1} \} \]

  \[ \Theta = \{ \theta'_1, ..., \theta'_{K+1} \}' \]

  \[ \Sigma = \{ \sigma^2_1, ..., \sigma^2_{K+1} \}' \]

  **Prior distributions on break parameters:**

  \[ \tau_0 = 0 \]

  \[ \tau_1 \sim U(1, T - (K + 2)) \quad \tau_{K+1} = T \]

  \[ \tau_i | \tau_{i-1} \sim U(\tau_i + 1, T - (K + 2 - i)) \]
Stephens (1994)

- **Gibbs sampler**: if AR process as before

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

with \[
\begin{align*}
\bar{\mu}_i &= \bar{\Sigma}_i [\sigma_i^{-2} \sum_{t=\tau_{i-1}+1}^{\tau_i} x_t y_t + \Sigma_0^{-1} \mu_0] \quad \forall i \in [1, K + 1] \\
\bar{\Sigma}_i &= [\sigma_i^{-2} \sum_{t=\tau_{i-1}+1}^{\tau_i} (x_t x_t') + \Sigma_0^{-1}]^{-1} \quad \forall i \in [1, K + 1]
\end{align*}
\]
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.: $\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.999 \\ 0.999 & 1 \end{pmatrix}\right)$

After 10000 draws

First 200 draws of the MCMC

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

![Diagram showing good and bad initial states](image)
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

\[ y_t | Y_{1:t-1}, s_t, \Theta, \Sigma \sim N(\theta'_{s_t} [1 \ y_{t-1}]', \sigma^2_{s_t}) \]

\[ s_t | s_{t-1}, P \sim p_{s_{t-1}, s_t} \]

• The set of parameters:

\[ \Theta = \{\theta'_1, ..., \theta'_{K+1}\}' \quad \text{Mean parameters} \]

\[ \Sigma = \{\sigma^2_1, ..., \sigma^2_{K+1}\}' \quad \text{Var. parameters} \]

\[ S_{1:T} = \{s_1, s_2, ..., s_T\} \quad \text{Discrete states} \]

\[ P \quad \text{Transition matrix} \]

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:

$s_t = i \quad \forall i \in [1, K + 1]$  

at time $t$, active state: $i$

Change-point configuration

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

No recurrent state!
Chib (1996 - 1998)

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

Markov-switching configuration

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

Recurrent states
Parsimonious model
Difficult to estimate (label switching problem)
Chib (1996 - 1998)

Markov-switching configuration

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

Change-Point configuration

**CP and MS models are mixture models with time-varying probabilities**
Chib (1996 - 1998)

- AR process of order 1
  \[
  y_t | Y_{1:t-1}, s_t \sim N(\theta'_s [1 \ y_{t-1}]', \sigma^2_{st})
  \\
  s_t | s_{t-1}, P \sim P_{s_{t-1}, s_t}
  \]

- 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 \( n_{i,1:j} = [n_{i,1} \ n_{i,2} \ ... \ n_{i,j}] \)

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
  \[
  s_T \xrightarrow{\xi} s_{T-1} \xrightarrow{\xi} \ldots \xrightarrow{\xi} s_1
  \]

  \[
  \pi(s_T|Y_{1:T}) \pi(s_{T-1}|Y_{1:T}, s_T) \ldots \pi(s_1|Y_{1:T}, S_{2:T})
  \]
Forward-Backward algorithm

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

- **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:**
  \[
  y_t|Y_{1:t-1}, s_t \sim N(\theta'_s x_t, \sigma^2_{s_t})
  \]
  with
  \[
  x_t = [1 \ y_{t-1} \ldots \ y_{t-p}]'
  \]

  then
  \[
  f(y_t|Y_{1:t-1}, s_t) = \left(2\pi\sigma^2_{s_t}\right)^{-\frac{1}{2}} e^{-\frac{(y_t-\theta'_s x_t)^2}{2\sigma^2_{s_t}}}
  \]
  **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: \( \pi(s_t | Y_{1:T}, s_{t+1:T}) \forall t \in [1, T-1] \)

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

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

\[
f(s_t | Y_{1:t}) \propto f(y_t | Y_{1:t-1}, s_t) f(s_t | Y_{1:t-1})
\]

\[
\propto f(y_t | Y_{1:t-1}, s_t) \sum_{i=1}^{K+1} f(s_t = i | s_{t-1}) f(s_{t-1} | Y_{1:t-1})
\]

\[
\propto f(y_t | Y_{1:t-1}, s_t) \sum_{i=1}^{K+1} p_{s_{t-1}, s_t=i} f(s_{t-1} | Y_{1:t-1})
\]

Starting point

\[
\begin{align*}
&f(s_1 = 1 | y_1) \\
&\vdots \\
&f(s_1 = K + 1 | y_1)
\end{align*}
\]

until $T$
Forward-Backward algorithm

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

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 \( F \in \mathbb{R}^{T,K+1} \)

\[
\begin{pmatrix}
  f(s_1 = 1|y_1) & \ldots & f(s_{K+1} = 1|y_1) \\
  f(s_2 = 1|Y_{1:2}) & \ldots & f(s_2 = K+1|Y_{1:2}) \\
  \vdots & \ddots & \vdots \\
  f(s_T = 1|Y_{1:T}) & \ldots & f(s_T = K+1|Y_{1:T})
\end{pmatrix}
\]

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 \( f(s_t = i|Y_{1:T}, S_{t+1:T}) = \frac{f(s_t = i|Y_{1:t})f(s_{t+1}|s_t = i)}{\sum_{j=1}^{K+1} f(s_t = j|Y_{1:t})f(s_{t+1}|s_t = j)} \)
Chib's Gibbs sampler

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

Prior on $P$: 

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

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

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

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

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

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

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

<table>
<thead>
<tr>
<th>No break</th>
<th>One break</th>
<th>Five breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi(\Theta, \Sigma</td>
<td>Y_{1:T}, K = 0) )</td>
<td>( \pi(\Theta, \Sigma, S_{1:T}, P</td>
</tr>
</tbody>
</table>

• At each time, compute the marginal quantity

<table>
<thead>
<tr>
<th>No break</th>
<th>One break</th>
<th>Five breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(Y_{1:T}</td>
<td>K = 0) )</td>
<td>( f(Y_{1:T}</td>
</tr>
</tbody>
</table>

• 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

\[ \pi(\Theta^*, \Sigma^*, P^* | Y_{1:T}) = \pi(P^* | Y_{1:T}) \pi(\Theta^* | Y_{1:T}, P^*) \pi(\Sigma^* | Y_{1:T}, P^*, \Theta^*) \]

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?
  
  \[ f(Y_{1:T} | K = k) = \int f(Y_{1:T} | \Theta, \Sigma, P) f(\Theta, \Sigma, P) d\Theta d\Sigma dP \]
  
  \[ = \int f(Y_{1:T} | \Theta, \Sigma, P) f(\Theta, \Sigma, P) \frac{q(\Theta, \Sigma, P)}{q(\Theta, \Sigma, P)} d\Theta d\Sigma dP \]
  
  \[ \approx \frac{1}{G} \sum_{i=1}^{G} \frac{f(Y_{1:T} | \Theta^i, \Sigma^i, P^i) f(\Theta^i, \Sigma^i, P^i)}{q(\Theta^i, \Sigma^i, P^i)} \]

Where \( \{\Theta^i, \Sigma^i, P^i\} \sim q(-) \)
US Monthly Inflation Rate

<table>
<thead>
<tr>
<th>#regimes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLL</td>
<td>-1424</td>
<td>-1409</td>
<td>-1391</td>
<td>-1386</td>
<td>-1384</td>
<td>-1388</td>
</tr>
</tbody>
</table>
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
  $$\Theta|\mu_0, \Sigma_0 \sim \prod_{i=1}^{K+1} N(\mu_0, \Sigma_0)$$
  $$\mu_0 \sim N(\underline{\mu}, \Sigma_{\mu})$$
  $$\Sigma_0 \sim IW(\nu, V)$$

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

- The hierarchical (random) parameters gather information from the different regimes
  
  If new regime: draw parameters from the hier. dist.
Predictions of structural breaks

- Hierarchical distributions

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

\[ \begin{aligned}
\mu_0 &\sim N(\mu_\mu, \Sigma_\mu) \\
\Sigma_0 &\sim IW(\nu, V)
\end{aligned} \]

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

\[ \theta_{K+2} \sim N(\mu_0, \Sigma_0) \]
\[ \theta_{K+3} \sim N(\mu_0, \Sigma_0) \]

...
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>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Posterior parameter estimates for the unconstrained AR(1) hierarchical HMC model with six break-points</strong></td>
</tr>
<tr>
<td>Regimes</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>S.E.</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>S.E.</td>
</tr>
<tr>
<td>AR(1) coefficient</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>S.E.</td>
</tr>
<tr>
<td>Variances</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>S.E.</td>
</tr>
<tr>
<td>Transition probability matrix</td>
</tr>
<tr>
<td>Mean duration</td>
</tr>
</tbody>
</table>

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

- Issue arises in MS specification (recurrent states)

\[ P = \begin{pmatrix}
   p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,K+1} \\
   p_{2,1} & p_{2,2} & p_{2,3} & \cdots & p_{2,K+1} \\
   \vdots & \vdots & \vdots & \ddots & \vdots \\
   p_{K+1,1} & p_{K+1,2} & p_{K+1,3} & \cdots & p_{K+1,K+1}
\end{pmatrix} \]

- Posterior distribution: invariant to the label of the states

Example:

\begin{align*}
\theta_1 &= 10, \theta_2 = 2 \\
\theta_1 &= 2, \theta_2 = 10
\end{align*}

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 symmetric to the labeling

   Example:

   \[
   \begin{align*}
   &\text{Regime 1} & \text{Regime 2} \\
   &y_t \sim N(10, 1) & y_t \sim N(2, 1) \\
   &1 & 100 & T \\
   &\text{Prior distributions} & & \\
   &\theta_1 \sim N(\mu_0, \Sigma_0) & & \\
   &\theta_2 | \theta_1 \sim N(\mu_0, \Sigma_0) \delta_{\theta_2 < \theta_1} & & \\
   \end{align*}
   \]

   \[\theta_1 = 2, \theta_2 = 10 \rightarrow \text{Impossible label}\]
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)


• Stephens (1994)


• Chib (1996)

References

• Chib (1998)


• Pesaran, Pettenuzzo and Timmermann (2006)

Other CP and MS specs

- Koop and Potter (2007) – CP models
  CP models without geometric durations
  Inference of breaks without marginal likelihood
  Extremely demanding!


- 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

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

Estimation by SMC

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


- Whiteley, Andrieu, Doucet (2013) - CP models
  Unknown 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


• Label-switching

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