Incorporating Structural Breaks in GARCH Models

_BI Economics Seminar_

*September 11, 2013*

Luc Bauwens
Université catholique de Louvain (Belgium)
Center for Operations Research and Econometrics
Thanks to my (co-)authors

- Jeroen Rombouts, Professor at ESSEC Business School
- Arnaud Dufays, Postdoc at CREST (Paris)

of the paper supporting this talk:

Marginal likelihood for Markov-switching and change-point GARCH models, forthcoming in the *Journal of Econometrics*. 
Motivation

- "Simple" GARCH models, like

\[ y_t = \epsilon_t \sigma_t, \quad \epsilon_t \sim N(0, 1), \]
\[ \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \]

estimated on long financial return series imply a strong persistence \((\alpha + \beta < 1 \approx 1)\) of the conditional variance.

- That spuriously strong persistence may be caused by changes of the unconditional variance level.

- Source of problem: the parameters of the GARCH equation are fixed throughout the entire sample.

⇒ Need for more flexible specifications.
S&P 500 index returns

Sample: May 20, 1999 to April 25, 2011 (3000 observations)
More flexible GARCH models

- Component models: Ding and Granger (1996); Engle and Lee (1999); Bauwens and Storti (2007) ...

- Smooth transition models: Gonzales-Rivera (1996) ...


- Mixture, Change-point (CP), and Markov-switching (MS) models.

See survey in Chapter 1 of *Handbook of Volatility Models and Their Applications*, Wiley (April 2012).
MS- and CP-GARCH models


- Prototype model:

  \[ y_t = \epsilon_t \sigma_t, \quad \epsilon_t \sim N(0, 1), \]
  \[ \sigma_t^2 = \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2, \]

  with \( s_t \) a discrete r.v. taking values in \( \{1, 2, \ldots, K + 1\} \).

- NB: \( \sigma_t^2 \) depends on \( S_t = (s_1, s_2, \ldots, s_{t-1}, s_t) \).
  We should write:
  \[ \sigma_t^2(S_t) = \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2(S_{t-1}). \]
MS- and CP-GARCH models

\( \{ s_t \} \) is a first-order Markovian process with transition matrix \( P \).

**Markov-switching case (recurrent regimes):**

\[
PM = \begin{pmatrix}
    p_{11} & p_{12} & \ldots & \ldots & p_{1K+1} \\
    p_{21} & p_{22} & p_{23} & \ldots & p_{2K+1} \\
    \ldots & \ldots & \ldots & \ldots & \ldots \\
    p_{K1} & p_{K2} & p_{K3} & \ldots & p_{K+1K+1}
\end{pmatrix}
\]

**Change-point case:**

\[
PC = \begin{pmatrix}
    p_{11} & 1 - p_{11} & 0 & \ldots & \ldots & 0 \\
    0 & p_{22} & 1 - p_{22} & \ldots & \ldots & 0 \\
    \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
    0 & 0 & 0 & \ldots & p_{KK} & 1 - p_{KK} \\
    0 & 0 & 0 & \ldots & 0 & \ldots & 1
\end{pmatrix}
\]
Which model: CP or MS?

- CP more robust to misspecification of the number of regimes.
  
  Intuition: CP with enough change points can mimic MS with small number of regimes.

- MS more parsimonious in GARCH parameters, less in $P$.

- Identification restrictions needed (label switching) in MS, not in CP.

- CP model inherently non-stationary, contrary to MS.

- Empirical issue to be decided by a model choice criterion or a statistical test.
ML estimation is impracticable

- The likelihood function for observation $t$ depends on the unobservable state variables $s_t$ from 1 to $t$. They must be integrated out by summation over all possible past paths of $S_t$ (path dependence problem).

- With $K + 1$ regimes and $t$ observations, there are $(K + 1)^t$ terms in the summation.

- Standard ML estimation is impracticable for sample sizes typically used in financial econometrics. ML with simulation: forthcoming paper by M. Augustyniak in CSDA.

- This problem is less important in CP models.

- It does not arise in the ARCH case: Hamilton and Susmel (JBES, 1994), Cai (JBES, 1994).
Path-dependence with two states

NB: \( \sigma_{t|1,2,...t}^2 = \sigma_t^2(S_t) \)

\[
\begin{align*}
\sigma_{1|1}^2 &= \omega_1 + \beta_1 \sigma_0^2 + \alpha_1 u_0^2 \\
\sigma_{2|1,2}^2 &= \omega_2 + \beta_2 \sigma_0^2 + \alpha_2 u_0^2 \\
\sigma_{2|1,1}^2 &= \omega_1 + \beta_1 \sigma_{1|1}^2 + \alpha_1 u_1^2 \\
\sigma_{2|1,2}^2 &= \omega_2 + \beta_2 \sigma_{1|1}^2 + \alpha_2 u_1^2 \\
\sigma_{2|2,1}^2 &= \omega_1 + \beta_1 \sigma_{1|2}^2 + \alpha_1 u_1^2 \\
\sigma_{2|2,2}^2 &= \omega_2 + \beta_2 \sigma_{1|2}^2 + \alpha_2 u_1^2
\end{align*}
\]
Bayesian estimation is practicable

- Bayesian estimation is feasible by MCMC methods.

- Bauwens, Preminger and Rombouts (EJ, 2010): algorithm for MS-GARCH model. Applicable to CP-GARCH. Problems:
  - Not efficient numerically (does not mix well);
  - Cannot compute the marginal likelihood (for model choice).

- He and Maheu (CSDA, 2010): CP-GARCH model. Problems:
  - Not applicable to MS-GARCH;
  - Numerically demanding;
  - May not converge in some cases.
Our contribution

- New and more efficient algorithm applicable to CP- and MS-GARCH models.

- Able to compute the marginal likelihood and perform model choice:
  Number of regimes?
  CP or MS?

- Applied to eleven time series, the MS-GARCH model preferred to CP-GARCH in all cases.

- C++ code available at https://sites.google.com/site/websiteofarnaudufays/
S&P 500 index: MS-GARCH

Returns with switches of 2 regime MS-GARCH model
S&P 500 index: CP-GARCH

Returns with switches of 3 regime CP-GARCH model
Model and parameters

- Prototype model:

\[ y_t = \epsilon_t \sigma_t, \quad \epsilon_t \sim N(0, 1), \]
\[ \sigma_t^2 = \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2, \]

with \( s_t \) a discrete r.v. taking values in \{1, 2, \ldots, K + 1\}.

- Parameters: GARCH, transition matrix, and states:

\[ \theta = (\omega_1, \ldots, \omega_{K+1}, \alpha_1, \ldots, \alpha_{K+1}, \beta_1, \ldots, \beta_{K+1}), \]
\[ \pi: \text{the non-redundant parameters of } P_M \text{ or } P_C, \]
\[ S_T = (s_1 s_2 \ldots s_T): \text{treated as parameters.} \]

- We proceed assuming a know value of \( K \), the number of breaks.
Data augmentation

- By treating $S_T$ as additional parameters, the problem is tractable.

- **Intuition**: if $S_T$ is observed, the likelihood function/data density is easy to compute (no need to integrate all possible past paths):

$$f(Y_T | \theta, S_T, \pi) \propto \prod_{t=1}^{T} (\sigma_t^{-1}) \exp \left( -\frac{y_t^2}{2\sigma_t^2} \right),$$

where $Y_T = (y_1, y_2, \ldots, y_T)$.

- Furthermore, inference about $S_T$ is useful as it provides indirectly estimates of break dates.
Posterior distribution

- Joint posterior:

\[ p(\theta, \pi, S_T|Y_T) \propto f(Y_T|\theta, S_T, \pi) p(S_T|\pi)p(\pi)p(\theta) \]

- \( \pi(S_T|\pi) \) results from the Markov chain assumption:

\[ p(S_T|\pi) = \prod_{t=1}^{T} p(s_t|s_{t-1}, \pi) \text{ where } p(s_t|s_{t-1}, \pi) \text{ is the transition probability to move from state } s_{t-1} \text{ to state } s_t. \]

- Prior on \( \ln[\theta/(1-\theta)] \) : \( N(\mu, 8I_{3(K+1)}) \), \( \mu = (\mu_\omega, \mu_\alpha, \mu_\beta)' \),

\[ \mu_\omega = -4t_{K+1}, \mu_\alpha = \ln\left(\frac{0.25}{0.75}\right)t_{K+1}, \mu_\beta = \ln\left(\frac{0.75}{0.25}\right)t_{K+1}. \]

- Prior on \( \pi \) chosen to facilitate posterior simulation: Dirichlet distribution.
MCMC algorithm for posterior distribution

- It is a Gibbs sampling algorithm with three blocks:
  1. Sample \( S_T \) from \( p(S_T|\theta, \pi, Y_T) \) → difficult part
  2. Sample \( \pi \) from \( p(\pi|S_T, Y_T, \theta) \) → analytically since Dirichlet
  3. Sample \( \theta \) from \( p(\theta|S_T, \pi, Y_T) \) → numerically but not difficult (we use a Metropolis-Hastings step)

- In Bauwens, Preminger and Rombouts (EJ, 2010), sampling \( S_T \) is broken into univariate sampling of each state variable given the other states (Gibbs sampling). This is easy numerically (discrete distributions), but numerically heavy and not mixing well.

- **Contribution of this paper**: a better algorithm, that samples \( S_T \) in one shot, thus mixing well.
Sampling $S_T$

- **Notations:** $S_t = (s_1, s_2, \ldots, s_t)$, $S_{t+1}^t = (s_{t+1}, s_{t+2}, \ldots, s_T)$.

- **We factorize** $p(S_T|\theta, \pi, Y_T)$ sequentially from the last date to the first one:

  
  $p(s_T|Y_T, \theta, \pi)p(s_{T-1}|s_T, Y_T, \theta, \pi) \ldots p(s_t|s_{t+1}, Y_T, \theta, \pi) \ldots p(s_1|s_2, Y_T, \theta, \pi)$

- **Sampling** is done sequentially from each univariate distribution from $t = T$ till $t = 1$ (forward filtering-backward sampling algorithm). **Computing** $p(s_t|s_{t+1}, Y_T, \theta, \pi)$ is far from trivial.

- **We adopt and adapt** an algorithm of Andrieu, Doucet and Hollenstein (ADH): Particle Markov chain Monte Carlo methods (JRSS B, 2010).

  ADH provide a way to incorporate a sequential Monte Carlo (SMC) algorithm inside a MCMC one.
Computing $s_t|S^{t+1}, Y_T, \theta, \pi$

\[
p(s_t|S^{t+1}, Y_T, \theta, \pi) = \frac{p(s_t|Y_t, \theta, \pi) f(Y^{t+1}, S^{t+1}|s_t, Y_t, \theta, \pi) f(Y_t|\theta, \pi)}{f(S^{t+1}, Y_T|\theta, \pi)}
\]

\[
\propto p(s_t|Y_t, \theta, \pi) f(Y^{t+1}, S^{t+1}|s_t, Y_t, \theta, \pi)
\]

\[
\propto p(s_t|Y_t, \theta, \pi) f(Y^{t+1}|S^t, Y_t, \theta, \pi) p(s_{t+1}|s_t, \pi)
\]

$p(s_t|Y_t, \theta, \pi)$ difficult to evaluate because of the path dependence, but evaluated sequentially by a conditional SMC.

$f(Y^{t+1}|S^t, Y_t, \theta, \pi)$ computationally demanding, but can be well approximated by considering the path of each particle.

We use 150 particles for CP, and 250 for MS. We tried different values.
SMC and PMCMC

- A standard SMC algorithm does not keep \( p(\theta, \pi, S_T | Y_T) \) invariant, but the conditional SMC does the job.

- We adapt an algorithm of ADH by including
  - the auxiliary particle filter of Pitt and Shephard (JASA, 1999) in the conditional SMC;
  - the backward sampling, as Godsill, Doucet and West (JASA, 2004).

- We consider also three other variants of our algorithm, see details in the paper.
Selection of number of breaks

- Use marginal likelihood as model choice criterion.

- Apply Bayesian inference conditional on $K$, as described previously, for $K = 1, 2, \ldots, K_{max}$.

- Select value of $K$ corresponding to highest MLL (marginal log likelihood).
The marginal likelihood/data density is defined by

$$f(Y_T) = \int f(Y_T|\lambda)p(\lambda)d\lambda.$$ 

For any function $t(\lambda)$ and density function $q(\lambda)$,

$$f(Y_T) = \frac{A_1}{A_2}$$

where

$$A_1 = \int t(\lambda)f(Y_T|\lambda)p(\lambda)q(\lambda)d\lambda$$

$$A_2 = \int t(\lambda)q(\lambda)p(\lambda|Y_T)d\lambda$$

Proof: substitute in $A_2$ the formula of $p(\lambda|Y_T)$ provided by Bayes theorem.
Meng and Wong (Statistica Sinica, 1996) propose to estimate $f(Y_T)$ by $\hat{A}_1/\hat{A}_2$, with

$$\hat{A}_2 = \frac{1}{G_2} \sum_{j=1}^{G_2} t(\lambda^j) q(\lambda^j)$$

where $\{\lambda^j\}_{j=1}^{G_2}$ is a set of $G_2$ posterior draws, and

$$\hat{A}_1 = \frac{1}{G_1} \sum_{i=1}^{G_1} t(\lambda^i) f(Y_T|\lambda^i)p(\lambda^i)$$

where $\{\lambda^i\}_{i=1}^{G_1}$ is a set of $G_1$ draws from $q(\lambda)$.

For $t(\lambda) = 1/q(\lambda)$, this is importance sampling. For $t(\lambda) = 1/p(\lambda|Y_T)$, it is reciprocal importance sampling.
We follow Meng and Wong who obtain
\[ t(\lambda) = \left[ p(\lambda|Y_T) + q(\lambda) \right]^{-1} \]
as an asymptotically optimal choice which minimizes the expected relative error of the estimator in the case of i.i.d draws from \( p(\lambda|Y_T) \) and \( q(\lambda) \).

In our case, \( \lambda = (\theta, \pi) \).

We specify \( q(\theta, \pi) \) as \( q(\theta)q(\pi) \). The two proposal distributions are respectively mixtures of normal and beta distributions constructed with posterior draws in order to cover the posterior support.

A similar mixture of normal distributions is used as proposal for sampling \( \theta \) in step 3 (MH step) of the Gibbs algorithm for the posterior distribution.
Bayes theorem:

\[ p(\theta, \pi | Y_T) = \frac{f(Y_T | \theta, \pi) p(\theta, \pi)}{f(Y_T)} \].

Hence

\[ f(Y_T) = \frac{f(Y_T | \theta, \pi) p(\theta, \pi)}{p(\theta, \pi | Y_T)} \],

which must hold for any \((\theta, \pi)\) in the posterior support.

Chib (1995) proposed to use the latter to compute \( f(Y_T) \) by evaluating the right-hand side at a high posterior density point, say \((\theta^*, \pi^*)\). He showed how to do this using the output of a Gibbs sampler generating draws from the posterior of \((\theta, \pi)\), in the presence of latent variables like \(S_T\) in our case.
Local formula/2

- $\pi^*$ and $\theta^*$ high density point.

- Prior $p(\theta^*, \pi^*) = p(\theta^*)p(\pi^*)$ easily computed.

- We have eliminated $S_T$ ...

- Data density $f(Y_T|\theta^*, \pi^*)$ computed by PMCMC sampler since it integrates out the state vector $S_T$.

- Posterior $p(\theta^*, \pi^*|Y_T)$ computed with an additional PMCMC sampler (to integrate out the state vector).

- Finally, in logarithm: MLL (marginal log likelihood):

$$\log f(Y_T) = \log p(\theta^*, \pi^*) + \log f(Y_T|\theta^*, \pi^*) - \log p(\theta^*, \pi^*|Y_T).$$
### CP-GARCH DGP (3 regimes) and estimates

<table>
<thead>
<tr>
<th>Regime</th>
<th>DGP values</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>( \omega )</td>
<td>0.20</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.80</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Break</td>
<td>1000</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>(31.7)</td>
<td>(8.5)</td>
</tr>
</tbody>
</table>

Estimates: posterior means and (standard deviations)
### Marginal log-likelihood values for simulated data of CP-GARCH

<table>
<thead>
<tr>
<th>Regimes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Change-Point</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td>-5463.22</td>
<td>-5451.42</td>
<td>-5438.43</td>
<td>-5442.10</td>
</tr>
<tr>
<td>Local</td>
<td>-5463.00</td>
<td>-5450.28</td>
<td>-5438.09</td>
<td>-5439.78</td>
</tr>
<tr>
<td><strong>Markov-switching</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td>-5463.29</td>
<td>-5448.95</td>
<td>-5442.05</td>
<td>-5445.63</td>
</tr>
<tr>
<td>Local</td>
<td>-5462.99</td>
<td>-5448.14</td>
<td>-5441.07</td>
<td>-5443.66</td>
</tr>
</tbody>
</table>
The parameter values of the 2-regime MS-GARCH DGP are the same as for Regimes 1 and 2 of the CP-DGP above.

Marginal log-likelihood values for 3000 simulated data of MS-GARCH

<table>
<thead>
<tr>
<th>Regimes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Change-Point</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td>-5879.88</td>
<td>-5862.75</td>
<td>-5848.05</td>
<td>-5851.07</td>
</tr>
<tr>
<td>Local</td>
<td>-5879.56</td>
<td>-5859.22</td>
<td>-5846.93</td>
<td>-5850.99</td>
</tr>
<tr>
<td><strong>Markov-switching</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td>-5879.89</td>
<td>-5843.55</td>
<td>-5849.04</td>
<td>-</td>
</tr>
<tr>
<td>Local</td>
<td>-5879.67</td>
<td>-5843.05</td>
<td>-5849.48</td>
<td>-</td>
</tr>
</tbody>
</table>
### S&P 500 index: model choice

**Marginal log-likelihood values for S&P 500 data**

<table>
<thead>
<tr>
<th>Regimes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Change-Point</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td>-4505.33</td>
<td>-4505.83</td>
<td>-4503.05</td>
<td>-4519.23</td>
</tr>
<tr>
<td>Local</td>
<td>-4504.95</td>
<td>-4505.93</td>
<td>-4502.97</td>
<td>-4516.16</td>
</tr>
<tr>
<td><strong>Markov-switching</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td>-4505.31</td>
<td>-4497.99</td>
<td>-4502.74</td>
<td>-</td>
</tr>
<tr>
<td>Local</td>
<td>-4505.08</td>
<td>-4496.04</td>
<td>-4497.73</td>
<td>-</td>
</tr>
</tbody>
</table>

Estimates of preferred models on slide 35.
S&P 500 index: MS model

Returns with switches from the 2 regime MS-GARCH model
S&P 500 index: vol. levels

Unconditional volatility

- LVOL CP-GARCH
- LVOL MS-GARCH
- LVOL SP-GARCH
\[ y_t / \tau_t = g_t \epsilon_t, \quad \epsilon_t \sim N(0, 1), \]

\[ g_t^2 = (1 - \alpha - \beta) + \alpha (y_{t-1} / \tau_t)^2 + \beta g_{t-1}^2 \]

\[ \tau_t^2 = \gamma \exp \left( \lambda_0 t + \sum_{i=1}^{k} \lambda_i [(t - t_{i-1})^+]^2 \right), \]

where \((\alpha, \beta, \gamma, \lambda_0, \ldots, \lambda_k)\) are parameters, 
\((t - t_i)^+ = \min(0, t - t_i)\) and \(\{t_0 = 0, t_1, t_2, \ldots, t_k = T\}\).

We used the BIC to determine the number of knots.
Estimates for S&P 500 daily returns

Sample period: May 20, 1999 to April 25, 2011 ($T = 3000$)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GARCH</th>
<th>CP-GARCH</th>
<th>MS-GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1.67</td>
<td>1.95</td>
<td>0.45</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.075</td>
<td>0.085</td>
<td>0.023</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.915</td>
<td>0.868</td>
<td>0.931</td>
</tr>
<tr>
<td>$\alpha + \beta$</td>
<td>0.990</td>
<td>0.953</td>
<td>0.954</td>
</tr>
</tbody>
</table>

$\sigma^2 = \omega / (1 - \alpha - \beta)$ is the (local) unconditional variance.
# Model choice for 11 series

<table>
<thead>
<tr>
<th>Series</th>
<th>Spline-GARCH</th>
<th>CP-GARCH</th>
<th>MS-GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>knots  log-BF</td>
<td>K+1 log-BF</td>
<td>K+1 log-BF nswitch</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>3  5.21</td>
<td>3  2.28</td>
<td>2  7.34   3</td>
</tr>
<tr>
<td>DJIA</td>
<td>3  2.99</td>
<td>1  0</td>
<td>2  4.7    3</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>3  3.20</td>
<td>1  0</td>
<td>2  1.94   7</td>
</tr>
<tr>
<td>NYSE</td>
<td>3  2.40</td>
<td>1  0</td>
<td>2  3.91   13</td>
</tr>
<tr>
<td>BAC</td>
<td>4  16.62</td>
<td>3  50.12</td>
<td>3  79.49  11</td>
</tr>
<tr>
<td>BA</td>
<td>4  9.10</td>
<td>2  8.9</td>
<td>2  11.48  6</td>
</tr>
<tr>
<td>JPM</td>
<td>3  8.82</td>
<td>3  5.17</td>
<td>3  7.22   9</td>
</tr>
<tr>
<td>MRK</td>
<td>5  48.78</td>
<td>5  215.39</td>
<td>3  335.23 56</td>
</tr>
<tr>
<td>PG</td>
<td>4  16.34</td>
<td>3  24.23</td>
<td>2  33.6   9</td>
</tr>
<tr>
<td>Metals</td>
<td>2  6.66</td>
<td>2  11.33</td>
<td>2  14.68  5</td>
</tr>
<tr>
<td>Yen/Dollar</td>
<td>1  -3.34</td>
<td>1  0</td>
<td>2  3.05   7</td>
</tr>
</tbody>
</table>

log-BF: log of Bayes factors wrt GARCH(1,1) model

- Bayes factors were calculated for each series using different GARCH models.
- The table shows the number of knots for Spline-GARCH and the log-BF values for CP-GARCH and MS-GARCH models.
- The last column indicates the number of switching points for MS-GARCH.
Kass and Raftery (JASA, 1995):

- If log-BF < 1: evidence (in favor of model having highest MLL) not worth than a bare mention.
- If 1 < log-BF < 3: evidence is positive.
- If log-BF > 3: evidence is strong.

Summary of comparisons for 11 series:

- MS versus CP: 9 "> 3" and 2 ∈ (1, 3).
- MS versus SPLINE: 5 "> 3", 4 ∈ (1, 3), and 2 ∈ (−3, −1).
- CP versus SPLINE: 5 "> 3", 2 "< −3", 3 ∈ (−3, −1) and 1 ∈ (−1, 0).
Efficiency of PMCMC over BPR

- **BDR**: states sampled jointly by PMCMC.
- **BPR**: states sampled one at a time (Gibbs for states).

Efficiency measured by the effective computing time needed to obtain one new "independent" draw:

\[
\frac{\text{time for } N \text{ posterior draws } \times \text{AC time}}{N}.
\]

This is computed
- for each break date (CP) or
- number of observations in a regime (MS),
then taking the maximum value.

**AC time (autocorrelation time):**

\[
1 + \sum_i |\rho_i|
\]

where \(\rho_i\) is the autocorrelation of order \(i\) of the posterior draws.
Autocorrelation times for **best models**

<table>
<thead>
<tr>
<th>Break</th>
<th>CP-GARCH</th>
<th>MS-GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 regimes</td>
<td>3 regimes</td>
</tr>
<tr>
<td></td>
<td>BPR</td>
<td>PMCMC</td>
</tr>
<tr>
<td>1</td>
<td>434.63</td>
<td>1.08</td>
</tr>
<tr>
<td>2</td>
<td>65.491</td>
<td>1.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Break</th>
<th>CP-GARCH</th>
<th>MS-GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 regimes</td>
<td>2 regimes</td>
</tr>
<tr>
<td></td>
<td>BPR</td>
<td>PMCMC</td>
</tr>
<tr>
<td>1</td>
<td>319.87</td>
<td>1.23</td>
</tr>
<tr>
<td>2</td>
<td>443.26</td>
<td>1.68</td>
</tr>
</tbody>
</table>
S&P 500 index: autocorrelations of draws

(a) CP Model-BPR-break 1

(b) CP Model-PMCMC-break 1
S&P 500 index: autocorrelations of draws

(c) MS Model-BPR-break 1

(d) MS Model–PMCMC-break 1
## Effective computing times per posterior draw

### CP-GARCH simulated data

<table>
<thead>
<tr>
<th>Number of Regimes</th>
<th>CP-GARCH</th>
<th>MS-GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMCMC</td>
<td>BPR</td>
</tr>
<tr>
<td>2</td>
<td>0.021</td>
<td>0.029</td>
</tr>
<tr>
<td>3</td>
<td>0.026</td>
<td>0.049</td>
</tr>
<tr>
<td>4</td>
<td>0.077</td>
<td>0.060</td>
</tr>
</tbody>
</table>

### S&P500 data

<table>
<thead>
<tr>
<th>Number of Regimes</th>
<th>CP-GARCH</th>
<th>MS-GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMCMC</td>
<td>BPR</td>
</tr>
<tr>
<td>2</td>
<td>0.017</td>
<td>0.041</td>
</tr>
<tr>
<td>3</td>
<td>0.017</td>
<td>0.046</td>
</tr>
<tr>
<td>4</td>
<td>0.017</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Computation time in minutes per effective posterior draw.
Number of particles

- We tried different values of the number of particles used in the PMCMC algorithm: $N = 25, 50, 100, 150, 250, 500$.

- Typically, the AC times decrease much in the beginning (when $N$ increases until 100-250) then decrease very slowly or become stable.

- There is a trade-off between effective CPU time and $N$, and we adopted safe values of $N$ (150 for CP, 250 for MS).
Current research

- Forecast evaluations to compare models.

- Extension to other GARCH models and other distributions for the error term.

- Extension to multivariate model (dynamic correlations).

- Including the determination of the number of regimes in the inference: