CORE DISCUSSION PAPER

2011/13

# MARGINAL LIKELIHOOD FOR MARKOV-SWITCHING AND CHANGE-POINT GARCH MODELS

Luc Bauwens[1], Arnaud Dufays[1], and Jeroen V.K. Rombouts[2]

November 22, 2011

**Abstract**

GARCH volatility models with fixed parameters are too restrictive for long time series due to breaks in the volatility process. Flexible alternatives are Markov-switching GARCH and change-point GARCH models. They require estimation by MCMC methods due to the path dependence problem. An unsolved issue is the computation of their marginal likelihood, which is essential for determining the number of regimes or change-points. We solve the problem by using particle MCMC, a technique proposed by Andrieu, Doucet, and Holenstein (2010). We examine the performance of this new method on simulated data, and we illustrate its use on several return series.

**Keywords:** Bayesian inference, Simulation, GARCH, Markov-switching model, Change-point model, Marginal likelihood, Particle MCMC

**JEL Classification:** C11, C15, C22, C58.

[1]Université catholique de Louvain, CORE, B-1348 Louvain-La-Neuve.

[2]Institute of Applied Economics at HEC Montréal, CIRANO, CIRPEE, and CORE (Université catholique de Louvain).

Research supported by the contract "Projet d'Actions de Recherche Concertées" 07/12-002 of the "Communauté francaise de Belgique", granted by the "Académie universitaire Louvain". Bauwens is also a member of ECORE, an association between ECARES and CORE. This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

# 1 Introduction

GARCH models with fixed parameters are used to model and predict the volatility of financial time series since the contributions of Engle (1982) and Bollerslev (1986). When estimating such models a common finding is that conditional variances are strongly persistent, especially for long time series. This high degree of persistence (very close to unit root type) has been questioned, see e.g. Noh, Engle, and Kane (1994). Several researchers, e.g. Diebold (1986) and Mikosch and Starica (2004), have argued that the nearly integrated behaviour of conditional variances is due to changes in the parameters of the GARCH process, which are overlooked if the model specification imposes fixed parameters.

An interesting way of making GARCH models more flexible is enriching them with a dynamic discrete latent state Markov process in such a way that the parameters can switch from one value to another.[1] These models are called Markov-switching (MS) GARCH models when the Markov chain is recurrent, see among others Francq and Zakoian (2008) and Bauwens, Preminger, and Rombouts (2010). Change-point (CP) GARCH models, see He and Maheu (2010), arise when the chain is not recurrent, a feature that makes these models non-stationary. Whether a MS- or CP-GARCH model is estimated, the number of possible states (or regimes) must be chosen. To do this, one can maximize the marginal likelihood which is the usual tool for model choice in Bayesian inference. However, the computation of the marginal likelihood for a MS- or CP-GARCH model, and more generally models subject to the path dependence problem, is an unsolved difficult problem.

In this paper, we solve this problem by applying a particle Markov chain Monte Carlo (PMCMC) method, a technique introduced by Andrieu, Doucet, and Holenstein (2010). This approach is particularly suitable for conducting inference in non-linear state space models as pointed out by Flury and Shephard (2011). The MS- and CP-GARCH models belong to this class. For a fixed number of regimes, a Gibbs sampling algorithm for Bayesian inference on the MS-GARCH model has been proposed by Bauwens, Preminger, and Rombouts (2010).

[1] Other more flexible GARCH models are component models, e.g. Ding and Granger (1996), smooth transition models, e.g. Gonzales-Rivera (1998), and mixture models, e.g. Haas, Mittnik, and Paolella (2004a). Markov-switching models that circumvent the path dependence problem are proposed by Gray (1996) and Haas, Mittnik, and Paolella (2004b), and non-stationary GARCH models by Engle and Rangel (2008), Baillie and Morana (2009) and Amado and Terasvirta (2011).

They sample the state variables individually, whereas in our new algorithm, called particle Gibbs sampler, they are sampled jointly. This makes a big difference in performance, due to the strong dependence between the state variables. Using PMCMC, it turns out that we can also go one step further and compute the marginal likelihood using either bridge sampling, as proposed by Meng and Wong (1996), or the method of Chib (1996), see also Chib and Jeliazkov (2001). Note that the marginal likelihood can be computed in MS-ARCH models, introduced by Hamilton and Susmel (1994) and Cai (1994), where the conditional variance depends only on past shocks. For example, Kaufman and Fruhwirth-Schnatter (2002) compute the marginal likelihood for a MS-ARCH model using Chib (1996), and mention that it cannot be extended to the MS-GARCH case due to the path dependence problem.

The path dependence problem occurs because the conditional variance at time $t$ depends on the entire sequence of regimes visited up to time $t$, due to the recursive nature of the GARCH process. Since the regimes are unobservable, one needs to integrate over all possible regime paths when computing the likelihood function. However, the number of possible paths grows exponentially with $t$, rendering the likelihood evaluation unfeasible. In the CP-GARCH model, the path dependence problem is less acute, since the number of regimes visited up to time $t$ increases at least linearly in $t$ but not exponentially. The path dependence is the reason why maximum likelihood estimation is very difficult, if not out of reach, for MS-GARCH models even for a given number of regimes. The possibility to compute the likelihood function by the PMCMC algorithm mentioned above is not useful for ML estimation by standard optimization algorithms because the likelihood function is approximated by simulation in such a way that it is not differentiable with respect to the model parameters, see Pitt, Silva, Giordani, and Kohn (2010) who develop a general framework for computing the marginal likelihood using SMC.

PMCMC combines the advantages of sequential Monte Carlo (SMC) and Markov chain Monte Carlo (MCMC). Particle filtering, a widely applied SMC method, provides a discrete approximation of a distribution of interest that contains latent variables, see for example Fernandez-Villaverde and Rudio-Ramirez (2007) and Johannes, Polson, and Stroud (2009). Andrieu, Doucet, and Holenstein (2010), ADH hereafter, make use of SMC to build high dimensional proposal distributions for MCMC samplers. We use a particle filter algorithm to sample the state variables jointly given the parameters of MS- and CP-GARCH models, and

we sample these parameters given the states. Thus we embed the particle filter in a Gibbs sampler, hence the name particle Gibbs sampler. We adapt the particle filter of ADH for the states in two ways in our sampler. First, we employ an auxiliary particle filter of Pitt and Shephard (1999) instead of a multinomial resampling step. Second, we sample backward, rather than forward, the full state vector using a smoothing approach similar to Godsill, Doucet, and West (2004). Moreover, we can also use the particle filter algorithm to compute the likelihood function for a given number of regimes since it integrates out the full state vector. Thanks to this, the computation of the marginal likelihood becomes feasible.

It is an empirical question whether a MS-GARCH model or a CP-GARCH model (or any other model) is better fitting a particular series. We apply the two types of models (MS and CP) to several series of returns over the period 1999-2011. For four US stock indices, MS-GARCH models with two regimes dominate CP-GARCH models. One regime of the MS models has a low unconditional volatility regime and the other has a high level. For individual stock returns and one commodity index, more regimes (MS) or breaks (CP) are selected and MS models are preferable in all cases, with small differences between marginal log-likelihood values. For the dollar/yen exchange rate, a MS-GARCH model with two regimes is favored.

In Section 2, we present the particle Gibbs algorithm we propose for posterior inference on the parameters of MS- and CP-GARCH models. In Section 3, we explain the two methods for computing the marginal likelihood. In Section 4, we illustrate the algorithm on simulated and real data. Conclusions are presented in the last section.

## 2 Inference for MS- and CP-GARCH models

We consider the model defined by

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

wihere $s_t$ is an integer random variable taking values in $[1, K+1]$. We define $Y_T = \{y_1, ..., y_T\}'$ and $S_T = \{s_1, ..., s_T\}'$ where $T$ denotes the sample size, and $\theta = (\omega_1, \ldots, \omega_{K+1}, \alpha_1, \ldots, \alpha_{K+1}, \beta_1, \ldots, \beta_{K+1})$. The latent state process $\{s_t\}$ is first order Markovian either with the transition

matrix

$$P_S = \begin{pmatrix} p_{11} & p_{12} & p_{13} & ... & p_{1K} & 1-\sum_{j=1}^{K} p_{1j} \\ p_{21} & p_{22} & p_{23} & ... & p_{2K} & 1-\sum_{j=1}^{K} p_{2j} \\ ... & ... & ... & ... & ... & ... \\ p_{K1} & p_{K2} & p_{K3} & ... & p_{KK} & 1-\sum_{j=1}^{K} p_{Kj} \\ p_{K+1,1} & p_{K+1,2} & p_{K+1,3} & ... & p_{K+1K} & 1-\sum_{j=1}^{K} p_{K+1,j} \end{pmatrix},$$

where $p_{ij} = P[s_t = j | s_{t-1} = i]$, or with the absorbing and non-recurrent transition matrix

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

The first transition matrix characterizes a Markov-switching model (MS-GARCH) with $K+1$ regimes and the second a change-point model (CP-GARCH) with $K$ breaks. Note that other distributional assumptions than the normal, a non-zero conditional mean, and other functional forms for the conditional variance $\sigma_t^2$ can be easily handled. In fact, a state dependent mean would make it easier to separate the regimes. A relevant empirical issue is the value $K$ and the choice between a model with recurrent states ($P_S$) or non recurrent states ($P_C$). The marginal likelihood is a standard Bayesian criterion to make this choice.

Estimation by maximum likelihood of the model parameters, consisting of $\theta$ and $P$, where $P$ denotes the unrestricted elements of $P_S$ or $P_C$, is unfeasible for realistic sample sizes because of the path dependence problem. In fact, this would require integration over the $(K+1)^T$ possible paths in the case of a MS-GARCH model and likewise the number of paths increase at least linearly in $T$ (but not exponentially) for a CP-GARCH model.

Bayesian inference is feasible by treating explicitly $S_T$ as a parameter, a technique called data augmentation. This is typically done within a Gibbs sampling algorithm that samples from the posterior distribution $f(\theta, P, S_T|Y_T)$ by iteratively drawing from three full conditional distributions:

1. $p(S_T|\theta, P, Y_T)$
2. $f(P|S_T, \theta, Y_T) = f(P|S_T)$

3. $f(\theta|S_T, P, Y_T) = f(\theta|S_T, Y_T)$.

Sampling from the last two distributions is standard. The full conditional distribution of $P$ is Dirichlet under a Dirichlet prior distribution assumption, and the full conditional distribution of $\theta$ can be simulated with an adaptive Metropolis-Hastings step the details of which are given in the Appendix. After convergence, the algorithm - called particle Gibbs in the sequel - generates a sample $\{S_T^i, P^i, \theta^i\}_{i=1}^{G_1}$ which is a dependent sample of $f(\theta, P, S_T|Y_T)$. In the next two subsections, we describe and explain how we implement step 1 with a conditional sequential Monte Carlo (SMC) algorithm.

## 2.1 Sampling the full state vector using a conditional SMC algorithm

Sampling the state vector $S_T$ is complex because of the path dependence problem. Bauwens, Preminger, and Rombouts (2010) sample each $s_t$ given the other, which gives a slowly converging and computationally demanding sampler. We next show how we can draw $S_T$ in one step using a SMC sampler that furthermore allows to compute the marginal likelihood of the data as explained in Section 3.

We define $S_t = \{s_1, ..., s_t\}$ and $S^{t+1} = \{s_{t+1}, ..., s_T\}$ and likewise for $Y_t$ and $Y^{t+1}$. We factorize $p(S_T|\theta, P, Y_T)$ as

$$p(s_T|Y_T, \theta, P)p(s_{T-1}|s_T, Y_T, \theta, P)\dots p(s_t|S^{t+1}, Y_T, \theta, P)\dots p(s_1|S^2, Y_T, \theta, P). \tag{2}$$

For the CP-GARCH model the first and last distributions are degenerate since $s_T = K + 1$ and $s_1 = 1$ with probability one. We explain next how to sample $S_T$ by focussing on the typical term $p(s_t|S^{t+1}, Y_T, \theta, P)$ which can be written as follows:

$$\begin{aligned} p(s_t|S^{t+1}, Y_T, \theta, P) &= \frac{p(s_t|Y_t, \theta, P)f(Y^{t+1}, S^{t+1}|s_t, Y_t, \theta, P)f(Y_t|\theta, P)}{f(S^{t+1}, Y_T|\theta, P)} \\ &\propto p(s_t|Y_t, \theta, P)f(Y^{t+1}, S^{t+1}|s_t, Y_t, \theta, P) \\ &\propto p(s_t|Y_t, \theta, P)f(Y^{t+1}|S^t, Y_t, \theta, P)p(s_{t+1}|s_t, P). \end{aligned} \tag{3}$$

The probabilities $p(s_t|Y_t, \theta, P)$ (for each $t$ and possible value of $s_t$) in (3) are complicated to evaluate because of the path dependence problem. They are computed by a conditional SMC algorithm defined below, which integrates out the state vector, while more explanations are provided in the next subsection. We denote by $w_t^i$ the normalized weights that are associated to $N$ particles $\{s_t^1, ..., s_t^N\}$ which represent possible realizations of $s_t$. We condition on the $s_t$

draw from the previous Gibbs iteration - hence the name conditional SMC - which becomes the first particle $s_t^1$. The fact that the previous state draw needs to survive ensures convergence to the stationary distribution. The SMC algorithm for $p(s_t|Y_t, \theta, P)$ for $t = 1, \ldots, T$ given uniform initial weights $w_0^i = 1/N$ and initial particles $s_0^i$ (equal to 0 for a change-point model and a uniform draw for a MS-model) is given by:

1. $\forall i \in [2, N]$, compute $g_t^i = w_{t-1}^i \sum_{j=1}^{K+1} p(s_t = j|s_{t-1}^i, P) f(y_t|F_{t-1}, \theta, P, s_t = j)$, $F_{t-1}$ denoting the data and particles until $t-1$, and the normalized weights $\Pi_{t|t-1}^i = g_t^i / \sum_{j=1}^N g_t^j$.

2. $\forall i \in [2, N]$, sample independently a label variable $A_{t-1}^i \sim \Pi_{t|t-1}$ such that $A_{t-1}^i \in [1, N]$.

3. $\forall i \in [2, N]$, sample a particle $s_t^i \sim p(s_t|s_{t-1}^{A_{t-1}^i}, P)$.

4. $\forall i \in [1, N]$, compute $\hat{w}_t^i = \frac{f(y_t|F_{t-1}, s_t^i, \theta, P)}{\sum_{j=1}^{K+1} f(y_t|F_{t-1}, \theta, P, s_t=j) p(s_t=j|s_{t-1}^{A_{t-1}^i})}$ and the normalized weights $w_t^i = \hat{w}_t^i / \sum_{j=1}^N \hat{w}_t^j$.

The weights $w_t^i$ serve to approximate the probability $p(s_t|Y_t, \theta, P)$ that appears in (3). Specifically $p(s_t = j|Y_t, \theta, P) = \sum_{i=1}^N w_t^i 1_{\{s_t^i = j\}}$. Notice that the algorithm is computationally demanding since $N$ particles are used for each $t$. The choice of $N$ is discussed in section 4.1.

The second term in (3) is approximated by considering the path of each particle: for each $q \in [1, K+1]$, we compute

$$\lambda_t^q = \frac{\sum_{i=1}^N f(Y^{t+1}|s_t^i, ..., s_T^i, F_{t-1}, \theta, P) 1_{\{s_t^i = q\}}}{\sum_{i=1}^N f(Y^{t+1}|s_t^i, ..., s_T^i, F_{t-1}, \theta, P)},$$

where each $f()$ is the product of Gaussian densities implied by (1).

To end the procedure, we iteratively sample an entire state vector $S_T$ starting from $T$:

1. Sample $k \sim w_T$. Set $b_T^k = k$ and $s_T = s_T^{b_T^k}$

2. $\forall t = T-1, ..., 2, 1$,

   - $\forall i \in [1, N]$ compute $\pi_t^i = w_t^i \lambda_t^{s_t^i} p(s_{t+1}|s_t^i, P)$ and the normalized weights $\Pi_t^i = \pi_t^i / \sum_{j=1}^N \pi_t^j$.
   - Sample $b_t^k \sim \Pi_t$ and set $s_t = s_t^{b_t^k}$.

The full vector $S_T$ is therefore sampled from $t = T, \ldots, 1$ as written in (2).

## 2.2 More details on the conditional SMC algorithm

Our SMC algorithm is valid according to the auxiliary particle filter (APF) theory, see Pitt and Shephard (1999) for details. This approach adds an auxiliary particle to ease the sampling of $s_t$. It introduces an integer random variable $k$ taking values in $[1, K+1]$ and defines (we drop the conditions $\theta, P$ for ease)

$$\begin{aligned} p(s_t, k|Y_t) &\propto f(y_t|Y_{t-1}, s_t)p(s_t|s_{t-1}^k)w_{t-1}^k \\ &\propto \frac{f(y_t|Y_{t-1}, s_t)p(s_t|s_{t-1}^k)w_{t-1}^k g(s_t, k|Y_t)}{g(s_t, k|Y_t)} \\ &\propto \frac{f(y_t|Y_{t-1}, s_t)p(s_t|s_{t-1}^k)w_{t-1}^k g(s_t|Y_t, k)g(k|Y_t)}{g(s_t, k|Y_t)}. \end{aligned}$$

The intuition behind the first line of the formula above is that the sum of $p(s_t, k|Y_t)$ over all values of $k$ is the probability $p(s_t|Y_t)$ which appears in (3). The idea of the next two lines is that the sampling of $s_t$ from the proposal distribution $g(s_t, k|Y_t)$ will be quite accurate if the proposal takes into account $y_t$. We take $g(s_t, k|Y_t) \propto w_{t-1}^k p(s_t|s_{t-1}^k)f(y_t|Y_{t-1}, s_t)$. Hence $g(k|Y_t) \propto w_{t-1}^k \sum_{j=1}^{K+1} p(s_t = j|s_{t-1}^k)f(y_t|s_t = j, Y_{t-1})$. Finally,

$$p(s_t, k|Y_t) \propto \frac{f(y_t|Y_{t-1}, s_t)}{\sum_{j=1}^{K+1} f(y_t|Y_{t-1}, s_t = j)p(s_t = j|s_{t-1}^k)} p(s_t|s_{t-1}^k)g(k|Y_t) \tag{4}$$

since $g(s_t|k, Y_t) = \frac{p(s_t|s_{t-1}^k)f(y_t|Y_{t-1}, s_t)}{\sum_{j=1}^{K+1} f(y_t|Y_{t-1}, s_t=j)p(s_t=j|s_{t-1}^k)}$.

The SMC algorithm described in four steps in the previous subsection comes from the formula (4). We start by sampling $k_i$ from $g(k|Y_t)$ (step 1), then we sample $s_t^i \sim p(s_t|s_{t-1}^{k_i})$ (step 3), and we compute the weight $\frac{f(y_t|Y_{t-1}, s_t^i)}{\sum_{j=1}^{K+1} f(y_t|Y_{t-1}, s_t=j)f(s_t=j|s_{t-1}^{k_i})}$ (step 4). The normalized weights provide in fact an approximation of the distribution $p(s_t|Y_t, \theta, P)$.

The sampler we develop is in line with the particle Gibbs sampler defined in ADH. To make the link clear, here are more details. Define the ancestor variable $A_t^k$ as the particle from which the particle $k$ at time $t$ is sampled, and the variable $b_t^k$ as the particle belonging to the path of the particle $k$ at time $t$. Set $b_T^k := k$ so that we have the backward recursion $b_t^k = A_t^{b_{t+1}^k}$. The particle Gibbs sampler extends the target distribution by incorporating all random variables generated by the conditional SMC, but ADH show that $\theta, P$ and $S_T$ are still distributed according to the distribution of interest $f(\theta, P, S_T|Y_T)$. The extended target distribution is $\tilde{f}(\theta, P, S_T, A_1, ..., A_{T-1}, \tilde{S}_1, ..., \tilde{S}_T, k)$ where $A_1, ..., A_{T-1}, \tilde{S}_1, ..., \tilde{S}_T, k$ are the set of random variables generated by the SMC algorithm ($A_t = \{A_t^1, ..., A_t^N\}$, $\tilde{S}_t = \{s_t^1, ..., s_t^N\}$ and $k$ denote

the selected particle at time $T$ in the SMC sequence). The justification of our algorithm is based on Theorem 5 in section 4.5 of ADH which implies that the designed particle Gibbs algorithm admits $f(\theta, P, S_T|Y_T)$ as invariant density. This holds for a conditional SMC which considers a resampling step using the current weights and samples $k$ from its full conditional under $\tilde{f}(\,)$ and deterministically tracing back the ancestral lineage of $S_T^k$. We deviate from this in our algorithm since we apply an auxiliary particle filter, and having sampled the particle $k$ we sample backward a new path $b_t^k$. However, as we explain next, the theorem still holds under these two adaptations embedded in our algorithm.

First, the auxiliary particle filter improves the resampling scheme with respect to the multinomial resampling. Following the discussion of R. Chen in ADH, the APF can be viewed as a change in the intermediate distribution and hence does not modify the theoretical properties of the standard particle filter or the particle Gibs.

Second, following the discussion of N. Whiteley in ADH, the particle Gibbs still works if we sample the particle $k$ and sample a new ancestral lineage of this particle. Indeed, we can show - this is related to the decomposition of $p(s_t|S^{t+1}, Y_T, \theta, P)$ presented in (3) - that

$$\begin{aligned}
&\tilde{f}(b_t^k|Y_T, \theta, P, S^{t+1}, A_1, .., A_{T-1}, \tilde{S}_1, .., \tilde{S}_T, b_{t+1}^k, .., b_T^k, k) \\
&\propto p(s_t^{b_t^k}|F_t, \theta, P) f(Y^{t+1}|F_t, S^{t+1}, s_t^{b_t^k}, \theta, P) p(s_t^{b_t^k}|s_{t+1}) \\
&\propto w_t^{b_t^k} f(Y^{t+1}|F_t, S^{t+1}, s_t^{b_t^k}, \theta, P) p(s_t^{b_t^k}|s_{t+1}).
\end{aligned}$$

The advantage of this backward sampling is that it enables the exploration of all possible ancestral lineages and not only those obtained during the forward conditional SMC sequence.

Remark that in our algorithm, we compute $\lambda_t^{s_t^{b_t^k}} \approx f(Y^{t+1}|s_t^{b_t^k}, F_t, \theta, P)$ which is not equal to $f(Y^{t+1}|F_t, S^{t+1}, s_t^{b_t^k}, \theta, P)$ but can be viewed as a good approximation. The computation of $f(Y^{t+1}|F_t, S^{t+1}, s_t^{b_t^k}, \theta, P)$ would be much more time consuming and avoiding it allows us to consider more particles for the conditional SMC.

# 3 Marginal likelihood

We use two ways to compute the marginal likelihood, a global method that relies on bridge sampling, as proposed by Meng and Wong (1996), and a local method based on the marginal likelihood identity of Chib (1995). The difficulty of computation of the likelihood $f(Y_T|\theta, P)$ is the main reason why the marginal likelihood has not been used. The SMC algorithm

constitutes an interesting alternative to obtain an unbiased estimation of the quantity, see Chib, Nardari, and Shephard (2000).

### 3.1 Bridge sampling

The marginal likelihood is defined as $f(Y_T) = \int f(Y_T|\theta, P)f(\theta, P)d\theta\, dP$. The bridge sampling idea is to estimate this integral by using the MCMC output and an importance sampling approach. For a given function $t(\theta, P)$ and a proposal density $q(\theta, P)$, we define

$$\begin{aligned} A_1 &= \int t(\theta, P)q(\theta, P)f(\theta, P|Y_T)d\theta\, dP \\ A_2 &= \int t(\theta, P)f(Y_T|\theta, P)f(\theta, P)q(\theta, P)d\theta\, dP. \end{aligned}$$

Meng and Wong (1996) highlight that $f(Y_T) = A_2/A_1$ and that the quantities $A_1$ and $A_2$ can be estimated by $\hat{A}_1 = \frac{1}{G_1}\sum_{j=1}^{G_1} t(\theta^j, P^j)q(\theta^j, P^j)$ with $\{\theta^j, P^j\}$ the $G_1$ posterior draws, and $\hat{A}_2 = \frac{1}{G_2}\sum_{j=1}^{G_2} t(\theta^j, P^j)f(Y_T|\theta^j, P^j)f(\theta^j, P^j)$, this time with $G_2$ draws $\{\theta^j, P^j\}$ from $q(\theta, P)$. The likelihood $f(Y_T|\theta^j, P^j)$ is computed ($G_2$ times) by the conditional SMC algorithm described in Section 2 . In fact,

$$f(Y_T|\theta, P) = f(y_1|\theta, P)\prod_{t=2}^{T} f(y_t|Y_{t-1}, \theta, P) \tag{5}$$

where $f(y_t|Y_{t-1}, \theta, P)$ can be estimated by, see Pitt, Silva, Giordani, and Kohn (2010),

$$\left(\frac{1}{N}\sum_{i=1}^{N}\hat{w}_t^i\right)\left(\sum_{i=1}^{N} g_t^i\right). \tag{6}$$

Notice that if $t(\theta, P) = 1/q(\theta, P)$, the method is equivalent to importance sampling, and to reciprocal importance sampling if $t(\theta, P) = 1/f(\theta, P|Y_T)$. We follow Meng and Wong (1996) who obtain $t(\theta, P) = (f(\theta, P|Y_T) + q(\theta, P))^{-1}$ as an asymptotically optimal choice which minimizes the expected relative error of the estimator in the case of i.i.d draws from $f(\theta, P|Y_T)$ and $q(\theta, P)$. The proposal distribution $q(\theta, P)$ is split into two independent blocks $q(\theta)$ and $q(P)$. 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 (see the Appendix) is used as proposal for sampling $\theta$ in step 3 of the particle Gibbs algorithm sketched in the beginning of Section 2. We refer the reader to Fruhwirth-Schnatter (2004) for more details on the implementation of bridge sampling and examples for mixture and Markov-switching models.

### 3.2 Chib's method

As proposed by Chib (1995), the marginal likelihood can also be computed as

$$f(Y_T) = \frac{f(\theta^*, P^*) f(Y_T|\theta^*, P^*)}{f(\theta^*, P^*|Y_T)} \tag{7}$$

where $P^*$ and $\theta^*$ can be any admissible value but is typically chosen to be a high density point like the mode, mean or median of the posterior distribution. The prior is easily computed and the likelihood $f(Y_T|\theta^*, P^*)$ is computed (once) by the SMC algorithm as in the previous subsection.

The evaluation of the posterior distribution $f(\theta^*, P^*|Y_T)$ is done in two parts. Since $f(\theta^*, P^*|Y_T) = f(P^*|Y_T, \theta^*) f(\theta^*|Y_T)$ we use in the first part that

$$f(P^*|Y_T, \theta^*) = \int f(P^*|Y_T, S_T) p(S_T|Y_T, \theta^*) dS_T \approx \frac{1}{G_3} \sum_{g=1}^{G_3} f(P^*|Y_T, S_T^g), \tag{8}$$

where $S_T^g$ is the sampled value of $S_T$ at the $g$-th iteration of an auxiliary Gibbs/PMCMC sampler where $\theta$ is kept fixed at $\theta^*$, and $G_3$ denotes the number of iterations after convergence. The auxiliary sampler iterates between $p(S_T|\theta^*, Y_T, P)$ and $f(P|S_T, Y_T)$.

For the second part, we use the method of Chib and Jeliazkov (2001) since we sample $\theta$ with a proposal distribution through a Metropolis step. The method uses the reversibility of the Markov chain generated by the PMCMC sampler to compute $f(\theta^*|Y_T)$. Let us denote by $\alpha(\theta', \theta^*|Y_T, P, S_T)$ the Metropolis-Hastings probability to move from $\theta'$ to $\theta^*$ and by $q(\theta', \theta^*|Y_T, P, S_T)$ the density of the proposal at $(\theta', \theta^*)$. The subkernel satisfies the local reversibility condition

$$\begin{aligned} &f(\theta^*|Y_T, S_T, P) \alpha(\theta^*, \theta'|Y_T, S_T, P) q(\theta^*, \theta'|Y_T, S_T, P) \\ &= f(\theta'|Y_T, S_T, P) \alpha(\theta', \theta^*|Y_T, S_T, P) q(\theta', \theta^*|Y_T, S_T, P). \end{aligned}$$

By multiplying both sides by $f(P, S_T|Y_T)$ and integrating over $(\theta', P, S_T)$, we get

$$f(\theta^*|Y_T) = \frac{\int\int\int \alpha(\theta', \theta^*|Y_T, S_T, P) q(\theta', \theta^*|Y_T, S_T, P) f(\theta', P, S_T|Y_T) d\theta' dP dS_T}{\int\int\int \alpha(\theta^*, \theta'|Y_T, S_T, P) q(\theta^*, \theta'|Y_T, S_T, P) f(P, S_T|Y_T, \theta^*) d\theta' dP dS_T}. \tag{9}$$

A Monte Carlo estimate of the numerator is given by

$$\frac{1}{G_1} \sum_{g_1=1}^{G_1} \alpha(\theta^{g_1}, \theta^*|Y_T, S_T^{g_1}, P^{g_1}) q(\theta^{g_1}, \theta^*|Y_T, S_T^{g_1}, P^{g_1}) \tag{10}$$

where $(\theta^{g_1}, P^{g_1}, S_T^{g_1})$ is the $g_1$-th draw of the particle Gibbs posterior sampler described in Section 2 and $G_1$ is the total number of draws. The denominator is estimated by

$$\frac{1}{G_3}\sum_{g=1}^{G_3}\alpha(\theta^*, \theta^g|Y_T, S_T^g, P^g) \tag{11}$$

where $(P^g, S_T^g)$ is the $g$-th draw of the auxiliary sampler (with $G_3$ draws) defined above for the first part. Given this draw, $\theta^g$ is generated from $q(\theta^*, \theta|Y_T, S_T^g, P^g)$, the proposal for $\theta$ conditioned on $(P^g, S_T^g)$. Hence $(P^g, S_T^g, \theta^g)$ is a draw from the joint distribution of $(\theta', P, S_T)$ defined by $q(\theta^*, \theta'|Y_T, S_T, P)f(P, S_T|Y_T)$, and (11) is an estimate of the expectation of $\alpha(\theta^*, \theta'|Y_T, S_T, P)$ with respect to that joint distribution.

The computation of the marginal likelihood can be summarized as follows:

1. Choose a high density point $(\theta^*, P^*)$ from the posterior particle Gibbs sample.
2. Compute the prior density value $f(\theta^*, P^*)$.
3. Launch a SMC algorithm to compute the likelihood $f(Y_T|\theta^*, P^*)$ using (5) and (6).
4. Estimate the numerator of (9) from the posterior sample, using formula (10).
5. Launch an auxiliary particle Gibbs sampler with fixed parameter $\theta^*$ and from the generated draws compute the denominator of (9) using formula (11). Also compute an estimate of $f(P^*|Y_T, \theta^*)$ using (8).
6. Collect all terms and compute $\log f(Y_T)$ from the right hand side of (7).

For Markov-switching models Fruhwirth-Schnatter (2004) highlights that Chib's marginal likelihood estimator is biased. We do not apply this correction because it is very difficult to quantify the bias. The marginal likelihood estimator à la Chib is a bridge sampling estimator corresponding to a specific non-optimal choice of $t(\theta, P)$, see Meng and Schilling (2002), Mira and Nicholls (2004) and Ardia, Basturk, Hoogerheide, and van Dijk (2010) for examples. However the bridge sampling estimator with optimal choice of $t(\theta, P)$ is derived asymptotically and assumes i.i.d draws of the posterior distribution, so it is interesting to provide an empirical comparison of the two estimators, as we do in the next section. We remark finally that Chib's estimator requires to launch $G_3$ auxiliary particle Gibbs samplers, and is therefore as time-consuming as the bridge sampling estimator that requires to launch $G_2$ SMC samplers, assuming that $G_2$ an $G_3$ are equal.

# 4 Illustrations

This section is divided in three parts. To be precise on the implementation of the sampler in the illustrations, we first describe the prior distributions, starting values and other parameters of the algorithm. Second, we illustrate the approach on simulated data, which allows us to check if the correct model is chosen by the marginal likelihood criterion and to investigate the posterior distributions of misspecified models. Third, we provide applications to daily returns of eleven return series.

## 4.1 Prior distributions, starting values and other parameters

We use standard prior distributions for this type of models. We assume independence between the transition matrix parameters P and the GARCH parameters $\theta$. Following Chib (1996) and Chib (1998), the prior on P is a Dirichlet distribution. The prior hyperparameters are given in Table 1. They imply a probability of 0.9991 to stay in a given regime, or an expected duration of 1111 days in a given regime, which is similar to He and Maheu (2010). The prior on the GARCH equation parameters is specified in terms of appropriate transformations of the elements of $\theta$ - see the note of Table 1- and is a multivariate normal distribution with a diagonal covariance matrix having large variances. The sensitivity of the results, i.e. selection of optimal number of regimes using the marginal likelihood, to the choice of the hyperparameters is discussed in Section 4.4.

Although the particle Gibbs algorithm should converge in principle for any starting point in the parameter space, a high density starting value for the parameters ensures a quicker convergence to the posterior distribution. For the MS- and CP-GARCH models considered here, we use the particle swarm optimization method, see Kennedy and Eberhart (1995), to find starting values that are likely to be close to the maximum likelihood estimate.

For every model, we perform 10,000 particle Gibbs iterations ($G_1$) after convergence acoding to the Geweke diagnostic (Geweke (1992)). The marginal likelihood is computed by bridge sampling with 1,000 draws ($G_2$) of the proposal distribution, and by Chib's method using the posterior median of each parameter, by running 600 auxiliary particle Gibbs iterations ($G_3$). Based on several experiments, we fix the number of particles ($N$) to 150 for CP- and 250 for MS-GARCH models, respectively. This number is very low compared to the several thousands of particles used for models with a continuous state vector or models that

Table 1: Hyperparameters of the prior distributions

| GARCH parameters ($\theta$) | |
|---|---|
| | Distribution : |
| $\mu = (\mu_\omega, \mu_\alpha, \mu_\beta)' \quad \Sigma = 8I_{3(K+1)}$ | |
| $\mu_\omega = (-4, \ldots, -4)$ | Normal$(\mu, \Sigma)$ |
| $\mu_\beta = (\ln(\frac{0.75}{0.25}), \ldots, \ln(\frac{0.75}{0.25}))$, $\mu_\alpha = (\ln(\frac{0.25}{0.75}), \ldots, \ln(\frac{0.25}{0.75}))$ | |

| Transition probabilities (P) | | |
|---|---|---|
| Model | | Distribution: |
| CP-GARCH | $\alpha = 1110.11, \beta = 1$ | Beta$(\alpha, \beta)$ |
| MS-GARCH | $\alpha = \begin{pmatrix} K \times 1110.11 & 1 & ... & ... & 1 \\ 1 & K \times 1110.11 & 1 & ... & 1 \\ ... & ... & ... & ... & ... \\ 1 & 1 & 1 & ... & K \times 1110.11 \end{pmatrix}$ | Dirichlet$(\alpha)$ |

GARCH parameters are mapped on the real line with $\omega \in ]0, +\infty] \rightarrow \ln \omega$, $\alpha \in ]0, 1[ \rightarrow \ln(\frac{\alpha}{1-\alpha})$, and $\beta \in ]0, 1[ \rightarrow \ln(\frac{\beta}{1-\beta})$. $K$ is the number of regimes minus 1.

use the particle filter for inference on all the parameters, as He and Maheu (2010) who use 300,000 particles for a CP-GARCH model.

Note that inference on the models described here requires nontrivial programming and many computations that can be time consuming. For example, given the configuration defined above, and for a sample size of 3000 observations, the computing time for estimating a MS- or CP-GARCH model (including the marginal likelihood) with $K = 2$ is of the order of three hours on a Intel Core 2 Duo 3Ghz with 3.48Gb RAM memory. This is about 200 times more than for the standard GARCH model. Executable codes are available on the web site of the corresponding author.

### 4.2 Illustrations with simulated data

We illustrate our algorithm and the marginal likelihood computation on two simulated data sets of 3000 observations. The first dataset is generated by a CP-GARCH model with two breaks and the second by a MS-GARCH model with two regimes. We compute the marginal

likelihood for the true number of regimes plus one to illustrate that the algorithm selects the true model, and we report some posterior information. A Monte Carlo study investigating the sampling properties of the "Bayesian estimator" is infeasible given the computation time this would imply.

### 4.2.1 CP-GARCH data

The true parameter values that we used to simulate 3000 observations are given in Table 3. A structural break occurs after 1000 observations and another one after 2000. This implies that the probabilities $p_{11}$ and $p_{22}$ are equal to 0.999. The persistence of the volatility processes, measured by $\alpha + \beta$, is 0.9 in the first and second regimes and 0.6 in the third regime. The unconditional variance jumps from 2 to 7 in the second regime and drops to 1 in the third regime.

Table 2: Marginal log-likelihood values for 3000 simulated data of CP-GARCH

| Regimes | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Change-Point | | | | |
| BS | -5463.22 | -5451.42 | **-5438.43** | -5442.10 |
| Chib | -5463.00 | -5450.28 | **-5438.09** | -5439.78 |
| Markov-switching | | | | |
| BS | -5463.22 | -5448.95 | *-5442.05* | -5445.63 |
| Chib | -5463.00 | -5448.14 | *-5441.07* | -5443.66 |

The parameters of the 3-regime CP-GARCH DGP are shown in Table 3.

The marginal likelihood values (in logarithms, MLL hereafter) computed for MS- and CP-GARCH models are given in Table 2. The differences between the values estimated by bridge sampling (BS) and by Chib's method are very small. The fact that both the global and local way of computing the marginal likelihood gives the same results indicates that we obtain the correct estimate with high probability. The CP-GARCH model with three regimes (i.e. two breaks) is correctly selected among all models, and consistently with the data generating process (DGP), the MS-GARCH model with three regimes is selected among MS-GARCH models. We observe that the MLL increases substantially from one to three regimes but decreases less strongly beyond the correct number of regimes. In fact, imposing one superfluous regime is less harmful than missing an existing one.

Tables 3 and 4 display posterior information about the parameters of the GARCH equations and of the transition matrix of all the MS- and CP-GARCH models for which we report the MLL values. When the misspecified model with one regime is estimated, we find that as expected the persistence is overestimated, i.e. 0.99, and the unconditional variance amounts to 2.85. The ignored latent state dynamics is partly picked up by the volatility dynamics.

The estimation of the misspecified one break CP-GARCH model finds a break at observation 2007 of the 3000 observations. This is no surprise since this is the biggest of the two breaks in the DGP in the sense that the unconditional variance drops from seven to one. The estimated parameters of the first regime are closest to the first regime true parameter values. The estimated break dates of the correctly specified two break model are 1046 and 2010 (with standard deviations 31 and 8) compared to the true values of 1000 and 2000 respectively. The corresponding volatility process parameter estimates are also reasonably close to the true values if we take into account the posterior standard deviations. The three break CP-GARCH model finds a spurious estimated regime starting at observation 2847, i.e almost at the end of the sample as expected since the third break has to occur in-sample by construction. The high standard deviation of 170 clearly indicates that this break date is highly uncertain, in contrast with what occurs for the other breaks. As expected, the estimated parameters of the more general three-regime MS-GARCH model are globally in line with the true parameters. For the MS-GARCH models, though more regime switches can in principle occur, the dates of regime switches are very close to those reported for the break models. Finally, three regimes of the over-fitted four regime MS-GARCH model have regime parameters close to the MLE given true states parameters, while the spurious regime has completely different parameters implying an unreasonably high unconditional variance.

The posterior means of the transition probabilities in Table 4 are close to the prior means. Actually, the prior information about these parameters is quite close to the data information, since expected durations of staying in a given regime are 1111 in the prior and 1000 in the DGP. We checked the robustness of our results by varying the beta hyper-parameters. Our conclusion is that, similar to He and Maheu (2010), an informative prior is necessary to ensure that the conditional SMC behaves well and then it does not affect the posterior distribution. For example, for the correctly specified model with a less informative prior ($p \sim \text{Beta}(100, 0.5)$, implying expected durations of 201 observations), the estimated break dates of 999 and 2029

Table 3: Results for 3000 simulated data of 3-regime CP-GARCH: GARCH parameters

| | DGP | | | | MLE given true states | | | |
|---|---|---|---|---|---|---|---|---|
| Regime | 1 | 2 | 3 | | 1 | 2 | 3 | |
| $\omega$ | 0.2 | 0.7 | 0.4 | | 0.29 | 0.71 | 0.31 | |
| $\alpha$ | 0.1 | 0.2 | 0.2 | | 0.14 | 0.16 | 0.16 | |
| $\beta$ | 0.8 | 0.7 | 0.4 | | 0.70 | 0.72 | 0.54 | |
| Break date | 1000 | 2000 | | | | | | |
| | Change-point $K=0$ | | | | Markov-switching $K=0$ | | | |
| Regime | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| $\omega$ | 0.05 | | | | 0.05 | | | |
| | (0.02) | | | | (0.02) | | | |
| $\alpha$ | 0.13 | | | | 0.13 | | | |
| | (0.02) | | | | (0.02) | | | |
| $\beta$ | 0.86 | | | | 0.86 | | | |
| | (0.02) | | | | (0.02) | | | |
| | Change-point $K=1$ | | | | Markov-switching $K=1$ | | | |
| Regime | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| $\omega$ | 0.12 | 0.31 | | | 0.19 | 0.67 | | |
| | (0.04) | (0.08) | | | (0.05) | (0.24) | | |
| $\alpha$ | 0.14 | 0.15 | | | 0.15 | 0.18 | | |
| | (0.02) | (0.05) | | | (0.03) | (0.04) | | |
| $\beta$ | 0.83 | 0.54 | | | 0.72 | 0.71 | | |
| | (0.03) | (0.10) | | | (0.06) | (0.05) | | |
| Break date | 2007 | | | | | | | |
| | (32.2) | | | | | | | |
| | Change-point $K=2$ | | | | Markov-switching $K=2$ | | | |
| Regime | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| $\omega$ | 0.34 | 0.78 | 0.32 | | 0.39 | 0.81 | 0.29 | |
| | (0.14) | (0.23) | (0.07) | | (0.14) | (0.24) | (0.08) | |
| $\alpha$ | 0.15 | 0.17 | 0.17 | | 0.15 | 0.17 | 0.15 | |
| | (0.03) | (0.04) | (0.05) | | (0.03) | (0.03) | (0.04) | |
| $\beta$ | 0.68 | 0.70 | 0.52 | | 0.65 | 0.70 | 0.56 | |
| | (0.12) | (0.06) | (0.09) | | (0.09) | (0.05) | (0.10) | |
| Break date | 1046 | 2010 | | | | | | |
| | (31.7) | (8.5) | | | | | | |
| | Change-point $K=3$ | | | | Markov-switching $K=3$ | | | |
| Regime | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| $\omega$ | 0.26 | 0.76 | 0.42 | 0.24 | 0.38 | 0.30 | 0.82 | 2.24 |
| | (0.09) | (0.22) | (0.18) | (0.09) | (0.06) | (0.15) | (0.08) | (0.22) |
| $\alpha$ | 0.13 | 0.19 | 0.10 | 0.41 | 0.15 | 0.14 | 0.17 | 0.03 |
| | (0.03) | (0.03) | (0.04) | (0.17) | (6E-4) | (0.03) | (0.04) | (0.03) |
| $\beta$ | 0.73 | 0.69 | 0.51 | 0.33 | 0.65 | 0.56 | 0.70 | 0.90 |
| | (0.06) | (0.05) | (0.16) | (0.14) | (0.08) | (0.09) | (0.05) | (0.11) |
| Break date | 1007 | 2011 | 2847 | | | | | |
| | (36.7) | (6.7) | (170.4) | | | | | |

Posterior means and standard deviations in parentheses. The break dates are the posterior modes of the state variables.

Table 4: Results for 3000 simulated data of 3-regime CP-GARCH : transition probabilities

| Regimes | Change-point | Markov-switching |
|---|---|---|
| $K=2$ | $\begin{pmatrix} 0.9994 & 0.0006 \\ 0 & 1 \end{pmatrix}$ | $\begin{pmatrix} 0.9990 & 0.0010 \\ 0.0007 & 0.9993 \end{pmatrix}$ |
| $K=3$ | $\begin{pmatrix} 0.9991 & 0.0009 & 0 \\ 0 & 0.9990 & 0.001 \\ 0 & 0 & 1 \end{pmatrix}$ | $\begin{pmatrix} 0.9990 & 0.0006 & 0.0004 \\ 0.0003 & 0.9994 & 0.0003 \\ 0.0006 & 0.0003 & 0.9991 \end{pmatrix}$ |
| $K=4$ | $\begin{pmatrix} 0.9991 & 0.0009 & 0 & 0 \\ 0 & 0.9990 & 0.001 & 0 \\ 0 & 0 & 0.9982 & 0.0018 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ | $\begin{pmatrix} 0.9991 & 0.0003 & 0.0003 & 0.0003 \\ 0.0002 & 0.9991 & 0.0005 & 0.0002 \\ 0.0002 & 0.0002 & 0.9991 & 0.0005 \\ 0.0004 & 0.0003 & 0.0002 & 0.9991 \end{pmatrix}$ |

Posterior means of the transition probabilities. The DGP parameters of the 3-regime CP-GARCH model are given in Table 3.

(with standard deviations 32 and 21) are close to those in Table 3.

#### 4.2.2 MS-GARCH data

We simulated 3000 observations of a two-regime MS-GARCH model with the same GARCH parameters for the first two regimes as in the CP-GARCH model, see Table 3. The transition probabilities are given by $p_{11} = 0.9999$ and $p_{22} = 0.9995$. They are chosen to be high in order to have only two regime switches, so that the CP-GARCH model can also cover this case without needing to estimate models with many breaks. For conciseness, we do not report the posterior results as in Table 3.

Table 5: Marginal log-likelihood values for 3000 simulated data of MS-GARCH

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

The DGP parameters of the 2-regime MS-GARCH model are shown in Table 3 (regimes 1 and 2 of DGP).

Table 5 presents the MLL values. The differences between the values by bridge sampling (BS) and Chib's method are again very small. The MS-GARCH model with two regimes is

correctly chosen as the best model. As expected, in the CP-GARCH class the three-regime model has the highest MLL.

### 4.3 Illustrations on financial time series

We first provide detailed results for MS- and CP-GARCH models fitted to S&P 500 daily index returns. Next, we provide results for ten other series. For the sake of comparison, we also estimate the spline GARCH model of Engle and Rangel (2008). The latter model is more flexible than the standard GARCH model since in addition to the usual GARCH dynamics it captures long run volatility movements by spline functions. It is defined as

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

where $(\alpha, \beta, \gamma, \lambda_0, \ldots, \lambda_k)$ are parameters, $(t-t_i)_+ = \min(0, t-t_i)$ and $\{t_0 = 0, t_1, t_2, \ldots, t_{k-1}\}$ are time indices (knots) partitioning the sample size $T$ in $k$ equally spaced intervals. For this model, the number of knots is chosen by the BIC criterion and the prior density to be integrable but fairly little informative since it is uniform on finite intervals for each parameter.

#### 4.3.1 S&P 500 index

We use a sample of 3000 daily percentage returns from May 20, 1999 to April 25, 2011. The time series is plotted in Figure 1 with estimated regime switches shown by vertical lines.

Table 6: Marginal log-likelihood values for S&P 500 data

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

The MLL estimates computed by bridge sampling and by Chib's method are given in Table 6 and they indicate that the two-regime MS-GARCH model fits the data best. There

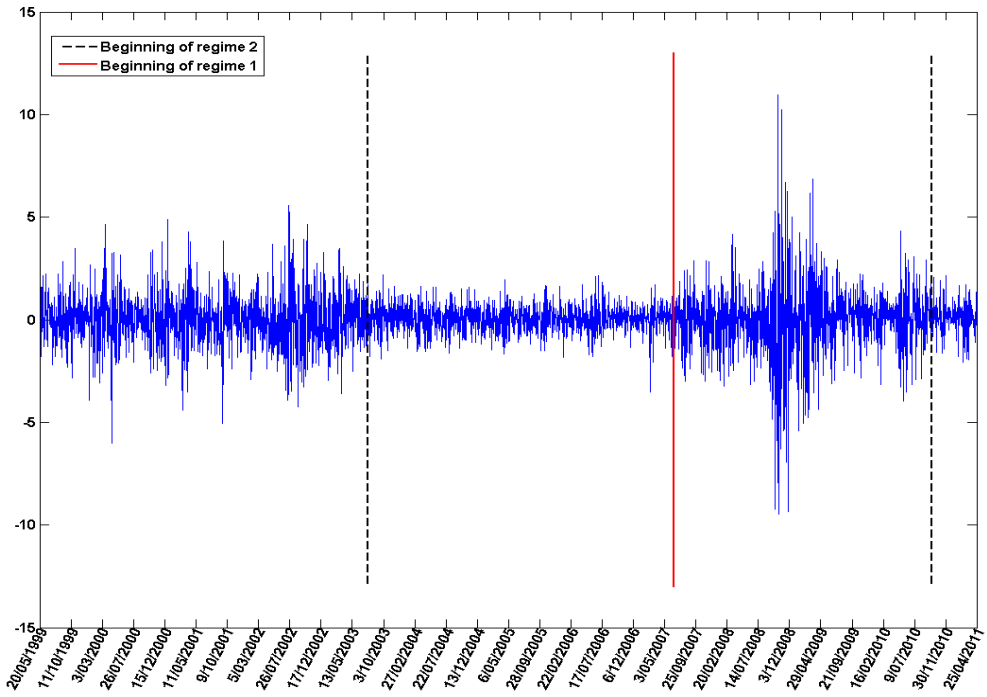


Figure 1: S&P 500 index returns with switches from the 2 regime MS-GARCH model

are three regime switches, occurring on July 22, 2003, June 15, 2007, and September 27, 2010, which make sense after inspecting Figure 1. These dates are the modes of the posterior draws of the state variables; estimation uncertainty as measured by posterior standard deviations is respectively 37, 17 and 20. The second best model, with a decrease in MLL of about 5, is the CP-GARCH model with two breaks (three regimes), at dates July 18, 2003 and June 14, 2007 (see the values in italics in the table).

To get an idea about the precision of the marginal likelihood estimators, we computed the MLL's ten times for all the models in Table 6 using different seeds. More than ten times would be desirable but too computationally intensive. It turns out that both the BS and Chib estimators seem to be fairly precise. In fact, from Table 7 we see that for the best MS-GARCH model (two regimes) the difference between the maximum and minimum MLL is 0.49 and 3.13 for the BS and Chib's estimators respectively. We obtain similar results (not reported) for the best CP-GARCH model (with two breaks), results not reported, i.e the difference is 0.40 and 1.08 for the BS and Chib's estimators respectively. However it becomes larger when the model exhibits at least one spurious regime, see e.g. the reulst in the last

column of Table 7.

Table 7: Minima and maxiima of ten marginal log-likelihood estimates for S&P 500 data

| Regimes | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Change-Point | | | | |
| BS-min | -4505.36 | -4506.02 | -4503.01 | -4519.23 |
| BS-max | -4505.28 | -4505.40 | -4502.61 | -4509.14 |
| Chib-min | -4505.11 | -4506.04 | -4503.20 | -4516.16 |
| Chib-max | -4504.97 | -4505.44 | -4502.12 | -4504.92 |
| Markov-switching | | | | |
| BS-min | -4505.36 | -4497.99 | -4505.01 | |
| BS-max | -4505.28 | -4497.50 | -4501.89 | |
| Chib-min | -4505.11 | -4496.94 | -4505.35 | |
| Chib-max | -4504.97 | -4493.81 | -4497.73 | |

Table 8 provides the posterior means for the single regime GARCH model and the best MS- and CP-GARCH models. The single regime GARCH model has an unconditional variance of 1.67, with a persistence of 0.99. The first regime in the CP-GARCH model has a higher unconditional variance of 1.95 with a lower persistence of 0.95, the second regime unconditional variance is equal to 0.45, with the same persistence as the first regime. Finally in the last regime, triggered in June 2007, the unconditional variance jumps to 2.75, with a persistence of 0.99 due to the relatively high posterior mean of 0.098 for $\alpha$. The two-regime MS-GARCH model has local unconditional variances of 2.32 and 0.46, with persistences of 0.98 and 0.93, respectively. This model alternates between these two regimes, and detects a switch back to the low volatility regime in September 2010. The CP-GARCH model does not infer a new episode of low volatility at the end of the sample, contrary to the MS-GARCH model.

The $\alpha$ and $\beta$ parameter estimates (posterior means of 0.073 and 0.902, respectively) for the best spline-GARCH model (which has three knots) are very close to the estimates for the standard GARCH model. Figure 2 provides a graphical comparison of the spline-GARCH and the CP and MS-GARCH models in terms of local unconditional variances and volatility persistence ($\alpha + \beta$). While the spline-GARCH has a smooth unconditional volatility function determined by the three knots, the MS- and CP-GARCH models have local constant levels, which for forecasting purposes may be more desirable. Visually, the short term volatilities

Table 8: Posterior means for S&P 500

| | GARCH | | | CP-GARCH | | | MS-GARCH | | |
|---|---|---|---|---|---|---|---|---|---|
| Regime | $\sigma^2$ | $\alpha$ | $\beta$ | $\sigma^2$ | $\alpha$ | $\beta$ | $\sigma^2$ | $\alpha$ | $\beta$ |
| 1 | 1.67 | 0.075 | 0.915 | 1.95 | 0.085 | 0.868 | 2.32 | 0.089 | 0.891 |
| | (0.51) | (0.009) | (0.011) | (0.32) | (0.020) | (0.031) | (0.512) | (0.012) | (0.015) |
| 2 | | | | 0.45 | 0.023 | 0.931 | 0.46 | 0.031 | 0.901 |
| | | | | (0.033) | (0.011) | (0.027) | (0.036) | (0.013) | (0.042) |
| 3 | | | | 2.75 | 0.098 | 0.890 | | | |
| | | | | (0.792) | (0.015) | (0.016) | | | |

The (local) unconditional variance $\sigma^2$ is computed as $\omega/(1-\alpha-\beta)$.

are very similar for the three models.

Finally, we also estimated the above models on the S&P 500 index starting at April, 1988 instead of May, 1999 which increases the sample size from 3000 to 5800 observations. The MS-GARCH model, with a MLL of -7840.31, is still the preferred model but now with three regimes instead of two regimes for the shorter series analyzed above. Similarly, the best CP-GARCH model has now five regimes instead of three with a marginal likelihood of -7857.99.

#### 4.3.2 Other series

To get more insight in the differences between MS- and CP-GARCH models, we provide MLL estimates for three other major US indices, five stocks, one exchange rate, and one commodity index. For each series, we estimated the models on data from May 20, 1999 to April 25, 2011 (3000 observations). Table 9 reports the MLL estimates of the best CP- and MS-GARCH models together with the single regime GARCH model, and the maximized log-likelihood values of the spline-GARCH and all other models. The reported MLL values are those obtained by bridge sampling, the values obtained by Chib's method are close to them and not reported to save space.

To compare the MLL values of two models, we use the informal rule of Kass and Raftery (1995): if the difference is smaller than 1, the evidence in favor of the model that has the highest values is "not worth than a bare mention", whereas if it is larger than 1, the evidence is positive, and strong if it exceeds 3. For the 11 series, the MLL values are higher for the MS-GARCH model than for the CP-GARCH model. The evidence is at least positive in 10 cases,

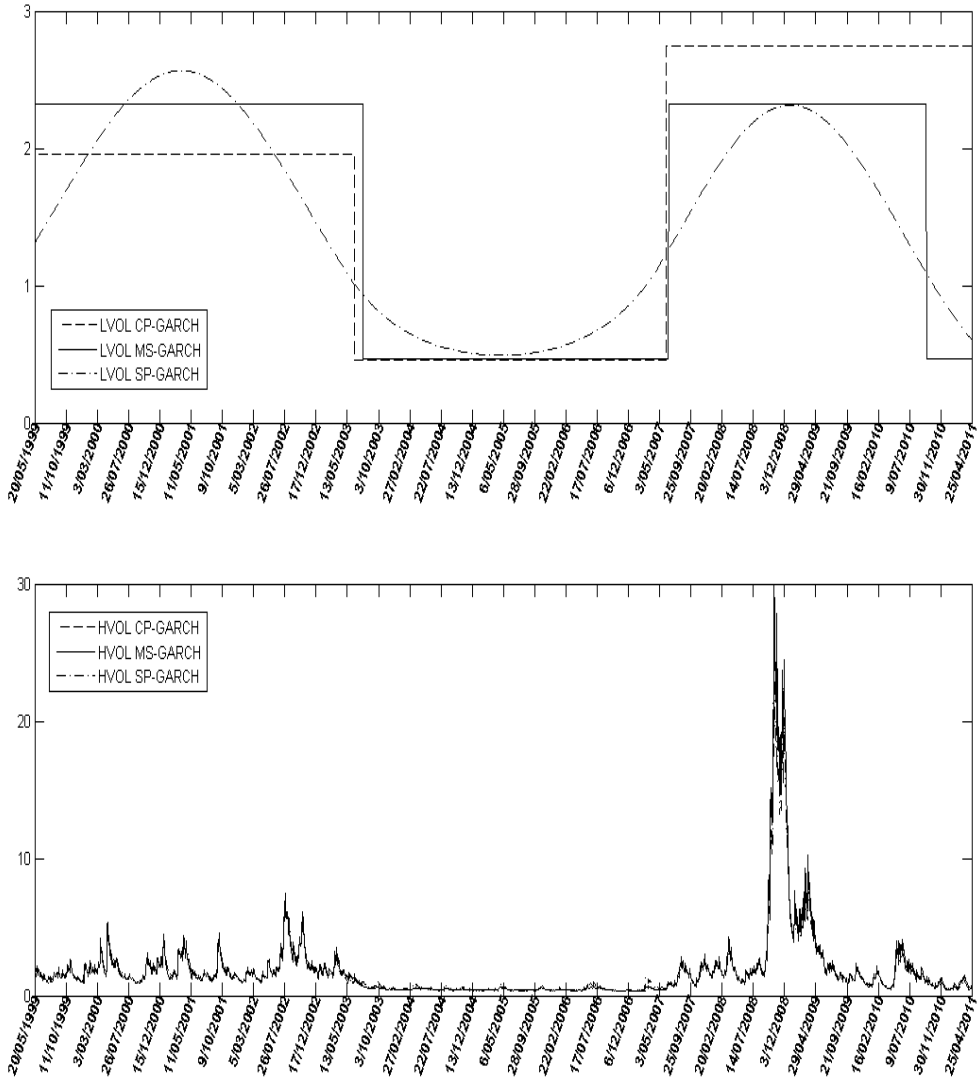


Figure 2: Unconditional volatility (top) and conditional volatility (bottom), S&P 500

Table 9: Marginal log-likelihoods (MLL) for various time series

| Series | Spline-GARCH | | | GARCH | | CP-GARCH | | | MS-GARCH | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | knots | log-lik | MLL | log-lik | MLL | K+1 | log-lik | MLL | K+1 | log-lik | MLL | nswitch |
| S&P 500 | 3 | -4477.36 | -4500.12 | -4494.55 | -4505.33 | 3 | -4476.70 | -4503.05 | 2 | -4478.59 | -4497.99 | 3 |
| NASDAQ | 3 | -5404.39 | -5426.64 | -5418.99 | -5429.84 | 1 | -5418.99 | -5429.84 | 2 | -5410.00 | -5429.58 | 5 |
| DJIA | 3 | -4307.91 | -4330.44 | -4322.79 | -4333.43 | 1 | -4322.79 | -4333.43 | 2 | -4309.52 | -4328.97 | 3 |
| NYSE | 3 | -4355.77 | -4378.22 | -4369.77 | -4380.62 | 1 | -4369.77 | -4380.62 | 2 | -4360.08 | -4377.5 | 4 |
| BAC | 4 | -6088.38 | -6110.77 | -6117.49 | -6127.39 | 3 | -6059.24 | -6085.62 | 3 | -6021.05 | -6049.83 | 9 |
| BA | 4 | -6140.58 | -6165.47 | -6163.93 | -6174.57 | 2 | -6147.13 | -6165.7 | 2 | -6141.35 | -6161.77 | 4 |
| JPM | 3 | -6370.73 | -6391.45 | -6388.83 | -6400.27 | 3 | -6367.75 | -6394.74 | 3 | -6363.1 | -6392.69 | 5 |
| MRK | 5 | -6136.08 | -6160.95 | -6198.78 | -6209.73 | 5 | -5928.36 | -5999.31 | 3 | -5889.78 | -5922.88 | 10 |
| PG | 4 | -4795.84 | -4825.68 | -4832.63 | -4842.02 | 4 | -4781.52 | -4816.14 | 2 | -4789.65 | -4809.59 | 9 |
| Metals | 2 | -5239.80 | -5260.78 | -5256.63 | -5267.44 | 3 | -5234.26 | -5258.68 | 2 | -5234.73 | -5253.15 | 5 |
| Yen/Dollar | 1 | -2966.94 | -2985.67 | -2972.94 | -2982.33 | 1 | -2972.94 | -2982.33 | 2 | -2957.62 | -2980.52 | 7 |

log-lik: Maximum of the log-likelihood values over all the MCMC draws; MLL: marginal log-likelihood value computed by bridge sampling; K+1: number of regimes; nswitch: number of regime switches. S&P 500: Standard and Poors 500 index; NASDAQ: Nasdaq Composite Index; DJIA: Dow Jones Industrial Average; NYSE: New York Stock Exchange Composite Index; BAC: Bank of America Corporation: BA: Boeing Co.; JPM: JPMorgan Chase & Co.; MRK: Merck & Co Inc.: PG: Procter & Gamble Co. Metals: WCFI Base Metals Sub-Index.

and strong in 8 of these cases. The standard one-regime GARCH model has even lower MLL values than the CP-GARCH model, except when they are identical (i.e. the CP-GARCH model has a single regime). The spline model has a higher MLL than the MS model in two cases (NASDAQ and JPM, with positive evidence), and a lower MLL in the other cases (with at least positive evidence in 8 cases, and strong in 5 of these). In brief, regarding in-sample fit, there is clear evidence in favor of the MS-GARCH model, i.e. recurrent regimes for the series we have analyzed, but the spline model might be considered as a useful alternative. Obviously, from this analysis it is unclear how the models differ in producing volatility forecasts out-of-sample. We checked the MS and CP-GARCH regime parameters that are prevailing at the end of the sample period (values unreported to save space) and we see pronounced differences both in the level and the dynamics of the (local) volatility process. Forecast comparisons are left for further research.

Note that the log-likelihood values of MS- and CP-GARCH models can be very close. For example, the S&P 500 log-likelihood values in Table 9 are respectively -4478.59 and -4476.70 – thus slightly higher for the CP model – but the MLL is higher for the MS-GARCH model due to the penalization of the more heavily parameterized CP-GARCH model (eleven parameters versus six). Similarly for the DJIA, the two regime MS-GARCH has a log-likelihood of -4309.52 while the three regime CP-GARCH has a value of -4307.57.

The number of regimes in the MS-GARCH models varies between one and three, and one and five in the CP-GARCH models. The four major indices have the same optimal number of regimes, i.e. two, for the MS-GARCH model. As can be seen in Table 9, the maximum number of MS regime switches over the index series is five, while it goes up to eleven for the individual series, which is a relatively high number in order to be replicated by a CP-GARCH model, especially knowing that some regimes have durations as small as forty-four days. Note that the above discussion on the best models may depend on the choice of the prior hyperparameters, as discussed next in Section 4.4.

### 4.4 Sensitivity of marginal likelihood to the prior distribution

It is well known, but perhaps too often neglected, that the marginal likelihood is sensitive to the choice of the prior distribution, see for example Kass and Raftery (1995) and Sinharay and Stern (2002). Chib's marginal likelihood identity, eq (7), particularly underlines the interde-

pendence between the marginal likelihood and the prior. The penalty for the introduction of new parameters does not have to be too strong or too small, in the sense that adding an extra regime should sufficiently improve the fit. This section illustrates how the marginal likelihood based model selection varies when using three different priors for the GARCH parameters. Prior 1 uses a uniform distribution for $\omega$, $\alpha$, and $\beta$. The other two priors use Gaussian distributions on the GARCH parameters transformed to the real line. Prior 2 is the one used in the paper so far, and Prior 3 is much more diffuse, see Table 10 for details. Each prior differently penalizes the marginal likelihood. The uniform distribution on finite (small) intervals nearly imposes any penalty for an extra regime. In contrast, Prior 3 strongly decreases the marginal likelihood if an additional regime is imposed.

We study the model selection by the marginal likelihood criterion for the two simulated data sets and the four US index series. Some results for the CP-GARCH and MS-GARCH models are shown in Table 11. Not surprisingly, the selection varies according to the choice of the prior. More precisely, Prior 1 sometimes overestimates the number of breaks, in particular it selects the wrong model for the MS-GARCH data. Alternatively, for time series with small evidence in favor of one specific model, Prior 3 selects the model with the smallest number of parameters. Prior 2 gives in-between results. We also observe that for the stock indices, the MS-GARCH models find breaks irrespective of the prior and therefore we find strong evidence of regime switches in at least two of the four index series.

To complement Table 11 we give another more visual illustration of the sensitivity of the MLL with respect to the prior. In particular, we focus on Prior 1 since only the interval of $\omega$ penalizes the MLL (the two other parameters, having a uniform prior on the unit interval, do not modify the MLL). Moreover as any uniform density cancels out in the Metropolis-Hastings acceptance probability, the denominator of equation (7), i.e. the posterior distribution, is not modified if the selected prior interval is large enough. The alteration of the MLL is then only due to the prior density. In order to show the dependence of the MLL on the prior, we let the upper bound $b_\omega$ increase from 10 to 1000 while keeping the lower bound $a_\omega$ fixed to 0. As can be seen in Figures 3 and 4, the optimal number of regimes may change with respect to the size of the prior interval. Nevertheless, the values of the upper bound of the uniform prior of $\omega$ has to be quite high (and very unrealistic for the type data we analyze) for the changes to occur. Thus this type of sensitivity is not a source of worry.

Table 10: Hyperparameters for the prior distributions

| | |
|---|---|
| Prior 1 : GARCH parameters $(\theta)$ | Distribution : |
| $a = (a_\omega, a_\alpha, a_\beta)' \quad b = (b_\omega, b_\alpha, b_\beta)'$ | |
| $a_\omega = (0, \ldots, 0) \quad b_\omega = (25, \ldots, 25)$ | $\mathrm{U}[a, b]$ |
| $a_\alpha = a_\beta = (0, \ldots, 0) \quad b_\alpha = b_\beta = (1, \ldots, 1)$ | |
| Prior 2 : GARCH parameters $(\theta)$ | Distribution : |
| $\mu = (\mu_\omega, \mu_\alpha, \mu_\beta)' \quad \Sigma = 8I_{3(K+1)}$ | |
| $\mu_\omega = (-4, \ldots, -4)$ | $\mathrm{Normal}(\mu, \Sigma)$ |
| $\mu_\beta = (\ln(\frac{0.75}{0.25}), \ldots, \ln(\frac{0.75}{0.25}))$, $\mu_\alpha = (\ln(\frac{0.25}{0.75}), \ldots, \ln(\frac{0.25}{0.75}))$ | |
| Prior 3 : GARCH parameters $(\theta)$ | Distribution : |
| $\mu = (\mu_\omega, \mu_\alpha, \mu_\beta)' \quad \Sigma = 100I_{3(K+1)}$ | |
| $\mu_\omega = (-4, \ldots, -4)$ | $\mathrm{Normal}(\mu, \Sigma)$ |
| $\mu_\beta = (\ln(\frac{0.75}{0.25}), \ldots, \ln(\frac{0.75}{0.25}))$, $\mu_\alpha = (\ln(\frac{0.25}{0.75}), \ldots, \ln(\frac{0.25}{0.75}))$ | |

GARCH parameters for Prior 2 and Prior 3 are mapped on the real line. One to one functions to map parameters are $\omega \in ]0, +\infty] \to \ln \omega$, $\alpha \in ]0, 1[ \to \ln(\frac{\alpha}{1-\alpha})$, and $\beta \in ]0, 1[ \to \ln(\frac{\beta}{1-\beta})$.

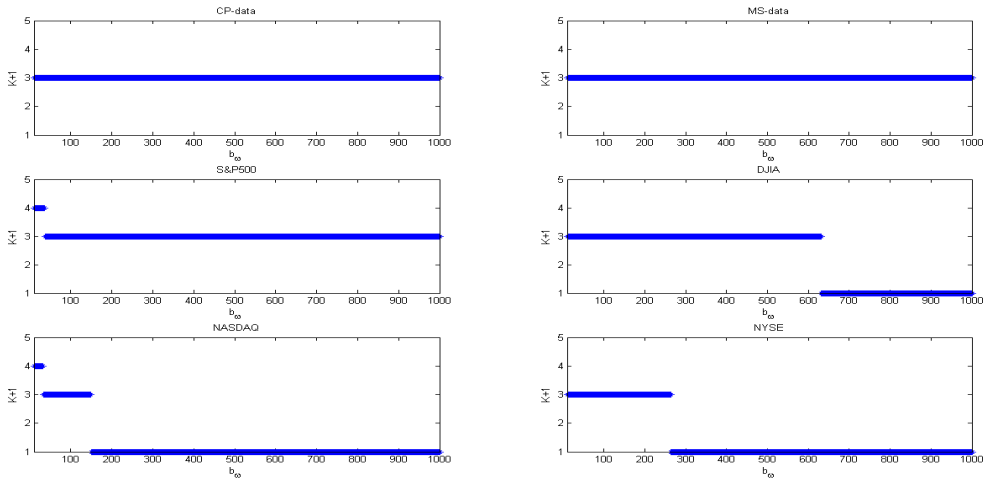


Figure 3: CP-GARCH: Model selection with respect to the upper bound $b_\omega$

Table 11: Marginal log-likelihoods for various priors

| Series | CP-GARCH | | | | | |
|---|---|---|---|---|---|---|
| | Prior 1 | | Prior 2 | | Prior 3 | |
| | K+1 | MLL | K+1 | MLL | K+1 | MLL |
| CP-data | 3 | -5435.78 | 3 | -5438.43 | 3 | -5447.94 |
| MS-data | 3 | -5837.46 | 3 | -5848.05 | 3 | -5856.81 |
| S&P 500 | 4 | -4493.14 | 3 | -4503.05 | 1 | -4508.94 |
| NASDAQ | 4 | -5423.00 | 1 | -5429.84 | 1 | -5433.31 |
| DJIA | 3 | -4324.15 | 1 | -4333.43 | 1 | -4337.11 |
| NYSE | 3 | -4373.01 | 1 | -4380.62 | 1 | -4384.2 |
| | MS-GARCH | | | | | |
| | Prior 1 | | Prior 2 | | Prior 3 | |
| | K+1 | MLL | K+1 | MLL | K+1 | MLL |
| CP-data | 3 | -5438.72 | 3 | -5442.05 | 3 | -5451.31 |
| MS-data | 3 | -5836.07 | 2 | -5843.55 | 2 | -5850.59 |
| S&P 500 | 2 | -4491.72 | 2 | -4497.99 | 2 | -4504.99 |
| NASDAQ | 2 | -5423.48 | 2 | -5429.58 | 1 | -5433.31 |
| DJIA | 3 | -4319.39 | 2 | -4328.97 | 2 | -4335.77 |
| NYSE | 3 | -4372.18 | 2 | -4377.50 | 1 | -4384.2 |

MLL values are computed by bridge sampling. K+1 is the number of regimes.

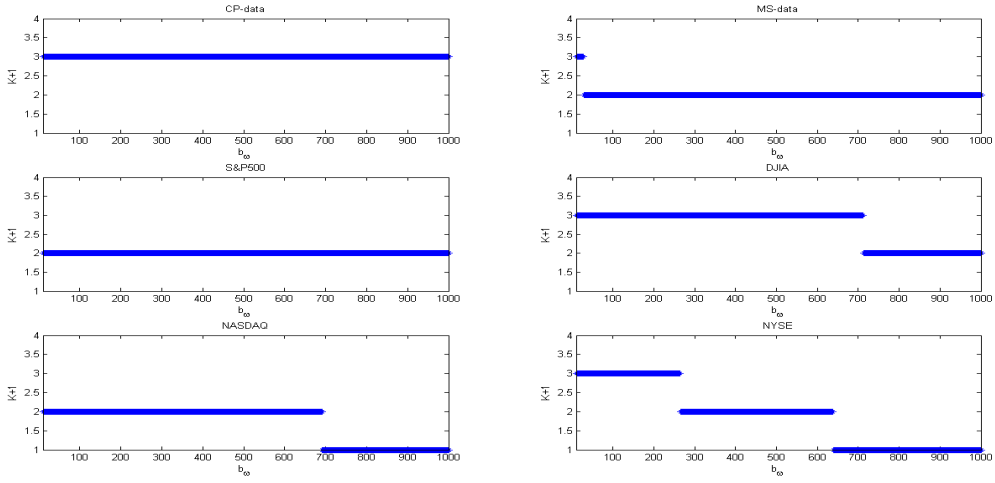


Figure 4: MS-GARCH: Model selection with respect to the upper bound $b_\omega$

# 5 Conclusion

MS- and CP-GARCH models are flexible alternatives to GARCH models with fixed parameters. We estimate them by Bayesian inference using data augmentation because of the path dependence problem. We choose the number of regimes or breaks by computing the marginal likelihood. We introduce an efficient method to do this, which was not feasible until our contribution, due to the challenge posed by such models in integrating the latent state variables that govern the parameter evolution between regimes. The algorithm belongs to the particle filter class and is intensive in computations but feasible as we are able to use a reasonable number of particles, due to the discrete nature of the state variables, and the fact that we do not use particles for the parameters of the volatility processes and of the transition matrix. We have illustrated the use of the method on several time series of financial returns, for which it seems that CP-GARCH and especially MS-GARCH models are useful for capturing changes in the dynamics and level of volatilities. Further research will be centered on forecast comparisons with these and competing models, and using the same framework for multivariate volatility models. Furthermore, other empirically relevant issues related to the optimal number of regimes are first the effect of relaxing the Gaussian innovation assumption, and second the impact of more complex volatility functions than the standard GARCH specification.

# Appendix: Sampling the conditional variance parameters

This appendix describes the sampling of $\theta$ in the Gibbs sampler of Section 2. We implement a Metropolis-Hastings step that samples from a mixture of five normally distributed components. The mixture is adapted during the burn-in period. The expectation and the variance-covariance matrix of the first component are computed using burn-in draws. This component behaves as an independent Metropolis-Hastings. For the other components we only specify the variance-covariance matrix. Besides the second component, variance-covariance matrixes only differ from a scaling parameter. The expectation is given by the current parameter of the Particle Gibbs as in a standard random-walk Metropolis-Hastings. The weights are given in Table 12 where $\mu$ and $\Sigma$ respectively stand for the posterior mean and the posterior variance-covariance matrix estimated using available draws, $I$ denotes the identity matrix and $\theta_{cur}$ is the current parameter of the particle Gibbs.

Table 12: Mixture weights of the proposal distribution

| Mixt. comp. | weight | distribution |
|---|---|---|
| 1 | 0.05 | $N(\mu, 0.01\Sigma)$ |
| 2 | 0.15 | $N(\theta_{cur}, 0.5I)$ |
| 3 | 0.15 | $N(\theta_{cur}, 0.05\Sigma)$ |
| 4 | 0.55 | $N(\theta_{cur}, 0.1\Sigma)$ |
| 5 | 0.10 | $N(\theta_{cur}, \Sigma)$ |

In the case of a MS-GARCH model, one can switch the label of the states without changing the likelihood. A way to deal with this problem is to use identification constraints. However, this is difficult to implement in a high dimensional parameter space and may generate a bias in the estimation of the posterior means, see Geweke (2007) for examples. Instead, we run an unconstrained sampler and apply a loss function on the posterior sample by considering all possible permutations. We minimize this loss function on the posterior sample which leads to the best permutation. The following idea of Marin, Mengersen, and Robert (2005) has been implemented: $\tau \in \Sigma_k$ stands for a possible permutation on the set of all possible permutations of $\{1, ..., k\}$ and we denote by $\tau(\theta, P, S_T) = \{(\theta_{\tau(1)}, ..., \theta_{\tau(k)}), (P_{\tau(1)}, ..., P_{\tau(k)}), (S_{T\tau(1)}, ..., S_{T\tau(k)})\}$ the corresponding permutation of the parameters $(\theta, P, S_T)$. Considering a posterior sample of size $M$, we apply the following scheme:

1. Find $(\theta, P, S_T)^{i*} = \arg\max_{i=1,...,M} f(Y_T|\theta^i, P^i, S_T^i)$
2. $\forall i \in \{1, ..., M\}$
   (a) Compute $\tau_i = \arg\min_{\tau \in \Sigma_k} \left\langle \tau(\theta, P, S_T)^i, (\theta, P, S_T)^{i*} \right\rangle$ where $\langle . \rangle$ stands for the canonical scalar product.
   (b) Set $(\theta, P, S_T)^i = \tau_i(\theta, P, S_T)$.

# References


Amado, C., and T. Terasvirta (2011): "Modelling volatility by variance decomposition," NIPE WP 01/2011.

Andrieu, C., A. Doucet, and R. Holenstein (2010): "Particle Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society B*, 72, 269–342.

Ardia, D., N. Basturk, L. Hoogerheide, and H. K. van Dijk (2010): "A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood," *Computational Statistics & Data Analysis*, forthcoming.

Baillie, R., and C. Morana (2009): "Modeling Long Memory and Structural Breaks in Conditional Variances: an Adaptive FIGARCH Approach," *Journal of Economic Dynamics and Control*, 33, 1577–1592.

Bauwens, L., A. Preminger, and J. Rombouts (2010): "Theory and Inference for a Markov-switching GARCH Model," *Econometrics Journal*, 13, 218–244.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Cai, J. (1994): "Markov Model of Unconditional Variance in ARCH," *Journal of Business and Economics Statistics*, 12, 309–316.

Chib, S. (1995): "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

——— (1996): "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 79–97.

——— (1998): "Estimation and comparison of multiple change-point models," *Journal of Econometrics*, 86, 221–241.

Chib, S., and I. Jeliazkov (2001): "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96, 270–281.

Chib, S., F. Nardari, and N. Shephard (2000): "Markov Chain Monte Carlo Methods for Generalized Stochastic Volatility Models," Nuffield College, Oxford University.

Diebold, F. (1986): "Comment on Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 51–56.

Ding, Z., and C. Granger (1996): "Modeling Volatility Persistence of Speculative Returns: A New Approach," *Journal of Econometrics*, 73, 185–215.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Engle, R., and J. Rangel (2008): "The Spline-GARCH Model for Low-Frequency Volatility and its Global Macroeconomic Causes," *Review of Financial Studies*, 21, 1187–1222.

Fernandez-Villaverde, J., and J. Rudio-Ramirez (2007): "Estimating Macroeconomic Models: a Likelihood Approach," *Review of Economic Studies*, 74, 1059–1087.

Flury, T., and N. Shephard (2011): "Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models," *Econometric Theory*, forthcoming.

Francq, C., and J. Zakoian (2008): "Deriving the autocovariances of powers of Markov-switching GARCH models, with applications to statistical inference," *Computational Statistics and Data Analysis*, 52, 3027–3046.

Fruhwirth-Schnatter, S. (2004): "Estimating Marginal Likelihoods for Mixture and Markov-switching Models Using Bridge Sampling Techniques," *Econometrics Journal*, 7, 143–167.

Geweke, J. (1992): "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," *Bayesian Statistics, Oxford University Press.*, 4, 169–193.

——— (2007): "Interpretation and Inference in Mixture Models: Simple MCMC works," *Computational Statistics and Data Analysis*, 51, 3259–3550.

Godsill, S. J., A. Doucet, and M. West (2004): "Monte Carlo Smoothing for Nonlinear Time Series," *Journal of the American Statistical Association*, 99, 156–168.

Gonzales-Rivera, G. (1998): "Smooth Transition GARCH Models," *Studies in Nonlinear Dynamics and Econometrics*, 3, 61–78.

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

Haas, M., S. Mittnik, and M. Paolella (2004a): "Mixed Normal Conditional Heteroskedasticity," *Journal of Financial Econometrics*, 2, 211–250.

——— (2004b): "A New Approach to Markov-Switching GARCH Models," *Journal of Financial Econometrics*, 2, 493–530.

Hamilton, J., and R. Susmel (1994): "Autoregressive Conditional Heteroskedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307–333.

He, Z., and J. Maheu (2010): "Real Time Detection of Structural Breaks in GARCH Models," *Computational Statistics and Data Analysis*, 54, 2628–2640.

Johannes, M., N. Polson, and J. Stroud (2009): "Optimal Filtering of Jump-Diffusions: Extracting Latent States from Asset Prices," *Review of Financial Studies*, 22, 2759–2799.

Kass, R., and A. Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Kaufman, S., and S. Fruhwirth-Schnatter (2002): "Bayesian analysis of switching ARCH models," *Journal of Time Series Analysis*, 23, 425–458.

Kennedy, J., and R. Eberhart (1995): "Particle Swarm Optimization," *Proceedings of IEEE International Conference on Neural Networks*, 4, 1942–1948.

Marin, J.-M., K. Mengersen, and C. P. Robert (2005): "Bayesian Modelling and Inference on Mixtures of Distributions," Handbook of Statistics 25, pp. 459–507. North Holland.

Meng, X., and S. Schilling (2002): "Warp Bridge Sampling," *Journal of Computational and Graphical Statistics*, 11, 552–586.

Meng, X.-L., and W. Wong (1996): "Simulating Ratios of Normalizing Constants via a Simple Identity : A theoretical Exploration," *Statistica Sinica*, 6, 831–860.

Mikosch, T., and C. Starica (2004): "Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects," *Review of Economics and Statistics*, 86, 378–390.

Mira, A., and G. Nicholls (2004): "Bridge Estimation of the Probability Density at a Point," *Statistica Sinica*, 14, 603–612.

Noh, J., R. Engle, and A. Kane (1994): "Forecasting Volatility and Option Prices of the S&P 500 Index," *Journal of Derivatives*, pp. 17–30.

Pitt, M., R. Silva, P. Giordani, and R. Kohn (2010): "Auxiliary particle filtering within adaptative Metropolis-Hastings sampling," arXiv:1006.1914v1.

Pitt, M. K., and N. Shephard (1999): "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94, 590–599.

Sinharay, S., and H. S. Stern (2002): "On the Sensitivity of Bayes Factors to the Prior Distributions," *The American Statistician*, 56, No 3, 196–201.