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

$$\begin{aligned} y_t &= \epsilon_t \sigma_t, \qquad \epsilon_t \sim N(0,1), \\ \sigma_t^2 &= \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

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

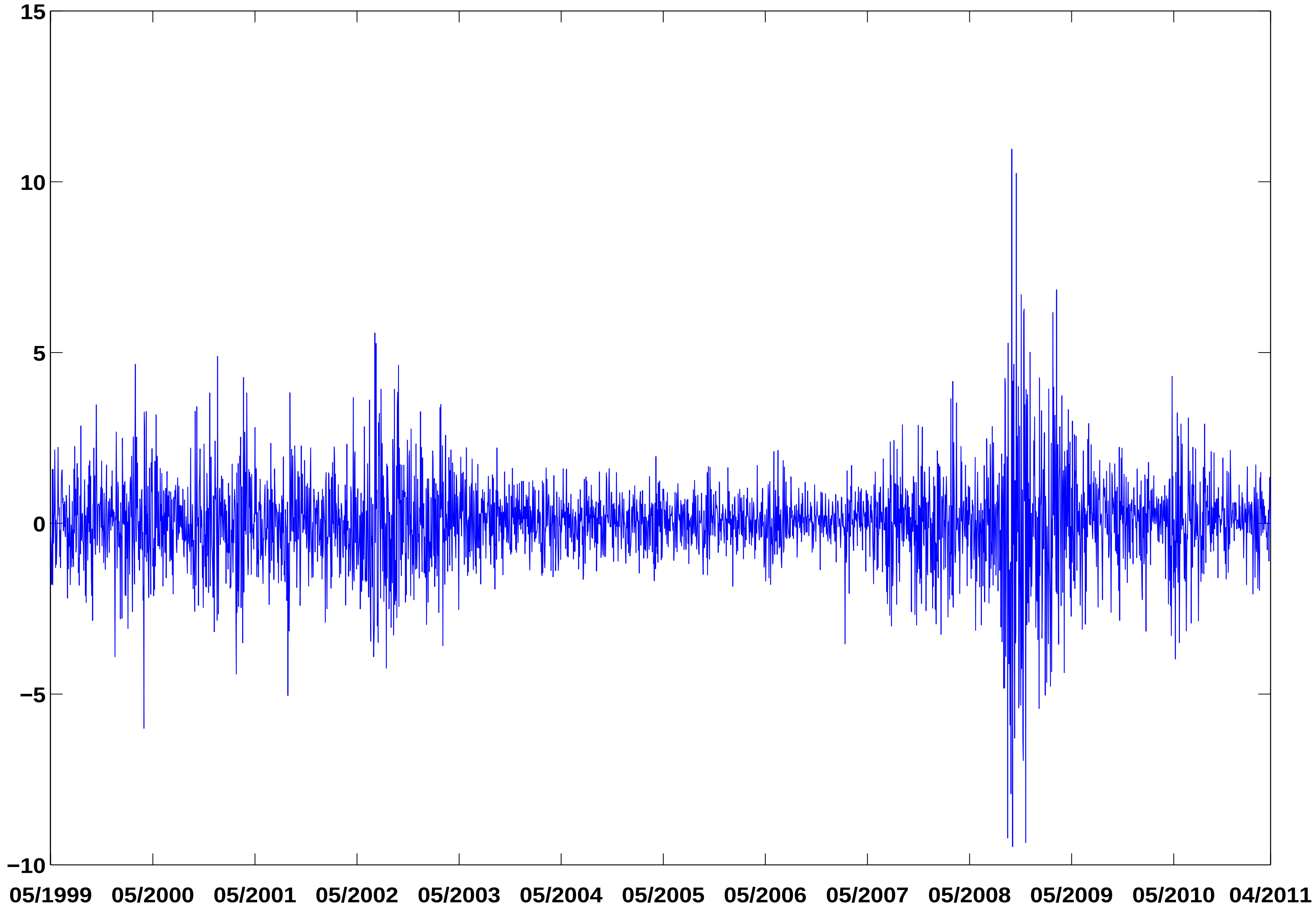
- That spuriously strong persistence may be caused by changes of the unconditional variance level.

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

$\Rightarrow$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) ...
- Smoothly changing level models: Dalhaus and Subba Rao (2006), Engle and Rangel (2008), Baillie and Morana (2009), Amado and Terasvirta (2008, 2011) ...
- 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

- Hamilton and Susmel (1994), Cai (1994), Gray (1996), Francq and Zakoian (2001, 2008), Haas et al. (2004), Bauwens, Preminger and Rombouts (2010)...

- Prototype model:

$$\begin{aligned} y_t &= \epsilon_t \sigma_t, \qquad \epsilon_t \sim N(0,1), \\ \sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2, \end{aligned}$$

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

- NB: σ_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

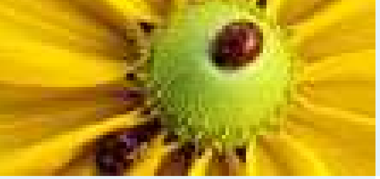
$\{s_t\}$ is a first-order Markovian process with transition matrix P.

Markov-switching case (recurrent regimes):

$$P_M = \begin{pmatrix} p_{11} & p_{12} & \dots & \dots & p_{1K+1} \\ p_{21} & p_{22} & p_{23} & \dots & p_{2K+1} \\ \dots & \dots & \dots & \dots & \dots \\ p_{K1} & p_{K2} & p_{K3} & \dots & p_{K+1K+1} \end{pmatrix}$$

Change-point case:

$$P_C = \begin{pmatrix} p_{11} & 1-p_{11} & 0 & \dots & \dots & 0 \\ 0 & p_{22} & 1-p_{22} & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & p_{KK} & 1-p_{KK} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$



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: $\sigma^2_{t|1,2,\ldots t} = \sigma^2_t(S_t)$

$$\sigma^2_0, u_0$$

$$\nearrow \quad \sigma^2_{1|1} = \omega_1 + \beta_1 \sigma^2_0 + \alpha_1 u^2_0$$

$$\nearrow \quad \sigma^2_{2|1,1} = \omega_1 + \beta_1 \sigma^2_{1|1} + \alpha_1 u^2_1$$

$$\searrow \quad \sigma^2_{2|1,2} = \omega_2 + \beta_2 \sigma^2_{1|1} + \alpha_2 u^2_1$$

$$\searrow \quad \sigma^2_{1|2} = \omega_2 + \beta_2 \sigma^2_0 + \alpha_2 u^2_0$$

$$\nearrow \quad \sigma^2_{2|2,1} = \omega_1 + \beta_1 \sigma^2_{1|2} + \alpha_1 u^2_1$$

$$\searrow \quad \sigma^2_{2|2,2} = \omega_2 + \beta_2 \sigma^2_{1|2} + \alpha_2 u^2_1$$

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

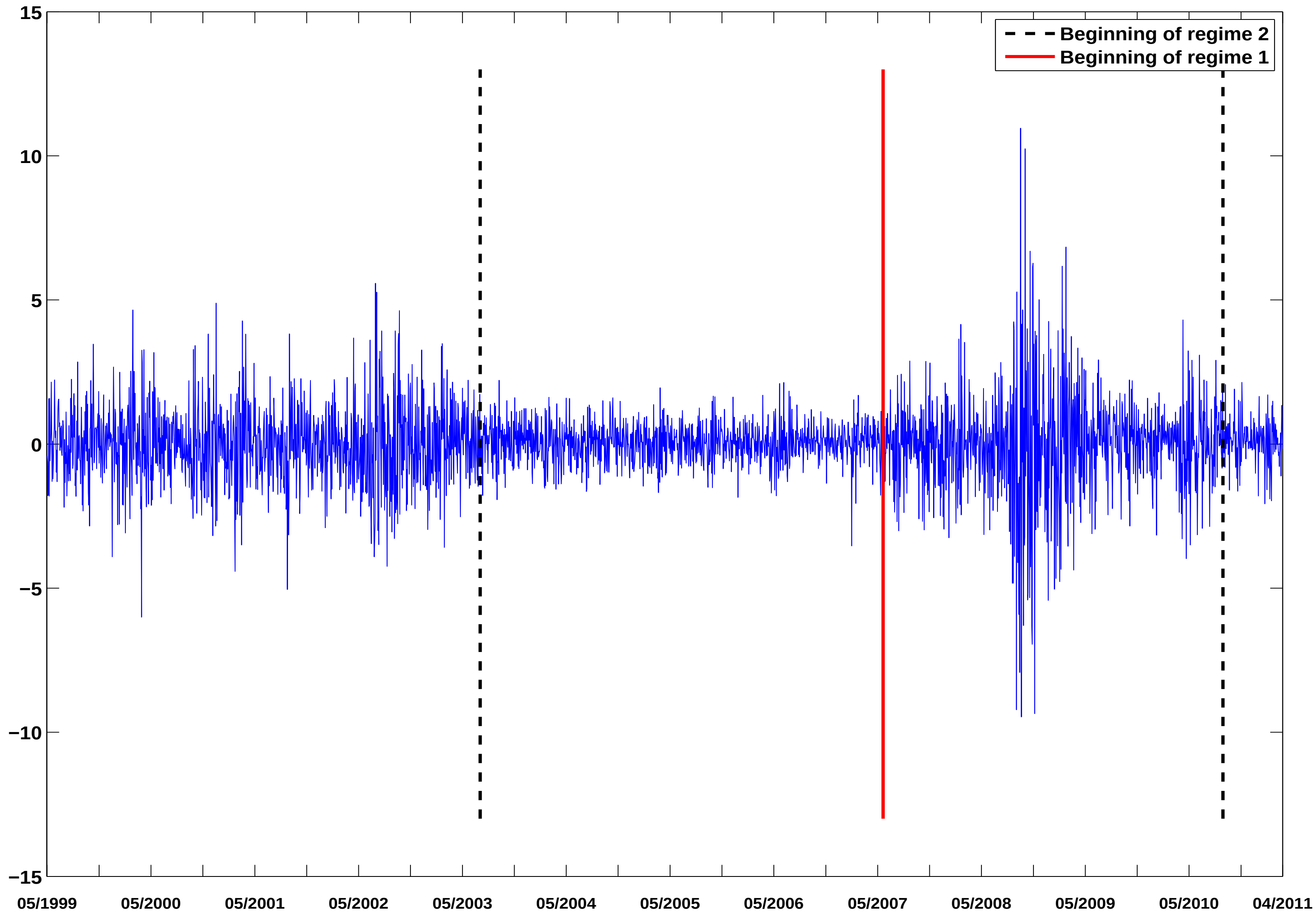
- New and more efficient algorithm applicable to CP- and MS-GARCH models.
- Able to compute the marginal likelihood and perform model choice:

  Number of regimes?

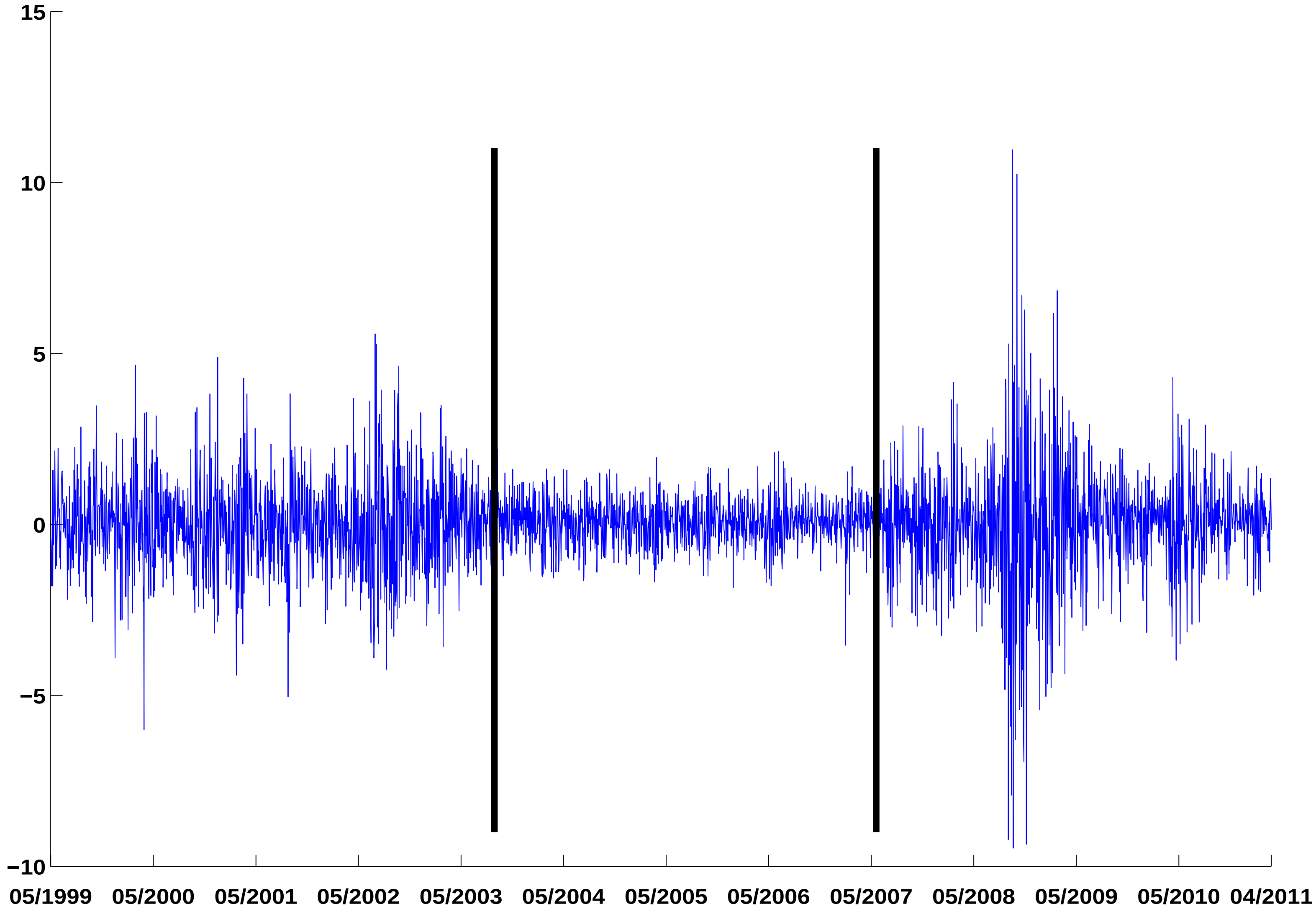
  CP or MS?
- Applied to eleven time series, the MS-GARCH model preferred to CP-GARCH in all cases.
- C++ code available at https://sites.google.com/site/websiteofarnauddufays/

S&P 500 index: MS-GARCH

Returns with switches of 2 regime MS-GARCH model



Returns with switches of 3 regime CP-GARCH model



Model and parameters

- Prototype model:

$$
\begin{aligned}
y_t &= \epsilon_t \sigma_t, \qquad \epsilon_t \sim N(0,1), \\
\sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2,
\end{aligned}
$$

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

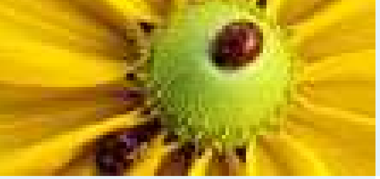
- Parameters: GARCH, transition matrix, and states:

  $\theta = (\omega_1, \ldots, \omega_{K+1}, \alpha_1, \ldots, \alpha_{K+1}, \beta_1, \ldots, \beta_{K+1})$,

  π: the non-redundant parameters of P_M or P_C,

  $S_T = (s_1 s_2 \ldots s_T)$: 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:

$$p(\theta, \pi, S_T|Y_T) \propto \underbrace{f(Y_T|\theta, S_T, \pi)}_{\text{data density}} \underbrace{p(S_T|\pi)p(\pi)p(\theta)}_{\text{prior densities}}$$

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

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

- Prior on $\ln[\theta\,./(1-\theta)]$: $N(\mu, 8I_{3(K+1)})$, $\mu = (\mu_\omega, \mu_\alpha, \mu_\beta)'$, $\mu_\omega = -4\iota_{K+1}$, $\mu_\alpha = \ln(\frac{0.25}{0.75})\iota_{K+1}$, $\mu_\beta = \ln(\frac{0.75}{0.25})\iota_{K+1}$.

- Prior on π 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 $p(S_T|\theta, \pi, Y_T) \rightarrow$ difficult part
  2. Sample π from $p(\pi|S_T, Y_T, \theta) \rightarrow$ analytically since Dirichlet
  3. Sample θ from $p(\theta|S_T, \pi, Y_T) \rightarrow$ 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: $S_t = (s_1, s_2, \ldots, s_t)$, $S^{t+1} = (s_{t+1}, s_{t+2}, \ldots, s_T)$.
- 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 $s_t|S^{t+1}, Y_T, \theta, \pi$

$$
\begin{aligned}
p(s_t|S^{t+1}, Y_T, \theta, \pi) &= \frac{p(s_t|Y_t, \theta, \pi) f(Y^{t+1}, S^{t+1}|s_t, Y_t, \theta, \pi) f(Y_t|\theta, \pi)}{f(S^{t+1}, Y_T|\theta, \pi)} \\
&\propto p(s_t|Y_t, \theta, \pi) f(Y^{t+1}, S^{t+1}|s_t, Y_t, \theta, \pi) \\
&\propto p(s_t|Y_t, \theta, \pi) f(Y^{t+1}|S^t, Y_t, \theta, \pi) p(s_{t+1}|s_t, \pi)
\end{aligned}
$$

$p(s_t|Y_t, \theta, \pi)$ 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).

Global formula/1

- 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 $q(\lambda)$,

$$f(Y_T) = A_1/A_2$$

where

$$\begin{aligned} A_1 &= \int t(\lambda)f(Y_T|\lambda)p(\lambda)q(\lambda)d\lambda \\ A_2 &= \int t(\lambda)q(\lambda)p(\lambda|Y_T)d\lambda \end{aligned}$$

Proof: substitute in A_2 the formula of $p(\lambda|Y_T)$ provided by Bayes theorem.

Global formula/2

- 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.

Global formula/3

- We follow Meng and Wong who obtain $t(\lambda) = [p(\lambda|Y_T) + q(\lambda)]^{-1}$ 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 θ in step 3 (MH step) of the Gibbs algorithm for the posterior distribution.

Local formula/1

- 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 (θ, π) 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 (θ^*, π^*). He showed how to do this using the output of a Gibbs sampler generating draws from the posterior of (θ, π), in the presence of latent variables like S_T in our case.

Local formula/2

- π^* and θ^* 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

	DGP values			Estimates		
Regime	1	2	3	1	2	3
ω	0.20	0.70	0.40	0.34	0.78	0.32
				(0.14)	(0.23)	(0.07)
α	0.10	0.20	0.20	0.15	0.17	0.17
				(0.03)	(0.04)	(0.05)
β	0.80	0.70	0.40	0.68	0.70	0.52
				(0.12)	(0.06)	(0.09)
Break	1000	2000		1046	2010	
				(31.7)	(8.5)	

Estimates: posterior means and (standard deviations)

CP-GARCH DGP: MLL

Marginal log-likelihood values for simulated data of CP-GARCH

Regimes	1	2	3	4
		Change-Point		
Global	-5463.22	-5451.42	-5438.43	-5442.10
Local	-5463.00	-5450.28	-5438.09	-5439.78
		Markov-switching		
Global	-5463.29	-5448.95	-5442.05	-5445.63
Local	-5462.99	-5448.14	-5441.07	-5443.66

MS-GARCH DGP: MLL

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

Regimes	1	2	3	4
Change-Point				
Global	-5879.88	-5862.75	-5848.05	-5851.07
Local	-5879.56	-5859.22	-5846.93	-5850.99
Markov-switching				
Global	-5879.89	-5843.55	-5849.04	-
Local	-5879.67	-5843.05	-5849.48	-

S&P 500 index: model choice

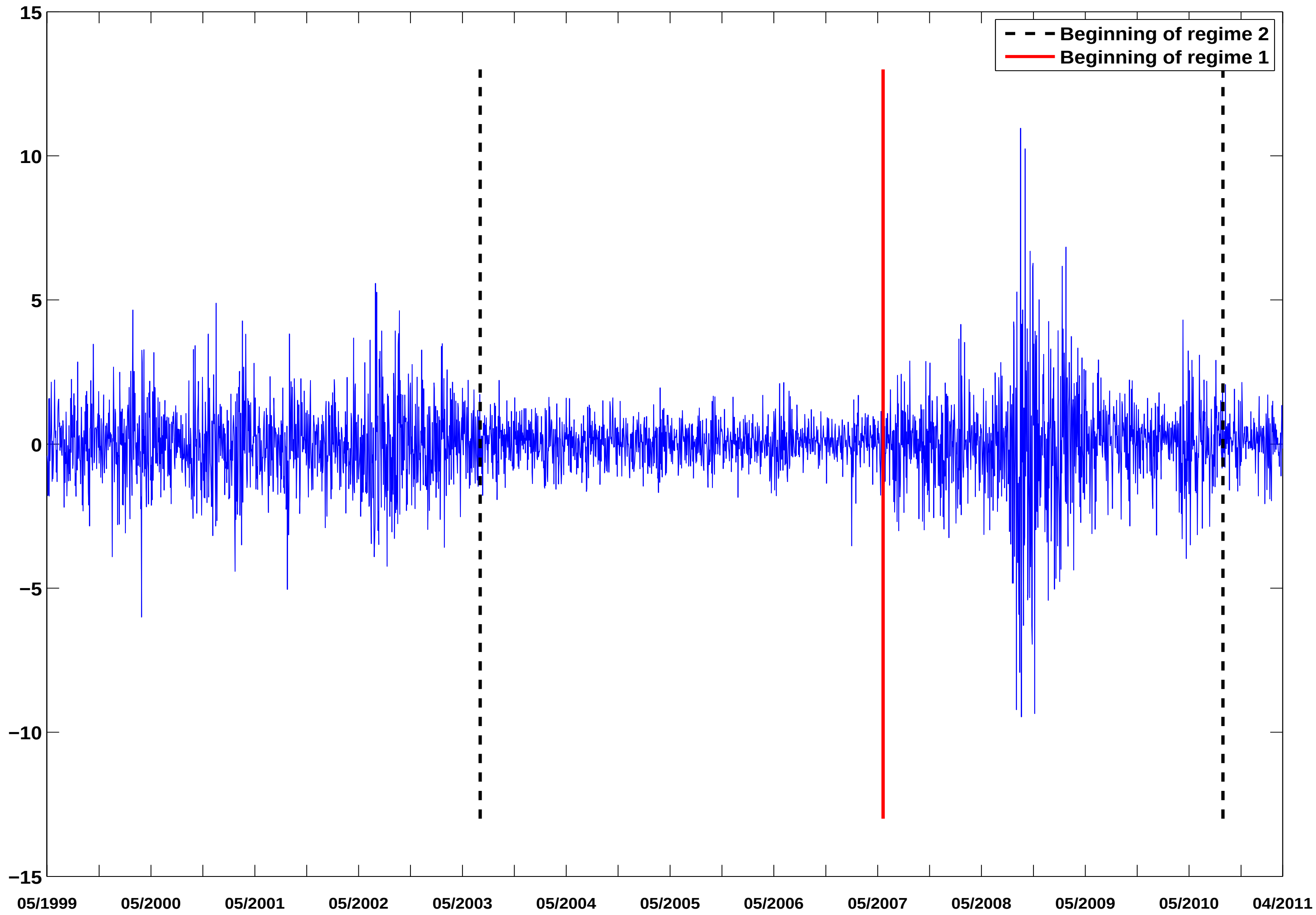
Marginal log-likelihood values for S&P 500 data

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

Estimates of preferred models on slide 35.

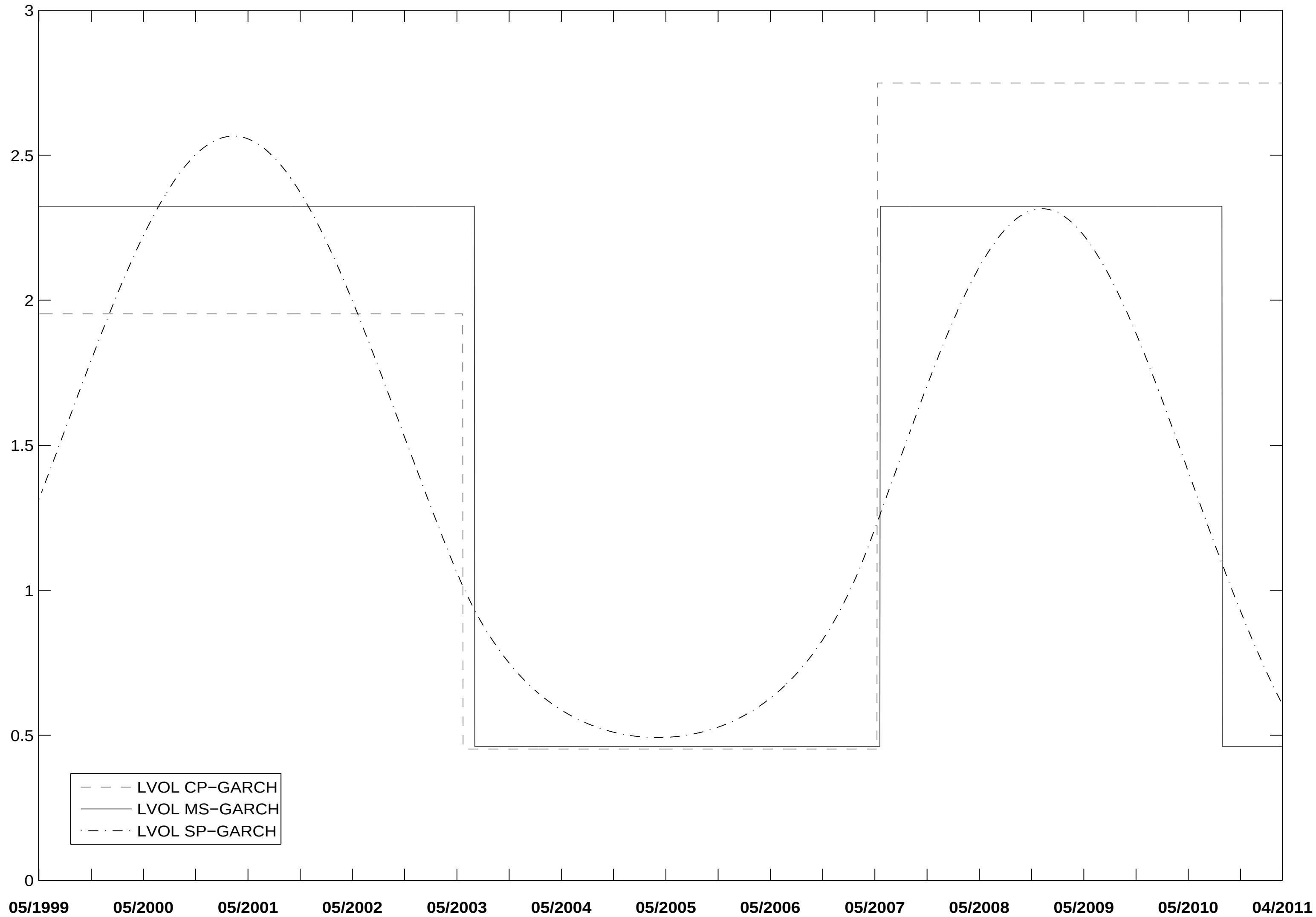
S&P 500 index: MS model

Returns with switches from the 2 regime MS-GARCH model



S&P 500 index: vol. levels

Unconditional volatility



The spline-GARCH (Engle and Rangel, 2008)

$$
\begin{aligned}
y_t/\tau_t &= g_t\epsilon_t, \quad \epsilon_t \sim N(0,1), \\
g_t^2 &= (1-\alpha-\beta) + \alpha(y_{t-1}/\tau_t)^2 + \beta g_{t-1}^2 \\
\tau_t^2 &= \gamma \exp\left(\lambda_0 t + \sum_{i=1}^{k} \lambda_i [(t-t_{i-1})_+]^2\right),
\end{aligned}
$$

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 ($T = 3000$)

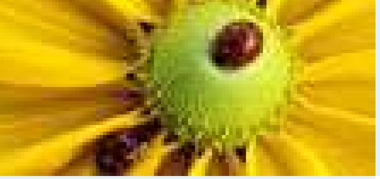
	GARCH	CP-GARCH Regime			MS-GARCH Regime	
Parameter		1	2	3	1	2
σ^2	1.67	1.95	0.45	2.75	2.32	0.46
α	0.075	0.085	0.023	0.098	0.089	0.031
β	0.915	0.868	0.931	0.890	0.891	0.901
$\alpha + \beta$	0.990	0.953	0.954	0.978	0.980	0.932

$\sigma^2 = \omega/(1 - \alpha - \beta)$ is the (local) unconditional variance.

Model choice for 11 series

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

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



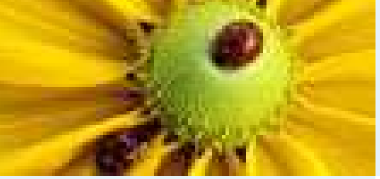
Rule of thumb for log-BF

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 $\in (1, 3)$.
- MS versus SPLINE: 5 "> 3", 4 $\in (1, 3)$, and 2 $\in (-3, -1)$.
- CP versus SPLINE: 5 "> 3", 2 "< -3", 3 $\in (-3, -1)$ and 1 $\in (-1, 0)$.



Efficiency of PMCMC over BPR

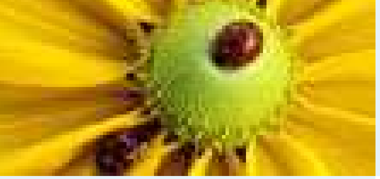
- BDR: states sampled jointly by PMCMC.
  BPR: states sampled one at a time (Gibbs for states).
- Efficiency measured by the effective computing time needed to obtain one new "independent" draw:

  (time for N posterior draws $\times$ 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 ρ_i is the autocorrelation of order i of the posterior draws.



Autocorrelation times for **best models**

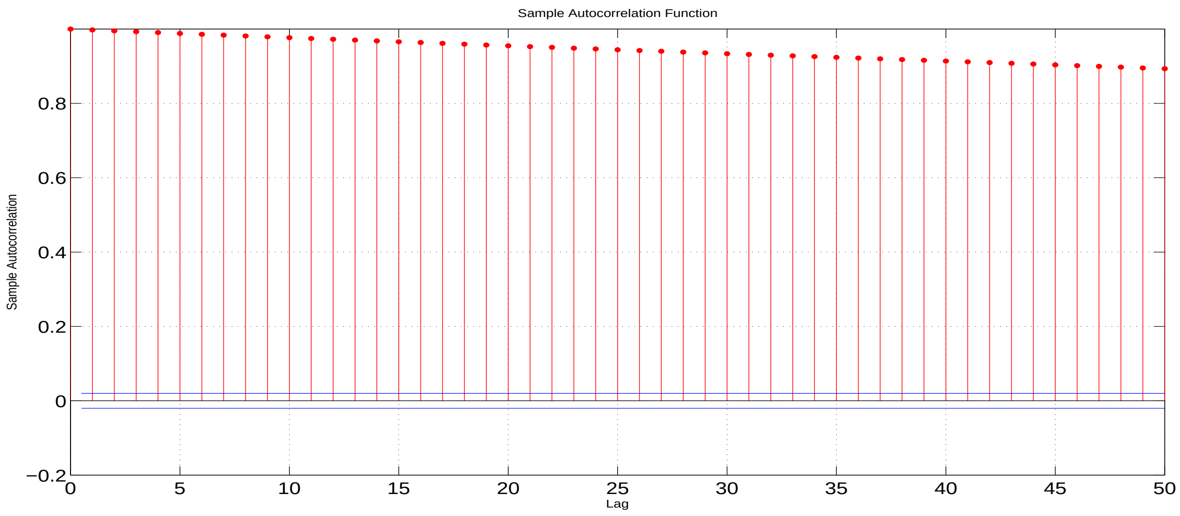
CP-GARCH simulated data

Break	CP-GARCH 3 regimes		MS-GARCH 3 regimes	
	BPR	PMCMC	BPR	PMCMC
1	*434.63*	1.08	*461.23*	*1.53*
2	65.491	*1.17*	430.15	1.45

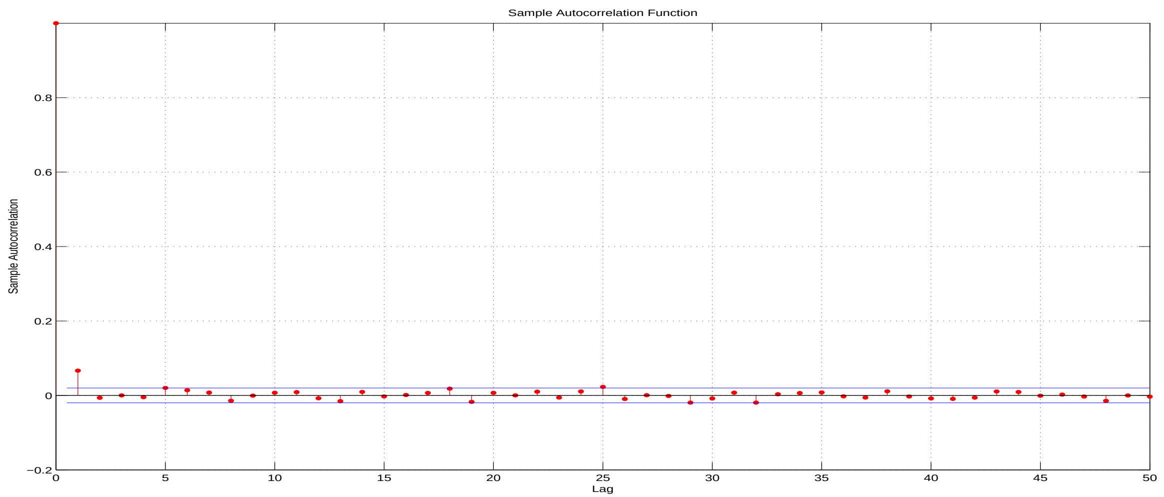
S&P 500 data

Break	CP-GARCH 3 regimes		MS-GARCH 2 regimes	
	BPR	PMCMC	BPR	PMCMC
1	319.87	1.23	*327.77*	*1.76*
2	*443.26*	*1.68*	—	—

S&P 500 index: autocorrelations of draws

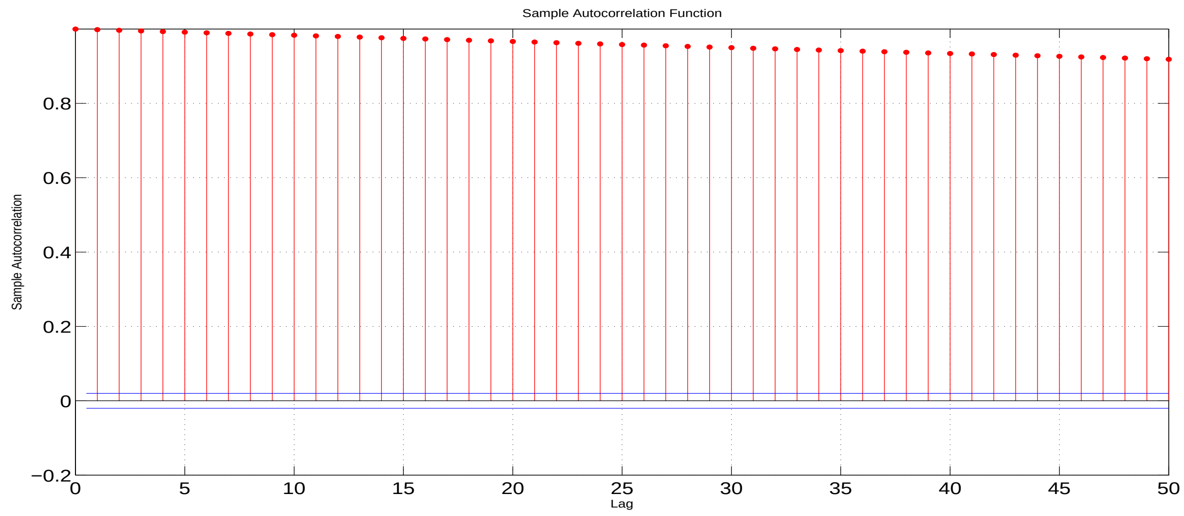


(a) CP Model-BPR-break 1

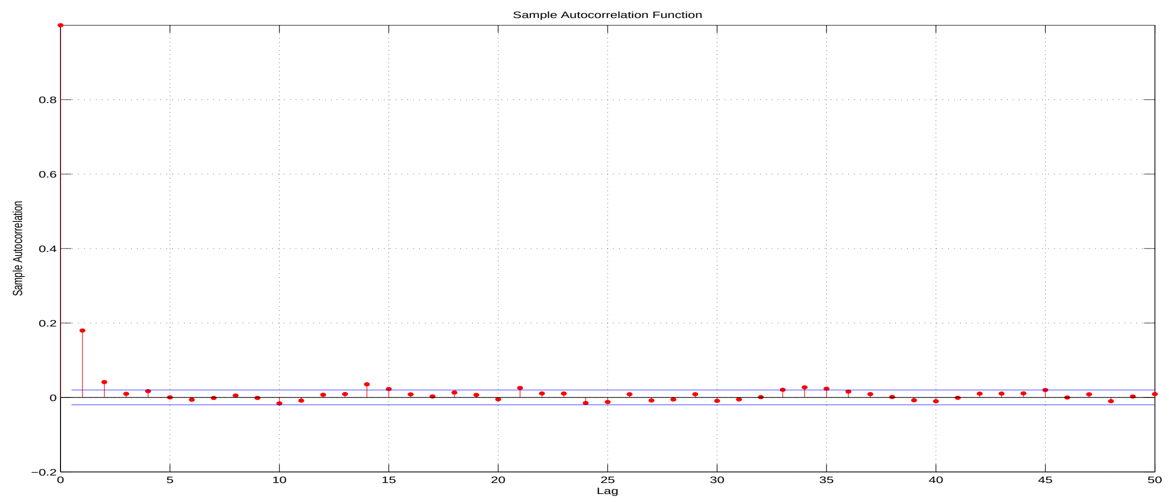


(b) CP Model-PMCMC-break 1

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

Number of Regimes	CP-GARCH PMCMC	CP-GARCH BPR	MS-GARCH PMCMC	MS-GARCH BPR
2	0.021	0.029	0.039	17.22
3	0.026	0.049	0.066	30.51
4	0.077	0.060	0.072	39.57

S&P500 data

Number of Regimes	CP-GARCH PMCMC	CP-GARCH BPR	MS-GARCH PMCMC	MS-GARCH BPR
2	0.017	0.041	0.067	18.85
3	0.017	0.046	0.099	27.72
4	0.017	0.054	—	—

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:

  Dufays, A. (2012), Infinite-state Markov-switching for dynamic volatility and correlation models, CORE DP 2012/43 (November 2012)