Relevant parameter changes in structural break models

A. Dufays

J. Rombouts

Forecasting from Complexity
April 27th, 2018
Outline

**Sparse Change-Point models**

1. Motivation
2. Model specification
   - Shrinkage prior
   - Break estimation
3. Application to time series
The sparse CP model

1. Motivation
Motivations

Empirical evidence of structural breaks in macroeconomic and financial time series

• Breaks due to changes in market sentiments, regulatory conditions, misspecification of the model, …

• Growing interests in flexible models dealing with breaks

Main motivations

1. Economic interpretations
   Better understanding of the time series dynamics
2. Forecasting
US GDP growth rate (1947-2014)

**Standard CP model**

\[ y_t = \alpha_i + \beta_i y_{t-1} + \epsilon_t \quad \text{with} \quad \epsilon_t \sim N(0, \sigma_i^2), \; i \in [1, 2] \]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate 1</th>
<th>Estimate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.05</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>AR term</td>
<td>0.44</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>1.38</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.04)</td>
</tr>
</tbody>
</table>

Chib, S. ‘Estimation and comparison of multiple change-point models’, *JoE*, **1998**, 86, 221-241
Motivations

**Limitation of Change-point models**

1. Optimal number of regimes computed by Marginal likelihood
   - Many useless estimations and uncontrolled penalty.
2. Every new regime increases the number of parameters
   - Over-parametrization.
3. Parameters related to short periods exhibit high uncertainties

**Sparse Change-point models**

1. Optimal number of regimes obtained in one estimation.
   - Controlled penalty
2. Controls the over-parametrization.
   - Only a few parameters change over time
US GDP growth rate (1947-2014)

**Standard CP model**

- Intercept: 1.05 (0.17), 0.56 (0.12)
- AR term: 0.44 (0.07), 0.55 (0.08)
- $\sigma^2$: 1.38 (0.16), 0.36 (0.04)

**Sparse CP model**

- Intercept: 0.66 (0.13), 0.63 (0.09)
- AR term: 0.54 (0.05), 0.54 (0.05)
- $\sigma^2$: 1.44 (0.17), 0.35 (0.04)

Dufays A.
The sparse model

2. Model specification
Model specification

To ease the presentation, we focus on a simple AR(1) model.

Let $y_{1:T} = \{y_1, \ldots, y_T\}$ be a univariate real valued time series of $T$ observations

$$y_t = \alpha_0 + \beta_0 y_{t-1} + \sigma_0 \varepsilon_t, \text{ where } \varepsilon_t \sim i.i.d. N(0, 1).$$

**Change-point specification:**

$$y_t = \alpha_k + \beta_k y_{t-1} + \sigma_k \varepsilon_t, \text{ for } t \in [\tau_k, \tau_{k+1}],$$

where $\tau_0 = 1 < \tau_1 < \ldots < \tau_{K+1} = T$ denote the break points.
Model specification

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

Test all the possibilities and choose according to the marginal likelihood

1. Too many posterior distributions to simulate
   - For CP models: 4 regimes and 4 parameters by regime: 256 models!

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

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

Keep the standard specification
but shrink irrelevant parameters to zero

1. Only one estimation is required.

2. Break is identified only if it improves the likelihood function.
Model specification

Reframing the model with first-differenced parameters:

\[ y_t = \alpha_0 + \sum_{i=1}^{k} \Delta \alpha_i + [\beta_0 + \sum_{i=1}^{k} \Delta \beta_i] y_{t-1} + [\sigma_0 + \otimes_{i=1}^{k} \Delta \sigma_i] \varepsilon_t \quad \text{for} \quad t \in [\tau_{k-1}, \tau_k] \]

The parameters in level are obtained by

\[ \alpha_k = \alpha_0 + \sum_{i=1}^{k} \Delta \alpha_i \]

Assuming we know the true break dates:

Boils down to a high-dimensional shrinkage problem:

\[
\left\{ \alpha_0, \Delta \alpha_1, \ldots, \Delta \sigma_{K_{\text{max}}} \right\} = \arg \max_{\theta} \sum_{t=2}^{T} \ln f_N(y_t | \mu_t(\theta, y_{t-1}), \sigma^2_t(\theta)) + l(\theta).
\]

Dufays A.
Model specification

What should be the best penalty function?

1. Sparsity: irrelevant parameters shrink to zero
2. Unbiasedness: to accommodate short regime

**Lasso or Ridge penalty functions lead to biased estimators**

We opt for a Hard thresholding penalty function like BIC:

$$l(\theta) = - \sum_{k=1}^{K_{\text{max}}} P_T \left[ 1 \{ \Delta \alpha_k \neq 0 \} + 1 \{ \Delta \beta_k \neq 0 \} + 1 \{ \Delta \sigma_k \neq 1 \} \right]$$

with $P_T = -0.5 \ln T$ if Bayesian information criterion.
Problem solved?

Talk is not over since the approach exhibits two major issues:

1. Likelihood function is not continuous anymore
   • Optimization very difficult and only in small dimensions.
2. We do not know the true break dates…

We carry out the estimation in the Bayesian framework.

1. Needs shrinkage prior distributions that mimic the BIC penalty
2. Needs to select the maximum number of breaks
   • MCMC not appropriate \(\rightarrow\) Estimation using a SMC sampler.
1. shrinkage prior distributions that mimic the BIC

We propose a spike and slab prior based on two uniform distributions:

The 2MU : a Mixture of two Uniform components

\[ \Delta \alpha_k \sim \omega \ U\left[\frac{-a}{2}, \frac{a}{2}\right] + (1 - \omega) U\left[\frac{-b}{2}, \frac{b}{2}\right] \]
Mixture of two uniform components

The prior is denoted $2MU(a, b, P)$

1. $a$ : bound of the narrow uniform component (related to no break).
2. $b$ : bound of the wide uniform component.

Example with $a = 1, b = 10$

$P = -5$

$$\log f(x) - \log f(0) = P$$

$P = -3$
Shrinkage prior on the model parameters

Each differenced parameter is driven by a 2MU distribution:

$$\Delta \alpha_k \sim 2MU(a,b,P)$$

Given the break dates, the posterior mode is given by

$$\left\{ \alpha_0, \Delta \alpha_1, \ldots, \Delta \sigma_{K_{\text{max}}} \right\} = \arg\max_{\theta} \sum_{t=2}^{T} \ln f_N(y_t | \mu_t(\theta, y_{t-1}), \sigma_t^2(\theta)) + l(\theta),$$

with

$$l(\theta) = \ln f(\alpha_0, \beta_0, \sigma_0) + \sum_{k=1}^{K_{\text{max}}} P[1 \{ \frac{a}{2} < |\Delta \alpha_k| < \frac{b}{2} \} + 1 \{ \frac{a}{2} < |\Delta \beta_k| < \frac{b}{2} \} + 1 \{ \Delta \sigma_k \in B \}].$$

The prior acts as a hard thresholding penalty function
2. How to deal with the break parameters?

So far, the break dates are assumed to be known.

• We infer the break dates using a Metropolis-Hastings algorithm.

But…

• Does not solve how to fix the maximum number of breaks.
• Difficult to set a prior on the break dates.

We exploit the frequentist literature on change-point detection to build an informative prior for the break dates.
Exploiting exact inference

Ng, Pan and Yau (2017) propose a simple procedure to build a set of potential break dates: \( J = \{\tilde{\tau}_1, \ldots, \tilde{\tau}_{K_{\text{max}}} \} \)

Under mild conditions (that apply in the current framework), they prove that, asymptotically,

- The number of potential break dates is never underestimated.
- Each true break date is in the h-neighbourhood of one potential break date with probability one.

However, results only valid asymptotically.

Moreover, they suggest a value of h equal to 100 observations.

Minimum regime duration of 100 observations.
Exploiting exact inference

We modify slightly their CP detection procedure to build a larger potential break date set: $J_1 = \{\tilde{\tau}_1, \ldots, \tilde{\tau}_{K_{\text{max}}} \}$ with $J \subset J_1$.

- The maximum number of break is set to $K_{\text{max}} = |J_1|$.
- The prior of each break parameter is set to a uniform distribution with bounds defined by the previous and the next potential break:
  
  $\tau_2 \sim U\left[\frac{\tilde{\tau}_1 + \tilde{\tau}_2}{2} - 1, \frac{\tilde{\tau}_2 + \tilde{\tau}_3}{2}\right]$
The Sparse CP model

3. Simulations
AR Data generating process

AR Model: $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t$ where $\epsilon_t \sim N(0, \sigma^2)$

<table>
<thead>
<tr>
<th>DGP 1 - AR(1)</th>
<th>DGP 2 - AR(2)</th>
<th>DGP 3 - AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Breaks</strong></td>
<td><strong>Sample size</strong></td>
<td><strong>Breaks</strong></td>
</tr>
<tr>
<td>{}</td>
<td>500</td>
<td>{514,768}</td>
</tr>
<tr>
<td><strong>Values</strong></td>
<td><strong>Regimes</strong></td>
<td><strong>Values</strong></td>
</tr>
<tr>
<td>{0}</td>
<td>1</td>
<td>{0}</td>
</tr>
<tr>
<td>{-0.7}</td>
<td>1</td>
<td>{0.9,1.69,1.32}</td>
</tr>
<tr>
<td>{}</td>
<td>(-)</td>
<td>{0,-0.81,-0.81}</td>
</tr>
<tr>
<td>{1}</td>
<td>1</td>
<td>{1,1}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DGP 4 - AR(1)</th>
<th>DGP 5 - AR(1)</th>
<th>DGP 6 - AR(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Breaks</strong></td>
<td><strong>Sample size</strong></td>
<td><strong>Breaks</strong></td>
</tr>
<tr>
<td>{50}</td>
<td>1024</td>
<td>{400,750}</td>
</tr>
<tr>
<td><strong>Values</strong></td>
<td><strong>Regimes</strong></td>
<td><strong>Values</strong></td>
</tr>
<tr>
<td>{0}</td>
<td>1</td>
<td>{0}</td>
</tr>
<tr>
<td>{0.75,-0.5}</td>
<td>2</td>
<td>{0.999}</td>
</tr>
<tr>
<td>{}</td>
<td>(-)</td>
<td>{}</td>
</tr>
<tr>
<td>{}</td>
<td>(-)</td>
<td>{}</td>
</tr>
<tr>
<td>{}</td>
<td>(-)</td>
<td>{}</td>
</tr>
<tr>
<td>{}</td>
<td>(-)</td>
<td>{}</td>
</tr>
<tr>
<td>{1}</td>
<td>1</td>
<td>{1,2.25,1}</td>
</tr>
</tbody>
</table>

Dufays A.
## AR Data generating process

<table>
<thead>
<tr>
<th># Regimes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>True dates</th>
<th>Estimated dates</th>
<th>St. dev. dates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGP 1</strong></td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>{514,768}</td>
<td>{498.81,770.82}</td>
<td>{35.26,18.71}</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>{514}</td>
<td>{515.98}</td>
<td>{73.01}</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0</td>
<td>94</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>99</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DGP 2</strong></td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>{514,768}</td>
<td>{498.81,770.82}</td>
<td>{35.26,18.71}</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>6</td>
<td>6</td>
<td>87</td>
<td>1</td>
<td>0</td>
<td>{514}</td>
<td>{515.98}</td>
<td>{73.01}</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0</td>
<td>94</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>99</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DGP 3</strong></td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>{400,612}</td>
<td>{398.63,610.65}</td>
<td>{6.10,4.13}</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DGP 4</strong></td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>{50}</td>
<td>{49.58}</td>
<td>{6.17}</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>1</td>
<td>98</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DGP 5</strong></td>
<td>95</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>{400,750}</td>
<td>{394.71,754.29}</td>
<td>{17.29,21.15}</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>96</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0</td>
<td>1</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0</td>
<td>1</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The sparse model

4. Empirical applications

4.1 Macro series
4.2 Realized volatility series
4.3 Financial returns.
4.1 Macro series
3-Month US Treasury Bill (1947-2002)


**CP-AR(1) model**
## 3-Month US Treasury Bill (1947-2002)

### PPT’s result : CP-AR model with 7 regimes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>0.021</td>
<td>0.252</td>
<td>0.017</td>
<td>0.220</td>
<td>0.412</td>
<td>0.246</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.208)</td>
<td>(0.067)</td>
<td>(0.161)</td>
<td>(0.521)</td>
<td>(0.211)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.002</td>
<td>0.895</td>
<td>1.006</td>
<td>0.969</td>
<td>0.958</td>
<td>0.968</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.071)</td>
<td>(0.020)</td>
<td>(0.026)</td>
<td>(0.045)</td>
<td>(0.27)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.023</td>
<td>0.256</td>
<td>0.015</td>
<td>0.260</td>
<td>2.558</td>
<td>0.161</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.068)</td>
<td>(0.003)</td>
<td>(0.031)</td>
<td>(0.671)</td>
<td>(0.027)</td>
<td>(0.005)</td>
</tr>
</tbody>
</table>

### Sparse CP-AR(1) model

<table>
<thead>
<tr>
<th>Dates</th>
<th>→ 12-2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>[0.02 0.07]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dates</th>
<th>→ 09-2000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[08-2000 11-2000]</td>
</tr>
</tbody>
</table>

| $\beta_1$ | 0.99 |
|           | [0.98 1.00] |
|           | 0.92 |
|           | [0.89 0.95] |

<table>
<thead>
<tr>
<th>Dates</th>
<th>→ 12-2002</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>[1.53 2.85]</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>[0.06 0.09]</td>
</tr>
</tbody>
</table>

| $\sigma^2$ | 0.02 |
|            | [0.02 0.03] |
|            | 0.18 |
|            | [0.13 0.26] |
|            | 0.02 |
|            | [0.01 0.03] |
|            | 0.27 |
|            | [0.23 0.32] |
|            | 0.67 |
|            | [0.07 0.10] |

Dufays A.
4.2 Realized volatility series
Sparse CP-HAR models

HAR model: standard model to forecast the realized variance.
• Stands for a constrained AR(22) model.

• **Do the parameters evolve over time?**

\[
y_t = (\beta_{0,1} + \sum_{i=2}^{s_t} \Delta \beta_{0,i}) + (\beta_{1,1} + \sum_{i=2}^{s_t} \Delta \beta_{1,i})y_{t-1} + (\beta_{2,1} + \sum_{i=2}^{s_t} \Delta \beta_{2,i})y_{t-1}^{(w)} + (\beta_{3,1} + \sum_{i=2}^{s_t} \Delta \beta_{3,i})y_{t-1}^{(m)} + \varepsilon_t \quad \sim \quad N(0, \sigma_1^2 \prod_{i=2}^{s_t} \Delta \sigma_i^2) \quad \text{for} \quad s_t = 2, \ldots, K + 1,
\]

• Application to 11 daily series from January 3, 2000 to August 5, 2015 (around 3900 observations)
## Standard CP-HAR models

### Best models according to a standard CP-HAR model

<table>
<thead>
<tr>
<th>Series</th>
<th># Regimes</th>
<th>Break dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEX</td>
<td>1</td>
<td>2001-10-31 2005-08-03 2005-08-26</td>
</tr>
<tr>
<td>CAC</td>
<td>4</td>
<td>2007-02-26 2007-03-06 2009-03-10 2009-09-30</td>
</tr>
<tr>
<td>DJI</td>
<td>5</td>
<td>2004-02-18 2005-07-06 2005-07-08</td>
</tr>
<tr>
<td>FTSE</td>
<td>4</td>
<td>2011-03-11 2011-03-24</td>
</tr>
<tr>
<td>Russell</td>
<td>6</td>
<td>2007-02-05 2009-02-10 2009-09-30</td>
</tr>
</tbody>
</table>

Many regimes even if we neglect the short ones.

Dufays A.
## Sparse CP-HAR model

<table>
<thead>
<tr>
<th>Series</th>
<th>Shrinkage CP-HAR ($K_{Max} = 8$)</th>
<th># Regimes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_0$ $\beta_1$ $\beta_2$ $\beta_3$ $\sigma^2$</td>
<td></td>
</tr>
<tr>
<td>AEX</td>
<td>1 1 1 1 4</td>
<td>1</td>
</tr>
<tr>
<td>CAC</td>
<td>1 1 1 1 5</td>
<td>4</td>
</tr>
<tr>
<td>DAX</td>
<td>1 1 1 1 5</td>
<td>1</td>
</tr>
<tr>
<td>DJI</td>
<td>1 1 1 1 4</td>
<td>5</td>
</tr>
<tr>
<td>EuroStoxx</td>
<td>1 1 1 1 8</td>
<td>6</td>
</tr>
<tr>
<td>FTSE</td>
<td>1 1 1 1 6</td>
<td>4</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>1 1 1 1 5</td>
<td>5</td>
</tr>
<tr>
<td>NIKKEI</td>
<td>1 1 1 1 5</td>
<td>4</td>
</tr>
<tr>
<td>Russell</td>
<td>2 2 1 1 4</td>
<td>6</td>
</tr>
<tr>
<td>SMI</td>
<td>1 1 1 1 6</td>
<td>6</td>
</tr>
<tr>
<td>SP500</td>
<td>1 1 1 1 4</td>
<td>4</td>
</tr>
</tbody>
</table>
S&P 500

\[ \beta_{1:T \mid y_1:T}^{(d)} \]

\[ \sigma_{1:T \mid y_1:T}^2 \]
4.3 Financial returns
GARCH(1,1) application

• 384 stocks that are in the S&P 500 index for the full January 2000 to October 2017 period (4,464 observations).

• We estimate the Sparse-GARCH(1,1) model on each stock

\[
\begin{align*}
y_t &= \sigma_t \eta_t \quad \text{with} \quad \eta_t \sim N(0, 1), \\
\sigma^2_t &= \omega + \alpha (y^2_{t-1} - \sigma^2_{t-1}) + \beta (\sigma^2_{t-1} - \omega) \\
\end{align*}
\]

unc. var. persistence

**Table 1** – Estimated probabilities (in %) of the most likely number of regimes per model parameter over the 384 series.

<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>(\omega)</td>
<td><strong>38.54</strong></td>
<td>35.68</td>
<td>20.05</td>
<td>5.21</td>
<td>0.52</td>
<td>0</td>
</tr>
<tr>
<td>(\alpha)</td>
<td><strong>73.18</strong></td>
<td>22.66</td>
<td>3.65</td>
<td>0.52</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(\beta)</td>
<td>37.24</td>
<td><strong>55.21</strong></td>
<td>6.51</td>
<td>1.04</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Dufays A.
GARCH(1,1) application

- Average of the posterior medians of the unc. var.
- Red: all sectors and black: financial sector
Conclusion

**Sparse Change-point models**

1. Detects the parameters that change from one regime to another.
2. Shrinks every irrelevant parameters toward zero.
   • One estimation and no need of the Marginal likelihood.

**Empirical contributions**

1. Could improve the interpretation of the presence of breaks.
2. Could test the sparsity of a specific model.
3. Good prediction performances.
Papers

1. ‘Relevant parameter changes in structural break models’
   Dufays, A. and Rombouts, J.
   - Models with path dependence issue: ARMA/GARCH
   - Univariate and inference by SMC

2. ‘Sparse change-point HAR models for realized volatility’
   (forthcoming in ER), Dufays, A. and Rombouts, J.
   - AR/HAR models
   - Univariate and inference by Gibbs sampler

3. ‘Sparse change-point VAR process’
   Dufays, A and Li, Z. and Rombouts, J. and Song, Y.
   - Multivariate AR/HAR models and inference by Gibbs sampler
   - Future breaks
Prior: Theoretical justifications

**Given the simple model:** 

\[ y_t = \mu + \Delta \mu 1_{t > \tau} + \epsilon_t \]

The MAP of \( \Delta \mu | y_{1:T}, \sigma^2, \mu, \tau \) is given by

\[
\operatorname{Arg.Min.} |\Delta \mu| < \frac{b}{2} \left\{ \frac{1}{2 \sigma^2} \sum_{t=\tau+1}^{T} (y_t - \mu - \Delta \mu)^2 - P1_{\left(\frac{a}{2} < |\Delta \mu| \leq \frac{b}{2}\right)} \right\}
\]

- \( L_0 \) type of penalization
- Intuitive way to fix the penalty parameter (BIC, DIC, …)

**If the penalty grows with the sample size:**

\[
\operatorname{Prob}[\text{Break detected} | y_{1:T}, \sigma^2, \mu, \tau] \xrightarrow{p} 1
\]

if the true parameter \( \Delta \mu \) lies in \([−b/2, b/2]\)\(\setminus\left[−\frac{a}{2}, \frac{a}{2}\right]\)

\[
\operatorname{Prob}[\text{Break detected} | y_{1:T}, \sigma^2, \mu, \tau] \xrightarrow{p} 0 \quad \text{Otherwise.}
\]
How to set the 2MU hyper-parameters?

- The short Uniform component:
  For every model parameters, the threshold is set to the difference between the median and the 5%-th quantile of the marginal posterior distribution of the model without structural breaks.

- The penalty $P$: Random variable around the BIC:

\[
\frac{P(M_2|y_{1:T})}{P(M_1|y_{1:T})} \geq \frac{0.95}{1 - 0.95} \\
\approx e^{-0.5(BIC_{M_2} - BIC_{M_1})} \geq \frac{0.95}{1 - 0.95} \\
\log f(y_{1:T}|\{\hat{\Theta}, \hat{\Sigma}\}_{M_2}) - \log f(y_{1:T}|\{\hat{\Theta}, \hat{\Sigma}\}_{M_1}) \geq \log \frac{0.95}{1 - 0.95} + \log T
\]
Motivations

**Standard model versus Change-point one**

\[ y_t = c + \beta y_{t-1} + \epsilon_t \]

No dynamic for the parameters

\[ y_t = c_i + \beta_i y_{t-1} + \epsilon_t \]

A Markov-chain drives the dynamic of the breaks
Sequential Monte Carlo Samplers

Sequence of distributions:

\[ f_\phi(x_{1:t} \mid y_{1:t}) = \frac{f(y_{1:t} \mid x_{1:t})^{\phi n}}{Z_\phi} f(x_{1:t}) \]

\[ f_0(x_{1:t} \mid y_{1:t}) = f(x_{1:t}) \]

\[ f_p(x_{1:t} \mid y_{1:t}) \propto f(y_{1:t} \mid x_{1:t}) f(x_{1:t}) \]

Initialization from the prior

1. Correction step
2. Re-sampling step
3. Mutation step
The sparse CP-AR model

4. Simulation
Simple DGP from Chan et al. (2014)

Estimation of the Sparse CP-AR(2) model with K=10

Dufays A.
Estimation of the Sparse CP-AR(2) model (K=10)

Quarterly Canadian Inflation 1961Q2-2012Q2

Marg. Prob. Of having a break

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.43</td>
<td>0.19</td>
<td>0.19</td>
<td>0.18</td>
<td>0.01</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1) term</td>
<td>0.40</td>
<td>0.13</td>
<td>0.42</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(2) term</td>
<td>0.19</td>
<td>0.69</td>
<td>0.11</td>
<td>0.01</td>
<td>0.00</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>0.11</td>
<td>0.38</td>
<td>0.39</td>
<td>0.10</td>
<td>0.01</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Estimation of the Sparse CP-AR(2) model

Quarterly Canadian Inflation 1961Q2-2012Q2

Dufays A.
Quarterly Canadian Inflation 1961Q2-2012Q2

**Sensitivity with respect to the penalty**

**Intercept**

**AR(1) parameter**
Quarterly Canadian Inflation 1961Q2-2012Q2

Sensitivity with respect to the penalty

AR(2) parameter

Variance

Dufays A.
Quarterly US GDP growth rate 1959-2011

Estimation of the Sparse CP-ARMA model

Intercept

AR(1) Coefficient

MA(1) Coefficient

Variance

Dufays A.
Quarterly US GDP growth rate 1959-2011

Sensitivity with respect to the penalty parameter

Intercept

AR parameter

Dufays A.
Quarterly US GDP growth rate 1959-2011

Sensitivity with respect to the penalty parameter

MA parameter

Variance

Dufays A.
The sparse model

2.1 Shrinkage prior
Shrinkage priors

Let us focus on one **first-differenced** parameter of the model: $\alpha_{ik}$

Recall that the parameter in level at regime $k$ and dimension $i$ is given by

$$\tilde{\alpha}_{ik} = (\tilde{\alpha}_{i1} + \sum_{i=2}^{k} \alpha_{ik})$$

What kind of shrinkage prior should be apply to $\alpha_{ik}$?

**Shrinkage priors in CP context must comply with**

1. **Unbiasedness:** typically an uninformative prior.

2. **Sparsity:** High density around zero.

   Mixture of Uniform distributions.
We propose a spike and slab prior based on two uniform distributions:

The 2MU: a Mixture of two Uniform components

\[ \alpha_{ik} \mid \sim \omega \ U \left[ \frac{-a}{2}, \frac{a}{2} \right] + (1 - \omega) U \left[ \frac{-b}{2}, \frac{b}{2} \right] \]
Mixture of two uniform components

The prior is denoted $2MU(a, b, P)$

1. $a$: bound of the narrow uniform component (related to no break).
2. $b$: bound of the wide uniform component.

Example with $a = 1, b = 10$

$$P = -5$$

$$\log f(x) - \log f(0) = P$$

$$P = -3$$
Shrinkage prior on the model parameters

Each differenced parameter is driven by a 2MU distribution:

\[ \alpha_{ik} \sim 2MU(a, b, P) \]

Given the break dates, the posterior mode is given by

\[
\{\alpha, \beta\} = \arg\max_{\{\alpha, \beta\}} \ln f_N(y_{1:T}, f(\alpha, Z), \Sigma(\beta)) + l(\alpha, \beta),
\]

with \( l(\alpha, \beta) = \ln \left( f(\alpha_1, \beta_1) \right) + \sum_{k=1}^{K_{\text{max}}+1} \left[ \sum_{i=1}^{K_1} P1\left\{ \frac{a}{2} < |\alpha_{ki}| < \frac{b}{2} \right\} \right] + \sum_{j=1}^{K_2} P1\{\beta_{kj} \in B\} \).

The prior acts as a hard thresholding penalty function
Shrinkage prior and its hyper-parameters

How to set the hyper-parameters?

For each differenced parameter: $\alpha_{ik} \sim 2MU(a_i, b, P_{ik})$

1. The bounds of the narrow component is parameter-dependent to account for the scale of each explanatory variable.

   Based on estimates of the model without break on segmented data.

2. The penalty value is parameter-dependent to account for its uncertainty in the posterior distribution:

   $\exp P_{ik} \equiv \rho_{ik} \sim U\left[\frac{1}{2}\rho_{\text{BIC}}, \frac{3}{2}\rho_{\text{BIC}}\right]$ where $\rho_{\text{BIC}} = e^{P_{\text{BIC}}}$ with $P_{\text{BIC}} = -\ln T$
Shrinkage prior: an example

\[ a_i = 1, \quad b = 10, \quad P_{\text{BIC}} = -3, \exp P_{ik} \sim U\left[\frac{1}{5}\rho_{\text{BIC}}, \frac{3}{2}\rho_{\text{BIC}}\right] \]
Penalty function

How to visualize the impact of the penalty value?

We consider a simple model with breaks only in the variance:

$$y_t = \mu + \epsilon_t, \text{ where } \epsilon_t \sim N(0, \sigma_0^2 \prod_{j=1}^{K} \Delta \sigma_j^2 1_{\{t > \tau_j\}}),$$

with $\Delta \sigma_j^2 \sim 2MU^+(a, b, P)$.

When $K = 1$, a break is detected if,

$$\ln\left[e^{2P - \eta} (\lambda \bar{\sigma}_{1: \tau_1}^2 + (1 - \lambda) \bar{\sigma}_{\tau_1+1:T}^2)\right] > \lambda \ln \bar{\sigma}_{1: \tau_1}^2 + (1 - \lambda) \ln \bar{\sigma}_{\tau_1+1:T}^2, \text{ with } \lambda = \frac{\tau_1}{T},$$

where $\bar{\sigma}_{t_1:t_2}^2$ is the empirical variance of $\{y_t - \bar{\mu}_{1:T}\}$ over the sample $[t_1, t_2]$. 
Range of break detection

Given a sample size, a penalty value and a value of \( \sigma_0^2 \)
We can simulate series and compute if a break would be detected:

\[
\Delta \sigma_1^2
\]

\[
\lambda = \tau / T
\]
2.2 Break parameters
How to deal with the break parameters?

So far, the break dates are assumed to be known.
- We infer the break dates using a Metropolis-Hastings algorithm.

But…
- Does not solve how to fix the maximum number of breaks.
- Difficult to set a prior on the break dates.

**We exploit the frequentist literature on change-point detection to build an informative prior for the break dates.**
Exploiting exact inference

Ng, Pan and Yau (2017) propose a simple procedure to build a set of potential break dates: \( J = \{\tilde{\tau}_1, \ldots, \tilde{\tau}_{K_{\max}}\} \)

Under mild conditions (that apply in the current framework), they prove that, asymptotically,

- The number of potential break dates is never underestimated.
- Each true break date is in the \( h \)-neighbourhood of one potential break date with probability one.

However, results valid asymptotically.

Moreover, they suggest a value of \( h \) equal to 100 observations.
Exploiting exact inference

We modify slightly their CP detection procedure to build a potential break date set: \( J_1 = \{ \tilde{\tau}_1, \ldots, \tilde{\tau}_{K_{\text{max}}} \} \) with \( J \subset J_1 \).

- The maximum number of break is set to \( K_{\text{max}} = |J_1| \).
- The prior of each break parameter is set to a uniform distribution with bounds defined by the previous and the next potential break:

\[
\tau_2 \sim U\left[ \frac{\tilde{\tau}_1 + \tilde{\tau}_2}{2} - 1, \frac{\tilde{\tau}_2 + \tilde{\tau}_3}{2} \right]
\]
Sparse CP models

3. Estimation
Penalized likelihood

With shrinkage prior, the shape of the posterior distribution can be highly multimodal.

- Exacerbated with L0-type of penalty function.

- MCMC is not well suited for highly multimodal distribution
  → Markov chain which implies that next state only depends on the current one.

We use a sequential monte carlo sampler for simulating the posterior distribution
MCMC issue with multimodal distributions

Convergence?

MCMC iterations : 0
Sequential Monte Carlo sampler

In a nutshell...

- The SMC sampler builds a sequence of slightly different distributions
  - Starts from the prior to the posterior distribution.
- Each distribution of the sequence is approximated using Importance sampling.
  - The proposal being the previous distribution of the sequence.
- Rejuvenation of the particles is done using an MCMC algorithm.

Particles start by being very diffuse and converge slowly to the posterior distribution
Sequential Monte Carlo Samplers

Sequence of distributions: \( \pi_n(x_n | y_{1:t}) = \frac{f(x_n | y_{1:t}) \phi(n) f(x_n)}{Z_n} \)

Tempering function such that \( \phi(1) = 0 \) and \( \phi(p) = 1 \).
\[
f_\phi(x_{1:t} | y_{1:t}) = \frac{f(y_{1:t} | x_{1:t})^\phi f(x_{1:t})}{Z_\phi}
\]

\[
f_0(x_{1:t} | y_{1:t}) = f(x_{1:t})
\]

\[
f_p(x_{1:t} | y_{1:t}) \propto f(y_{1:t} | x_{1:t}) f(x_{1:t})
\]

Initialization from the prior

1. Correction step
2. Re-sampling step
3. Mutation step
Sequential Monte Carlo Sampler

SMC sampler vs MCMC

Tempering : 0

MCMC iterations : 0

Dufays A.
Bayesian inference

Univariate AR/MA or GARCH models

- **Sampling from the posterior distribution**: SMC sampler
- **Marginal likelihood**: SMC sampler

Univariate/multivariate AR/HAR models

- **Sampling from the posterior distribution**: Gibbs sampler
- **Marginal likelihood**: Stepping-Stone algorithm.