UNIVERSITE CATHOLIQUE DE LOUVAIN

Faculté des Sciences Economiques, Sociales, Politiques et de Communication

Département des Sciences Economiques

Center for Operations Research and Econometrics

Modeling structural changes in volatility

Arnaud Dufays

Thèse présentée en vue de l'obtention

du grade de docteur en sciences économiques et de gestion.

Jury:

Thesis director :

Professor Luc Bauwens (Université Catholique de Louvain)

President of the jury:

Professor Sébastien Van Bellegem (Université Catholique de Louvain)

Other jury members :

Professor Jeroen Rombouts (ESSEC Business School, Paris)

Professor Herman van Dijk (Erasmus Universiteit, Rotterdam)

Doctor Yukai Yang (Université Catholique de Louvain)

Louvain-la-Neuve, Juillet 2013

Acknowledgments

This research project would not have been possible without the support of many people. I thus take this opportunity to express my gratitude to those who have been instrumental in the successful completion of this thesis. First and foremost, I would like to warmly acknowledge the tremendous supervision of Professor *Luc Bauwens*. During four years, He gave me confidence in my research, delivered helpful advices and encouraged me when the research motivation was lacking. Professor *Jeroen Rombouts* has also been a guide to carry out this thesis. Needless to mention that he had been a source of inspiration and that he was able to canalized me especially for the Chapter 2. I gratefully thank him for having given me the opportunity to visit H.E.C. Montreal.

A special thank goes to my doctoral committee composed by professor *Herman van Dijk*, professor *Jeroen Rombouts*, doctor *Yukai Yang* and professor *Sébastien Van Bellegem* for their time spent in reading this manuscript.

I highly appreciated all the moments spent with *Bruno De Backer* who worked with me on the Chapter 3. His accurate mind led to several memorable discussions on mathematical details that finally do not appear in this final version.

Professor *Jean-Pierre Urbain* kindly accepted to support me in my job research and so attended to one or two never-ending presentations on Change-point and Markov-switching models in order to get some insights of what I did during these four years. I really appreciated his helpful comments and his encouragements.

I would like to acknowledge *Marie-Hélène Chassagne* who, as a Ph.D. administrator, was always available to help me when I encountered several administrative difficulties at the end of my thesis.

A Ph.D. student though mastering some theoretical and practical aspects still behaves like a student living an eternal exam preparation period. I will remember the support that I

received from my beloved *family* and from my lovely girlfriend *Isabelle De Clerck* during these years. The wonderful time spent to do a Ph. D. thesis goes along with difficult choices and disheartening moments. They always backed me even in the hardiest times.
This research has been carried out thanks to two financial supports : the contract "Projet d'Actions de Recherche Concertées" 07/12-002 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain" and the F.S.R. support also granted by the "Académie universitaire Louvain".

Contents

1

Introduction

Interpretations and forecasts of relevant economic variables in order to clarify the economic situation at different levels (region, country, ...) constitute the core of works in economics. To improve their effectiveness, the economists rely on models in which they translate their beliefs into a list of assumptions. A good model fairly reproduces the stylized facts exhibited by the series while preserving its specification as simple as possible (i.e. the Ockham's razor principle). We thus expect that models will at least provide good descriptions of past observations and reliable predictions. These forecasts are essential for good decision making, for example with respect to fiscal and monetary policy, public spending and investment, but also financial decisions related to asset allocations and hedging of risky investments.

The last few years, marked economically by the financial and the Euro-zone debt crisis, have emphasized the need of efficient tools to accurately forecast economic time series. Specifying dynamic models that quickly account for contemporaneous changes has become one of the mainstream research topics. The present thesis belongs to this literature and focuses on dynamic models that allow for abrupt switches in the model parameters.

Although financial time series exhibit complex dynamics, it turns out that they share several statistical properties. This observation led researchers to investigate the different financial stylized facts with the idea of gathering them into a financial model. Modern finance started in the beginning of the twentieth century with Bachelier (1900). In his Ph.D thesis, he suggested modeling government bond prices as random variables. Bond prices were assumed to follow a random walk model with Brownian motion and constant

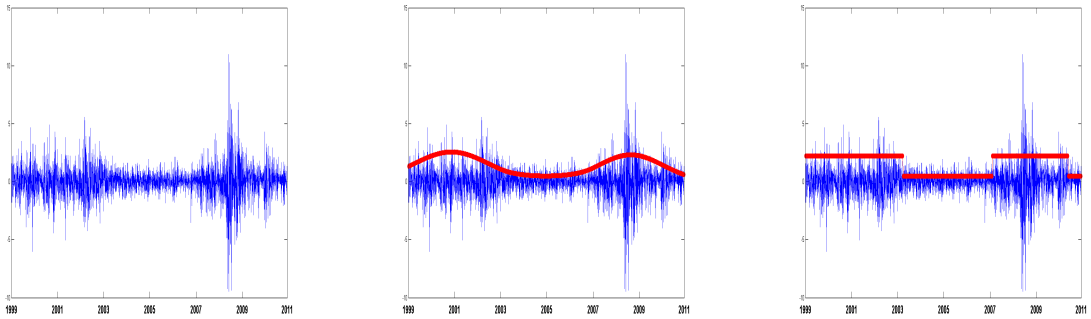
variance. This revolutionary model has been left unimproved for sixty years and has generated relevant enhancements in option pricing theory (Black and Scholes (1973)), in market theory (Malkiel and Fama (1970)) and in portfolio optimization (Markowitz (1959)). In 1963, a new model based on Levy processes was proposed by Mandelbrot (1963). His study on the cotton prices emphasized the fat tail behavior of the series. He thus strongly rejected the Brownian motion assumption. Not surprisingly this observation was not a specific feature of the cotton price series and researchers started to highlight similar features for many other financial series. As data became more and more available, another assumption of the Bachelier's model was about to be ruled out. Obviously the constant variance was not supported by the data. Moreover the variance seemed to cluster in high and low volatility and auto-correlations in absolute returns were observed. The last two observations imply that financial time series are realized from not independent processes as it was initially assumed. Engle (1982) and Bollerslev (1986) managed to build an observation-driven model, called (G)ARCH (Generalized Auto-Regressive Conditional Heteroskedastic), that unified all these stylized facts, was tractable and was moreover economically interpretable. Nowadays this model and its numerous variants are still widely used. In concrete terms, if a time series is denoted by $Y_T = \{y_1, y_2, ..., y_T\}$, the GARCH model in its simplest form is as follows

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \quad \text{with} \quad \mathrm{E}(\epsilon_t) = 0, \quad \mathrm{V}(\epsilon_t) = 1 \\ \sigma_t^2 &= \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned}$$

where ϵ_t is an unobservable random variable belonging to an i.i.d. process; E(.) and V(.) respectively stand for the expectation and the Variance operators and $\{\omega, \alpha, \beta\}$ denote the model parameters.

Another relevant financial stylized fact has recently emerged : The long run volatility evolves over time and generates the non-stationarity of many time series (see, e. g., Diebold (1986), Lamoureux and Lastrapes (1990), Hillebrand (2005)). GARCH models with static parameters are no longer sufficient and should be adapted over time.

Changes in volatility may be slow and progressive or quick and abrupt. Ignoring these breaks by assuming constant parameters in econometric models typically leads to forecasts that are far from realisations (see for example Stock and Watson (1996), Pesaran and Timmermann (2004)). As a consequence, an important econometric challenge is to detect as soon as possible a structural break. The slow variation of parameters is typically dealt with a non-parametric approach such as spline or kernel estimation. Engle and Rangel (2008) or Amado and Teräsvirta (2008) successfully incorporate smooth transition of parameters in their models. Abrupt changes are coped with Markov-switching (MS) when switches are recurrent or change-point (CP) models (Hamilton (1989), Chib (1998)) when they are not. Figure 1.1 displays the $S\&P$ 500 daily percentage returns spanning from May 20, 1999 to April 25, 2011. The time frame is large enough to observe the low frequency variance variation. To highlight this slow movement, graphics 1.1b and 1.1c show the estimated unconditional variance over the time series of a smooth transition model (i.e. the spline-GARCH model of Engle and Rangel (2008)) and of an abrupt one (MS-GARCH model, see chapter 2). The present work focuses on MS and CP framework applied to volatility models and adopts the Bayesian framework to estimate them.



(a) $S\&P$ 500 percentage returns

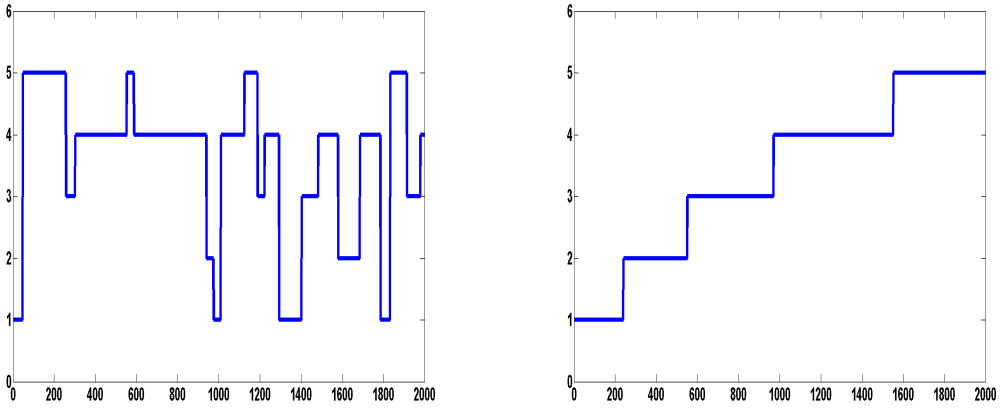
(b) Smooth transition in the unconditional variance (spline-GARCH model)

(c) Abrupt changes in the unconditional variance (MS-GARCH model)

Figure 1.1: $S\&P$ 500 percentage returns and low frequencies of the $S\&P$ 500 volatility computed from the spline-GARCH model (smooth transition) and from the MS-GARCH model (abrupt switches)

The first Bayesian CP approach dates back to the seminal paper of Chernoff and Zacks (1964). They allowed for a change in the mean of a normal distribution. The switch

was specified as a random walk. By that time no simulation methods were available so they achieved analytical formulas for At Most One Change in the mean (AMOC estimator). Thirty years later, Carlin, Gelfand, and Smith (1992a) relied on a novel simulation method, named Markov-chain Monte-Carlo (MCMC) (Geman and Geman (1984), Gelfand and Smith (1990)), to estimate the posterior distribution of parameters (of a linear regression) in presence of one break. Their framework considered the break date τ as a random parameter. Sampling τ from its full conditional distribution was feasible due to the discrete nature of the variable. Stephens (1994a) extended their work by allowing for multiple change-points. He preserved the former specification but introduced multiple break dates $\tau_1, ..., \tau_K$. Stephens' MCMC estimation sampled each break date conditional to the others resulting in a depreciation of the MCMC mixing properties. Chapter 3 details a CP estimation of GARCH models based on this multiple break date specification. Meanwhile Markov-switching models widely progressed in the economic literature. Introduced by Hamilton (1989) to study non-stationary data and business cycles, the MS model specifies recurrent regimes whereas the CP model assumes non-recurrent regimes. MS specification consists in enriching the model with a dynamic discrete latent state Markov process (say $S_T = \{s_1, ..., s_T\}$ where T denotes the sample size) in such a way that the parameters can abruptly switch from one value to another. The Markov process is determined by a full transition probability matrix P, a specification that highly increases the number of model parameters. MS estimation relies on the forward algorithm (Rabiner (1989)) that integrates the latent variable S_T and returns the likelihood value given other model parameters. In 1998, Chib (1998) brought a real turning point in the CP literature by constraining the MS transition matrix P to only account for CP dynamics. By data augmentation (see e.g. Tanner and Wong (1987)), he derived an MCMC procedure capable of updating the latent variable in one block. The MCMC mixing properties were highly improved. Nowadays the Chib's method dominates the Bayesian CP and MS literature whereas Hamilton's estimation prevails in the frequentist one. CP and MS specifications depending on a latent process are enlarged to univariate and multivariate GARCH models in Chapter 2 and 4. Figure 1.2 depicts one possible state vector S_T where the number of states is five for the MS and the CP specification. It highlights the constraints on the CP dynamic.



Markov-switching Change-point

Figure 1.2: One possible Markov-switching and Change-point paths that govern the dynamic of the model parameters.

Likelihood estimation of MS or CP models relies on a recursive computation of the likelihood function over time. Concretely, a likelihood function $f(Y_T|\theta)$ where θ denotes a set of parameters, can always be decomposed as

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

Estimation of MS- or CP-GARCH models is numerically unfeasible either by the method of maximum likelihood[1] or by the Bayesian approach, given 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 the likelihood function at each time period t over all possible regime paths. The forward-backward algorithm of Chib (1996, 1998) makes the number of computations to be performed by the computer grow exponentially with the length of the time series. Moreover, subsequent improvements of the Chib's (1996) algorithm by Koop and Potter (2007) and Giordani and Kohn (2008) do not solve the problem. The table 1.1 documents the number of

[1]Recently, Augustyniak (2013) proposes a ML estimation based on the single-move of Bauwens, Preminger, and Rombouts (2010).

possible paths for a Markov-switching model with 2 regimes. The number of possible paths explodes even for extremely short time series.

T	1	2	3	4	8	12	16
Poss. paths	2	4	8	16	256	4096	65536

Table 1.1: Number of possible paths with respect to the length of the series for a MS model accounting for 2 regimes.

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 for MS-GARCH models even for a given number of regimes. MS-GARCH models that avoid the path dependence problem were proposed by Gray (1996), Klaassen (2002) and Haas, Mittnik, and Paolella (2004). Nowadays the issue is usually circumvented either by applying the model of Haas, Mittnik, and Paolella (2004), who suggest switching from one stationary GARCH process to another stationary one instead of moving from one state (explosive or not) to another or by estimating the model of Klaassen (2002), which gets rid of the path dependence problem by approximating the conditional variance with a mixture. Two recent papers tackle the issue of path dependence. He and Maheu (2010) use a Sequential Monte Carlo *on-line* method, which is more useful for forecasting than for in-sample detecting purposes. On the contrary, Bauwens, Preminger, and Rombouts (2010) develop *off-line* MCMC methods and employ a single-move strategy to update the latent process S_T resulting in highly auto-correlated posterior draws.

The thesis firstly concentrates on new MCMC methods for CP- and MS-GARCH models. Chapter 2 solves the path dependence problem by using Particle MCMC, a technique proposed by Andrieu, Doucet, and Holenstein (2010). We estimate CP- and MS-GARCH models and compare their in-sample fit with the spline-GARCH model of Engle and Rangel (2008). Chapter 3 focuses on CP-GARCH model and recurrent regimes. Relying on Stephens (1994a), we propose a multi-move MCMC that renders CP estimation of popular models (i. e. auto-regressive models, auto-regressive moving average models and GARCH models) much faster than the standard forward-backward algorithm. We

develop a new generic discrete proposal distribution for a Metropolis-Hastings algorithm and prove the ergodicity of the MCMC. The last chapter 4 is devoted to univariate and multivariate CP- and MS-GARCH models specified by a hierarchical Dirichlet process. A novel Metropolis-Hastings procedure enlarges the class of Klaassen (2002) and Haas, Mittnik, and Paolella (2004) models.

Choosing the optimal number of regimes can be delicate. The thesis investigates two different strategies to derive the optimal number of regimes. Chapter 2 and 3 rely on the marginal log-likelihood (MLL) which requires the posterior distribution estimation of competing models, followed by its computation. These chapters document multiple ways of estimating it such as the Importance sampling, the Bridge sampling (Meng and Wong (1996), Fruhwirth-Schnatter (2004)) and the Chib's methods (Chib (1995) and Chib and Jeliazkov (2001)). A recent review on the MLL computation has been brought out by Ardia, Basturk, Hoogerheide, and van Dijk (2010). The sticky infinite hidden Markov chain model (IHMM), proposed by Fox, Sudderth, Jordan, and Willsky (2007), allows to bypass these repeated computations by treating the number of regimes as an unknown parameter. The framework that is applied in Chapter 4 relies on a Markov-chain with a potentially infinite number of states (regimes). The structure encompasses the CP and the MS specifications as special cases.

2

Marginal Likelihood Computation for Markov Switching and Change-point GARCH Models[1]

2.1 Introduction

The Chapter solves the path dependence 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). 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.

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. 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),

[1] Joint work with Luc Bauwens and Jeroen Rombouts. This Chapter is a significant revision of Bauwens, Dufays, and Rombouts (2011).

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.

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.

We follow the standard practice in econometrics for model comparison. We first pick the number of regimes and then move on to inference. However, in the context of models without path dependence appealing alternatives have been proposed to prevent fixing the number of regimes in advance. First, there is the reversible jump MCMC algorithm of Green (1995) which allows a change in the dimension of the MCMC by applying a more general Metropolis-Hasting step. Second, the sticky infinite hidden Markov-Chain of Fox, Sudderth, Jordan, and Willsky (2008), based on the hierarchical Dirichlet process of Teh, Jordan, Beal, and Blei (2006), implements a non-parametric transition matrix that can handle an infinite number of regimes; see Chapter 4 for an implementation of this alternative in the GARCH context. Third, Carlin and Chib (1995) develop a MCMC algorithm that encompasses all the parameters for each number of regimes. These three alternatives of sampling both the number of breaks and parameters are potentially applicable using the PMCMC approach and are challenging topics for future research.

The rest of the Chapter is organized as follows. In Section 2.2, we present the particle Gibbs algorithm that we propose for posterior inference on the parameters of MS- and CP-GARCH models. In Section 2.3, we explain the two methods for computing the marginal likelihood. In Section 2.5, we illustrate the algorithm on simulated and real data. Furthermore, we discuss the sensitivity to the choice of the priors, the mixing properties of the sampler and provide a comparison with alternative algorithms. Conclusions are presented in the last section.

2.2 Inference for MS- and CP-GARCH models

We consider the model defined by

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2 \\ \epsilon_t &\sim N(0,1), \end{aligned} \tag{2.1}$$

where 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 tran-

sition 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 σ_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 θ 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 θ 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.

2.2.1 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 2.3.

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

$$p(s_T|Y_T, \theta, P)p(s_{T-1}|s_T, Y_T, \theta, P) \dots p(s_t|S^{t+1}, Y_T, \theta, P) \dots p(s_1|S^2, Y_T, \theta, P). \quad (2.2)$$

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:

$$\begin{aligned} p(s_t|S^{t+1}, Y_T, \theta, P) &= \frac{p(s_t|Y_t, \theta, P)f(Y^{t+1}, S^{t+1}|s_t, Y_t, \theta, P)f(Y_t|\theta, P)}{f(S^{t+1}, Y_T|\theta, P)} \\ &\propto p(s_t|Y_t, \theta, P)f(Y^{t+1}, S^{t+1}|s_t, Y_t, \theta, P) \\ &\propto p(s_t|Y_t, \theta, P)f(Y^{t+1}|S^t, Y_t, \theta, P)p(s_{t+1}|s_t, P). \quad (2.3) \end{aligned}$$

The probabilities $p(s_t|Y_t,\theta,P)$ (for each t and possible value of s_t) in (2.3) are complicated to evaluate because of the path dependence problem. This problem can be alleviated by applying SMC techniques. In particular, our SMC algorithm adds an auxiliary particle, see Pitt and Shephard (1999) for details, to ease the sampling of s_t. 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. These weights serve to approximate the probability $p(s_t|Y_t,\theta,P)$ that appears in (2.3), more specifically $p(s_t=j|Y_t,\theta,P)=\sum_{i=1}^N w_t^i 1_{\{s_t^i=j\}}$ with $1_{\{\}}$ the indicator function. The particle filter introduces an integer random variable η taking values in $[1,N]$ and defines (we drop the conditions θ,P for ease)

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

The intuition behind the first line of the formula above is that the sum of $p(s_t,\eta|Y_t)$ over all values of η is the probability $p(s_t|Y_t)$ which appears in (2.3). The idea of the next two lines is that the sampling of s_t from the proposal distribution $g(s_t,\eta|Y_t)$ will be quite accurate if the proposal takes into account y_t. We take $g(s_t,\eta|Y_t)\propto w_{t-1}^\eta p(s_t|s_{t-1}^\eta)f(y_t|Y_{t-1},s_t)$. Hence $g(\eta|Y_t)\propto w_{t-1}^\eta\sum_{j=1}^{K+1}p(s_t=j|s_{t-1}^\eta)f(y_t|s_t=j,Y_{t-1})$. Finally,

$$p(s_t,\eta|Y_t)\propto\frac{f(y_t|Y_{t-1},s_t)}{\sum_{j=1}^{K+1}f(y_t|Y_{t-1},s_t=j)p(s_t=j|s_{t-1}^\eta)}p(s_t|s_{t-1}^\eta)g(\eta|Y_t) \tag{2.4}$$

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

To ensure convergence to the stationary distribution, Andrieu, Doucet, and Holenstein (2010) use the full path of each SMC particle (i.e. lineage) and show that the S_T draw and its lineage from the previous Gibbs iteration needs to survive in the SMC algorithm. This is called the conditional SMC. Define the ancestor variable A_t^k as the particle from which the particle k at time t is sampled, and the lineage 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 b_t^k variable represents next the lineage of the previous s_t draw.

The conditional SMC can be computed for $p(s_t|Y_t,\theta,P)$ for $t=1,\ldots,T$, assuming we

have $\{s_1^{b_1^k}, \ldots, , s_T^{b_T^k}\}$ and 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) as:

1. $\forall i \in [1, N]$, 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 $\tilde{\omega}_t^i = g_t^i / \sum_{j=1}^N g_t^j$.

2. $\forall i \in [1, N] \setminus b_t^k$, sample independently a label variable $A_{t-1}^i \sim \tilde{\omega}_t$ such that $A_{t-1}^i \in [1, N]$.

3. $\forall i \in [1, N] \setminus b_t^k$, 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 SMC algorithm is derived from the formula (2.4). We start by sampling η_i from $g(\eta|Y_t)$ (step 1), then we sample $s_t^i \sim p(s_t|s_{t-1}^{\eta_i})$ (step 3), and we compute the weight $\frac{f(y_t|Y_{t-1}, s_t^i)}{\sum_{j=1}^{K+1} f(y_t|Y_{t-1}, s_t=j) f(s_t=j|s_{t-1}^{\eta_i})}$ (step 4). The normalized weights provide in fact an approximation of the distribution $p(s_t|Y_t, \theta, P)$. Note 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 sampler we develop is in line with the particle Gibbs sampler defined in ADH. The particle Gibbs sampler extends the target distribution by incorporating all random variables generated by the conditional SMC, but ADH show that θ, 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 multinomial resampling step and samples k from its full conditional under $\tilde{f}(.)$ and deterministically tracing back the ancestral lineage of S_T^k. While the ADH algorithm is directly applicable to MS- and CP-GARCH models,

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, as explained below. 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 Gibbs.

- 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 (2.3) - 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.

To end the procedure, we iteratively sample backward an entire state vector S_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 $\tilde{\pi}_t^i = \pi_t^i / \sum_{j=1}^N \pi_t^j$.
   - Sample $b_t^k \sim \tilde{\pi}_t$ and set $s_t = s_t^{b_t^k}$.

The probability $\lambda_t^{s_t^i}$, i.e. the second term in (2.3), is approximated by considering the path of each particle. For each $q \in [1, K+1]$, we compute

$$\lambda_t^q = \frac{\sum_{i=1}^N f(y_{t+1}, ..., y_{t+\tau}|s_t^i, s_{t+1}^i, ..., s_{t+\tau}^i, F_t, \theta, P) 1_{\{s_t^i = q\}}}{\sum_{i=1}^N f(y_{t+1}, ..., y_{t+\tau}|s_t^i, s_{t+1}^i, ..., s_{t+\tau}^i, F_t, \theta, P)},$$

where each $f(.)$ is the product of Gaussian densities implied by (2.1). The interval length τ is computed by solving the equation $\beta_i^{\tau_i} = 0.001$ for $i = 1, ..., K+1$ (β being the autoregressive coefficient of the GARCH equation) and by taking the maximum value of τ_i. The full vector S_T is therefore sampled from $t = T$ until $t = 1$ as written in (2.2). Remark that in this backward step of our algorithm, we compute $\lambda_t^{s_t^{b_t^k}}$ 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, see Section 2.5.5 for illustrations. 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.

Note that other SMC algorithms exist. For example, Whiteley, Andrieu, and Doucet (2011) build on the PMCMC theory of Andrieu, Doucet, and Holenstein (2010) and develop an interesting algorithm for a change-point model based on a latent discrete variable. They do inference on the break dates instead of the state vector itself (which is a one-to-one mapping). Their procedure consists in incorporating a SMC algorithm with a deterministic resampling step inside a MCMC algorithm to draw the break dates. This improvement (as well as the gain in memory storage) is particularly important when dealing with very long time series. However, their algorithm is not able to manage the path dependence problem, hence it is not applicable to the MS-GARCH model. In fact, their deterministic resampling step would require an evaluation of all new possible paths at each MCMC iteration. Nevertheless, the idea of a deterministic resampling step is worth exploring in further research on regime switching GARCH models. Furthermore, as we show in Section 4.5, their interesting idea of working with break dates leads to another algorithm in the context of the CP-GARCH model.

2.3 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).

2.3.1 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,

$$f(Y_T|\theta, P) = f(y_1|\theta, P)\prod_{t=2}^{T} f(y_t|Y_{t-1}, \theta, P) \tag{2.5}$$

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

$$\left(\frac{1}{N}\sum_{i=1}^{N}\hat{w}_t^i\right)\left(\sum_{i=1}^{N} g_t^i\right). \tag{2.6}$$

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 θ 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.

2.3.2 Chib's method

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

$$f(Y_T) = \frac{f(\theta^*, P^*) f(Y_T|\theta^*, P^*)}{f(\theta^*, P^*|Y_T)} \tag{2.7}$$

where P^* and θ^* 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

$$f(P^*|Y_T, \theta^*) = \int f(P^*|Y_T, S_T) p(S_T|Y_T, \theta^*) dS_T \approx \frac{1}{G_3} \sum_{g=1}^{G_3} f(P^*|Y_T, S_T^g), \tag{2.8}$$

where S_T^g is the sampled value of S_T at the g-th iteration of an auxiliary Gibbs/PMCMC sampler where θ is kept fixed at θ^*, 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 θ 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 θ' to θ^* and by $q(\theta', \theta^*|Y_T, P, S_T)$ the density of the proposal at (θ', θ^*). 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 (θ', P, S_T), we get

$$f(\theta^*|Y_T) = \frac{\int\int\int \alpha(\theta', \theta^*|Y_T, S_T, P) q(\theta', \theta^*|Y_T, S_T, P) f(\theta', P, S_T|Y_T) d\theta' dP dS_T}{\int\int\int \alpha(\theta^*, \theta'|Y_T, S_T, P) q(\theta^*, \theta'|Y_T, S_T, P) f(P, S_T|Y_T, \theta^*) d\theta' dP dS_T}. \tag{2.9}$$

A Monte Carlo estimate of the numerator is given by

$$\frac{1}{G_1}\sum_{g_1=1}^{G_1}\alpha(\theta^{g_1},\theta^*|Y_T,S_T^{g_1},P^{g_1})q(\theta^{g_1},\theta^*|Y_T,S_T^{g_1},P^{g_1}) \tag{2.10}$$

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 2.2 and G_1 is the total number of draws. The denominator is estimated by

$$\frac{1}{G_3}\sum_{g=1}^{G_3}\alpha(\theta^*,\theta^g|Y_T,S_T^g,P^g) \tag{2.11}$$

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, θ^g is generated from $q(\theta^*,\theta|Y_T,S_T^g,P^g)$, the proposal for θ conditioned on (P^g,S_T^g). Hence (P^g,S_T^g,θ^g) is a draw from the joint distribution of (θ',P,S_T) defined by $q(\theta^*,\theta'|Y_T,S_T,P)f(P,S_T|Y_T)$, and (2.11) 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 (θ^*,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 (2.5) and (2.6).
4. Estimate the numerator of (2.9) from the posterior sample, using formula (2.10).
5. Launch an auxiliary particle Gibbs sampler with fixed parameter θ^* and from the generated draws compute the denominator of (2.9) using formula (2.11). Also compute an estimate of $f(P^*|Y_T,\theta^*)$ using (2.8).
6. Collect all terms and compute $\log f(Y_T)$ from the right hand side of (2.7).

For Markov-switching models Fruhwirth-Schnatter (2004) highlights that Chib's marginal likelihood estimator is biased. The reason is that the posterior distribution is invariant to the labeling of the states, and therefore the marginal likelihood computation requires to explore all possible labels. In fact, the marginal likelihood can also be written as $f(Y_T)=(K+1)!\int_{\mathcal{L}_1}f(Y_T|\theta,P)f(\theta,P)d\theta\,dP$ with $\mathcal{L}_1$ the subspace for the first labeling.

In practice, sampling from $\mathcal{L}_1$ only is difficult to impose and therefore we do not apply this correction. Furthermore, the bias in the log marginal likelihood is not higher than $\log(K+1)!$, which is small for small values of K as we use in this Chapter.

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.

2.4 Posterior Predictive Distributions

Generating h step ahead out-of-sample forecasts is easily achieved in the MCMC estimation procedure. Bayesian methods treat the predictions as random variables and draw the entire posterior predictive distribution $y_{T+1}, ..., y_{T+h}|Y_T$. In our context, the simulation of this distribution is feasible by adding a step in the MCMC scheme and sampling from the full conditional distribution $y_{T+1}, ..., y_{T+h}|Y_T, \theta, P, S_{T+h}$. For CP forecasts, the regime from T to $T+h$ is necessarily $K+1$ due to the specific transition matrix. However for MS specifications, since recurrent states are allowed, we need to simulate the state up to $T+h$. It leads to two different simulation methods.

1. CP forecasts are drawn from the full conditional distributions as follows :

$$\begin{aligned} y_{T+1}, ..., y_{T+h} &\sim f(y_{T+1}, ..., y_{T+h}|Y_T, \theta, P, S_T, s_{T+1} = K+1, ..., s_{T+h} = K+1) \\ &\sim \prod_{t=T+1}^{T+h} f(y_t|Y_{t-1}, \theta, P, S_{1:T}, s_{T+1} = K+1, ..., s_{T+h} = K+1) \\ &\sim \prod_{t=T+1}^{T+h} N(0, \sigma_t^2) \end{aligned}$$

where the variance σ_t^2 in the last equation is computed by the recursive equation of the GARCH model since we condition to all the model parameters and all the previous observations.

2. Drawing the MS predictions are slightly different since we have to sample the future states. Despite this additional sampling, it does not complicate the procedure :

$$\begin{aligned} y_{T+1}, s_{T+1}, ..., y_{T+h}, s_{T+h} &\sim f(y_{T+1}, s_{T+1}, ..., y_{T+h}, s_{T+h}|Y_T, \theta, P, S_T) \\ &\sim \prod_{t=T+1}^{T+h} f(y_t|Y_{t-1}, \theta, P, S_t) f(s_t|s_{t-1}, P) \\ &\sim \prod_{t=T+1}^{T+h} N(0, \sigma_t^2) P_{s_t, s_{t-1}} \end{aligned}$$

2.5 Illustrations

This section is divided in five 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. Thirdly, we provide applications to daily returns of eleven return series. Fourthly, we discuss and illustrate the sensitivity of the marginal likelihood to the prior distribution. Finally, we discuss and illustrate the performance of the PMCMC sampler.

2.5.1 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 θ. Following Chib (1996) and Chib (1998), the prior on P is a Dirichlet distribution. The prior hyper-parameters are given in Table 2.1. 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 θ - see the note of Table 2.1- 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 2.5.4.

Table 2.1: Hyperparameters of the prior distributions

GARCH parameters (θ)		Distribution :
	$\mu = (\mu_\omega, \mu_\alpha, \mu_\beta)' \quad \Sigma = 8I_{3(K+1)}$	
	$\mu_\omega = (-4, \ldots, -4)$	Normal(μ, Σ)
	$\mu_\beta = (\ln(\frac{0.75}{0.25}), \ldots, \ln(\frac{0.75}{0.25}))$, $\mu_\alpha = (\ln(\frac{0.25}{0.75}), \ldots, \ln(\frac{0.25}{0.75}))$	
Transition probabilities (P)		
Model		Distribution:
CP-GARCH	$\alpha = 1110.11, \beta = 1$	Beta(α, β)
MS-GARCH	$\alpha = \begin{pmatrix} K \times 1110.11 & 1 & \ldots & \ldots & 1 \\ 1 & K \times 1110.11 & 1 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & 1 & \ldots & K \times 1110.11 \end{pmatrix}$	Dirichlet(α)

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.

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 accoding 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). We fix the number of particles (N) to 150 for CP-GARCH and 250 for MS-GARCH models, respectively. These numbers are obtained by comparing the autocorrelation times for $N = 25, 50, 100, 150, 250$ and 500 for several financial time series series used in this Chapter. For series like the S&P 500, N could be even as small as 100 for both the MS and CP-GARCH models. These number are very low compared to the

several thousands of particles used for models with a continuous state vector or models that 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. Inference for the models has been programmed in C++. Executable codes, data, and some illustrations are available on the web site of Luc Bauwens.

For MS-GARCH models, the posterior distribution is invariant to the label of the state vector, the so-called label switching problem (e. g. Fruhwirth-Schnatter (2004)). In order to report label dependent summary statistics such as the posterior means or the standard deviations of each regime, we use a loss function to sort the posterior samples simulated by the MCMC algorithm in one specific label ordering. For interested readers, more details are given in the appendix.

2.5.2 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.

2.5.2.1 CP-GARCH data

The true parameter values that we used to simulate 3000 observations are given in Table 2.3. 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 2.2: 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 2.3.

The marginal likelihood values (in logarithms, MLL hereafter) computed for MS- and CP-GARCH models are given in Table 2.2. 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 2.3 and 2.4 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

Table 2.3: Results for 3000 simulated data of 3-regime CP-GARCH: GARCH parameters

	DGP				MLE given true states			
Regime	1	2	3		1	2	3	
ω	0.2	0.7	0.4		0.29	0.71	0.31	
α	0.1	0.2	0.2		0.14	0.16	0.16	
β	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
ω	0.05				0.05			
	(0.02)				(0.02)			
α	0.13				0.13			
	(0.02)				(0.02)			
β	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
ω	0.12	0.31			0.19	0.67		
	(0.04)	(0.08)			(0.05)	(0.24)		
α	0.14	0.15			0.15	0.18		
	(0.02)	(0.05)			(0.03)	(0.04)		
β	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
ω	0.34	0.78	0.32		0.39	0.81	0.29	
	(0.14)	(0.23)	(0.07)		(0.14)	(0.24)	(0.08)	
α	0.15	0.17	0.17		0.15	0.17	0.15	
	(0.03)	(0.04)	(0.05)		(0.03)	(0.03)	(0.04)	
β	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
ω	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)
α	0.13	0.19	0.10	0.41	0.15	0.14	0.17	0.03
	(0.03)	(0.03)	(0.04)	(0.17)	(0.04)	(0.03)	(0.04)	(0.03)
β	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 2.4: 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 2.3.

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 2.4 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 (with standard deviations 32 and 21) are close to those in Table 2.3.

2.5.2.2 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 2.3. 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 2.3.

Table 2.5: 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 2.3 (regimes 1 and 2 of DGP).

Table 2.5 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.

2.5.3 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.

2.5.3.1 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 2.1 with estimated regime switches shown by vertical lines.

Table 2.6: 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 2.6 and they indicate that the two-regime MS-GARCH model fits the data best. There are three regime switches, occurring on July 22, 2003, June 15, 2007, and September 27, 2010, which make sense after inspecting Figure 2.1. These dates are the modes of the posterior draws of the state variables; estimation uncertainty as measured by posterior

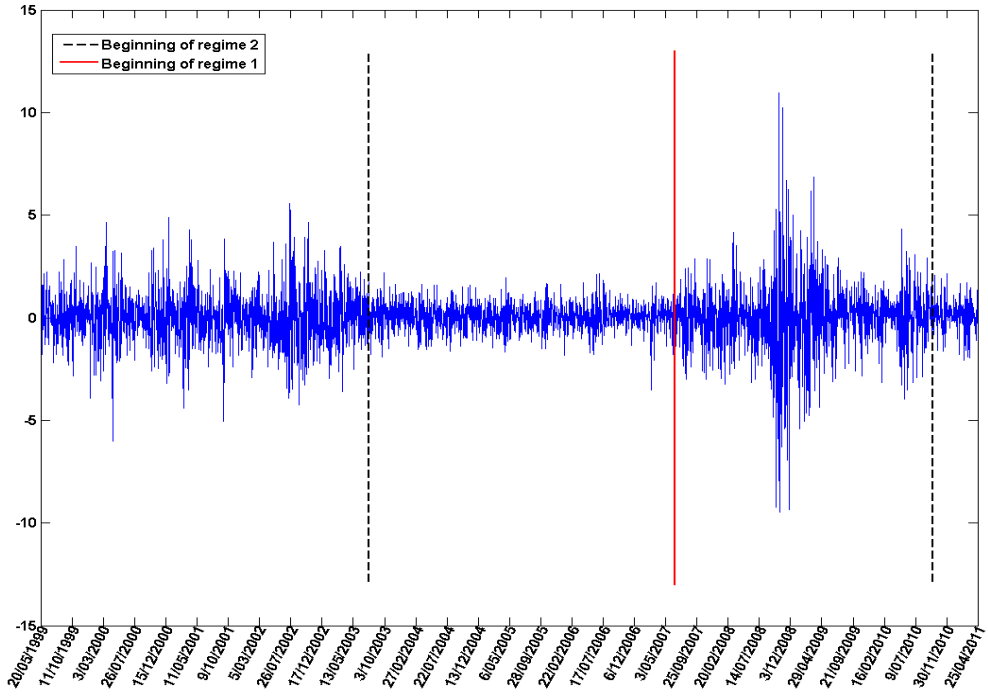


Figure 2.1: S&P 500 index returns with switches from the 2 regime MS-GARCH model

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 ten times for the models in Table 2.6 (up to three regimes) 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. From Table 2.7 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. For the best CP-GARCH model (with three regimes), the difference is 0.40 and 1.08 for the BS and Chib's estimators, respectively.

Table 2.8 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.

Table 2.7: Minima and maxima of ten marginal log-likelihood estimates for S&P 500 data

Regimes	1	2	3
Change-Point			
BS-min	-4505.36	-4506.02	-4503.01
BS-max	-4505.28	-4505.40	-4502.61
Chib-min	-4505.11	-4506.04	-4503.20
Chib-max	-4504.97	-4505.44	-4502.12
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

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 α. 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.

Table 2.8: Posterior means for S&P 500

	GARCH			CP-GARCH			MS-GARCH		
Regime	σ^2	α	β	σ^2	α	β	σ^2	α	β
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 σ^2 is computed as $\omega/(1-\alpha-\beta)$.

The α and β 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 2.2 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 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.

2.5.3.2 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 2.9 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 logarithm of the Bayes factor (log-BF) is smaller than 1, the evidence in favor of the model that has the highest value 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 log-BF values are higher for the MS-GARCH model t han for the CP-GARCH model. The evidence is positive in all cases, and strong in 8 of these cases. The standard one-regime GARCH model has even lower log-BF 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 log-BF than the MS model in two cases (NASDAQ and JPM, with positive evidence), and a lower log-BF in the other cases (with at least positive evidence in 10 cases, and strong in 8 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,

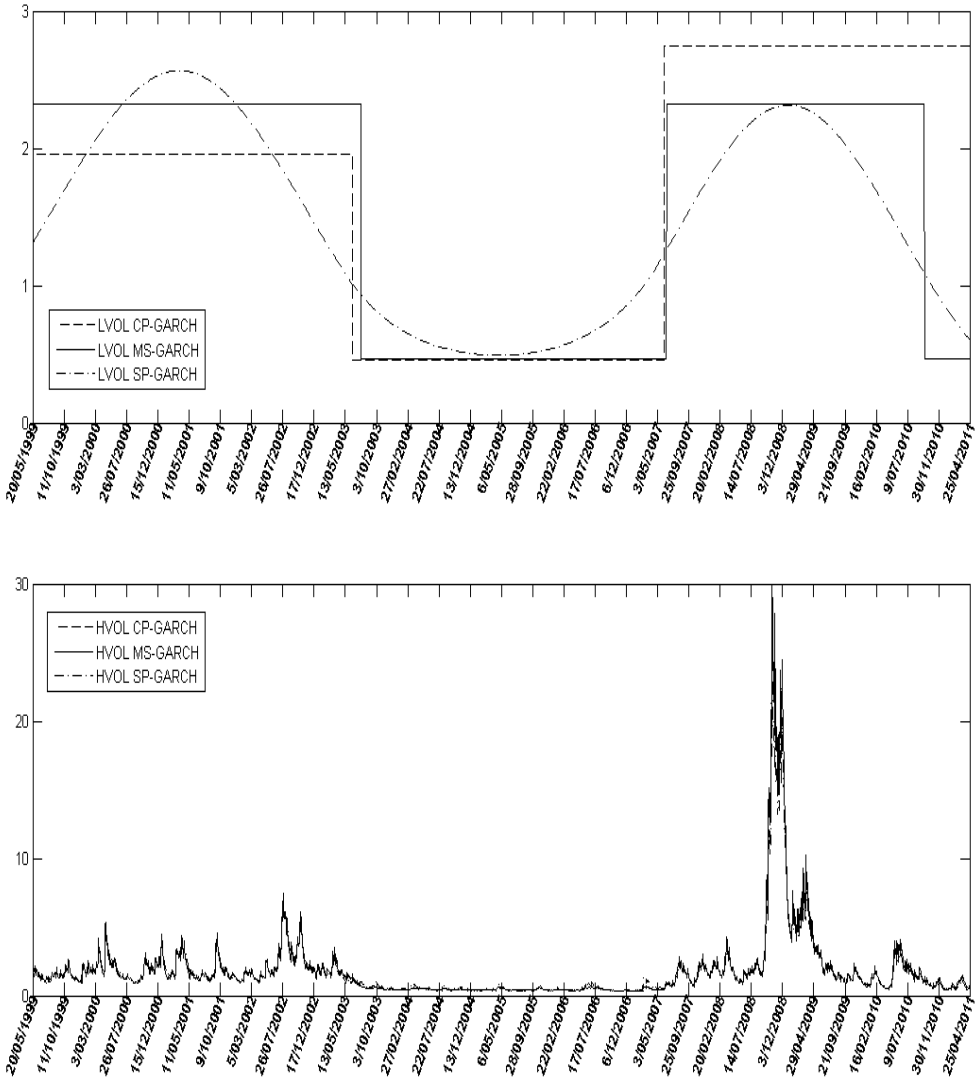


Figure 2.2: Unconditional volatility (top) and conditional volatility (bottom), S&P 500

Table 2.9: Marginal log-likelihoods (MLL) for various time series

Series	Spline-GARCH			GARCH		CP-GARCH			MS-GARCH			
	knots	log-lik	log-BF	log-lik	MLL	K+1	log-lik	log-BF	K+1	log-lik	log-BF	nswitch
S&P 500	3	-4477.36	5.21	-4494.55	-4505.33	3	-4476.70	2.28	2	-4478.59	7.34	3
DJIA	3	-4307.91	2.99	-4322.79	-4333.43	1	-4322.79	0	2	-4307.84	4.7	3
NASDAQ	3	-5404.39	3.20	-5418.99	-5429.84	1	-5418.99	0	2	-5407.31	1.94	7
NYSE	3	-4355.77	2.40	-4369.77	-4380.62	1	-4369.77	0	2	-4355.14	3.91	13
BAC	4	-6088.38	16.62	-6117.49	-6127.39	3	-6036.22	50.12	3	-6004.31	79.49	11
BA	4	-6140.58	9.10	-6163.93	-6174.57	2	-6145.80	8.9	2	-6138.95	11.48	6
JPM	3	-6370.73	8.82	-6388.83	-6400.27	3	-6365.39	5.17	3	-6355.66	7.22	9
MRK	5	-6136.08	48.78	-6198.78	-6209.73	5	-5906.74	215.39	3	-5789.39	335.23	56
PG	4	-4795.84	16.34	-4832.63	-4842.02	3	-4787.49	24.23	2	-4780.89	33.6	9
Metals	2	-5239.80	6.66	-5256.63	-5267.44	2	-5236.85	11.33	2	-5228.87	14.68	5
Yen/Dollar	1	-2966.94	-3.34	-2972.94	-2982.33	1	-2972.94	0	2	-2954.46	3.05	7

log-lik: Maximum of the log-likelihood values over all the MCMC draws; MLL: marginal log-likelihood value computed by Bridge sampling; log-BF: log Bayes factor with respect to the standard GARCH model; 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.

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 2.9 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 -4307.84 while the three regime CP-GARCH has a value of -4307.57.

The number of regimes in the MS-GARCH models varies between two 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 2.9, the maximum number of MS regime switches over the index series is thirteen, while it goes up to fifty-six 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 2.5.4.

2.5.4 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[1], see for example Kass and Raftery (1995) and Sinharay and Stern (2002). Chib's marginal likelihood identity, eq (2.7), particularly underlines the interdependence 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

[1] An alternative to the marginal likelihood is the predictive marginal likelihood since it has the advantage of being less impacted by the prior distribution. However the predictive marginal likelihood only focuses on the predictive ability of the model on a small subset of the observations. Although this quantity is very interesting when we care about forecast ability of different models, it can be misleading for choosing the optimal number of regimes exhibited by a time series. Indeed, the choice should be based on the entire sample.

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 ω, α, and β. The other two priors use Gaussian distributions on the GARCH parameters transformed to the real line. Prior 2 is the one used in the Chapter so far, and Prior 3 is much more diffuse, see Table 2.10 for details. Each prior differently penalizes the marginal likelihood. The uniform distribution on finite (small) intervals hardly imposes any penalty for an extra regime. In contrast, Prior 3 strongly decreases the marginal likelihood if an additional regime is imposed.

Table 2.10: Hyperparameters for the prior distributions

Prior 1 : GARCH parameters (θ)	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 (θ)	**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 (θ)	**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 $\beta \in]0, 1[\to \ln(\frac{\beta}{1-\beta})$.

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 2.11. 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 2.11 we give a more visual illustration of the sensitivity of the

Table 2.11: 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.

MLL with respect to the prior. In particular, we focus on Prior 1 since only the interval of ω 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 (2.7), 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_ω increase from 10 to 1000 while keeping the lower bound a_ω fixed to 0. As can be seen in Figures 2.3 and 2.4, 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 ω has to be quite high (and unrealistic for the type data we analyze) for the changes to occur. For example in the CP-GARCH case, the smallest value is already 33 (NASDAQ) and it becomes even as large as 642 (DJIA). Thus this type of sensitivity is not a source of worry.

Finally, we study the effect of the prior on P in the same way as for the GARCH parameters above. We change the hyperparameter in Table 2.1 from 1110.11 to 500 and

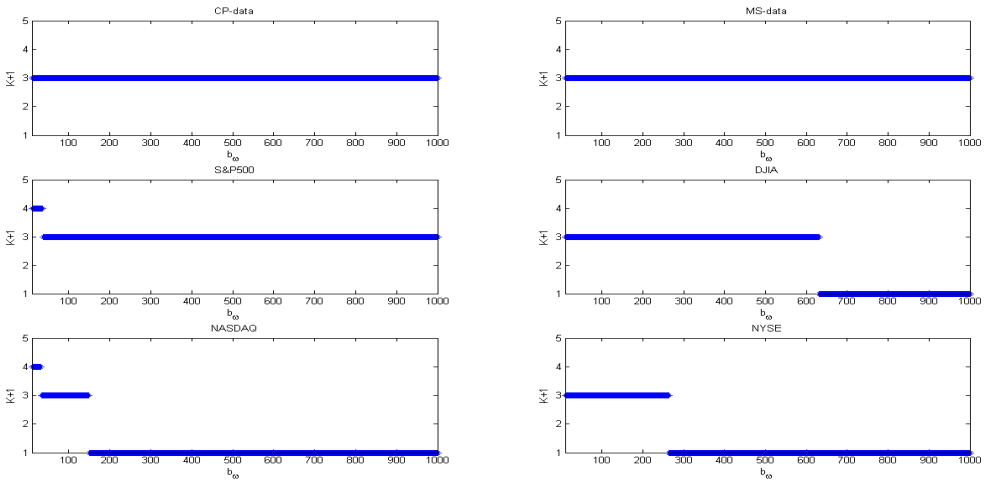


Figure 2.3: CP-GARCH: Model selection with respect to the upper bound b_ω

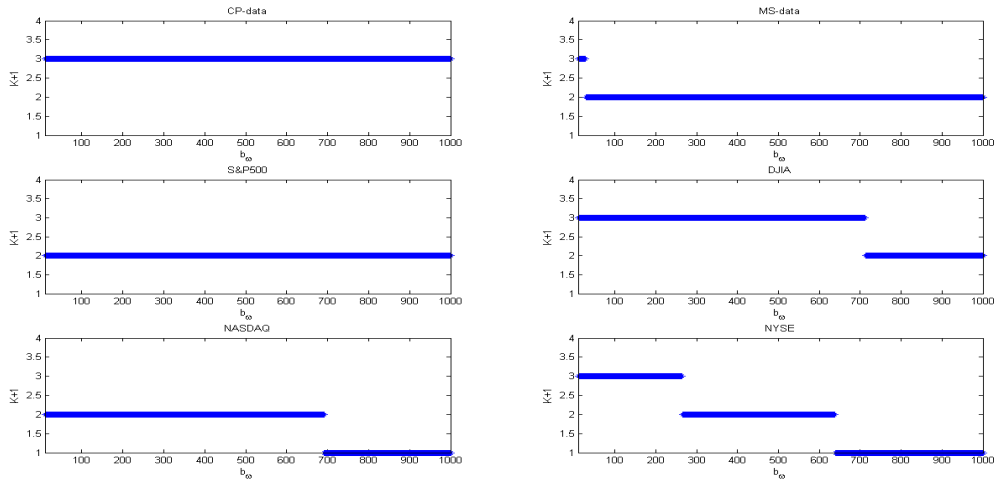


Figure 2.4: MS-GARCH: Model selection with respect to the upper bound b_ω

1500. The corresponding expected durations of staying in a given regime are respectively 501, 1111 and 1501 days. The results, not reported here, show for all series very small differences (all smaller than one) in the log marginal likelihood values over these priors, and this for both the MS- and CP-GARCH models. The optimal models, as selected by the highest marginal likelihood, stay the same.

2.5.5 Performance of the PMCMC sampler

In this subsection, we provide more details on the numerical properties of our sampler. First, we compare the PMCMC sampler that draws the states jointly with a sampler that draws states individually. Second, we investigate autocorrelation times for different SMC algorithms. Finally, we show the quality of the approximation that we use in the backward step of our algorithm.

2.5.5.1 Sampling states jointly versus individually

As mentioned in the introduction, the main alternative approach for a fixed number of regimes is the Gibbs sampling algorithm proposed by Bauwens, Preminger, and Rombouts (2010), denoted BPR hereafter, for the MS-GARCH model. That algorithm samples the state variables individually, whereas in our new algorithm, called particle Gibbs sampler or particle MCMC (PMCMC), they are sampled jointly. This makes substantial differences as we illustrate below. For both the MS-GARCH and CP-GARCH) models, we compare the PMCMC and BPR algorithms for drawing the state variables in two ways. For the sake of a fair comparison, the BPR and PMCMC algorithms differ only in the way the sampling of the states is implemented.

First, we compute the autocorrelation time, i.e. the number of draws from the MCMC to obtain one effectively new independent draw from the posterior distribution, and defined as $1 + 2\sum_{i=1}^{\infty} |\rho_i|$ where ρ_i is the autocorrelation of order i between the posterior draws. Table 2.12 reports for the BPR and PMCMC algorithms the autocorrelation times for the optimal MS-GARCH and CP-GARCH models estimated on the simulated CP-GARCH data and S&P 500 data of Section 2.5. From this, it is clear that the PMCMC autocorrelation time is in each comparison drastically lower than the BPR autocorrelation time.

Table 2.12: Autocorrelation times for the BPR and PMCMC algorithms

CP-GARCH data				
Break	CP-GARCH		MS-GARCH	
	BPR	PMCMC	BPR	PMCMC
1	434.6325	1.0796	461.2361	1.5284
2	65.4907	1.1681	430.1539	1.4501
S&P 500 data				
Break	CP-GARCH		MS-GARCH	
	BPR	PMCMC	BPR	PMCMC
1	319.8695	1.5309	327.7704	1.2257
2	443.2646	1.6758	—	—

Autocorrelation time computed by batch means, see e.g. Geyer (1992). The BPR and PMCMC algorithms differ only in the way the sampling of the states is implemented (i.e. the block of the GARCH parameters is fixed).

As second comparison, we summarize in Table 2.13 the performance of the samplers in terms of CPU time (in minutes) to obtain one effectively new independent posterior draw. Related to the CP-GARCH simulated data, we see that for all the CP-GARCH models, both the PMCMC and the BPR algorithms have similar performance. Hence, for a fixed number of breaks, the BPR algorithm is sufficient for CP-GARCH models. However, BPR is unable to compute the marginal likelihood. For the MS-GARCH models, the difference in performance is substantial. For example, for three regimes the CPU times for PMCMC and BPR are respectively 0.0661 and 30.5090 minutes. The conclusions related to the S&P 500 data in the second panel of Table 2.13 are the same as for the simulated data.

2.5.5.2 Other SMC algorithms

The idea of Whiteley, Andrieu, and Doucet (2011) of working with break dates as parameters can lead to another algorithm in the context of the CP-GARCH model considered in this Chapter. Consider the variable $\Delta_T = \{\delta_1, \delta_2, \ldots, \delta_T\}$ with δ_t indicating the duration of the most recent regime up to time t. We have the usual decomposition

$$p(\Delta_T \mid Y_T) = p(\delta_T \mid Y_T)p(\delta_{T-1} \mid Y_T, \delta_T) \ldots p(\delta_1 \mid Y_T, \Delta^2), \tag{2.12}$$

Table 2.13: CPU time in minutes to obtain one effective posterior draw

	CP-GARCH simulated data			
Regimes	CP-GARCH		MS-GARCH	
	PMCMC	BPR	PMCMC	BPR
2	0.0208	0.0288	0.0391	17.2207
3	0.0257	0.0489	0.0661	30.5090
4	0.0768	0.0602	0.0723	39.5739
	S&P 500 data			
Regimes	CP-GARCH		MS-GARCH	
	PMCMC	BPR	PMCMC	BPR
2	0.0165	0.0413	0.0666	18.8514
3	0.0168	0.0455	0.0994	27.7195
4	0.0169	0.0540	—	—

Computation time in minutes per effective posterior draw.

of which the typical term is

$$p(\delta_t \mid Y_T, \Delta^{t+1}) = \frac{f(Y_t)p(\delta_t \mid Y_t)f(Y^{t+1} \mid Y_t, \Delta^t)p(\Delta^{t+1} \mid \delta_t)}{p(Y_T, \Delta^{t+1})} \tag{2.13}$$

$$\propto p(\delta_t \mid Y_t)f(Y^{t+1} \mid Y_t, \Delta^t)p(\delta_{t+1} \mid \delta_t). \tag{2.14}$$

The factor $p(\delta_t \mid Y_t)$ can be computed with a forward step as follows:

$$p(\delta_t \mid Y_t) \propto f(y_t \mid Y_{t-1}, \delta_t) \int p(\delta_t \mid \delta_{t-1})p(\delta_{t-1} \mid Y_{t-1})\, d\delta_{t-1}. \tag{2.15}$$

The advantage of working with durations δ_t instead of the state vector is that Equation (2.14) does not need to be evaluated at each t, but only for $\delta_{t+1} = 1$. This computational advantage allows one to evaluate exactly $f(Y^{t+1} \mid Y_t, \Delta^t)$ for small N and small number of regimes. On the other hand, the support of δ_t, i.e. $(1, 2, \ldots, t-1)$, implies using more particles compared to working with the states s_t where the support is only $(1, \ldots K+1)$.

As an illustration, we implement this "duration" algorithm for the simulated CP-GARCH data of Section 2.5. For completeness, we also adapt the particle Metropolis-Hastings (PMH) and the ADH particle Gibbs (PGibbs) algorithms of Andrieu, Doucet, and Holenstein (2010) for the CP-GARCH models. The latter two algorithms have no backward step. Table 2.14 displays the autocorrelation times for the PMH, PGibbs and the "duration" algorithms for 150 and 1000 particles, and where only the states are sam-

pled. For the sake of comparison, we also add our PMCMC algorithm. We see clearly that using 150 particles for these algorithms result in draws which are more persistent, especially so for the PGibbs algorithm. When the number of particles is increased to 1000, all the algorithms perform quite similarly.

Table 2.14: Autocorrelation times for different SMC algorithms

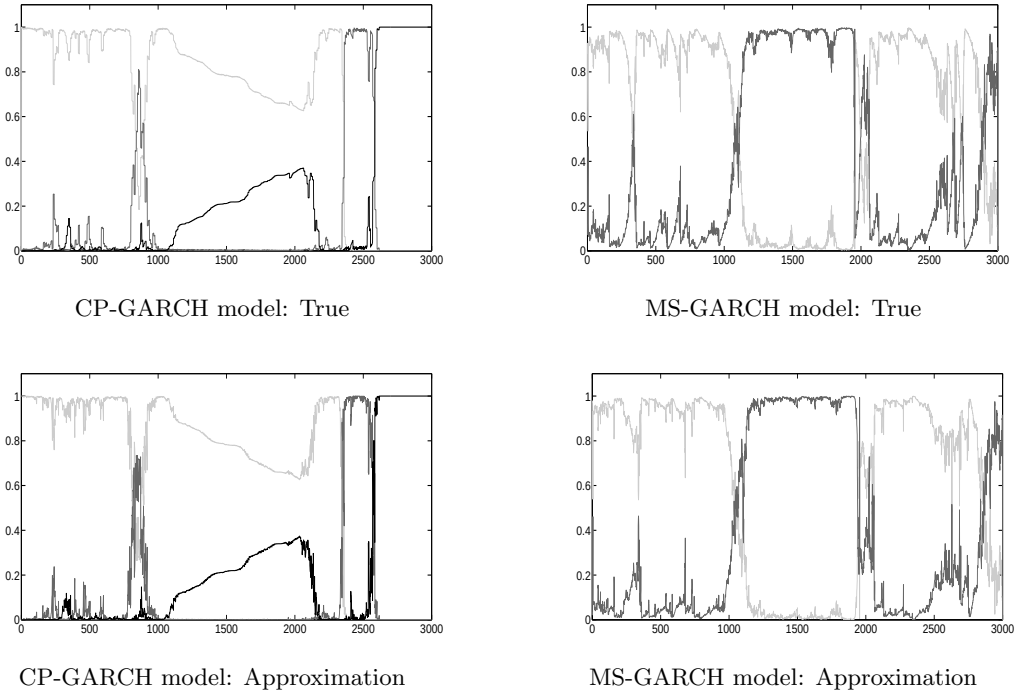
	PMH		PGibbs		Duration		PMCMC	
Break/N	150	1000	150	1000	150	1000	150	1000
1	5.70	3.72	53.12	2.34	7.15	2.36	1.08	1.09
2	3.61	1.66	8.39	1.72	8.73	1.94	1.25	1.17

Auto-correlation times using the simulated CP-GARCH data.

Note that the performance of the different algorithms depends on the resampling procedure (multinomial, stratified, etc.). Moreover, as mentioned by Fearnhead in the discussion of Andrieu, Doucet, and Holenstein (2010), the mixing properties of the algorithm can be improved by decreasing the number of resampling steps in the conditional SMC procedure. In our comparisons of the algorithms, we have used identical resampling steps so that the algorithms only differ in the way the states are drawn.

2.5.5.3 Quality of the approximation

We explain in Section 2.2.1 that $\lambda_t^{s_t^i}$ has to be approximated for computational reasons. The accuracy of this approximation is illustrated here for the S&P 500 data, though results for the other series are very similar. Figure 2.5 presents the average of 100 draws of $\lambda_{s_t}^{b_t^k}$ for the true values and the approximate values. The corresponding computing time to obtain the 100 draws is 500 and 0.6 minutes, respectively. The approximation is visually accurate.



CP-GARCH model: True

MS-GARCH model: True

CP-GARCH model: Approximation

MS-GARCH model: Approximation

Figure 2.5: One possible Markov-switching and Change-point paths that govern the dynamic of the model parameters.

2.6 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.

Thanks to the particle MCMC approach, any high dimensional conditional distribution appearing in an MCMC scheme could potentially be approximated by a (conditional) Sequential Monte Carlo algorithm. It therefore opens the possibility of specifying complex models that mimic what we observe in time series nowadays. For instance, further research could be devoted to GARCH models that allow for smooth time varying parameters subject to some abrupt switches. Their estimation though challenging could be carried out by particle MCMC methods.

Appendix: Sampling the conditional variance parameters

This appendix describes the sampling of θ 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 2.15 where μ and Σ respectively stand for the posterior mean and the posterior variance-covariance matrix estimated using available draws, I denotes the identity matrix and θ_{cur} is the current parameter of the particle Gibbs.

Table 2.15: 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 posterior density. 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 (θ, 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 $(\theta, P, S_T)^i = \tau_i(\theta, P, S_T)$.

3

A Bayesian Method of Change-Point Estimation with Recurrent Regimes: Application to GARCH Models[1]

3.1 Introduction

This Chapter develops a Discrete-DREAM (D-DREAM hereafter) algorithm to sample the break dates (i.e. structural breaks) separatly from the model parameters in a Metropolis algorithm. The algorithm is said to be "block-at-a-time" since we sample first the parameters of the model (which constitute the first block) and then the break dates (second block). Although both blocks could be sampled using a simple Random Walk Metropolis (RWM) algorithm, we resort to slightly more elaborated techniques to highly improve the MCMC mixing properties. The D-DREAM algorithm is a combination of the DREAM algorithm of Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009), which can sample parameters having a continuous support, and a discretised version of it, which can sample parameters having a discrete support.[2] The D-DREAM algorithm has the advantage to accommodate path dependence issues since break dates are treated as parameters and not anymore subject to a hidden Markov chain of transition probabilities, as in Chib (1996).

Although our main goal is to estimate Change-Point models, we can also deal with recurrent regimes. More precisely, in this setup we can easily allow for the regimes to

[1] Joint work with Luc Bauwens and Bruno De Backer. This Chapter is a significant revision of Bauwens, Dufays, and De Backer (2011).

[2] The $DREAM_{(D)}$ algorithm of Vrugt and ter Braak (2011) is an example of discretised version of the DREAM algorithm. In that respect, it should be noted that we worked independently from Vrugt and ter Braak (2011) but reached the same MCMC algorithm.

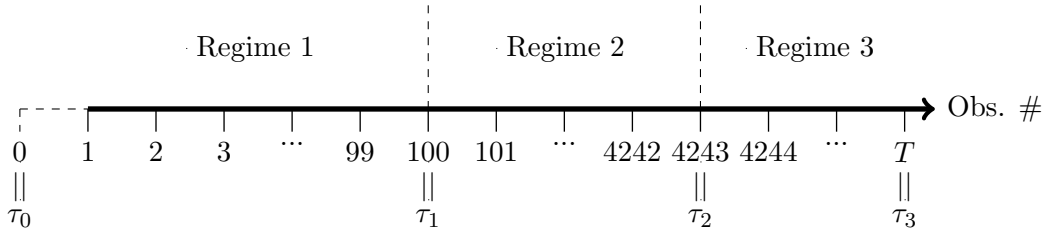
be active in several distinct periods in the sample. In short, it suffices to constrain the model parameters of non-subsequent regimes to be the same and to sample the two blocks as usual. Consequently, a contribution of this Chapter is to provide a framework for the analysis of processes halfway between Change-Point and Markov-switching specifications. The marginal likelihood criterion can be used to decide on the unknown number of breaks as well as on whether returning regimes are present in the sample, or whether the data are closer to pure Change-Point models.

The Chapter is organized as follows. In section 3.2, we present our algorithm in the context of Change-Point GJR-GARCH models. In section 3.3, we show how to deal with recurrent regimes in our framework. In section 3.4, we explain how to determine the unknown number of breaks and unknown ordering of regimes by computing Bayes factors. Simulations show the efficiency of the algorithm in section 3.5. In section 3.6, we apply the algorithm to daily data for the S&P 500 index and other financial time series. Section 3.7 shows that the detection of break dates and recurrent regimes can be useful in forecasting exercises, though further research is needed in this topic. Section 3.8 concludes.

3.2 D-DREAM Algorithm

We denote by $Y_T = (y_1, ..., y_T)'$ a time series of T observations. Instead of considering a $T \times 1$ vector of states as in Chib (1996), we focus on a smaller vector of break dates $\bar{\Gamma} = (\tau_0, \tau_1, \tau_2, ..., \tau_K, \tau_{K+1})'$, where $\tau_0 = 0$, $\tau_{k-1} < \tau_k$, $\tau_{K+1} = T$, and the number of break dates K is considered known for the moment. For instance, the break date τ_2 indicates the end of regime 2. The break date parameters hence constitute a discrete vector. Figure 3.1 illustrates how the break dates split a time series of length T into 3 regimes when $K = 2$.

Figure 3.1: Example of a time series with 3 regimes



For simplicity, we present the D-DREAM algorithm by considering a Change-Point GJR-GARCH(1,1), or CP-GJR-GARCH(1,1), specification though the algorithm could easily be adapted to many other time series models. A CP-GJR-GARCH(1,1) model is characterized by the following set of equations:

$$\begin{aligned} y_t &= \mu_k + \epsilon_t, \\ \epsilon_t &= \sigma_t \eta_t, \ with \ \eta_t \sim i.i.d. \, N(0,1), \\ \sigma_t^2 &= \omega_k + \alpha_k \epsilon_{t-1}^2 + \gamma_k \epsilon_{t-1}^2 \mathbb{1}_{\epsilon_{t-1}<0} + \beta_k \sigma_{t-1}^2, \end{aligned}$$

if $t \in]\tau_{k-1}, \tau_k]$, with $k = \{1, 2, ..., K+1\}$ denoting the regime. We constrain some parameters spaces to accelerate the sampling from the posterior distributions but we allow for large supports so as not to bias our results. More specifically, we assume that, $\forall k$, $\mu_k \in \Re$, $\omega_k \in [0, 10]$, α_k and $\gamma_k \in [-0.2, 1]$, $\beta_k \in [0, 1.2]$. We do not impose any stationarity condition. To ensure the positivity of the variance, the model is slightly modified in a way to generate very small (unrealistic) positive variance when the parameters should lead to

a negative variance. By doing so, parameters implying a negative variance are rejected without having to truncate the prior distribution. We detail our priors in subsection 3.2.4.

Let $\Theta = (\theta_1, \theta_2, ..., \theta_{K+1})'$, where $\theta_k = (\mu_k, \omega_k, \alpha_k, \gamma_k, \beta_k)$ contains the parameters of regime k, denote the set of parameters characterising the CP-GJR-GARCH(1,1) model. Our MCMC scheme simply consists in drawing iteratively from the conditional posterior of the model parameters Θ, $\pi(\Theta|Y_T, \Gamma)$, and from the conditional posterior of the break dates $\Gamma = (\tau_1, \tau_2, ..., \tau_K)'$, $\pi(\Gamma|Y_T, \Theta)$. Note that these two conditional densities have no closed form.

3.2.1 Sampling the Model Parameters Θ

The full conditional distribution of Θ does not have a closed form but we can nevertheless resort to a Metropolis-Hastings algorithm to obtain draws from it. However, in a Metropolis-Hastings algorithm convergence towards the limiting distribution can be very slow if the proposal distribution is not appropriate. Therefore, Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009) propose to run multiple chains in parallel and let these chains interact to propose new values for the parameters, without resorting to a fixed proposal distribution. They show that their Differential Evolution Adaptative Metropolis (DREAM) algorithm significantly improves the efficiency of the MCMC simulations, compared with a random walk Metropolis algorithm and other adaptative schemes, by automatically tuning the scale and orientation of the proposal distribution. They show in particular that the DREAM algorithm is especially useful for high-dimensional and multi-modal problems.

Following the DREAM algorithm, the proposal of a new location for a given chain is constructed from the current locations of other chains. Focusing on the i^{th} chain ($i = 1, 2, ..., N$) at the $j + 1^{th}$ MCMC iteration, the procedure can be described in 2 steps:

1. Propose a new Θ_i^{t+1}, denoted by Z_i hereafter, according to the following rule (omitting superscripts referring to iterations for simplicity):

$$Z_i = \Theta_i + \gamma(\delta, K)(\sum_{g=1}^{\delta} \Theta_{r_1(g)} - \sum_{h=1}^{\delta} \Theta_{r_2(h)}) + \zeta, \tag{3.1}$$

with $\forall g, h = 1, 2, ..., \delta$, $i \neq r_1(g)$, $r_2(h)$; $r_1(.)$ and $r_2(.)$ stand for random inte-

gers uniformly distributed on the support $[1, N]_{-i}$ and it is required that $r_1(g) \neq r_2(h)$ when $g = h$; and $\zeta \sim N(0, \eta_\Theta^2 I)$. Notice that if $\gamma(\delta, K) = 0$ the algorithm boils down to a RWM scheme. However, since we run multiple chains in parallel, we take advantage of the locations of other chains by introducing the term $\gamma(\delta, K)(\sum_{g=1}^{\delta} \Theta_{r_1(g)} - \sum_{h=1}^{\delta} \Theta_{r_2(h)})$ and ζ becomes a small perturbation term present to ensure convergence of the algorithm. Consequently, η_Θ^2 is a small value compared with the width of the target distribution and we set $\gamma(\delta, d) = \frac{2.38}{\sqrt{2\delta d}}$, where d is the number of elements in Θ_i (and is equal to $5(K+1)$ in the case of GJR-GARCH(1,1) models).[1] Note that the fixed number of pairs of chains, δ, must be smaller than or equal to $\frac{N-1}{2}$.

2. Accept the proposed Z_i according to the acceptance probability ratio of the Metropolis algorithm $\alpha(\Theta_i, Z_i) = \min\{\frac{f(Z_i|\Gamma_i, Y_T)}{f(\Theta_i|\Gamma_i, Y_T)}, 1\} = \min\{\frac{f(Y_T|Z_i, \Gamma_i)p(Z_i|\Gamma_i)}{f(Y_T|\Theta_i, \Gamma_i)p(\Theta_i|\Gamma_i)}, 1\}$.

It is important to realize that, in large dimensional spaces, it might not be optimal to update the entire vector of parameters as a whole. Therefore, we choose in the simulations and applications below to randomly break the large vector Θ_i into sub-vectors of smaller dimensions at each MCMC iteration and to apply the two steps above to each sub-vector in turn with the appropriate conditioning parameters. As initially proposed by Chib and Ramamurthy (2010), randomly choosing the size of the block that will be updated has several advantages over a one-block strategy such as avoiding some circumstances where the MCMC algorithm can be stuck and allowing for more efficient exploration of the posterior support. However these improvements come at a price since for each block, a Metropolis ratio must be computed. It therefore increases the computational time. As a last remark, notice that since the support of proposal (3.1) is not constrained, we have to map the constrained space of the model parameters into the real space.[2]

3.2.2 Sampling the Break Dates Γ

The sampling procedure of the break date vector Γ is similar to the sampling of Θ and can be described in the same two steps. The only difference is that we need to round the

[1] We follow Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009) for the choice of $\gamma()$. Note that if $\delta = 1$, then $\gamma(d) = \frac{2.38}{\sqrt{2d}}$ is optimal for normal posteriors as regards acceptance rates for $N \to \infty$, as shown in ter Braak (2006).

[2] If $x \in [x_{min}, x_{max}]$, $log(\frac{x - x_{min}}{x_{max} - x})$ in $\Re$.

elements of the proposed vector of break dates Z_i so as to have integers. Proposal (3.1) then becomes:

$$Z_i = \Gamma_i + round[\gamma(\delta, d')(\sum_{g=1}^{\delta} \Gamma_{r_1(g)} - \sum_{h=1}^{\delta} \Gamma_{r_2(h)}) + \xi], \tag{3.2}$$

with $round[.]$ meaning that we take the nearest integer and $\xi \sim N(0, \eta_\Gamma^2 I)$.

Proposition 1 ensures the convergence and ergodicity of the chains in the D-DREAM algorithm.

Proposition 1. Consider a Change-Point model with a given number of breaks K. Then, $\forall i = 1, ..., N$, let $X_i = (\Theta_i', \Gamma_i')'$. The D-DREAM algorithm yields a Markov chain that is ergodic with unique stationary distribution, such that

$$(X_1, ..., X_N) \sim \tilde{\pi}(X_1, ..., X_N | Y_T) = \pi(X_1|Y_T) \times ... \times \pi(X_N|Y_T).$$

Proof. See appendices.

Why does the D-DREAM algorithm work ? Following the ergodicity proof of the D-DREAM algorithm detailed in the appendices, we have that the stationary distribution of the generated Markov-Chain (by the MCMC algorithm) factorizes into N independent posterior distributions of interests where N is the number of parallel MCMC chains. At each MCMC iteration, the states of the individual chains (X_i, $\forall i = 1, ..., N$) are therefore independent. For the sake of clarity, let us focus on the continuous parameter Θ and assume that the first two posterior moments exist and are denoted by $\bar{\mu} = E(\Theta|Y_T)$ and $\bar{\Sigma} = \text{Cov}(\Theta|Y_T)$. According to (3.1), once all the independent chains have converged to the posterior distribution, we have on average that the current parameter of any independent chain tends to $\bar{\mu}$ and the difference $(\sum_{g=1}^{\delta} \Theta_{r_1(g)} - \sum_{h=1}^{\delta} \Theta_{r_2(h)})$ converges to $2\delta\bar{\Sigma}$. Since the variance parameter η_θ^2 is extremely small, for large N, the proposal (3.1) looks like $Z_i = \Theta_i + \gamma(\delta, K)\eta$ with $\eta \sim N(0, 2\delta\bar{\Sigma})$. The DREAM algorithm thus behaves like a standard random walk (RW) algorithm and RW theory can be used to optimally set the user-defined value $\gamma(\delta, K)$. Following Roberts and Rosenthal (2001), the parameter $\gamma(\delta, K)$ is set to $\frac{2.38}{\sqrt{2\delta d}}$.

3.2.3 Advantages of the D-DREAM Algorithm

In addition to accommodating path dependence issues and allowing for recurrent regimes (see section 3.3), the D-DREAM sampling scheme has some additional practical advantages compared with existing methods that sample break dates. First, the D-DREAM algorithm is relatively simple due to the fact that we consider break dates directly as parameters and not a more complex structure entailing a hidden Markov chain for the probabilities of jumping to the next regime. Consequently, the D-DREAM algorithm is a mere Metropolis sampling procedure which does not require the use of filtering methods such as the forward-backward scheme as in Chib (1996) or the PMCMC algorithm of Chapter 2.

Second, the computational time needed to simulate posterior draws using the D-DREAM algorithm is dramatically cut compared with the other papers cited above. Although we show in Chapter 2 that the PMCMC scheme is as fast as the MCMC method of Bauwens, Preminger, and Rombouts (2010) when considering effective posterior draws, their algorithm consists of sequential Monte Carlo simulations run at every iteration of an MCMC scheme. To illustrate, the S&P 500 simulation in section 3.6.1 takes 15 minutes to be screened by the D-DREAM algorithm and 15 hours by the PMCMC algorithm.[3]

Third, our framework is more flexible than existing models as any distribution can be assumed for the break date parameters. For instance, a Geometric distribution could be chosen as in Chib (1996), or a Poisson distribution as in Koop and Potter (2007). Moreover, our framework also allows for hierarchical prior distributions for the GARCH parameters. One could specify, for instance, a common distribution for all regimes as in Pesaran, Pettenuzzo, and Timmermann (2006) or an evolving distribution depending on the parameters of the previous regime as in Koop and Potter (2007).

Fourth, the D-DREAM algorithm estimates multiple break points simultaneously, as opposed to a "one-at-a-time" (or sequential) estimation strategy. Stephens (1994b), who extends the work of Carlin, Gelfand, and Smith (1992b) to multiple change points, considers such a sequential strategy. The disadvantage of sequential strategies is that the

[3]The computer is an AMD Athlon 2Ghz with four cores.

locations of the breaks might not coincide in practice with the locations of the breaks when estimated simultaneously (for further discussions from a frequentist point of view, see also Bai and Perron 1998, 2003). Besides, as Chib (1998, p.224) puts it:

> The first [weakness of this approach] arises from the fact that the change points [...] are simulated one at a time from the [...] full conditional distributions [...] rather than from the joint distribution [...]. Sampling the latter distribution directly, say through a Metropolis-Hastings step, is not practical because of the difficulty of developing appropriate proposal generators for the change points. Liu, Wong, and Kong (1994) in theoretical work and Carter and Kohn (1994), Chib and Greenberg (1995a) and Shephard (1994) in empirical work have shown that the mixing properties of the MCMC output is considerably improved when highly correlated components are grouped together and simulated as one block. Conversely, MCMC algorithms that do not exploit such groupings tend to be slow to converge.

In that respect, note the D-DREAM algorithm samples the break dates as one block and that a symmetric generic proposal distribution is automatically generated by the chain interactions. Note also that the simplicity of the D-DREAM sampling procedure follows a structure that needs to be coded only once.

3.2.4 Priors, Starting Points, and Burn-in

We choose rather standard and uninformative priors for the model parameters, as shown in Table 3.1. The prior for the μ's parameters, the mean of the series in the different regimes, is a normal distribution centered at 0 and with variance 10 and the priors for the GARCH parameters are uniform distributions.[4] Regarding the break dates, we choose a prior that complies with the modelling constraints. More precisely, the chronological order of the break dates needs always to be respected and, for any $k \in \{1, 2, ..., K\}$, no break date can lie outside the interval $[1, T-1]$. We work with the regimes durations, defined $\forall k \in [1, K+1]$ as $d_k = \tau_k - \tau_{k-1}$, and choose independent negative binomial prior

[4]Since the D-DREAM algorithm takes draws from the posteriors of the parameters mapped into the real space $\Re$, we need to take into account the Jacobian of the mapping function when evaluating the prior density at a given point.

distributions, as in Koop and Potter (2007).[5]

Table 3.1: Prior Distributions

	CP-GJR-GARCH(1,1) Parameters	
$\forall k \in [1, K+1]$	$\mu_k \sim N(0, 10)$	$\omega_k \sim Uniform[0, 10]$
	α_k and $\gamma_k \sim Uniform[-0.2, 1]$	$\beta_k \sim Uniform[0, 1.2]$
	Duration Parameters	
$\forall k \in [1, K]$	$d_k = \tau_k - \tau_{k-1} \sim NB_{tr}(\frac{5(T-\tau_{k-1})}{2(T-\tau_{k-1}-5)}, 1 - \frac{5}{T-\tau_{k-1}})\ \mathbb{1}_{[1, T-\tau_{k-1}-(K+1)+k]}$	
	$d_{K+1} = T - \tau_K$	

Table 3.1. Priors for the regime durations imply that the chronological order of the break dates is always respected and that no break date lie outside the interval [1,T-1], otherwise the prior is zero and the draw of the break date vector Γ is rejected. If a drawn break date is not the final one it will leave enough space for the other break dates before the end of the sample. NB_{tr} denotes the truncated negative binomial distribution. The parameters r and p are the usual parameters of the negative binomial distributions (not truncated) such that, $\forall k \in [1, K]$, $E[d_k] = \frac{T-\tau_{k-1}}{2}$ and $Var[d_k] = \frac{(T-\tau_{k-1})^2}{10}$.

The starting point of an MCMC algorithm is another relevant practical issue since choosing a highly likely point with respect to the posterior distribution ensures faster convergence. Typically, MCMC algorithms start at the MLE but it cannot be readily computed in the case of structural breaks.[6] Therefore, in order to find good starting values, we randomly generate many break date vectors Γ and draw many sets of parameters θ_k's for each generated break date vector. The initial points of the N chains are the N randomly generated vectors of break dates and associated parameters with the highest global likelihoods.

We determine the length of the burn-in period using the criterion in Gelman and Rubin (1992) ($\hat{R} < 1.2$ for all parameters). At each iteration of the burn-in phase we replace outlier chains by the chain with the highest likelihood (plus a small random term) to ease convergence. Outlier chains are defined similarly as in Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009) as the ones with a data log-likelihood smaller than the third quartile minus 2 times the inter-quartile range (Q3-Q1).

[5] Although the D-DREAM algorithm draws from the posteriors of the break dates, we can directly evaluate the prior density of durations since the Jacobian of the transformation is one.

[6] See Augustyniak (2013) for more on MLE estimation of structural break GARCH models.

3.3 Dealing with Recurrent Regimes

The D-DREAM algorithm as described up to now can only deal with Change-Point specifications. Hence, a natural extension is to allow for recurrent regimes, like in a Markov-switching framework. Therefore, we consider a permutation variable λ which is a $(K+1)\times 1$ vector indicating which regime is active between two break dates. More precisely, at each MCMC iteration all possible combinations in λ are considered and the algorithm draws randomly one combination (section 3.4 will detail the drawing procedure).[7] As an example, assume that the number of regimes is four, all possible λ's, generally denoted by $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, are as follows: (1,2,3,4), (1,2,1,4), (1,2,3,1), (1,2,3,2), (1,2,1,2).[8] The first possibility (1,2,3,4) is a pure Change-Point specification whereas the others are recurrent regime specifications. For instance, in the case of the combination $(1, 2, 1, 4)$ the first regime is said to be recurrent as it is active before the first break date and between the second and third break dates.

The introduction of the λ parameter somewhat complicates the sampling of the model parameters Θ. Imagine that at a given iteration t we draw a combination λ^t from its full conditional density $\pi(\lambda^t|Y_T, \Theta^{t-1}, \Gamma^{t-1}) \propto f(Y_T|\Theta^{t-1}, \Gamma^{t-1}, \lambda^t)p(\Theta^{t-1}, \Gamma^{t-1}|\lambda^t)p(\lambda^t)$ and propose a whole set of parameters Θ^t (and break dates Γ^t) as described in section 3.2, that is, without yet taking into account that there might be recurrent regimes. If λ^t implies, for instance, that the third regime is actually the same as the first, then the data likelihood does not depend on set of parameters θ_3^t and the proposal (Θ^t) could be accepted while the model parameters of the inactive third regime are unrealistic. Therefore, following this strategy, the sampling of the set of model parameters of likely inactive regimes is highly inefficient, and Bayes factors will put too much weight on specifications with recurrence with respect to pure Change-Point specifications if the true specification includes recurrent regimes.

An obvious alternative sampling strategy is to "freeze" the model parameters of inactive regimes until the next iteration where the parameters are actually submitted to the

[7] Sampling directly from the parameter λ is much faster than running the algorithm considering all possible combinations one by one and then computing marginal likelihoods.

[8] All other combination possibilities would imply label switching problems. For example, the combination $(1, 3, 1, 3)$ is by no way different from labeling the regimes $(1, 2, 1, 2)$.

acceptance/rejection step of the Metropolis algorithm. In other words, we could keep the model parameters of inactive regimes at their previous values. Although this improves dramatically the mixing efficiency of the algorithm, the MCMC is no longer ergodic due to the absorbing state.

By introducing the combination parameter λ, and by showing in the next section how to accurately obtain draws from its posterior, we hope to provide a formulation that encompasses both Change-Point and Markov-switching models.

3.4 Model Selection

From Section 3.2 and 3.3, assuming that the number of lags in the CP-GJR-GARCH model is taken as given, the algorithm still requires the user to determine the number of breaks, K, and the combination of regimes, λ, which are often unknown. In a purely Bayesian approach, we need to calculate the probabilities of the CP-GJR-GARCH models for any values of K and λ. In this section, we first consider the case of a given number of breaks K, and we show how to compute Bayes factors of the models differing by the ordering of the regimes by sampling directly the combination parameter λ. The sampling technique used is based on Carlin and Chib (1995). In a second step, we show how to compute the marginal likelihood of the pure Change-Point models. Eventually, we are able to recover the marginal likelihoods of all models differing by the number of breaks or combination of regimes.

3.4.1 Bayes Factors to Determine the Recurrence of Regimes

Let's first asssume that the number of breaks is given, so that we can focus our attention on the combination of regimes. Then, we only need to sample from the combination parameter λ. Recall, however, that inactive regimes imply a dimensional problem in the sampling of the model parameters Θ (models with more recurrent regimes need less parameters than models with less recurrent regimes or no recurrent regime at all). Nevertheless, Carlin and Chib (1995) suggest estimating Bayes factors by including all models of interests in an MCMC sampling scheme and by allowing the Markov chains to jump from model to model. In fact, Carlin and Chib's (1995) method simply requires the introduction

of pseudoprior distributions needed in our context since the prior of Θ depends on the combination parameter λ, as explained below.

Strictly following Carlin and Chib's (1995) sampling scheme, pseudoprior distributions should be thought of as linking densities between parameters associated with a given combination parameter λ and another combination parameter $\tilde{\lambda} \neq \lambda$. That is, when a chain selects a combination λ, the parameters of the regimes associated with the combination $\tilde{\lambda}$ should be drawn from their pseudopriors. Formally, we have the following full conditional distributions for $\Theta^{K,j}$ (the set of parameters associated with the number of breaks K and the combination $\lambda^{K,j}$):

$$\pi(\Theta^{K,j}|\lambda, \Gamma^K, Y_T) \propto \begin{cases} f(Y_T|\Theta^{K,j}, \Gamma^K, \lambda)p(\Theta^{K,j}|\lambda)p(\Gamma^K)p(\lambda) & \lambda = \lambda^{K,j}, \\ p(\Theta^{K,j}|\lambda), & \lambda \neq \lambda^{K,j}, \end{cases}$$

where $p(\lambda)$ is the prior of λ, and $p(\Theta^{K,j}|\lambda \neq \lambda^{K,j})$ is the pseudoprior (or linking density).

However, the recurrent regimes offer the particularity that the same model parameters are present in the various $\Theta^{K,j}$'s. Hence, instead of drawing from the pseudopriors the entire set Θ of model parameters associated with a "not-drawn" $\tilde{\lambda}$, we take the advantage of this particularity and plug directly the accepted draw of active regimes in all $\Theta^{K,j}$'s. Then, we draw from the pseudopriors only the subsets of parameters θ that correspond to inactive regimes and plug them again in the other $\Theta^{K,j}$'s. We set pseudoprior distributions as independent normal distributions with the posterior mean and variance of the model parameters, which are estimated at the end of a first burn-in phase from the posterior draws of the pure Change-Point model, and we choose a uniform prior distribution on the possible combination parameters λ's.

Finally, the draw of λ^K is done according to the following probabilities: $\forall j = \{1, ..., J\}$,

$$\pi(\lambda^K = \lambda^{K,j}|\Theta^K, \Gamma^K, Y_T) = \frac{f(Y_T|\Theta^K, \Gamma^K, \lambda^{K,j})p(\Theta^K|\lambda^{K,j})p(\lambda^{K,j})}{\sum_{s=1}^{J} f(Y_T|\Theta^K, \Gamma^K, \lambda^{K,s})p(\Theta^K|\lambda^{K,s})p(\lambda^{K,s})}, \tag{3.3}$$

where Θ^K is the full set of model parameters associated with the number of breaks K, and $p(\Theta^K|\lambda^{K,j})$ is the joint prior density of the model parameters evaluated at $\Theta^{K,j}$ multiplied by the joint pseudoprior density of the model parameters of non-active regimes in $\lambda^{K,j}$.

To sum up, at a given iteration and for a given chain, the algorithm starts by drawing

the combination parameter λ according to (3.3). Then, a proposition is made for the model parameters of the active regimes according to (3.1) and the parameters of inactive regimes are drawn from their pseudopriors. When all G_1 iterations are run the posterior probability of having one specific permutation $\lambda^{K,j}$ is estimated by $\frac{\sum_{g_1=1}^{G_1} \mathbb{1}_{\lambda_{g_1}=\lambda^{K,j}}}{G_1} \approx \pi(\lambda^{K,j}|Y_T)$. From the next subsection it will be clear that our strategy to compute the marginal likelihood of all models builds upon reliable estimates of the Bayes factors of the pure Change-Point models. For a given K, in order to obtain a reliable estimate of $\pi(\lambda^{CP}|Y_T)$, the posterior probabily of the pure Change-Point specification, we need to ensure that the pure Change-Point model is visited by the MCMC algorithm. This can be done by determining an accurate prior probability $p(\lambda^{CP})$ after a burn-in phase. A simple and effective way to do that is by comparing the data likelihood implied by the different models (pure Change-Point and recurrent regime specifications) and offset the differences by playing on $\mathrm{p}(\lambda^{CP})$.

3.4.2 Bayes Factors to Determine the Number of Breaks

Now that we can compute Bayes factors of models with different regimes orderings, we need to find a way to compute the marginal likelihood (or Marginal Log-Likelihood, MLL) of the pure Change-Point models to compare the posterior probabilities of models differing by their number of regimes and regime ordering.

The first method we consider is the Bridge sampling technique of Meng and Wong (1996) (see also Fruhwirth-Schnatter 2004). Using the integral symbol to integrate out Γ (like in the continuous case, for notational simplicity), the marginal likelihood can be written as $f(Y_T) = \int\int f(Y_T|\Theta,\Gamma)p(\Theta,\Gamma)d\Theta d\Gamma$. Taking a function $t(\Theta,\Gamma)$ and a proposal distribution $q(\Theta,\Gamma)$, we can define the following two quantities:

$$\begin{aligned} A &= \frac{1}{G_1}\sum_{g_1=1}^{G_1} t(\Theta^{g_1},\Gamma^{g_1})q(\Theta^{g_1},\Gamma^{g_1}), \\ A_1 &= \frac{1}{G_2}\sum_{g_2=1}^{G_2} t(\Theta^{g_2},\Gamma^{g_2})f(Y_T|\Theta^{g_2},\Gamma^{g_2})p(\Theta^{g_2},\Gamma^{g_2}), \end{aligned}$$

where G_1 and G_2 are large. The sequence $\{\Theta^{g_1},\Gamma^{g_1}\}_{g_1=1}^{G_1}$ is sampled from $\pi(\Theta,\Gamma|Y_T)$, which has already been done by the MCMC algorithm, and $\{\Theta^{g_2},\Gamma^{g_2}\}_{g_2=1}^{G_2}$ is sampled from $q(\Theta,\Gamma)$. Meng and Wong (1996) show that the marginal likelihood can be computed using the ratio of these two quantities as $\frac{A_1}{A} \approx_a f(Y_T)$. We use the asymptotically optimal choice of $t(\Theta,\Gamma)$ derived in Meng and Wong (1996) that minimizes the expected relative error of the estimator for i.i.d. draws from $f(\Theta,\Gamma|Y_T)$ and $q(\Theta,\Gamma)$: $t(\Theta,\Gamma) = \frac{1}{\pi(\Theta,\Gamma|Y_T)+q(\Theta,\Gamma)}$. The proposal distribution $q(\Theta,\Gamma)$ is given in appendices.

We also consider an importance sampling approach, which simply consists in setting $t(\Theta,\Gamma) = \frac{1}{q(\Theta,\Gamma)}$.[9]

[9]Note that one could also consider an original application of Carlin and Chib (1995). The approach would require the estimation of a pure Change-Point model with respect to another, "simple" model the marginal likelihood of which would be known. Since after having applied Carlin and Chib's (1995) method we obtain the ratio of marginal likelihoods and since we know the marginal likelihood of one of the two models, it is easy to compute the MLL of the other model.

3.5 Simulations

In this section, we resort to Monte Carlo simulations to show that the D-DREAM algorithm performs well in practice. Two Monte Carlo experiments are performed. The first one aims to show that the algorithm is able to detect parameter changes in the volatility dynamics, even if the changes are small. The second one aims to show that the algorithm can deal with recurrent regimes.

3.5.1 Monte Carlo Simulation 1: Changes in Parameters

We simulate 1000 data points with 32 different Data Generating Processes (DGPs) of CP-GJR-GARCH(1,1) models, most of which entail a break in the middle of the sample (500th observation). The 16 first DGPs deal with breaks in the ω parameter of the volatility dynamics. The aim is to show that the D-DREAM algorithm can detect changes in the level of the unconditional variance. The 16 other DGPs deal with breaks in the β parameter. The goal is to show that the algorithm can also detect changes in the persistence of the dynamics. We simulate 100 time series for each DGP and after each simulation we estimate the marginal likelihood of 2 competing models: a model without structural break and a model with one change point.[10] We draw 1000 posterior values for the parameters per chain after convergence and the number of chains is equal to $5 \times \#regimes + 1$.[11] The parameters δ, η_Θ, and η_Γ of equations (3.1) and (3.2) are respectively set to 3, 10^{-6}, and 10^{-6}. We choose to sample sub-vectors of Θ and Γ by drawing randomly the size of the sub-vectors (comprised between 1 and the size of the full vector) at the beginning of each MCMC iteration. These values are kept unchanged in the rest of the Chapter. We also use the priors presented in Table 3.1.

Tables 3.2 and 3.3 display the results with respect to ω and β, respectively. It can be seen from Table 3.2 that the D-DREAM algorithm selects the right model most of the time. However, focusing on the DGPs that entail a structural break, the smaller the difference between the ω parameters of the two regimes or the greater the persistence of the volatility dynamics, the more difficult it is to favour the model with the structural break. For instance, for a relatively large parameter change of 0.4 ($\omega_1 = 0.1$ and $\omega_2 = 0.5$)

[10] We compute the MLLs by Bride sampling.
[11] Note that there are 5 parameters per regime to estimate.

and a persistence ($\alpha + \beta + \gamma/2$) of 0.9 (unchanged from regime 1 to regime 2) the MLL of the model with one structural break is always found to be larger. Now, if the parameter change is decreased to only 0.2 ($\omega_2 = 0.3$), the model with one Change-Point is chosen only 51 times (out of 100). If, instead of modifying the parameter change, the persistence is increased to 0.95 the model with Change-Point is chosen only 77 times. It is also worth noting that break dates are more accurately and precisely estimated if the change in parameter is great and/or persistence is low. For example, for a change in ω of 0.4 and a persistence of 0.8, the average posterior mean of the break date is 501.42 (recall that the true break date is located at observation 500) and the 97.5%-2.5% percentile range is 27.32 (days), which is rather accurate and precise. On the contrary, for a change in ω of 0.2 and a persistence of 0.95, the average posterior mean of the break date is 490.62 and the 97.5%-2.5% percentile range is 161.86.

Table 3.2: Monte Carlo Simulations of Breaks in ω

	Persistence	ω_2			
		0.1 (no break)	0.3	0.5	0.7
	A	# models with one break selected			
	0.7	0	100	100	99
	0.8	0	93	100	100
	0.9	1	51	100	98
	0.95	0	34	77	94
	B	Log avg. BF			
	0.7	-8.75	10.75	18.05	21.10
	0.8	-8.63	7.29	14.45	16.94
	0.9	-6.63	1.18	9.58	12.17
	0.95	-6.50	-1.04	4.01	7.18
$\omega_1 = 0.1$	C	Avg. post. break means			
	0.7	604.94	501.61	500.62	500.38
	0.8	675.05	500.21	501.42	501.31
	0.9	589.03	501.47	503.98	500.47
	0.95	508.81	490.62	505.53	500.17
	D	Avg. 95% range post. break distrib.			
	0.7	490.35	46.31	18.70	13.87
	0.8	544.07	56.01	27.32	24.66
	0.9	334.63	129.32	35.92	26.13
	0.95	241.36	161.86	71.23	43.37

Table 3.2. The Table shows the performance of the D-DREAM algorithm after 100 simulations of 16 DGPs. Depending on the DGP, a break can occur in the parameter ω after 500 observations (half of series lengths). The persistence $(\alpha + \beta + \gamma/2)$ is shown in the second column and is unchanged in the second regime. Panel A displays the number of times the MLL of the model with one break, computed via Bridge sampling, is estimated to be greater than the MLL of the model without Change-Point. Panel B shows the logarithms of the averages of the Bayes Factors (BF), that is, the MLL of the model with one break minus the MLL of the model without Change-Point. Panel C displays the average posterior means of the break points. Panel D shows the percentile range between the 97.5% percentile and the 2.5% percentile (in days).

Results regarding changes in the β parameter need careful interpretation. It can be seen from Table 3.3 that small variations in the β parameter are more difficult to detect when β is small. For instance, when β is 0.6 and increases to 0.7 in regime 2 the model without structural break is preferred on average. This can be seen from the average of the logarithm of the Bayes factors which is negative (-4.46). On the other hand, when β is 0.8 and increases to 0.9 the average of the logarithm of the Bayes factor is 3.70 and the model with a structural break is chosen most of the time (82 times out of 100). In order to understand this result, notice that the parameter β has a nonlinear impact on

the unconditional variance as well as on the autocorrelation function of the conditional variance. Consequently, for an equal increase, the larger the original beta (β_1), the larger the impact on the unconditional variance and on the autocorrelation of the conditional variance, and the easier it is to detect a change in regime.

Table 3.3: Monte Carlo Simulations of Breaks in β

	β_1	β_2			
		0.6	0.7	0.8	0.9
	A	# Models with one break selected			
	0.6	1	0	94	100
	0.7	2	2	80	100
	0.8	90	57	0	82
	0.9	99	96	78	6
	B	Log avg. BF			
$\omega = 0.1$	0.6	-6.65	-4.46	5.46	13.91
$\alpha = 0.05$	0.7	-4.31	-6.05	1.51	10.22
$\gamma = 0.05$	0.8	6.86	1.41	-5.48	3.70
	0.9	14.33	9.76	4.05	-3.46
	C	Avg. post. break means			
	0.6	547.92	538.08	505.77	501.61
	0.7	624.87	650.45	508.98	503.82
	0.8	497.25	454.01	507.55	489.27
	0.9	498.13	497.82	494.72	631.99
	D	Avg. 95% range post. break distrib.			
	0.6	523.45	254.90	71.50	24.48
	0.7	271.50	432.05	181.24	36.25
	0.8	72.38	183.76	479.43	90.94
	0.9	15.25	25.23	62.20	281.23

Table 3.3. The Table shows the performance of the D-DREAM algorithm after 100 simulations of 16 DGPs. Depending on the DGP, a break can occur in the parameter β after 500 observations (half of series lengths). The parameters ω, α, and γ are set to 0.1, 0.05, and 0.05, respectively and are unchanged in the second regime. Panel A displays the number of times the MLL of the model with one break, computed via Bridge sampling, is estimated to be greater than the MLL of the model without Change-Point. Panel B shows the logarithms of the averages of the Bayes Factors (BF), that is, the MLL of the model with one break minus the MLL of the model without Change-Point. Panel C displays the average posterior means of the break points. Panel D shows the percentile range between the 97.5% percentile and the 2.5% percentile (in days).

3.5.2 Monte Carlo Simulation 2: Recurrent Regimes

We illustrate further the flexibility of the D-DREAM algorithm by showing that it can efficiently deal with recurrent regimes, as explained in section 3.3. We simulate time series of 4000 observations according to a CP-GJR-GARCH(1,1) model with 3 change points and 1 recurrent regime. The DGP is displayed in Table 3.4. The values of the different parameters are chosen to be representative of financial time series and to switch from quiet periods to more volatile ones.

Table 3.4: DGP With Recurrent Regime

Parameters	Regime 1	Regime 2 and 4	Regime 3
μ	0	0	0
ω	0.400	0.300	0.150
α	0.125	0.050	0.075
γ	0.050	0.100	0.050
β	0.550	0.850	0.650
Break Dates	1000	2000	3000

The Monte Carlo study consists of 50 simulated series. For each simulated series we compute the MLL of CP-GJR-GARCH(1,1) models with $K = 0, 1, ..., 4$, considering all possible regime combinations. The result is striking: in all 50 simulations the right model is always selected according to the MLL criterion (3 breaks and the last regime equal to the second). In addition to this, we show that the two techniques we use to compute the MLL, namely Bridge sampling and Importance sampling, are equivalent. Table 3.5 displays for all K the differences between the averages (over the 50 simulations) of the MLLs of the pure Change-Point models. It appears that the differences grow with the number of regimes but remain small and insignificant.

Table 3.5: MLL Estimations

	K				
	0	1	2	3	4
MLL averages differences	-0.00	-0.07	-0.11	-0.05	-0.79
(std. dev.)	(0.01)	(0.08)	(0.23)	(0.25)	(1.09)

Table 3.5. The Table displays the differences between the averages of the MLL estimation via Bridge sampling and Importance sampling. Standard deviations after the 50 simulations are in parentheses.

The average of the posterior means of the model parameters and break dates of the right specification are reported in Table 3.6. The values are close to the true values of the DGP and the structural break dates are particularly well estimated.

Table 3.6: Averages of Posterior Means of Right Specification

Parameters	Regime 1	Regime 2 and 4	Regime 3
μ	0.01	-0.01	0.01
	(0.04)	(0.07)	(0.03)
ω	0.42	0.37	0.19
	(0.16)	(0.17)	(0.08)
α	0.12	0.04	0.08
	(0.06)	(0.03)	(0.04)
γ	0.05	0.11	0.04
	(0.06)	(0.04)	(0.05)
β	0.53	0.84	0.58
	(0.15)	(0.04)	(0.16)
Break Dates	997.70	2000.08	2999.86
	(12.12)	(6.45)	(6.51)

Table 3.6. The Table displays the averages, after 50 simulations, of the posterior means of the model parameters and break dates of the right specification. Standard deviations of the 50 posterior means are reported between parentheses. This Table must be compared with Table 3.4.

3.6 Empirical Application to Financial Time Series

This section documents results on financial time series. We start by focusing on the S&P 500 index and then briefly comment on the results for six other time series.

3.6.1 S&P 500

Our time series for the (weighted) Standard & Poor's 500 index consists of daily returns from June 20, 1991 to April 25, 2011 (5000 observations) and is downloadable from Datastream.

Focusing first on the number of regimes and the recurrence of states, it is shown in Table 3.7 that the MLL of a CP-GJR-GARCH(1,1) model is clearly maximized for a specification with four structural breaks and with the fourth regime equal to the second and the fifth equal to the third.[12] The posterior means of the break dates are the following: January 30, 1992; November 20, 1996; September 17, 2003; and June 5, 2007.

[12]The MLLs of pure Change-Point models were computed via Bridge sampling.

Table 3.7: S&P 500 Highest MLLs per Number of Regime

# Regimes	1	2	3	4	5	6
λ	$\{1\}$	$\{1,2\}$	$\{1,2,3\}$	$\{1,2,1,2\}$	$\{1,2,3,2,3\}$	$\{1,2,1,4,2,4\}$
MLL	-6725.53	-6728.50	-6724.68	-6706.45	**-6705.30**	-6709.19

The Table displays the highest MLLs among the possible regime combinations for each number of regimes considered. The row 'λ' indicates the regime combination that yields the highest MLL for each number of regimes. The bold value indicates the highest MLL.

These results seem plausible in light of figure 3.2 that plots the S&P 500 return series, along with the posterior means and 95% confidence bands of the structural break dates. Note in particular that, according to the figure, the volatility of the series seems indeed similar between November 1996 and September 2003 and after June 2007. It is nevertheless surprising that the regime of the last crisis is not different from the previous regimes. Instead, we find that the dynamics of the S&P 500 after June 2007 are comparable to the dynamics of the series at the end of the last century and beginning of the 21^{st} century. Note however that the latter period was also marked by exceptional events like the Russian default, the LTCM collapse, the "dot-com" bubble, or the attack on the World Trade Center. These events contributed to a large extent to add volatility in stock prices and therefore, given our priors, the D-DREAM algorithm does not find it worth creating a new regime for the financial crisis.

Table 3.8 shows the posterior means of the model parameters and break dates of the best specification according to the MLL criterion. As could be expected, we switch from the quiet regimes 2 and 4 (unconditional variance of 0.39 and persistence of 0.91) to the more volatile regimes 3 and 5 (unconditional variance of 2 and persistence of 0.98). Note also that the parameter γ of the leverage effect is greatly positive and significant in most regimes.

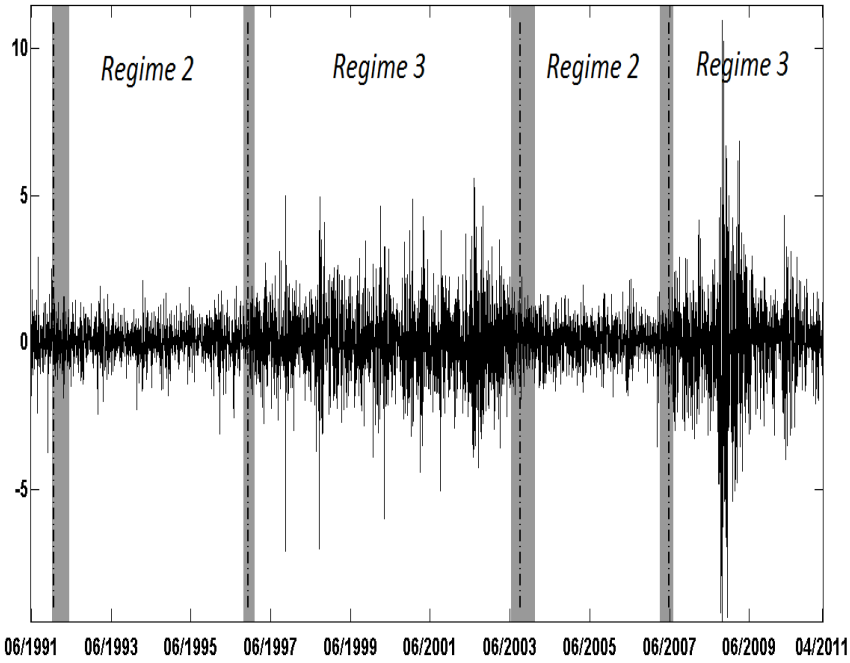


Figure 3.2: S&P 500 with Posterior Means of Structural Breaks and 95% Confidence Intervals

Table 3.8: S&P 500 Posterior Means

Parameters	Regime 1	Regime 2	Regime 4	Regime 3	Regime 5
μ	0.064	0.037		-0.001	
	(0.063)	(0.012)		(0.019)	
ω	0.447	0.037		0.045	
	(0.110)	(0.007)		(0.007)	
α	-0.010	-0.045		-0.027	
	(0.069)	(0.010)		(0.007)	
γ	0.218	0.162		0.189	
	(0.165)	(0.021)		(0.015)	
β	0.283	0.869		0.910	
	(0.137)	(0.019)		(0.011)	
Break Dates	30-Jan-92	20-Nov-96	05-Jun-07	17-Sep-03	—
	(29.47)	(17.91)	(25.16)	(36.40)	—
Durations in days	156	1217	978	934	—

Table 3.8. The Table displays the posterior means of the model parameters and break dates of the best CP-GJR-GARCH specification for the S&P 500 series. Standard deviations of the posterior draws are reported between parentheses. Durations of regimes, expressed in days, are computed with the posterior means of break dates as landmarks.

3.6.2 Other Financial Series

We apply the algorithm with a CP-GJR-GARCH(1,1) model to six other financial time series of daily returns spanning the same time period as the S&P 500 (from June 20, 1991 to April 25, 2011). The dataset contains series for the (weighted) Dow Jones Industrial Average (DJIA) index, the (weighted) New York Stock Exchange (NYSE) composite index, gold prices ("Gold", Gold Bullion London Bullion Market USD/Troy Ounce), the West Face Capital Inc. Base Metals sub-index ("Metals"), share prices for Boeing (BA: Boeing Co.) and JP Morgan (JPM: JP Morgan Chase & Co.).[13] All series are downloadable from Datastream.

Table 3.9 displays the main characteristics of the best specification (K, λ, break dates) for each series according to the MLL criterion.[14] We find that all series contain structural breaks and that five of them entail at least one recurrent regime. This finding should be placed in perspective with the results of Chapter 2 where we apply the PMCMC algorithm to similar series and find that Markov-switching models fit the data better than pure Change-Point models. Notice that the DJIA and NYSE time series are similar to the S&P 500 and thus, yield similar break points compared with those of the S&P 500. Note also that an interesting result we find is that the leverage parameter γ is significantly negative in three regimes for the series on gold prices, and in two regimes for the "Metals" series (descriptive statistics of the posterior distributions for all series are presented in appendices). This corroborates the results of Carpantier and Dufays (2012) who fit a GJR-GARCH model to many commodity series and (almost) systematically find an inverse leverage effect. Hence, it seems that their results are robust to the inclusion of structural breaks in the econometric model.

[13] We do not consider that dividends are reinvested in shares.
[14] The MLLs of pure Change-Point models were computed via Bridge sampling.

Table 3.9: CP-GJR-GARCH(1,1) Specifications with Highest MLL

Series	Number of Breaks K	Regime Combination λ
DJIA	3	{1,2,1,2}
Break Dates:	15-Nov-96, 08-Aug-03, 07-Jun-07	
NYSE	4	{1,2,3,2,3}
Break Dates:	28-Jan-92, 26-Dec-96, 08-Sep-03, 17-Jan-07	
Gold	3	{1,2,3,1}
Break Dates:	22-Nov-96, 27-Aug-01, 14-Nov-01	
Metals	3	{1,2,3,4}
Break Dates:	19-Jul-93, 27-May-03, 11-Nov-06	
Boeing	3	{1,2,3,1}
Break Dates:	04-Jun-98, 01-Dec-98, 29-Aug-02	
JP Morgan	2	{1,2,1}
Break Dates:	08-Sep-03, 13-Apr-07	

Table 3.9. The Table displays the best specification for each series according to the MLL criterion as well as the break dates.

3.7 Forecasting : The Contribution of Models with Recurrent States

Structural break models do not necessarily outperform models with fixed coefficients regarding their forecasting ability. Structural break models indeed need a certain number of observations once a new regime is detected to estimate the parameters accurately. If the estimates are unprecise, structural break models can actually do worse at forecasting than fixed parameters models based on a rolling or recursive window. Therefore, we compare the forecasting accuracy of GJR-GARCH and CP-GJR-GARCH models based on the predictive marginal likelihood (PML) criterion. The idea of the PML criterion is to compare the likelihood of observation y_{t+1} of several models after estimating these models with the information available up to time t. Hence, we need to evaluate the likelihood $f(y_{t+1}|Y_t)$ which can be computed by averaging the likelihoods $f(y_{t+1}|\Theta_{g_1}, \Gamma_{g_1}, Y_t)$, where Θ_{g_1} and Γ_{g_1} denote the g_1^{th} of G_1 posterior draws of the model parameters and break dates (see Del Negro and Schorfheide 2011).

We first focus on a broad forecasting exercise at horizons $h = \{1, 2, 5, 10, 20, 40, 50\}$ based on all time series. We start to forecast the series in the middle of the sample

(2500^{th} observation) and add fifty observations to the recursive window before the next forecasting evaluation (resulting in 49 evaluations since all series contain approximately 5000 observations). At each of the 49 evaluations, we run the D-DREAM algorithm with a CP-GJR-GARCH(1,1) model for $K = \{0, 1, 2, ..., 6\}$, allowing for recurrence (for $K \geq 2$), and without recurrence (for $K \geq 0$). Then we compare the forecast accuracy of three different classes of models: (1) a fixed parameter GJR-GARCH(1,1) model ($K = 0$); (2) a class of models allowing for structural breaks and recurrent regimes; (3) a class of models allowing for structural breaks but not reccurent regimes. For the class of models 2 and 3, we forecast the series with the best specification according to the MLL criterion (the MLL selects K and λ for the class of models 2, and only K for the class of models 3).[15]

Table 3.10 shows the cumulative average log-PMLs, over all evaluations, of the three classes of models for all forecast horizons h. The results might look disappointing at first sight. For a total of 49 cases (7 forecast horizons $\times$ 7 series), the fixed parameter model beats the structural break models allowing for recurrent regimes 31 times (63% of the time), and the structural break models not allowing for recurrent regimes 46 times (94%). Note that results are especially disappointing for break models not allowing for recurrent regimes which is due to the fact that recurrent regimes dampen the lack of observations after a break. We need however to mitigate these results. We start by emphasizing two points: first, we can see from Table 3.10 that all values are very close to each other; second, the forecast ability of structural break models allowing for regime recurrence seems to be superior for long forecast horizons ($h \geq 20$). Then, we recall that the finding that fixed parameters models perform as well as Change-Point models regarding their forecasting accuracy is not new. In a similar forecasting exercise, Bauwens, Koop, Korobilis, and Rombouts (2011) find that fixed parameters auto-regressive (AR) models often beat structural break AR models and outperform them on average according to the PML criterion. Interestingly, they find the opposite result according to the RMSE criterion.[16] Finally, note that we do not work with models for the mean equations (abstracting from the intercept) but for the volatility dynamics. Hence, forecast improvements are mainly obtained from the forecast conditional variance, which makes forecast likelihoods harder

[15]The MLLs of pure Change-Point models are computed via Bridge sampling.

[16]Notice that in the context of volatility models it does not make sense to use a criterion like the mean absolute error, root mean squared error, or Mincer and Zarnowitz's (1969) R^2.

to enhance by taking structural breaks into account.

Table 3.10: Average Cumulative Log-PMLs

Horizon h	1	2	5	10	20	40	50
			S&P 500				
GJR-GARCH(1,1)	-1.33	**-2.65**	**-7.22**	**-14.38**	-29.99	-66.00	**-84.14**
CP-GJR-GARCH(1,1) with rec	**-1.31**	-2.66	-7.25	-14.42	**-29.96**	**-65.98**	-84.32
CP-GJR-GARCH(1,1) without rec	-1.36	-2.72	-7.44	-14.72	-30.69	-67.87	-86.49
			DJIA				
GJR-GARCH(1,1)	**-1.32**	**-2.69**	**-7.03**	**-13.70**	**-28.64**	-63.53	-80.96
CP-GJR-GARCH(1,1) with rec	-1.34	-2.72	-7.14	-13.74	-28.47	**-62.33**	**-79.68**
CP-GJR-GARCH(1,1) without rec	-1.39	-2.91	-7.39	-14.19	-29.46	-64.67	-82.47
			NYSE				
GJR-GARCH(1,1)	**-1.34**	**-2.72**	**-7.20**	**-14.20**	**-29.80**	-68.00	-87.48
CP-GJR-GARCH(1,1) with rec	-1.35	-2.74	-7.33	-14.27	-29.83	**-67.16**	**-86.13**
CP-GJR-GARCH(1,1) without rec	-1.37	-2.80	-7.40	-14.53	-30.45	-68.44	-87.82
			Gold				
GJR-GARCH(1,1)	**-1.42**	**-2.93**	**-7.28**	-14.28	-29.59	-60.65	-77.93
CP-GJR-GARCH(1,1) with rec	-1.45	-2.96	-7.32	**-14.26**	**-29.55**	**-60.32**	**-77.28**
CP-GJR-GARCH(1,1) without rec	-1.51	-2.98	-7.44	-14.48	-30.07	-60.99	-78.19
			Metals				
GJR-GARCH(1,1)	-1.77	**-3.59**	**-8.83**	**-18.40**	**-37.09**	-77.87	-97.93
CP-GJR-GARCH(1,1) with rec	**-1.75**	-3.62	-8.98	-18.47	-37.35	**-77.07**	**-96.52**
CP-GJR-GARCH(1,1) without rec	-1.76	-3.60	-8.96	-18.49	-37.35	-79.46	-98.94
			Boeing				
GJR-GARCH(1,1)	-2.05	-3.96	**-9.86**	**-19.42**	-41.51	**-83.73**	**-104.41**
CP-GJR-GARCH(1,1) with rec	**-2.03**	-3.96	-9.94	-19.65	**-41.43**	-84.04	-104.71
CP-GJR-GARCH(1,1) without rec	-2.05	**-3.94**	-9.9	-19.56	-41.45	-84.31	-105.01
			JP Morgan				
GJR-GARCH(1,1)	-1.95	**-4.07**	-10.27	**-20.58**	**-41.31**	**-86.04**	**-108.53**
CP-GJR-GARCH(1,1) with rec	-1.96	-4.10	**-10.19**	-20.62	-41.69	-88.28	-111.04
CP-GJR-GARCH(1,1) without rec	**-1.94**	**-4.07**	-10.25	-20.60	-41.74	-109.12	-132.17

Table 3.10. The Table shows the averages, over the 49 forecast evaluations, of the cumulative log-PMLs of the three classes of models at different forecast horizons. The bold values indicate the highest log-PMLs for each series and forecast horizon. "with rec" means that we allow for recurrent regimes. "without rec" means that we do not allow for recurrent regimes.

In order to further mitigate the disappointing results, we carry out a small forecasting exercise on the S&P 500 comparable to Liu and Maheu (2008). We forecast data at one given point in time, namely September 21, 2009. We choose that date as it is located after the beginning of the financial crisis and gives enough observations to estimate the parameters of the last regime accurately. Table 3.11 shows the cumulative log-PMLs for the three classes of models. This time, we find that the model with recurrent regime outperforms the other two.

Table 3.11: Cumulative Log-PMLs for the S&P 500

Horizon	1	2	5	10	20	40	50
GJR-GARCH(1,1)	**-1.03**	-2.54	-7.96	-18.69	-31.84	-74.80	-89.24
CP-GJR-GARCH(1,1) with rec	-1.11	**-2.53**	**-7.61**	**-17.06**	**-30.73**	**-66.93**	**-81.35**
CP-GJR-GARCH(1,1) without rec	-1.05	-2.52	-8.00	-18.78	-32.25	-75.77	-90.37

Table 3.11. The Table shows the cumulative log-PMLs for a forecasting excercise on the S&P 500 series carried out on September 21, 2009.

3.8 Conclusion

Change-Points are generally accounted for following the method of Chib (1996) by introducing a hidden state variable in the model that indicates the regime which the observations belong to. However, this method suffers from the problem of path dependence.

In this Chapter, we introduce an MCMC algorithm that considers break dates as parameters, which avoids the path dependence problem, and we prove the convergence of the algorithm. The number of breaks, which is often unknown, is determined thanks to the marginal likelihood criterion which we compute in two different ways. Moreover, we allow for the regimes to be active in several distinct periods in the sample.

We illustrate the efficiency and rapidity of the algorithm by means of simulations and apply it to financial time series to highlight the presence of structural breaks. Indeed, all seven series studied exhibit structural breaks and six entail recurrent regimes. In particular, we find evidence of four structural breaks and two recurrent regimes in the daily series of returns of the S&P 500 index.

A forecast exercise on seven financial time series has been carried out in order to empirically assess the forecast ability of CP-GJR-GARCH models with and without recurrent regimes. Documented results show that these models, although they often perform slightly worst than the standard GJR-GARCH model, can produce better forecasts in some circumstances. These counter-intuitive results doubtfully arise due to the large parameter set of the CP model. Indeed, standard CP specifications assume that all the model parameters evolve when it faces a structural break. However it is highly likely that only a small subset of parameters changes in this situation. Allowing for recurrent states is one way to decrease the CP parameter set. Further research will be devoted to even more

limit the CP parameter set and the impact on predictions will be investigated.

Appendices

Proof of Proposition 1

In order to prove Proposition 1, we need to show the convergence of the two conditional transition kernels and then resort to the product of kernels principle (see Chib and Greenberg 1995b). The proof of convergence in Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009) only requires minor adjustments to the present case to prove the convergence for the Θ-block given Γ. We give here the proof for the Γ-block given Θ only.[17] It consists of four steps and is based on the fact that the chains do not converge separately of each other but rather converge together towards a stationary distribution.

- Step 1. Recalling conditional detailed balance.

  In this first step, we recall the conditional detailed balance condition required for the existence of a stationary distribution in a N-component Metropolis algorithm. In order to have the stationary distribution as target distribution one needs to show that if $(\Gamma_1^{(t)}, ..., \Gamma_N^{(t)}) \sim \tilde{\pi}(.)$, with $\tilde{\pi}(.)$ the joint pdf of $(\Gamma_1, ..., \Gamma_N)$ at iteration t, then $(\Gamma_1^{(t+1)}, ..., \Gamma_N^{(t+1)}) \sim \tilde{\pi}(.)$ at iteration $t+1$. That is, one needs the conditional detailed balance condition to be satisfied for each and every chain i such that, $\forall i$ (omitting Θ_i and Y_T from the conditioning variables):

$$\begin{aligned}&\tilde{\pi}(\Gamma_i^{(t)}|\Gamma_1^{(t+1)}, ..., \Gamma_{i-1}^{(t+1)}, \Gamma_{i+1}^{(t)}, ..., \Gamma_N^{(t)}) \times K_i(\Gamma_i^{(t+1)}|\Gamma_1^{(t+1)}, ..., \Gamma_{i-1}^{(t+1)}, \Gamma_i^{(t)}, \Gamma_{i+1}^{(t)}, ..., \Gamma_N^{(t)}) \\ &= \tilde{\pi}(\Gamma_i^{(t+1)}|\Gamma_1^{(t+1)}, ..., \Gamma_{i-1}^{(t+1)}, \Gamma_{i+1}^{(t)}, ..., \Gamma_N^{(t)}) \times K_i(\Gamma_i^{(t)}|\Gamma_1^{(t+1)}, ..., \Gamma_{i-1}^{(t+1)}, \Gamma_i^{(t+1)}, \Gamma_{i+1}^{(t)}, ..., \Gamma_N^{(t)}),\end{aligned}$$

  with $K_i(.|.)$ the jumping kernel.

- Step 2. Proving the conditional detailed balance condition.

  In the second step, we prove the conditional detailed balance of the D-DREAM algorithm. The D-DREAM algorithm satisfies the detailed balance condition for the Γ_i-block *with respect to* $\pi(.)$ if

$$\begin{aligned}&\pi(\Gamma_i^{(t)}) K_i(\Gamma_i^{(t+1)}|\Gamma_1^{(t+1)}, ..., \Gamma_{i-1}^{(t+1)}, \Gamma_i^{(t)}, \Gamma_{i+1}^{(t)}, ..., \Gamma_N^{(t)}) \\ &= \pi(\Gamma_i^{(t+1)}) K_i(\Gamma_i^{(t)}|\Gamma_1^{(t+1)}, ..., \Gamma_{i-1}^{(t+1)}, \Gamma_i^{(t+1)}, \Gamma_{i+1}^{(t)}, ..., \Gamma_N^{(t)}).\end{aligned}$$

[17] A less detailed proof can also be found in Vrugt and ter Braak (2011).

Recall that the proposal Z_i for an update of chain i proceeds by random selection of pairs of other chains. Hence, the kernel of this update is a mixture of kernels which maintains conditional detailed balance with respect to $\pi(.)$ if each of its components does. Since ξ follows a symmetric distribution and the pairs $\{\sum_{g=1}^{\delta}\Gamma_{r_1(g)}, \sum_{h=1}^{\delta}\Gamma_{r_2(h)}\}$ and $\{\sum_{h=1}^{\delta}\Gamma_{r_1(h)}, \sum_{g=1}^{\delta}\Gamma_{r_2(g)}\}$ are equally likely, this condition is fulfilled by accepting the proposal with probability $\min\{\frac{f(Y_T|Z_i)p(Z_i)}{f(Y_T|\Gamma)p(\Gamma)},1\}$.[18] This also holds true with the *round*[.] operator since it does not alter the symmetry of the proposal distribution and with the binomial cross-over scheme used to modify only selected dimensions of Γ_i. [19]

- Step 3. Showing that the joint is the product of the marginals.

  In the third step, we show that the joint distribution of the break date vectors of the N chains is the product of the marginal distributions of the vector of break dates of each chain. As showed in step 2, the kernel $K_i(.|.)$ for updating chain i satisfies the detailed balance condition of step 1 *with*, $\forall i = 1, ...N$, $\tilde{\pi}(\Gamma_i^{(t)}|\Gamma_1^{(t+1)}, ..., \Gamma_{i-1}^{(t+1)}, \Gamma_{i+1}^{(t)}, ..., \Gamma_N^{(t)}) = \pi(\Gamma_i^{(t)})$. Since the identity in step 1 is satisfied with the marginal distribution $\pi(\Gamma_i^{(t)})$, the D-DREAM algorithm is an N-component Metropolis algorithm with joint stationary distribution $\tilde{\pi}(\Gamma_1, ..., \Gamma_N)$, such that

$$\tilde{\pi}(\Gamma_1, ..., \Gamma_N) = \pi(\Gamma_1) \times ... \times \pi(\Gamma_N).$$

- Step 4. Proving ergodicity and uniqueness.

  In the fourth step, we prove the ergodicity of the chains and convergence towards a unique stationary distribution. In order to prove convergence towards a unique stationary distribution, it is sufficient to show that the chains are irreducible and aperiodic. Irreducibility is guaranteed by the unbounded support of the distribution of ξ in proposal (3.2) that necessarily covers the entire support [2,T], and aperiodicity is ensured by the random walk component generated by the same ξ. For ergodicity, it remains to show that the chains have a finite expected return time which is trivially

[18] As discussed above, the two possible issues in the proposal - that a break date falls outside the sample length [2,T] or that two break dates are inverted - are discarded by our prior distribution choice since the likelihood of the proposal Z_i would then be equal to zero.

[19] Note that the *round[.]* operator must also insure symmetry for the particular case where one of the proposed date has a fractional part equal to .5, although it will virtually never happen in practice. The proposed parameter would then be rounded with probability .5 to the upper integer and .5 to the closest smaller integer.

observed given the discreteness of Γ and the random terms in proposal (3.2).

In order to complete the proof one needs to combine the proof for the Γ-block above with the proof for the Θ-block available in Vrugt, ter Braak, Diks, Robinson, Hyman, and Higdon (2009) and recall the product of kernels principle. This principle states that the product of the transition kernels has an invariant density. The reader is referred to Chib and Greenberg (1995b) for further explanations.

Proposal Distribution for the Computation of the Marginal Likelihood via Bridge Sampling

We consider independence between the two proposal distributions Θ and Γ. Following Ardia, Hoogerheide, and van Dijk (2009) we work with a mixture of distributions. The proposal $q(\Theta)$ consists of two mixtures of normal distributions. The expectations, the variance-covariance matrices and the weights are determined by the K-Harmonic means algorithm based on the posterior draws but the variance-covariance matrices are multiplied by four in order to cover the entire posterior distribution. The proposal $q(\Gamma)$ is driven by a poisson distribution as follows:

$$\begin{aligned} q(\Gamma) &= \prod_{i=1}^{K} q(\tau_i) \\ q(\tau_i) &= \bar{\tau_i} - \tilde{\lambda} + Po(\tilde{\lambda}), \end{aligned}$$

where $\bar{\tau_i}$ refers to the rounded posterior means of the distributions $\tau_i|Y_T$ and $\tilde{\lambda}$ is equal to four times the standard deviations of the same distributions. The proposal distribution implies an expectation equals to $\bar{\tau_i}$ and a variance of $\tilde{\lambda}$. Recall that the Jacobian of the transformation from the break dates to durations is one.

Posterior Means of Parameters of the Best CP-GJR-GARCH(1,1) Model for the other series

Table 3.12: Posterior means of parameters of the best CP-GJR-GARCH(1,1) model on DOW JONES

Parameters	Regime 1	Regime 3	Regime 2	Regime 4
μ	0.046		0.015	
	(0.013)		(0.020)	
ω	0.054		0.036	
	(0.015)		(0.006)	
α	-0.007		-0.009	
	(0.008)		(0.007)	
γ	0.105		0.164	
	(0.020)		(0.015)	
β	0.829		0.906	
	(0.040)		(0.010)	
Break Dates	15-Nov-1996	07-Jun-2007	08-Aug-2003	—
	(29.93)	(21.12)	(34.76)	—

Table 3.12. The Table displays the posterior means of the model parameters and break dates of the best CP-GJR-GARCH specification for the DJIA series. Standard deviations are reported between parentheses.

Table 3.13: Posterior means of parameters of the best CP-GJR-GARCH(1,1) model on NYSE

Parameters	Regime 1	Regime 2	Regime 4	Regime 3	Regime 5
μ	0.059	0.047		-0.002	
	(0.060)	(0.012)		(0.018)	
ω	0.357	0.029		0.044	
	(0.093)	(0.006)		(0.006)	
α	0.025	-0.019		-0.014	
	(0.080)	(0.010)		(0.007)	
γ	0.185	0.148		0.178	
	(0.159)	(0.024)		(0.017)	
β	0.297	0.862		0.897	
	(0.137)	(0.021)		(0.011)	
Break Dates	28-Jan-1992	26-Dec-1996	17-Jan-2007	08-Sep-2003	—
	(29.215)	(24.198)	(18.73)	(64.838)	—

Table 3.13. The Table displays the posterior means of the model parameters and break dates of the best CP-GJR-GARCH specification for the NYSE series. Standard deviations are reported between parentheses.

Table 3.14: Posterior means of parameters of the best CP-GJR-GARCH(1,1) model on "Gold"

Parameters	Regime 1	Regime 4	Regime 2	Regime 3
μ	0.024		-0.039	0.108
	(0.009)		(0.018)	(0.093)
ω	0.001		0.088	0.277
	(0.000)		(0.018)	(0.096)
α	0.073		0.206	0.593
	(0.006)		(0.024)	(0.215)
γ	-0.059		-0.159	0.730
	(0.006)		(0.023)	(0.198)
β	0.958		0.721	0.142
	(0.003)		(0.042)	(0.074)
Break Dates	22-Nov-1996	—	27-Aug-2001	14-Nov-2001
	(2.59)	—	(13.15)	(12.66)

Table 3.14. The Table displays the posterior means of the model parameters and break dates of the best CP-GJR-GARCH specification for the series on gold prices. Standard deviations are reported between parentheses.

Table 3.15: Posterior means of parameters of the best CP-GJR-GARCH(1,1) model on "Metals"

Parameters	Regime 1	Regime 2	Regime 3	Regime 4
μ	-0.048	0.002	0.159	0.004
	(0.026)	(0.016)	(0.038)	(0.049)
ω	0.277	0.157	0.010	0.364
	(0.043)	(0.040)	(0.007)	(0.125)
α	0.252	0.061	0.103	0.053
	(0.062)	(0.015)	(0.015)	(0.021)
γ	-0.141	0.046	-0.110	0.058
	(0.043)	(0.020)	(0.015)	(0.030)
β	0.241	0.733	0.960	0.827
	(0.095)	(0.059)	(0.010)	(0.044)
Break Dates	19-Jul-1993	27-May-2003	07-Nov-2006	—
	(13.51)	(84.23)	(27.33)	—

Table 3.15. The Table displays the posterior means of the model parameters and break dates of the best CP-GJR-GARCH specification for the series METALS. Standard deviations are reported between parentheses.

Table 3.16: Posterior means of parameters of the best CP-GJR-GARCH(1,1) model on Boeing

Parameters	Regime 1	Regime 4	Regime 2	Regime 3
μ	0.050		-0.099	0.052
	(0.022)		(0.218)	(0.071)
ω	0.041		1.551	1.458
	(0.008)		(0.501)	(0.395)
α	0.013		-0.127	0.129
	(0.005)		(0.019)	(0.034)
γ	0.054		0.104	0.064
	(0.009)		(0.017)	(0.046)
β	0.947		0.928	0.609
	(0.007)		(0.052)	(0.077)
Break Dates	04-Jun-1998	—	01-Dec-1998	29-Aug-2002
	(4.83)	—	(14.48)	(82.92)

Table 3.16. The Table displays the posterior means of the model parameters and break dates of the best CP-GJR-GARCH specification for the series Boeing. Standard deviations are reported between parentheses.

Table 3.17: Posterior means of parameters of the best CP-GJR-GARCH(1,1) model on JP Morgan

Parameters	Regime 1	Regime 3	Regime 2
μ	0.031		0.056
	(0.027)		(0.029)
ω	0.069		0.089
	(0.011)		(0.044)
α	0.013		0.020
	(0.005)		(0.013)
γ	0.097		0.084
	(0.010)		(0.038)
β	0.930		0.854
	(0.006)		(0.056)
Break Dates	08-Sep-2003	—	13-Apr-2007
	(32.48)	—	(37.29)

Table 3.17. The Table displays the posterior means of the model parameters and break dates of the best CP-GJR-GARCH specification for the series JP Morgan. Standard deviations are reported between parentheses.

4

Infinite-State Markov-switching for Dynamic Volatility and Correlation Models[1]

4.1 Introduction

As previously highlighted choosing the number of regimes in an MS or CP model can be done, in Bayesian inference, by maximizing the marginal likelihood with respect to the number of regimes. Doing this for MS- and CP-GARCH models requires the estimation of the model for a given number of regimes, followed by the marginal likelihood computation itself. This procedure is repeated several times, up to a maximum number of regimes, which makes the search for the best model very time-consuming. The sticky infinite hidden Markov chain model (IHMM), proposed by Fox, Sudderth, Jordan, and Willsky (2007)), allows us to bypass these repeated computations by treating the number of regimes as an unknown parameter. It relies on a Markov-chain with a potentially infinite number of states (regimes) and is suited for series exhibiting persistence. The structure encompasses the MS and the CP specifications as special cases. Its building blocks are the Dirichlet process (Ferguson (1973) and Sethuraman (1994)) and the hierarchical Dirichlet process (Teh, Jordan, Beal, and Blei (2006)). The sticky IHMM has already been applied in fields such as genetics (Beal and Krishnamurthy (2006)), visual recognition (Kivinen, Sudderth, and Jordan (2007)), and economics (Jochmann (2010) and Song (2011)) for models *without* path dependence. The main contribution of this Chapter is to develop Bayesian inference for sticky infinite hidden Markov-GARCH and DCC models, thus for models *with* path

[1]This Chapter is a slight revision of Dufays (2012).

dependence.

In concrete terms, our contribution is twofold. As mentioned in the general introduction, the path dependence problem is usually circumvented either by applying the model of Haas, Mittnik, and Paolella (2004) or of Klaassen (2002). We generalize these two models by embedding a Metropolis-Hastings algorithm that allows us to estimate the MS-GARCH model. We additionally illustrate that these new algorithms outperform the existing MCMC alternatives in term of computational time, while the mixing properties remain competitive with respect to the Particle MCMC algorithm (PMCMC) of Chapter 2. Our method renders inference on MS- and CP-GARCH models almost as fast as inference on MS- and CP-ARCH models without adding much complexity in the computational structure. Moreover, even if our algorithms can be used to compute the marginal likelihood using Chib's formula (see Chib and Jeliazkov (2001)), so that we can maximize this criterion in order to select a specific number of regimes, we avoid such a time-consuming approach. This leads to our second contribution: we use an MCMC algorithm to determine directly the number of regimes as well as the specification (MS or CP), by incorporating the sticky IHMM into our MS-GARCH model. Currently this approach is limited to models without path dependence (Jochmann (2010) and Song (2011)). Thus we extend the scope of the IHMM modeling framework to richer models, such as GARCH (instead of ARCH).

The Chapter is organized as follows. In section 4.2 and 4.3, we present the infinite hidden Markov-switching GARCH model (IHMS-GARCH) and the Bayesian estimation of its parameters. In Section 4.4, we discuss the two new algorithms in the light of existing MCMC alternatives and illustrate them on simulated data. We highlight that the IHMS-GARCH model accurately estimates CP- and MS-GARCH models. Section 4.5.1 provides detailed results on the S&P500 index daily series and we compare our results with those of Chapter 2 for the MS-GARCH model. A forecast exercise on several financial time series is reported. We eventually apply the new method to the DCC model in Section 4.6. Conclusions are presented in the last section. A brief review of the IHMM modeling framework is provided in Appendix 4.D.

4.2 Model definition

In this section, we develop a Markov-switching framework with an undetermined number of regimes for univariate and multivariate GARCH models. The model rests on the sticky infinite hidden Markov model (sticky IHMM) that is shortly presented in Appendix 4.D. Before stating the model in subsection 4.2.2, we briefly define the Dirichlet process as well as its useful stick-breaking representation since they are directly used in our specification.

4.2.1 Dirichlet process and its stick-breaking representation

The Dirichlet process G, denoted by $G \sim DP(\eta, G_0)$, where the parameter G_0 is called the base distribution and the scalar $\eta \in \Re^+$ the concentration parameter, has been introduced by Ferguson (1973). It can be seen as an extension of the Dirichlet distribution to continuous spaces.

The Dirichlet process with base distribution G_0 is the unique distribution over the support Θ of G_0 (where $\Theta \in \Re^d$ and d stands for the dimension of the set Θ), such that the relation

$$\{G(A_1), G(A_2), ..., G(A_n)\} \sim Dir(\eta G_0(A_1), ..., \eta G_0(A_n))$$

holds for every natural number n and every n-partition $\{A_1, A_2, ..., A_n\}$ of Θ. The notation $\text{Dir}(a_1, ..., a_n)$ corresponds to a Dirichlet distribution with parameters $a_i, \quad i = 1, ..., n$ (see Balakrishnan and Nevzorov (2005)).

Sethuraman (1994) demonstrated that the DP has a stick-breaking representation. It is based on two independent sequences of i.i.d random variables $\{\pi_k\}_{k=1}^{\infty}$ and $\{\Theta_k\}_{k=1}^{\infty}$ (where $\pi_k \in \Re$ and Θ_k is a d-dimensional vector) and is constructed by the following formulas (δ_{Θ_k} is the probability measure concentrated at Θ_k) :

$$\begin{aligned} \beta_k &\sim Beta(1, \eta), \qquad & \Theta_k &\sim G_0, \\ \pi_k &= \beta_k \prod_{l=1}^{k-1}(1 - \beta_l), \qquad & G &= \sum_{k=1}^{\infty} \pi_k \delta_{\Theta_k}, \end{aligned}$$

which ensures that $G \sim DP(\eta, G_0)$.

The distribution over π is sometimes written $\pi \sim GEM(\eta)$[1] or $\pi \sim Stick(\eta)$.

4.2.2 The model

To ease the discussion we explain the model for the univariate case and in particular for the Markov-switching GARCH(1,1) model (MS-GARCH). Nevertheless this modeling approach is applicable to other models exhibiting path dependence. For instance, in Section 4.6 we provide the extension to a MS Dynamic Conditional Correlation models (MS-DCC).

Let $Y_T = \{y_1, ..., y_T\}'$ be a time series where T denotes the sample size. The infinite hidden MS-GARCH(1,1) model (IHMS-GARCH) consists in the following set of equations :

$$
\begin{aligned}
y_t &= \sigma_t \epsilon_t && (4.1)\\
\sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \sigma_{t-1}^2 && (4.2)\\
\epsilon_t &\sim N(0,1) && (4.3)\\
s_t | s_{t-1} = i, p_i &\sim p_i = \{p_{i1}, p_{i2}, p_{i3}, ..., ..., ...\} && (4.4)\\
p_i | \pi, \lambda, \kappa &\sim DP(\lambda + \kappa, \frac{\lambda \pi + \kappa \delta_i}{\lambda + \kappa}) && (4.5)\\
\pi | \eta &\sim \text{Stick}(\eta) && (4.6)\\
\{\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}\} | \mu, \Sigma &\sim N(\mu, \Sigma) && (4.7)\\
\mu &\sim N(\underline{\mu}, \underline{\Sigma}) && (4.8)\\
\Sigma^{-1} &\sim \text{Wishart}(\underline{V}, \underline{v}) && (4.9)
\end{aligned}
$$

where δ_i denotes the probability measure concentrated at i.

- **Equations (1) to (3)** define a standard Markov-switching model with GARCH parameters of the regime s_t equal to $\{\omega_{s_t}, \alpha_{s_t}, \beta_{s_t}\}$. The random variable s_t takes integer values in $[1, \infty]$ and denotes the current regime. We define the state vector $S_T = \{s_1, ..., s_T\}'$.

[1] GEM refers to Griffiths, Engen and McCloskey (see Pitman (2006))

- **Equations (4) to (6)** specify the first hierarchical structure of the model which is the sticky IHMM. We assume that the latent state process $\{s_t\}$ is first order Markovian with the transition matrix

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & ... \\ p_{21} & p_{22} & p_{23} & ... \\ ... & ... & ... & ... \\ p_{i1} & p_{i2} & p_{i3} & ... \\ ... & ... & ... & ... \end{pmatrix},$$

where $p_i = \{p_{i1}, p_{i2}, p_{i3}, ..., ..., ...\}$ is the transition probability distribution of moving from state i to another state (including the state i). This transition matrix characterizes a MS model with infinite regimes. The sticky parameter κ captures the persistence in the time series by setting more weights to the self-transition p_{ii} since $E(p_{ii}|\pi, \lambda, \kappa) = \frac{\lambda\pi_i+\kappa}{\lambda+\kappa}$. The (infinite dimensional) vector π is driven by a stick-breaking process ($\pi \sim Stick(\eta)$). We denote the set of random Dirichlet parameters $H_{Dir} = \{\eta, \lambda+\kappa, \rho\}$ where $\rho = \frac{\kappa}{\lambda+\kappa}$. Their prior distributions are detailed in Section 4.4.

- **Equations (7) to (9)** describe the second hierarchical structure, which bears on the GARCH parameters. We assume that the parameters are driven by a Normal distribution so we map our GARCH parameters on the real line. The one-to-one transformation[1] is denoted by $\{\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}\}$. We also define the set $\Theta = \{\tilde{\omega}_1, \ldots, \tilde{\omega}_\infty, \tilde{\alpha}_1, \ldots, \tilde{\alpha}_\infty, \tilde{\beta}_1, \ldots, \tilde{\beta}_\infty\}$ which includes all the GARCH parameters of the model and the set $\Theta_i = \{\tilde{\omega}_i, \tilde{\alpha}_i, \tilde{\beta}_i\}$ which contains all the relevant GARCH parameters of the regime i. The hierarchical structure takes full advantage of volatility parameters from past regimes. It enhances the proposed parameters of new regimes since if a new state is born, we draw related GARCH parameters from $N(\mu, \Sigma)$ whose expectation and variance-covariance matrix are updated by taking into account parameters of previous regimes.

[1] $\tilde{\omega}_i = \log(\frac{\omega_i}{5-\omega_i}), \tilde{\alpha}_i = \log(\frac{\alpha_i+1}{1-\alpha_i}), \tilde{\beta}_i = \log(\frac{\beta_i}{1.5-\beta_i})$.

The same kind of setting (4)-(9) has already been proposed by Song (2011) for an autoregressive model. Other distributional assumptions than the normal or other models than GARCH(1,1) can be handled.

4.3 Estimation by Bayesian inference

Bayesian inference is feasible by treating explicitly S_T as a parameter. We also augment our parameter set by an auxiliary variable $U_T = \{u_1, ..., u_T\}$ to deal with the infinite structure. Based on the slice sampler (Neal (2003)) this technique, the beam sampler from Van Gael, Saatci, Teh, and Ghahramani (2008), will be detailed in the next subsection. Our sampling scheme iteratively draws from each full conditional distribution of Table 4.1.

1. $f(S_T|\Theta, P, H_{Dir}, \pi, U_T, Y_T)$
2. $f(U_T|\Theta, P, H_{Dir}, \pi, S_T, Y_T)$
3. $f(P|\Theta, H_{Dir}, \pi, U_T, S_T, Y_T)$
4. $f(\Theta|\mu, \Sigma, P, H_{Dir}, \pi, U_T, S_T, Y_T)$
5. $f(\mu, \Sigma|\Theta, P, H_{Dir}, \pi, U_T, S_T, Y_T)$
6. $f(H_{Dir}|\Theta, P, \pi, U_T, S_T, Y_T)$
7. $f(\pi|\Theta, P, H_{Dir}, U_T, S_T, Y_T)$

Table 4.1: IHMS-GARCH Gibbs sampler

Updating the state vector S_T from its full conditional distribution is the most challenging part of the Gibbs sampler. The other full conditional distributions have already been detailed in the literature. The last two distributions and $f(P|S_T, \Theta, H_{Dir}, \pi, Y_T)$ constitute the sticky IHMM of which the sampling has been described in Fox, Sudderth, Jordan, and Willsky (2007). The hierarchical structure μ, Σ has the usual Normal-Wishart prior in order to get conjugate posterior distributions. Drawing from the full conditional of Θ is standard. In this Chapter, we use an adaptive Metropolis method with delayed rejection (see Haario, Saksman, and Tamminen (2001) and Mira (2001)). The entire MCMC sampler is detailed in Appendix 4.A. We now concentrate on the sampling of a complete state vector.

4.3.1 Sampling the state vector S_T

Updating each state s_t separately given the others (Bauwens, Preminger, and Rombouts (2010)) produces poor mixing properties due to the dependence of the states. We therefore propose another strategy for sampling the state vector. The method relies on the forward-

backward algorithm of Chib (1996). However a straightforward application is infeasible due to the infinite number of regimes and to the path dependence problem. We use the beam sampler to circumvent the first issue. The second one will be tackled by combining an approximation of the MS-GARCH model with a Metropolis-Hastings step. It is worth noticing that we could also use the Particle MCMC algorithm of Chapter 2 conjugated with the beam sampler but our approach is less time consuming and easier to implement.

Beam sampler

The random variables U_T have been embedded to ease the update of S_T. They act as a slice sampling to truncate the infinite summation that appears in the forward-backward algorithm into a finite one. The methodology leaves invariant the full posterior distribution. In the initial paper (Van Gael, Saatci, Teh, and Ghahramani (2008)) the distribution of $u_t|s_t, s_{t-1}, P$ is uniform : $U[0, p_{s_{t-1},s_t}]$. To sample an entire vector U_T, we use the decomposition :

$$f(U_T|S_T, P) = f(u_T|s_T, s_{T-1}, P)f(u_{T-1}|s_{T-1}, s_{T-2}, P)...f(u_2|s_2, s_1, P)f(u_1|s_1, P)$$

where $f(u_1|s_1, P) = \frac{\delta_{\{0\leq u_1\leq p_{s_1,s_1}\}}}{p_{s_1,s_1}}$.

The size of the finite set directly affects the mixing properties of the sampler. The probability of staying in a same state for the next period is generally high for time series due to their persistence. This stylized fact can drastically decrease the size of possible paths allowed by the beam sampler at each MCMC iteration. Consequently the draws of the forward-backward algorithm can become highly dependent. To avoid this problem we use a modified uniform distribution that concentrates more probabilities on small values of u_t. The modified density is as follows

$$\begin{aligned} f(u_t|s_t, s_{t-1}, P) &= \frac{k}{p_{s_{t-1},s_t}} \quad \text{If } 0 \leq u_t \leq k_1 p_{s_{t-1},s_t} \\ &= \frac{k_2}{p_{s_{t-1},s_t}} \quad \text{If } k_1 p_{s_{t-1},s_t} < u_t \leq p_{s_{t-1},s_t} \end{aligned}$$

where k, k_1, k_2 are constants. In the empirical exercise we respectively set $k = 20$ and $k_1 = 0.01$. The last constant k_2 is derived from $\int_0^{p_{s_{t-1},s_t}} f(u_t|s_t, s_{t-1}, P)du_t = 1$.

MS-GARCH approximation

The forward-backward algorithm fails when it faces a GARCH model due to the path dependence induced by the lag of the conditional variance. We propose a Metropolis-Hastings approach that avoids the path dependence problem. We first sample an entire state vector from an approximate GARCH model that allows us to apply the forward-backward algorithm and the proposal is accepted or rejected according to the Metropolis-Hastings ratio that preserves the required balance. Although an approximate model is used to sample the state vector, the posterior distribution is not altered thanks to the Metropolis-Hastings step. Two approximate models are discussed.

- MS-GARCH models are well approximated by the model of Klaassen (2002) :

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \tilde{\sigma}_{t-1,s_t}^2 \end{aligned} \tag{4.10}$$

  where $\tilde{\sigma}_{t-1,s_t}^2 = E[\sigma_{t-1}^2 | Y_{t-1}, s_t, U_T, \Theta, P]$. We derive the computation of $\tilde{\sigma}_{t-1,s_t}^2$ in Appendix 4.B. Under this specification, the likelihood $f(y_t|Y_{t-1}, \Theta, s_t)$ is computable.

- Relying on the research of Haas, Mittnik, and Paolella (2004), we also consider the following approximation of the MS-GARCH model :

$$\begin{aligned} y_t &= \sigma_{t,s_t} \epsilon_t \\ \sigma_{t,1}^2 &= \omega_1 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1,1}^2 \\ \sigma_{t,2}^2 &= \omega_2 + \alpha_2 y_{t-1}^2 + \beta_2 \sigma_{t-1,2}^2 \\ &\dots \\ \sigma_{t,K}^2 &= \omega_K + \alpha_K y_{t-1}^2 + \beta_K \sigma_{t-1,K}^2 \end{aligned}$$

  where K stands for the maximum number of regimes. The approximation gets rid of the path dependence problem since each state at any time is related to a known variance.

Procedure for updating the state vector

A draw of S_T is recursively obtained from the proposal distribution as follows (letting $S^i = \{s_i, ..., s_T\}$ and omitting the condition to the sets of parameters Θ,P,H_{Dir} and π) :

$$q(S_T|Y_T, U_T) = q(s_T|Y_T, U_T)q(s_{T-1}|Y_T, s_T, U_T)q(s_{T-2}|Y_T, S^{T-1}, U_T)...q(s_1|Y_T, S^2, U_T)$$

Since the MS-GARCH approximation rules out the path dependence problem, each conditional density $q(s_t|Y_T, U_T, S^{t+1})$ is proportional to

$$\begin{aligned} q(s_t|Y_T, U_T, S^{t+1}) &\propto q(s_t|Y_t, U_t)f(u_{t+1}|s_t, s_{t+1})f(s_{t+1}|s_t) \\ &\propto q(s_t|Y_t, U_t)(k\delta_{\{0\leq u_{t+1}\leq k_1 p_{s_t,s_{t+1}}\}} + k_2\delta_{\{k_1 p_{s_t,s_{t+1}} < u_{t+1}\leq p_{s_t,s_{t+1}}\}}) \end{aligned}$$

and $q(s_t|Y_t, U_t)$ is computed by forward looking :

$$\begin{aligned} q(s_t|Y_t, U_t) &\propto f(y_t|s_t, u_t)\sum_{i=1}^{\infty} f(u_t|s_t, s_{t-1})f(s_t|s_{t-1} = i)q(s_{t-1} = i|Y_{t-1}, U_{t-1}) \\ &\propto f(y_t|s_t, u_t)\sum_{i=1}^{\infty}(k\delta_{\{0\leq u_t\leq k_1 p_{s^i_{t-1},s_t}\}} + k_2\delta_{\{k_1 p_{s^i_{t-1},s_t} < u_t\leq p_{s^i_{t-1},s_t}\}})q(s^i_{t-1}|Y_{t-1}, U_{t-1}) \end{aligned}$$

where s^i_{t-1} stands for $s_{t-1} = i$ and the infinite sum of the last equation is handled thanks to the beam sampler. It becomes a finite one because only some states satisfy the constraint $\{0 \leq u_t \leq p_{s_{t-1},s_t}\}$.

The new state vector S'_T is accepted according to the Metropolis-Hastings ratio :

$$\alpha(S_T, S'_T|Y_T, \Theta, P, U_T, H_{Dir}, \pi) = \min\{1, \frac{f(S'_T|Y_T, U_T, \Theta, P, H_{Dir}, \pi)q(S_T|Y_T, U_T, \Theta, P)}{f(S_T|Y_T, U_T, \Theta, P, H_{Dir}, \pi)q(S'_T|Y_T, U_T, \Theta, P)}\}$$

The proposal distribution is not always a good approximation of the full conditional one. It occasionally leads to stick the algorithm at a fixed state vector. In order to avoid this situation we sample the state vector by randomized blocks instead of drawing an entire one in a single piece. At each MCMC iteration we randomly set the size of the

block. The method has been proposed in a different context by Chib and Ramamurthy (2010). In our empirical exercise the block size randomly varies from fifty observations to the whole sample size.

It is worth emphasizing that the proposed sampler does not need the IHMM to operate. Applications where the number of regimes K is a priori fixed (for instance Bauwens, Preminger, and Rombouts (2010), Henneke, Rachev, Fabozzi, and Nikolov (2011)) could also benefit from the current method. The number of regimes could then be determined using the marginal likelihood computed by the Chib's formula (Chib and Jeliazkov (2001)).

4.4 Illustration on artificial data

The algorithm is illustrated on simulated data in this section. First we document our prior choices and some practical issues for the MCMC implementation. We devote the next three subsections to detail results on simulated series from five different data generating processes (DGP). We start by describing our simulation strategy that includes the chosen DGP. The MCMC mixing properties are then compared with those of existing MCMC alternatives. The last subsection reports summary statistics of the posterior distributions of several simulated data. A discussion on the MCMC mixing properties of the two approximate models lies in Appendix 4.C. Based on a Monte Carlo study, the Klaassen model turns out to better approximate the MS-GARCH model than the Haas, Mittnik and Paolella model.

4.4.1 Starting point, priors, burn-in and label switching

As it is shown in Table 4.2, we use standard prior distributions for the model. We set the same prior distributions on the Dirichlet process parameters as Fox, Sudderth, Jordan, and Willsky (2007) and Jochmann (2010). The hyper-parameters have been chosen to reflect the persistence of high frequency time series. The expectation of $\lambda + \kappa$ and ρ are respectively set to 1000 and 0.9994. The choice of the persistence is close to the one set in Chapter 2. GARCH parameters of each regime are independently driven by Normal distributions.

Prior Distributions of the Dirichlet processes

$\eta \sim G(10, \frac{1}{2})$ $\quad$ $\lambda + \kappa \sim G(1000, 1)$ $\quad$ $\rho = \frac{\kappa}{\lambda+\kappa} \sim Beta(10000, 6)$

Prior Distributions of the GARCH parameters

For each regime i : $\{\tilde{\omega}_i, \tilde{\alpha}_i, \tilde{\beta}_i\} \sim \mathrm{N}(\mu,\Sigma)$

Hierarchical parameter : μ			Hierarchical parameter : Σ		
μ	$\sim$	$\mathrm{N}(\underline{\mu},\underline{\Sigma})$	Σ^{-1}	$\sim$	$\mathrm{W}(\underline{V},\underline{v})$
$\underline{\mu}$	$=$	$\{\log(\frac{0.5}{4.5}), \log(\frac{1.1}{0.9}), \log(\frac{0.7}{0.8})\}$	$\underline{V}$	$=$	$\frac{1}{5\underline{v}} I_3$
$\underline{\Sigma}$	$=$	I_3	$\underline{v}$	$=$	50

Table 4.2: Prior distributions. The d-dimensional identity matrix is denoted by I_d.

The starting point of an MCMC algorithm is also a relevant practical issue. Typically, MCMC algorithms start at the ML estimates but these cannot be computed in the presence of structural breaks. Our MCMC starting point for each simulation is the ML estimate of a GARCH(1,1) model without breaks. The starting values of λ, κ and η are set to 1.2,1000 and 3, respectively.

The posterior distribution is invariant to the labels of regimes. As a consequence a label of one regime can switch to another one during the MCMC simulation. If this label switching happens, summary statistics that are label dependent, such as the posterior means of the parameters, are misleading. Some methods have been proposed to alleviate the permutation in the MCMC by imposing some constraints on parameters or to build a sample of coherent labels at the end of the MCMC simulation by maximizing a loss function. In this Chapter we circumvent the problem by using statistics that are invariant to label switching. For instance, instead of showing posterior means of each regime, we display posterior means over time ($\approx E(\Theta_t|Y_T)$) which do not depend on the state label. We thus allow for switching of labels during the simulation as advocated by Geweke (2007).

For assessing the MCMC convergence we use Geweke's diagnostic (Geweke (1992)) on some of the GARCH parameters over time : $\Theta_t|Y_T$. We select ten parameters $\Theta_t|Y_T$ equally spaced in time in order to cover the whole sample and we apply Geweke's diagnostic to them. Once the MCMC has converged for these ten variables we save the next 50000 samples as draws of the posterior distribution.

The IHMS-GARCH program coded in c++ is available for Windows platform on Arnaud Dufays' website.

4.4.2 Simulation

The IHMM encompasses the Markov-switching and the Change-point models. We thus revisit some simulations of the CP and the MS literature. We consider the same simulations as He and Maheu (2010) (HM), Bauwens, Preminger, and Rombouts (2010)(BPR), Bauwens, Dufays, and Rombouts (2011) (BDR) (see Chapter 2), and a series without break. The DGP of He and Maheu consists in a CP model with three regimes. The BPR simulation is a MS model with two regimes. Eventually BDR consider a Change-point and a Markov-switching DGP. The last series does not exhibit any break for testing the IHMS-GARCH ability to estimate a standard specification. All the simulated DGP are summarized in Table 4.3. Each series has 3000 observations except the BPR simulation (1500 observations).

Name	Type	Regimes	Break point	ω	α	β
DGP_{HM}	CP	3	$\{1000, 2000\}$ obs.	$\{0.2; 0.6; 0.1\}$	$\{0.1\}$	$\{0.8\}$
DGP_{BDR1}	CP	3	$\{1000, 2000\}$ obs.	$\{0.2; 0.7; 0.4\}$	$\{0.1; 0.2; 0.2\}$	$\{0.8; 0.7; 0.4\}$
Name	Type	Regimes	Tr. Matrix	ω	α	β
DGP_{BPR}	MS	2	$P = \begin{pmatrix} 0.98 & 0.02 \\ 0.04 & 0.96 \end{pmatrix}$	$\{0.3; 2\}$	$\{0.35; 0.1\}$	$\{0.2; 0.6\}$
DGP_{BDR2}	MS	2	$P = \begin{pmatrix} 0.9999 & 0.0001 \\ 0.0005 & 0.9995 \end{pmatrix}$	$\{0.6; 0.4\}$	$\{0.1; 0.2\}$	$\{0.8; 0.4\}$
$DGP_{noBreak}$	—	1	—	$\{0.5\}$	$\{0.2\}$	$\{0.7\}$

Table 4.3: Data Generating Processes of the five simulated series.

The CP specification of HM assumes structural breaks in the unconditional variance while the persistence parameters (α and β) remain constant over regimes. The (local) unconditional variance (i. e. $\frac{\omega_{s_t}}{1-\alpha_{s_t}-\beta_{s_t}}$) before the first break is equal to 2 and then increases to 6. It decreases to 1 at the end of the sample. It tries to mimic a financial market that switches from quiet to more volatile periods. On the contrary each parameter of the DGP in BDR varies across regimes. The unconditional variance evolves from 2 to 0.67 with an intermediate state where it is equal to 7.

The two Markov-switching specifications differ among others things by their transition matrix. The BPR DGP assumes an expected duration of remaining in the first and second states of 50 and 25 observations respectively whereas the transition probabilities of the BDR MS model are very low (with duration of 10000 and 2000 observations).

4.4.3 Comparison with other algorithms

We first compare our sampling method of the state vector with two other MCMC samplers (BPR for Bauwens, Preminger, and Rombouts (2010) and PMCMC for the algorithm in Chapter 2). Bauwens, Preminger, and Rombouts (2010) draw each state one by one which leads to a highly autocorrelated samples. It requires many MCMC iterations to explore the entire support of the posterior distribution. The second method samples the entire state vector in one block using a Sequential Monte Carlo (SMC) algorithm within the MCMC. The mixing properties are by far improved but at a computational cost. The complexity order of the SMC is $O(N^2T)$ where N stands for the number of particles. Recall that we have chosen $N = 250$ for an MS model and $N = 150$ for a CP model. We expect that the mixing properties of our MH method lies in between the two of them since we randomize the size of the block we sample and we use a M-H step. However the computation time is drastically reduced compared to the SMC algorithm. Indeed the computational burden is equivalent to the forward-backward algorithm ($O(K^2T)$ where K denotes the number of regimes). To compare the different methods we launch the three algorithms on simulated series from the four DGP that exhibit structural breaks. We fix the number of regimes and the volatility parameters at the MLE given the true state vector. We use CP models for CP DGP and MS settings for the other ones. Finally we store 10 000 posterior draws for each simulation. Table 4.4 displays the maximum autocorrelation time posterior draws of a state variable.[1] The current algorithms (HMP-MH for Haas, Mittnik and Paolella Metropolis-Hastings and Kl-MH for Klaassen Metropolis-Hastings) display much better autocorrelation times than the BPR approach for all the simulated data. The comparison with the PMCMC is more delicate. The PMCMC method always exhibits slightly better autocorrelation times than the current algorithms but is by far more computationally demanding (see Table 4.5). The HMP-MH autocorrelation times are always higher than the Kl-MH ones. It corroborates the Monte Carlo results of Appendix 4.C.

[1]The autocorrelation time is computed by batch means (see Geyer (1992)) and is defined as $1+2\sum_{i=1}^{\infty}\rho_i$ where ρ_i is the autocorrelation coefficient of order i between the posterior draws of a state variable.

Name	Type	BPR	PMCMC	HMP-MH	Kl-MH
DGP_{HM}	CP	450.96	**1.94**	38.94	2.88
DGP_{BDR1}	CP	478.51	**1.30**	4.80	2.70
DGP_{BPR}	MS	116.96	**1.18**	16.11	4.98
DGP_{BDR2}	MS	477.49	**1.50**	5.81	2.28

Table 4.4: Autocorrelation time of posterior structural break draws. The minimum autocorrelation times are in bold. For MS model, the autocorrelation time is computed on the number of observations in each regime.

Table 4.5 shows the elapsed time for MCMC simulations. All the simulations have been executed on the same computer and the programs only differ in the way of sampling S_T. While our method is competitive with BPR for CP models, it clearly becomes the fastest method for MS models.

Name	Type	BPR	PMCMC	HMP-MH	Kl-MH
DGP_{HM}	CP	1	160.9	3.6	3.4
DGP_{BDR1}	CP	1	161.1	3.7	3.4
DGP_{BPR}	MS	104.9	178	3.2	2.9
DGP_{BDR2}	MS	413.7	351.3	6.3	6.3

Table 4.5: Elapsed time in minutes for a MCMC simulation of 10000 draws

4.4.4 Results on simulated data

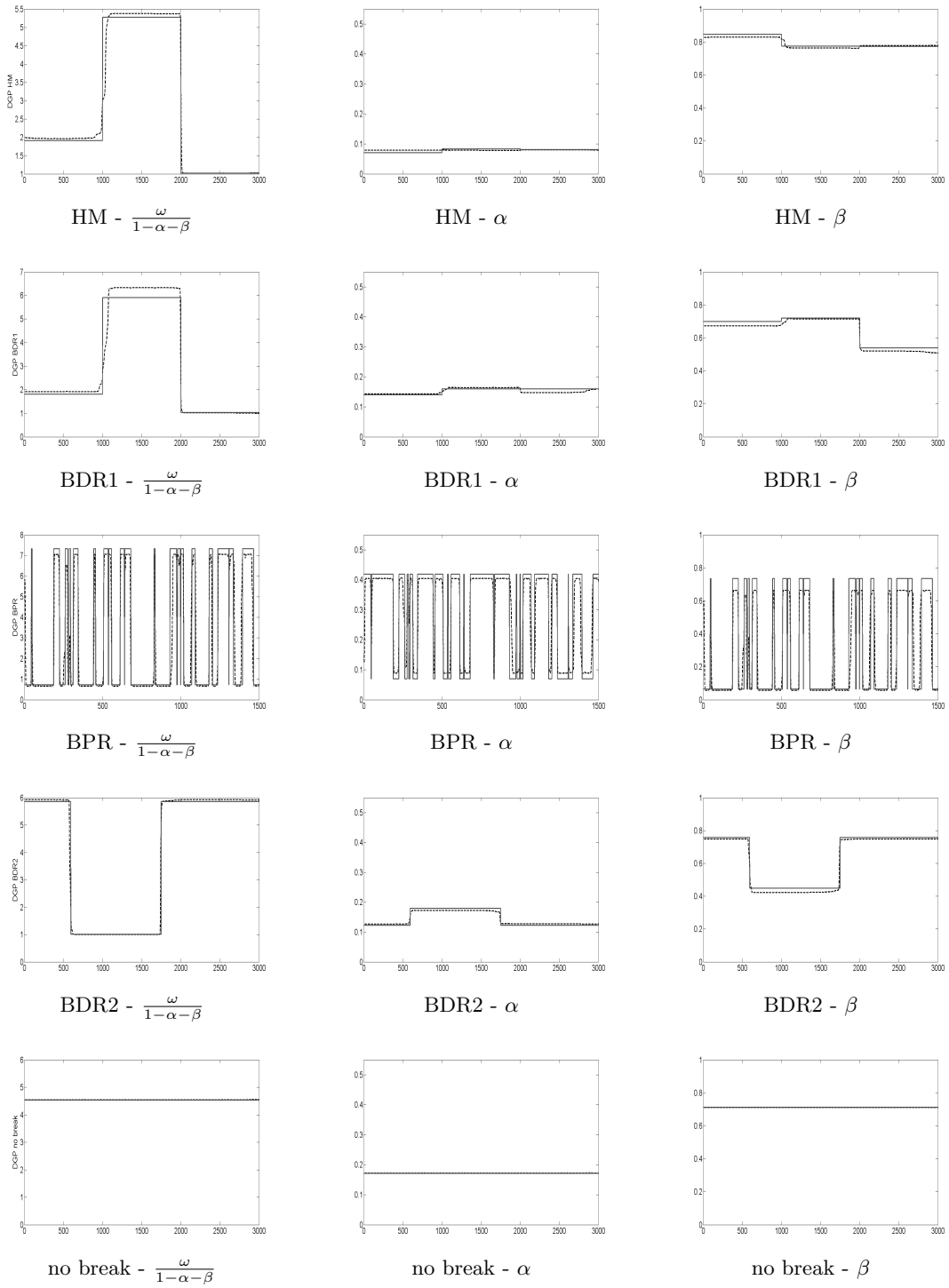
Throughout the Chapter we only report results derived from the Kl-MH algorithm since it provides better mixing properties than the HMP-MH algorithm. Nevertheless all the simulations have also been run with the HMP-MH algorithm. Results which are comparable could be available on request.

We present results of the IHMS-GARCH model on the different simulated data generated from the DGP of Table 4.3. Table 4.6 displays the posterior probability of having a specific number of regimes and for each simulation, probabilities are maximized at the true one. Also it is worth noticing that we never underestimate the number of regimes. However more regimes are sometimes counted but the algorithm quickly comes back to the true setting. As the likelihood does not decrease by adding more parameters, it is not surprising to observe such a pattern.

# of regimes	1	2	3	4	5
DGP_{HM}	0	0	**0.9812**	0.0176	0.0012
DGP_{BDR1}	0	0	**0.9152**	0.0812	0.0036
DGP_{BPR}	0	**0.9652**	0.0340	0.0008	0
DGP_{BDR2}	0	**0.9164**	0.0824	0.0012	0
$DGP_{noBreak}$	**0.9964**	0.0036	0	0	0

Table 4.6: Posterior probabilities of the number of regimes for five simulated series generated from DGP displayed in Table 4.3. The true number of regimes is bolded.

Figure 4.1 shows the posterior means of the parameters and the maximum likelihood estimates given the true states over time for each simulation. The IHMS-GARCH model closely tracks the MLE and sharply identifies the break points. It seems capable to reproduce Change-point and Markov-switching behaviors. Finally the acceptance rate for a new state vector does not decrease below 70 percent for all the simulated data.



HM - $\frac{\omega}{1-\alpha-\beta}$ HM - α HM - β

BDR1 - $\frac{\omega}{1-\alpha-\beta}$ BDR1 - α BDR1 - β

BPR - $\frac{\omega}{1-\alpha-\beta}$ BPR - α BPR - β

BDR2 - $\frac{\omega}{1-\alpha-\beta}$ BDR2 - α BDR2 - β

no break - $\frac{\omega}{1-\alpha-\beta}$ no break - α no break - β

Figure 4.1: ML estimates given the true states (solid grey lines) compared to the posterior means (dashed black lines). Results for each simulated data are presented in row in the same order as in Table 4.3 (i.e. HM, BDR1, BPR, BDR2 and no break). The first column of graphics displays the (local) unconditional variance ($\frac{\omega_t}{1-\alpha_t-\beta_t}|Y_T$) whereas the two others respectively show the parameters $\alpha_t|Y_T$ and $\beta_t|Y_T$

4.5 Illustration on financial time series

We have shown that the algorithm performs well for artificial data. We now turn to illustrate the IHMS-GARCH model on empirical time series. Detailed results on the $S\&P$ 500 daily index are documented in subsection 4.5.1. We next report a short forecast exercise on several financial time series that stresses the relevance of taking into account structural breaks. In Appendix 4.E.1, we provide posterior results for three commodities, namely Brent, Gold and Silver. These estimations are helpful to devolatilize the series in order to carry out the DCC estimation in Section 4.6.

4.5.1 $S\&P$ 500 daily index

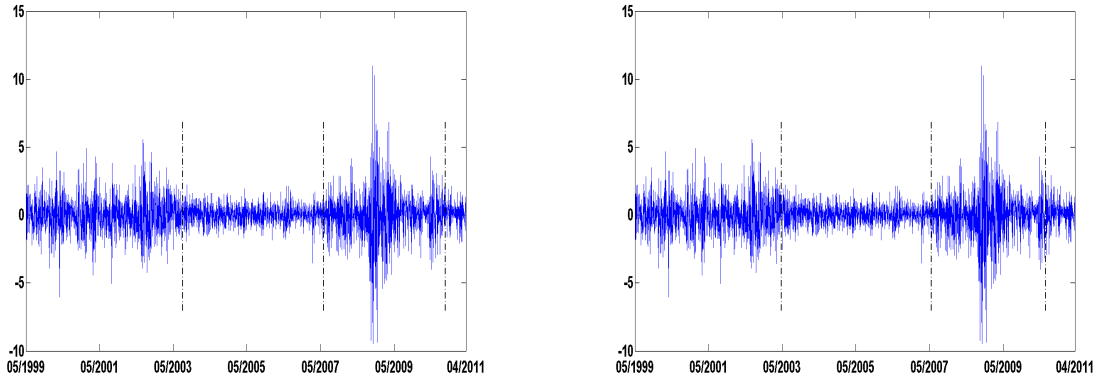
In this section we revisit the empirical exercise of Chapter 2. Recall that we have used a MS and a CP model on the $S\&P$ 500 daily percentage returns from May 20, 1999 to April 25, 2011 (3000 observations). We further have chosen the optimal model and the number of regimes with the marginal likelihood and have found evidence in favor of a Markov-switching model with two regimes.

The IHMS-GARCH posterior distribution covers five different numbers of regimes (2 to 7). The most observed one is 2. Table 4.7 displays the posterior distribution of the number of regimes.

# of regimes	1	2	3	4	5	6	7
Prob.	0	**0.6046**	0.2075	0.1455	0.0224	0.0196	0.0004

Table 4.7: Posterior probabilities of the number of regimes for the $S\&P$ 500 daily index

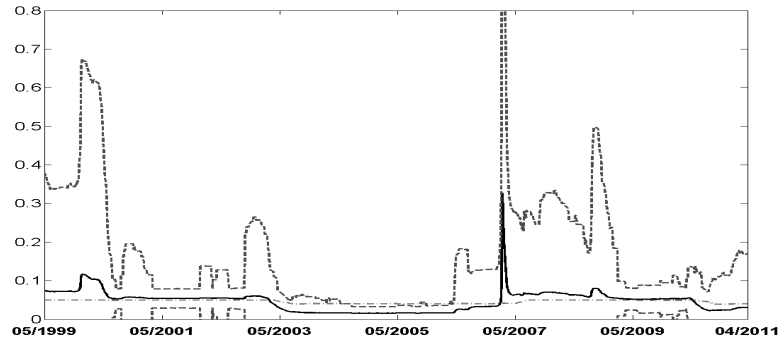
Figure 4.2 displays the time series with the estimated mode of the state vector at the most likely number of regimes. The regime switches occur at the same period as the best MS-GARCH model of section 2.5.3 in Chapter 2.



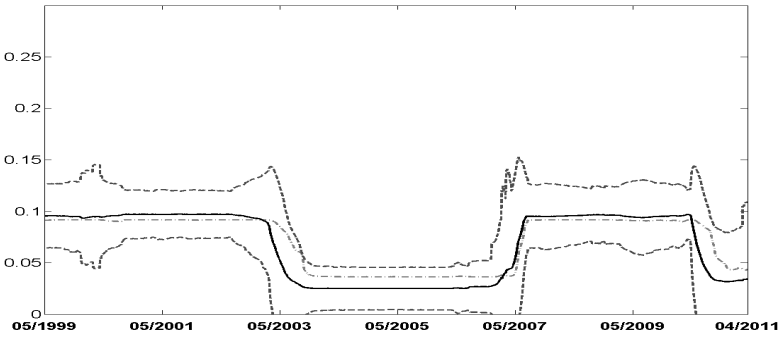
BDR - PMCMC IHMS-GARCH

Figure 4.2: The left graphic shows the $S\&P$ 500 daily index with structural breaks estimated by the PMCMC algorithm of BDR in vertical lines. The right one displays the same time series with structural breaks estimated by the IHMS-GARCH model.

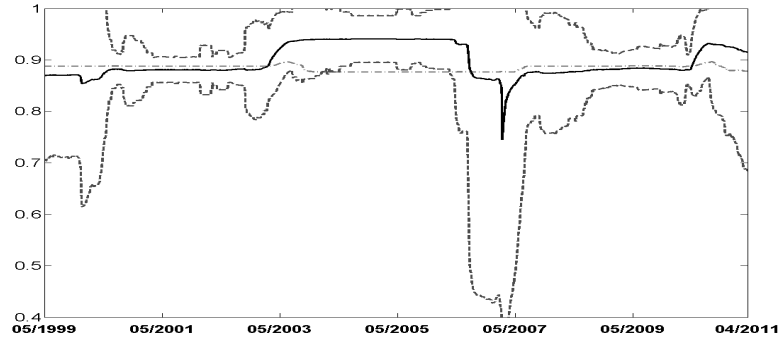
As label invariant statistic we show the posterior means of the parameters over time in Figure 4.3 below. We easily identify the regime switches on the graphic. A spike in the unconditional volatility occurs at February 27, 2007. The crisis had just begun to affect the financial sector five days before when HSBC, the world's largest bank at that time, laid off its US mortgage head for the loss of 10.5 billion dollar. It stresses the flexibility of the IHMS-GARCH model that accommodates extreme values. Also, the persistence of the volatility $(\alpha + \beta)$ is in some period smaller than 0.9. It is another evidence that capturing structural breaks decreases the persistence exhibited by financial time series. PMCMC posterior means are also reported and they only show small deviations from the estimated posterior means of the IHMS-GARCH model.



S&P 500 - $\frac{\omega}{1-\alpha-\beta}$



S&P 500 - α



S&P 500 - β

Figure 4.3: Posterior means of IHMS-GARCH parameters (black lines) with 95% percent confidence interval (dashed lines) for the *S&P* 500 daily index compared to the posterior means of PMCMC parameters (dash-dot grey lines) over time. The graphic at the top displays the (local) unconditional variance ($\frac{\omega_t}{1-\alpha_t-\beta_t}|Y_T$) while the two others respectively show the parameters $\alpha_t|Y_T$ and $\beta_t|Y_T$

4.5.2 Forecasting exercise

The IHMS-GARCH model exhibits relevant advantages compared to other MS-GARCH models regarding the forecasting approach. The number of regimes is inferred within the MCMC procedure. The predictions are therefore computed over different number of regimes without employing the Bayesian model averaging that requires the estimation of the marginal likelihood. Moreover the IHMS-GARCH model naturally takes into account new structural breaks in the forecasts since the probability of entering in a state that has never been visited so far is greater than zero. The parameters of the new regimes are then sampled from the hierarchical distribution on Θ which gathers information from all the previous regimes.

The IHMS-GARCH ability of forecasting the returns is explored on several financial time series. The exercise, comparable to Liu and Maheu (2008), consists in computing the iterated log predictive likelihood for a certain horizon at one given point.[2] The Table 4.8 reports the average of the cumulative sum of the log-predictive likelihood for four series. The IHMS model can behave as a CP or a MS specifications. Three different strategies are thus proposed for computing the predictions. First we condition to the last observed state and forecast the series as a change-point model. The MS model is mimicked by avoiding the creation of new regimes in the predictions. The last strategy does not modify the model and so allows for new regimes in the forecast period (IHMS in the Table). For the sake of comparison the average of the cumulative sum of the log-predictive likelihood of the GARCH model is also reported (GARCH(1,1) in Table 4.8).

Taking structural breaks into account clearly leads to more deviations in the forecasts but also to higher log-predictive likelihood posterior means. Following these prediction examples the IHMM CP model produces the best out-of sample fit with the lowest standard deviations among models exhibiting structural breaks. This observation is not astonishing since the transition probabilities are only computed using the past switching pattern without considering any economic information. The strategy inevitably gives birth to unrealistic switches of states in the forecasts that increase the deviation of the predictions and contribute to their deteriorations. Also the hierarchical structure on Θ could gather

[2]The given time varies with the financial time series.

Horizon (day)	1	2	4	8	12	16	20	40
				S&P 500 - May 20, 1999 to February 20, 2008				
IHMM	-1.68	-3.01	-6.06	-13.38	-19.99	-30.17	-44.19	-78.96
st. dev.	0.03	0.05	0.10	0.35	1.52	6.82	15.55	25.17
IHMM CP	**-1.68**	**-3.01**	**-6.06**	-13.38	-19.96	**-30.06**	**-43.92**	**-78.41**
st. dev.	0.03	0.05	0.12	0.40	0.62	1.70	4.18	5.79
IHMM MS	-1.68	-3.01	-6.06	-13.38	-19.96	-30.09	-44.01	-78.53
st. dev.	0.03	0.05	0.13	0.37	0.54	1.81	4.24	5.86
GARCH(1,1)	-1.68	-3.02	-6.06	**-13.37**	**-19.93**	-30.27	-45.00	-79.80
st. dev.	0.00	0.02	0.06	0.16	0.20	1.06	3.30	4.18
				GOLD - May 20, 1999 to July 14, 2008				
IHMM	-1.62	-2.90	-20.02	-27.49	-44.90	-53.10	-60.13	-98.08
st. dev.	0.04	0.07	3.26	3.68	7.16	7.95	8.37	11.94
IHMM CP	**-1.62**	-2.90	**-19.98**	**-27.43**	**-44.70**	**-52.85**	**-59.84**	**-97.39**
st. dev.	0.04	0.07	3.14	3.48	6.28	6.82	7.06	9.01
IHMM MS	-1.62	-2.90	-20.00	-27.46	-44.84	-53.02	-60.04	-97.90
st. dev.	0.04	0.07	3.20	3.57	6.79	7.48	7.83	10.81
GARCH(1,1)	-1.64	**-2.82**	-24.63	-32.60	-54.82	-63.91	-71.32	-112.81
st. dev.	0.00	0.03	2.20	2.44	5.65	6.33	6.67	8.87
				BAC - May 20, 1999 to June 06, 2008				
IHMM	-5.38	-7.34	-11.78	-28.93	-44.93	-113.47	-143.22	-227.78
st. dev.	1.02	0.99	1.14	6.13	11.14	56.78	73.83	108.39
IHMM CP	**-5.37**	**-7.33**	**-11.77**	**-28.70**	**-44.33**	**-109.02**	**-136.85**	**-215.47**
st. dev.	0.99	0.96	1.09	5.05	8.58	39.33	50.28	71.80
IHMM MS	-5.38	-7.34	-11.78	-28.98	-44.98	-113.35	-142.97	-227.45
st. dev.	1.04	1.01	1.18	6.76	12.26	59.63	76.98	112.56
GARCH(1,1)	-5.94	-7.81	-12.17	-30.21	-46.65	-118.72	-148.84	-230.93
st. dev.	0.00	0.02	0.05	0.98	1.94	12.55	16.33	23.59
				BA - May 20, 1999 to January 16, 2007				
IHMM	-1.26	-2.52	-5.31	-11.64	-16.87	-23.49	-29.06	-58.54
st. dev.	0.05	0.11	0.20	0.34	0.62	0.83	1.22	4.44
IHMM CP	**-1.26**	**-2.52**	**-5.31**	**-11.64**	**-16.85**	**-23.47**	**-29.02**	**-58.34**
st. dev.	0.05	0.10	0.19	0.32	0.55	0.67	0.88	1.94
IHMM MS	-1.26	-2.52	-5.31	-11.64	-16.86	-23.48	-29.04	-58.46
st. dev.	0.05	0.11	0.20	0.34	0.61	0.77	1.05	2.63
GARCH(1,1)	-1.48	-2.97	-6.11	-12.90	-19.00	-26.02	-32.39	-65.59
st. dev.	0.00	0.06	0.18	0.44	0.93	1.25	1.79	4.44

Table 4.8: Posterior means of the cumulative summation of the log-predictive likelihood and their corresponding standard deviations for different financial time series. Highest values are bolded. Description of the time series : S&P 500: Standard and Poors 500 index;GOLD (Gold Bullion LBM USD/Troy ounce); BAC: Bank of America Corporation; BA: Boeing Company.

extra information from the economic situation in order to draw more appropriate potential new regimes in the forecasts.

Regarding the predictions of the GARCH(1,1) model, it barely proposes better fit than the IHMM counterparts. To be precise, only three posterior means of the predictive likelihood densities out of twenty four are marginally the highest ones. On the contrary, the IHMM forecasts can staggeringly outperform the predictions of the GARCH(1,1) in some cases. These examples emphasize the usefulness of models including structural breaks at least for specific periods of time.

4.6 Multivariate extension

The presented methodology is not limited to the univariate GARCH model. As an example we incorporate an infinite hidden MS framework in the Dynamic Conditional Correlation model (IHMS-DCC). We choose the DCC model of Engle (2002) instead of another correlation model because it is generally estimated in two steps which is convenient in our context. The estimation procedure firstly extracts the conditional variance of each financial return using a volatility model and then estimates the conditional correlation dynamic on the devolatilized variables. It therefore allows to combine structural breaks in the volatility process and in the conditional correlation dynamic without having too many parameters to be estimated in one step.

Let $\mathbf{Y}_T = \{\mathbf{y}_1, ..., \mathbf{y}_T\}$ be a set of devolatilized variables where T denotes the sample size and $\mathbf{y}_t$ is a d-dimensional vector. The IHMS-DCC model is defined by equations (11)-(13) together with the hierarchical distributions defined by equations (4) to (9) in Section 4.2:

$$\mathbf{y}_t = \epsilon_t \quad \text{where} \quad \epsilon_t \sim N(0, R_t) \tag{4.11}$$

$$R_t = \text{diag}\{Q_t\}^{-\frac{1}{2}} Q_t \text{diag}\{Q_t\}^{-\frac{1}{2}} \tag{4.12}$$

$$Q_t = Q^*_{s_t}(1 - \alpha_{s_t} - \beta_{s_t}) + \alpha_{s_t}\epsilon_{t-1}\epsilon'_{t-1} + \beta_{s_t}Q_{t-1} \tag{4.13}$$

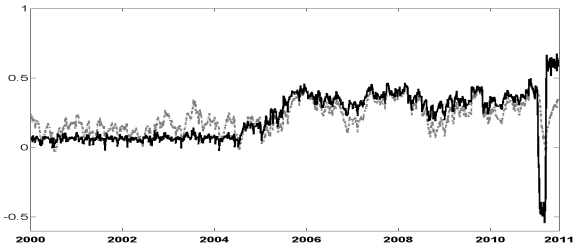
The constraint conditions are a definite positive matrix $Q^*_{s_t}$, $\alpha_{s_t} \in [0, 1]$ and $\beta_{s_t} \in [0, 1]$ for each regime to insure definite positive matrices.

The estimation procedure is very similar to the one exposed in Section 4.2. The parameters α and β are sampled by Metropolis steps while each matrix $Q^*_{s_t}$ is targeted to the sample covariance matrix of the regime. Sampling the matrices Q^* are feasible but it somehow complicates the MCMC sampler and is not of principal interest for the present discussion. We estimate the multivariate model on three commodity devolatilized percentage returns, namely BRENT, GOLD and SILVER (see Appendix 4.E.1 for the first step). The acceptance rate for the state vector amounts to 90.7 percent. Table 4.9 provides the posterior distribution of the number of regimes.

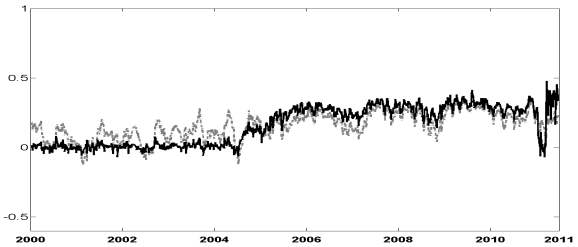
# of regimes	1	2	3	4	5	6	7	8
Prob.	0	0.0149	**0.2984**	0.2592	0.2419	0.0920	0.0860	0.0076

Table 4.9: Posterior probabilities of the number of regimes for the devolatilized daily commodities (IHMS-DCC model)

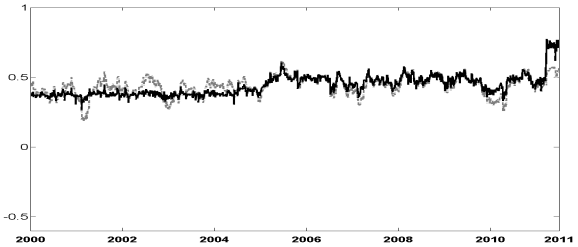
We display on figure 4.4 the posterior mean of each correlation over time. For the sake of comparison, we report the standard DCC correlations also based on the IHMS-GARCH devolatilized returns. The two dynamics are very similar although the standard DCC exhibits more volatile correlations during the period starting from 2000 to mid 2005 than the IHMS-DCC model. Over the same period of time, the IHMS-DCC correlations of the first two graphics ((a) and (b)) also show a lower level than the standard DCC correlations.



(a) Corr. BRENT - GOLD



(b) Corr. BRENT - SILVER



(c) Corr. GOLD - SILVER

Figure 4.4: IHMS-DCC(1,1) Correlation (in black) compared to DCC(1,1) correlations (in grey) for different commodities. Descriptions of the time series : BRENT (Crude Oil-Brent Current Month FOB USD/BBL), GOLD (Gold Bullion LBM USD/Troy ounce) and SILVER (Silver Fix LBM Cash Cents/Troy ounce)

The graphic of the IHMS-DCC parameters over time lies in Appendix 4.E.2. As a summary, the posterior mean of the persistence parameter $\alpha + \beta$ stabilizes around 0.63 over the period starting from 2000 to mid 2004. Afterwards it rises quickly to reach its highest value (0.96) at which it stays until the year 2010. At the end of the sample it sharply decreases to 0.38.

4.7 Conclusion

In this Chapter, we first propose an IHMS-GARCH model that allows for a potentially infinite number of regimes. Our Gibbs sampler relies on the sticky infinite hidden Markov model that has already been applied to autoregressive models but never to volatility models. Furthermore it accommodates the path dependence problem by using a novel Metropolis-Hastings method. As a result an entire state vector is sampled in a few blocks and with small autocorrelations. A comparison of the algorithm with existing MCMC alternatives confirms that the sampler outperforms the others in term of computational time although the mixing properties remain as good as the best known MCMC method. The IHMM encompasses the CP and the MS models. Some simulations in the Chapter show that the IHMS-GARCH model accurately estimates the two different specifications. We also detail results for the S&P500 from May 20, 1999 to April 25, 2011 (3000 observations). The number of regimes that is needed oscillates between 2 and 7. We highlight strong similarities with the MS-GARCH results of Chapter 2. A forecast exercise on four financial time series stresses the relevance of considering breaks at least for specific periods of time. The last section is devoted to a multivariate GARCH model in order to show the ability of the current algorithm to deal with more general configurations. We successfully estimate an IHMS-DCC model.

Appendices

4.A IHMS-GARCH Gibbs sampler : Implementation

Before developing the implementation of the sampler, we summarize some useful notations. The sum are denoted by dots. For instance $\sum_a x_{a,b} = x_{.,b}$ and $\sum_a \sum_b x_{a,b} = x_{.,.}$. The vector $\{x_1, x_2, ..., x_r\}$ is briefly denoted by $x_{1:r}$. Vectors are in row and the transpose operator is designated by $'$. The number K stands for the number of regimes. Some confusion can rise about the density function of the Gamma distribution. In the Chapter we always use the following one :

$$X \sim G(k, \theta) \quad \text{if} \quad f(x|k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{\frac{-x}{\theta}}.$$

After finding a starting point (see subsection 4.4.1), we iterate until convergence between the following steps.

1. Sampling U_T from $f(U_T|S_T, P)$: for t=2,...,T, sample $u_t \sim \tilde{U}[0, p_{s_{t-1},s_t}]$ and $u_1 \sim \tilde{U}[0, p_{s_1,s_1}]$ where $\tilde{U}[a, b]$ denotes the modified Uniform distribution described in 4.3.1 on the interval (a,b).

2. Generate new required states : while$(\max\{p_{i,K+1}\}_{i=1}^{K}) > \min(\{u_t\}_{t=1}^{T})$:

   - Sample $p_{K+1,1:K+1} \sim \text{Dir}(\lambda\pi_{1:K+1} + \kappa\delta_{K+1})$
   - Break the last stick of π :
     (a) Draw $\xi \sim \text{Beta}(1, \eta)$
     (b) Set $\pi_{K+2} = (1 - \xi)\pi_{K+1}$ and $\pi_{K+1} = \xi\pi_{K+1}$
   - Increase the dimension of each vector p_i : for i=1,...,K+1
     (a) Draw $\xi_i \sim \text{Beta}(\lambda\pi_{K+1} + \kappa 1_{\{i=K+1\}}, \lambda\pi_{K+2})$
     (b) Set $p_{i,K+2} = (1 - \xi_i)p_{i,K+1}$ and $p_{i,K+1} = \xi_i p_{i,K+1}$
   - Draw $\Theta_{K+1} \sim N(\mu, \Sigma)$
   - Set K = K+1.

3. Sampling S_T from $f(S_T|\Theta, P, H_{Dir}, \pi, U_T, Y_T) = f(S_T|\Theta, P, U_T, Y_T)$: see Section 4.3.1

4. According to the new vector S_T, remove the unvisited states and adapt K, π, Θ, P.

5. Sampling P from $f(P|S_T, \Theta, H_{Dir}, \pi, Y_T)$: for i=1,...,K, sample $p_{i,1:K+1} \sim \text{Dir}(\lambda\pi_1 + n_{i,1}, ..., \lambda\pi_i + \kappa + n_{i,i}, ..., \lambda\pi_{K+1})$ where $n_{i,j}$ denotes the number of transition from state i to j observed in the state vector S_T.

6. Sampling H_{Dir} from $f(H_{Dir}|\pi, \Theta, P, S_T, Y_T)$:

   (a) Introduce auxiliary variables :

   - Sampling m : For j=1,...,K, and k=1,...,K. Set $m_{j,k} = 0$. For i=1,...,$n_{j,k}$ sample $x_i \sim \text{Bernoulli}(\frac{\lambda\pi_k + \kappa 1_{\{j=k\}}}{i-1+\lambda\pi_k + \kappa 1_{\{j=k\}}})$ and increment $m_{j,k} = 0$ if $x_i = 1$.
   - Sampling r : For j=1,...,K. $r_j \sim \text{Binomial}(m_{j,j}, \frac{\rho}{(1-\rho)\pi_j + \rho})$ where $\rho = \frac{\lambda}{\lambda+\kappa}$

- set $\bar{m}_{j,k} = m_{j,k}$ if $j \neq k$ and $\bar{m}_{j,k} = m_{j,k} - r_j$ if $j = k$
- set $\bar{K} = 0$, for k=1,...,K, if $\bar{m}_{.,k} > 0$ then increment $\bar{K}$

(b) Sampling λ and κ

- Sample auxiliary variables : for i=1,...,K, $q_i \sim \text{Beta}(\lambda + \kappa + 1, n_{i,.})$ and $s_i \sim \text{Bernoulli}(\frac{n_{i,.}}{n_{i,.}+\lambda+\kappa})$
- Sample $\rho = \frac{\kappa}{\lambda+\kappa} \sim \text{Beta}(\rho_{\text{hyp1}} + r_., \rho_{\text{hyp2}} + m_{.,.} - r_.)$ where ρ_{hyp1} and ρ_{hyp2} denotes the hyper-parameters of ρ (see Table 4.2)
- Sample $\lambda + \kappa \sim \text{G}(a_{\text{hyp}} + m_{.,.} - s_., (\frac{1}{b_{\text{hyp}}} - \log q_.)^{-1})$ where a_{hyp} and b_{hyp} denotes the hyperparameters of $\lambda + \kappa$ (see Table 4.2)
- set $\lambda = (1 - \rho)(\lambda + \kappa)$ and $\kappa = \rho(\lambda + \kappa)$

(c) Sampling η

- Sample auxiliary variables : $\tilde{q} \sim \text{Beta}(\eta+1, \bar{m}_{.,.})$ and $\tilde{s} \sim \text{Bernoulli}(\frac{\bar{m}_{.,.}}{\bar{m}_{.,.}+\eta})$
- Sample $\eta \sim \text{G}(\eta_{\text{hyp1}}+\bar{K}-\tilde{s}, \{\frac{1}{\eta_{\text{hyp2}}}-\log \tilde{q}\}^{-1})$ where η_{hyp1} and η_{hyp2} denotes the hyper-parameters of η (see Table 4.2)

7. Sampling π from $f(\pi|H_{Dir}, \Theta, P, S_T, Y_T) \sim \text{Dir}(\bar{m}_{.,1}, \bar{m}_{.,2}, ..., \bar{m}_{.,K}, \eta)$.

8. Sampling Θ from $f(\Theta|S_T, \mu, \Sigma, Y_T)$ by delayed rejection adaptive Metropolis algorithm (see Haario, Saksman, and Tamminen (2001) and Mira (2001)).

9. Sampling μ from $f(\mu|\Sigma, \Theta) \sim N(\bar{\mu}, \bar{\Sigma})$ where $\bar{\mu} = \bar{\Sigma}(\sum_{i=1}^{K} \Sigma^{-1}\Theta_i + \underline{\Sigma}^{-1}\underline{\mu})$ and $\bar{\Sigma} = (K\Sigma^{-1} + \underline{\Sigma}^{-1})^{-1}$

10. Sampling Σ^{-1} from $f(\Sigma^{-1}|\mu, \Theta) \sim \text{Wishart}(\{\sum_{i=1}^{K}(\Theta_i-\mu)(\Theta_i-\mu)'+\underline{V}^{-1}\}^{-1}, \underline{v}+K)$

4.B Klaassen's approximations

First we derive the useful expression $f(s_{t-1}|s_t, Y_{t-1}, U_T, \Theta, P)$. This distribution will be used in the approximations derived next.

$$\begin{aligned} f(s_{t-1}|s_t, Y_{t-1}, U_T, \Theta, P) &= \frac{f(s_t, s_{t-1}, U^t|Y_{t-1}, U_{t-1}, \Theta, P)}{f(s_t, U^t|Y_{t-1}, U_{t-1}, \Theta, P)} \\ &= \frac{f(s_{t-1}|Y_{t-1}, U_{t-1}, \Theta, P) f(s_t|s_{t-1}, P) f(u_t|s_t, s_{t-1}, P) f(U^{t+1}|s_t, P)}{f(s_t, U^t|Y_{t-1}, U_{t-1}, \Theta, P)} \\ &= \frac{f(s_{t-1}|Y_{t-1}, U_{t-1}, \Theta, P) f(s_t|s_{t-1}, P) f(u_t|s_t, s_{t-1}, P)}{\sum_i^\infty f(s_{t-1}^i|Y_{t-1}, U_{t-1}, \Theta, P) f(s_t|s_{t-1}^i, P) f(u_t|s_t, s_{t-1}^i, P)} \\ &= \frac{f(s_{t-1}|Y_{t-1}, U_{t-1}, \Theta, P) \delta_{\{0 \le u_t \le p_{s_{t-1}, s_t}\}}}{\sum_i^\infty f(s_{t-1}^i|Y_{t-1}, U_{t-1}, \Theta, P) \delta_{\{0 \le u_t \le p_{s_{t-1}^i, s_t}\}}} \end{aligned}$$

where s_{t-1}^i stands for $s_{t-1} = i$ and we assume $f(u_t|s_t, s_{t-1}, P) \sim U[0, p_{s_{t-1}, s_t}]$.

4.B.1 Approximate GARCH : computation of the modified variance

We remind that $\tilde{\sigma}_{t-1,s_t}^2 = E[\sigma_{t-1}^2|Y_{t-1}, s_t, U_T, \Theta, P]$.

$$\begin{aligned} E[\sigma_{t-1}^2|Y_{t-1}, s_t, U_T, \Theta, P] &= \sum_{i=1}^{\infty} \sigma_{t-1,s_{t-1}^i}^2 f(s_{t-1}^i|s_t, Y_{t-1}, U_T, \Theta, P) \\ &= \sum_{i=1}^{\infty} (\omega_{s_{t-1}^i} + \alpha_{s_{t-1}^i} y_{t-2}^2 + \beta_{s_{t-1}^i} \tilde{\sigma}_{t-2,s_{t-1}^i}^2) f(s_{t-1}^i|s_t, Y_{t-1}, U_T, \Theta, P) \end{aligned}$$

where s_{t-1}^i stands for $s_{t-1} = i$.

4.B.2 Approximate DCC : computation of the modified Correlation matrix

We omit the condition to the sets of parameters Θ and P for saving space.

$$\begin{aligned} E[Q_{t-1}|\mathbf{Y}_{t-1}, s_t, U_T] &= \sum_{i=1}^{\infty} Q_{t-1,s_{t-1}^i} f(s_{t-1}^i|s_t, \mathbf{Y}_{t-1}, U_T) \\ &= \sum_{i=1}^{\infty}((Q^*_{s_{t-1}}(1-\alpha_{s_{t-1}^i}-\beta_{s_{t-1}^i})+\alpha_{s_{t-1}^i}\epsilon_{t-2}\epsilon'_{t-2}+\beta_{s_{t-1}^i}\tilde{Q}_{t-2,s_{t-1}^i}) \\ &\qquad f(s_{t-1}^i|s_t, \mathbf{Y}_{t-1}, U_T) \end{aligned}$$

4.C Comparison of the two MS-GARCH approximations

Section 4.3.1 details two algorithms to infer the parameters of an MS-GARCH model. These two algorithms differ on the approximation to the MS-GARCH model. A more accurate approximation will lead to a higher acceptance rate as well as a lower autocorrelation between posterior draws. In order to differentiate the two algorithms, we carry out a Monte Carlo study based on simulated data of 1000 observations and analyze the mixing properties. For each simulation we compute the autocorrelation time[3] as well as the required time for sampling one effective posterior draw.[4] We consider four different MS-GARCH models exhibiting a break in ω with one or multiple switches (CP and MS in the Table 4.10). The data generated processes differ on their persistence ($\alpha+\beta$) that are assumed to be equal across regimes. The Monte Carlo study consists in eight hundred different simulated data (one hundred per DGP per algorithm). The main interest of this study lies in the ability of sampling the state vector. These MCMC simulations therefore fix the GARCH parameters at their MLE and only draw the state vector given these parameters. The Table 4.10 displays the difference average of the maximum autocorrelation time and the difference average of the effective time for the two algorithms (Kl-MH for the model of Klaassen (2002) minus HMP-MH for the model of Haas, Mittnik, and Paolella (2004)).

[3]The autocorrelation time is computed by batch means (see Geyer (1992)) and is defined as $1+2\sum_{i=1}^{\infty}\rho_i$ where ρ_i is the autocorrelation coefficient of order i between the posterior draws of a state variable.

[4]following the formula : autocorrelation time x elapsed time for N posterior draws divided by N

Break in ω	$\omega_1 = 0.1$		$\omega_2 = 0.7$	
Persistence	0.7	0.8	0.9	0.95
	90 % CI of the difference between elapsed times for one effective draw			
CP	[-0.23;0.02]	[-0.54;0.02]	[-0.75;0.01]	[-1.10;0.01]
MS	[-0.85;0.15]	[-1.05;0.01]	[-1.19;0.05]	[-1.35;0.29]
	90 % CI of the difference between autocorrelation times			
CP	[-20.91;1.06]	[-49.03;1.65]	[-68.34;0.75]	[-90.24;0.11]
MS	[-79.05;6.55]	[-96.63;-1.10]	[-111.96;-1.08]	[-109.71;10.15]

Table 4.10: The differences are as follows Kl-MH minus HMP-MH where Kl-MH denotes 'Klaassens Metropolis-Hastings' and HMP-MH stands for 'Haas, Mittnik and Paolella Metropolis-Hastings'. A negative value provides evidence in favor of the Kl-MH algorithm.

Although almost all confidence intervals include zero, the distributions are skewed to the left whatever the persistence and the number of breaks. The left limit also increases if the persistence and/or the number of switches grow while no systematic rule can be depicted from the table for the right limit. These two observations leads us to believe that the Klaassen model provides a better approximation to the MS-GARCH model than the specification of Haas, Mittnik, and Paolella (2004). Another evidence is given in the simulation exercise (see 4.4.3). The result is not surprising since the former model keeps track of the preceding variances when a switch in the state occurs. Despite these comments, the HMP model could still be preferred in the IHMM context. Indeed the beam sampler combined with the Klaassen model somewhat complicates the sampling of the state vector. These computational difficulties do not arise with the HMP model.

4.D The sticky infinite hidden Markov model

The sticky infinite hidden Markov model is based on Dirichlet processes and hierarchical Dirichlet processes. The Section 4.2 defines the Dirichlet process and its stick-breaking representation. We go further by reviewing the concept of hierarchical Dirichlet process and the sticky infinite hidden Markov model.

4.D.1 The Hierarchical Dirichlet process

The infinite hidden Markov Model assumes an infinite number of states. It models a Markov chain that can move from one state to any other state and it should be the case that each state is linked to others states by the same set of states. For instance the state one should always be related to the same parameter Θ_1. The hierarchical Dirichlet process has been designed on this purpose. The hyper-parameters of the hierarchical Dirichlet process (HDP), (Teh, Jordan, Beal, and Blei (2006)) consist of the base distribution G_0 and concentrated parameters $\eta \in \Re^+$ and $\lambda \in \Re^+$. The HDP is defined as follows

$$G|\eta, G_0 \sim DP(\eta, G_0) \quad \text{and} \quad G_j|\lambda, G \sim DP(\lambda, G) \quad \forall j = 1, \ldots, n$$

So $G_j|G \perp G_i$ if $i \neq j$. As G is a random probability measures over Θ (the support of the base distribution G_0), the hierarchical process defines a set of random probability measures G_j, one for each group, over Θ. The stick-breaking representation of a HDP can be formulated as follows :

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\Theta_k} \quad \text{and} \quad G_j = \sum_{k=1}^{\infty} p_{jk} \delta_{\Theta_k} \quad \text{where} \quad \forall j = 1, \ldots, n$$

where $\Theta_k \sim G_0$, $\pi = \{\pi_k\}_{k=1}^{\infty} \sim Stick(\eta)$ are mutually independent, δ_{Θ_k} is the probability measure concentrated at Θ_k and $\{p_{jk}\}_{k=1}^{\infty}|\lambda, \pi \sim DP(\lambda, \pi)$ (as shown in Teh, Jordan, Beal, and Blei (2006)). Notice that by definition of the DP, each G_j ($\forall j \in \{1, ..., n\}$) has the same support which is the support of G. This property of the HDP is essential to develop an infinite hidden Markov model.

The hidden Markov-switching model is driven by two stochastic processes. On one hand a Markov-chain determines a discrete state vector $\{s_1, ..., s_T\}$ and on the other hand the observations follow a specific distribution conditioned to the state vector and the parameters of each regime ($y_t|s_t, \{\Theta_k\}_{k=1}^{\infty} \sim F(\Theta_{s_t})$). The hierarchical Dirichlet process can build this kind of structure with an infinite number of state (and of Dirichlet processes) and is shortly stated in Table 4.11.

1. Dirichlet process : $G = \sum_{k=1}^{\infty} \pi_k \delta_{\Theta_k} \sim DP(\eta, G_0)$	
$\pi \sim \text{Stick}(\eta)$	Stick-breaking representation of the Dirichlet Process
$\Theta_k \sim G_0$	Θ_k : Parameters of the model related to the state k
2. Hierarchical Dirichlet processes : $G_j \vert G = \sum_{k=1}^{\infty} p_{jk} \delta_{\Theta_k} \sim DP(\lambda, G)$	
$p_j = \{p_{jk}\}_{k=1}^{\infty} \sim \text{DP}(\lambda, \pi)$	Each row of the transition matrix is driven by a DP
3. Markov-switching model	
$s_t \vert s_{t-1}, \{p_j\}_{j=1}^{\infty} \sim p_{s_{t-1}}$	First order Markovian with transition matrix $\{p_j\}_{j=1}^{\infty}$
$y_t \vert s_t, \{\Theta_k\}_{k=1}^{\infty} \sim F(\Theta_{s_t})$	Each state shares the same support (of G_0)

Table 4.11: Infinite hidden Markov Model (IHMM)

4.D.2 The sticky parameter

Persistence of regimes is a well-known stylized fact of time series. However the IHMM probability transition matrix does not exhibit any persistence (i.e. $E[p_{jk}|\lambda, \pi] = \pi_k \quad \forall j$ (see Table 4.11)). The IHMM transition actually does not differ between a self-transition and a transition to another state, an unrealistic feature for time series.

Fox, Sudderth, Jordan, and Willsky (2007) have developed a IHMM framework which excludes a high probability posterior with rapid switching. They called it 'the sticky HDP-HMM' or 'the sticky IHMM'. They specify a new parameter κ for self-transition bias and set a separate prior on this parameter. Their specification is as follows :

$$\begin{aligned} \pi|\eta &\sim Stick(\eta) \\ \forall j = 1, \ldots, n \quad p_j|\lambda, \pi, \kappa &\sim DP(\lambda + \kappa, \frac{\lambda\pi + \kappa\delta_j}{\lambda + \kappa}) \end{aligned}$$

An amount $\kappa > 0$ is added to the j^{th} component of the (infinite) vector $\lambda\pi$. The new parameter implies a higher probability of staying in the same state in the next period than the original model (i.e. $E[p_{kk}|\lambda, \kappa, \pi] = \frac{\lambda\pi_k + \kappa}{\lambda + \kappa}$). Note that if $\kappa = 0$, we come back to the former specification (i.e the original IHMM of Table 4.11).

4.E Multivariate extension

4.E.1 First step : Devolatilization

We briefly provide results on three other financial time series. These estimations are helpful for the multivariate correlation model of section 4.6 that requires devolatilized time series as inputs. We shortly summarize in Table 4.12 the posterior distributions of three commodity percentage returns from July 3, 2000 to December 30, 2011 on a daily basis (3000 observations). Detailed results are available on request.

Series	Number of regimes			Local Unc. Var.			Local Persistence ($\alpha + \beta$)		
	Min.	Mode	Max.	Min.	Mean	Max.	Min.	Mean	Max.
BRENT	2	3	8	2.86	5.48	31.81	0.65	0.90	0.97
GOLD	4	5	11	0.26	1.61	8.18	0.09	0.54	0.97
SILVER	5	6	11	0.45	5.84	130.25	0.21	0.65	0.97

Descriptions of the time series : BRENT (Crude Oil-Brent Current Month FOB USD/BBL), GOLD (Gold Bullion LBM USD/Troy ounce) and SILVER (Silver Fix LBM Cash Cents/Troy ounce)

Table 4.12: Summaries of posterior distributions

Figure 4.5 shows the devolatilized commodity returns with respect to the GARCH(1,1) and the IHMS-GARCH(1,1) models. The graphic emphasizes the lack of flexibility of the GARCH(1,1) model since some values are extreme compared to the standard Normal distribution that they should follow. On the contrary the IHMS-GARCH(1,1) devolatilized returns does not exhibit so much extreme values.

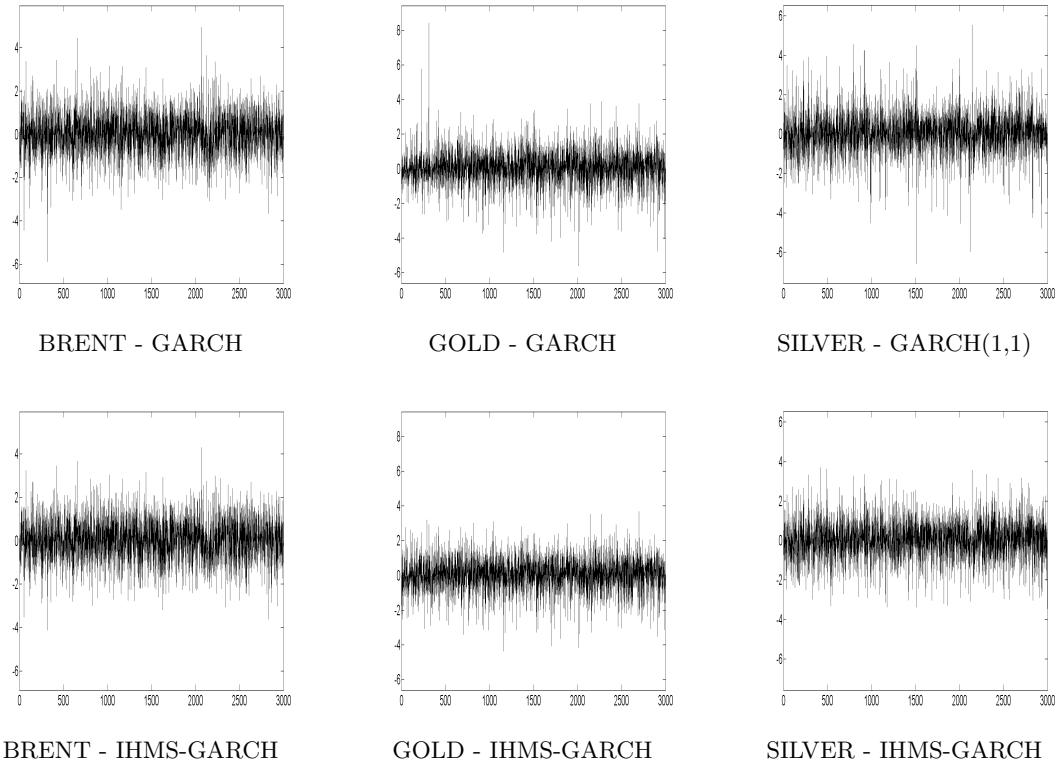
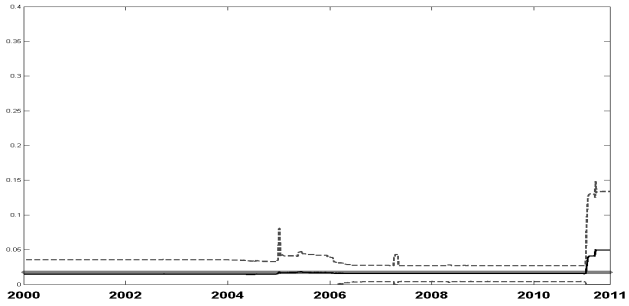


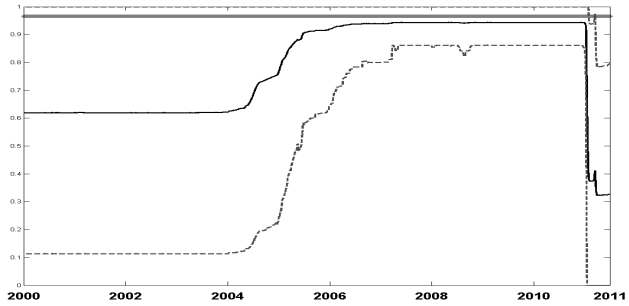
Figure 4.5: Devolatilized financial time series with respect to the standard GARCH(1,1) (top) and to the IHMS-GARCH(1,1) (bottom) models.

4.E.2 More MS-DCC results

The graphic 4.6 displays the IHMS-DCC parameters over time along with their respective standard deviations and the posterior mean of the standard DCC parameters. The parameter $\alpha_t|Y_T$ does not change over time except at the end of the sample. It is close to the value of the standard DCC parameter. On the other hand the parameter $\beta_t|Y_T$ exhibits structural breaks. It sharply increases in 2004 to reach its highest level (0.96) in 2007. At the end of the time series its value drops to 0.33.



$\alpha_t|Y_T$



$\beta_t|Y_T$

Figure 4.6: Posterior means of IHMS-DCC parameters (black lines) with 95% percent confidence interval (dashed lines) for three commodities (Brent,Gold and Silver) compared to the posterior means of standard DCC parameters (grey lines) over time. The graphic at the top displays the parameter $(\alpha_t|Y_T)$ while the other shows the parameter $\beta_t|Y_T$

5

Conclusion

The estimation of econometric models ignoring the presence of possible structural breaks can result in multiple drawbacks. Importantly, it can produce poor forecasts as a result of the bias in the estimates and often give the spurious impression of a nearly integrated behaviour of the data (see, e.g., Diebold (1986), Lamoureux and Lastrapes (1990), Hillebrand (2005)). In order to tackle these issues, the present thesis discusses numerous aspects of abrupt switching models in the context of financial time series. More precisely, it concentrates on structural breaks in the volatility process.

Since structural break inferences in volatility can be challenging, three new estimation methods have been developed and widely documented. The first one, based on a Sequential Monte Carlo algorithm embedded to an MCMC, provides *off-line* inferences of CP- and MS-GARCH models. The method, though computationally demanding, allows for the Marginal Log-Likelihood (MLL) estimation that can be used to select the optimal number of regimes. A sensitivity analysis with respect to the prior density has been carried out and empirically demonstrates the MLL ability in selecting the right number of regimes. The Chapter 3 details a generic MCMC method to infer on CP and recurrent regime models. It constitutes a competing alternative to the standard method consisting in embedding the model with a latent Markov-chain hidden variable. Simulation exercises confirm the MLL ability in selecting the right number of regimes. The Chapter 4 generalizes CP and MS specifications by using the hierarchical Dirichlet process framework. A novel estimation, not specific to Dirichlet processes, is derived from the models of Klaassen (2002)

and Haas, Mittnik, and Paolella (2004). Recall that although the manuscript focuses on univariate and multivariate GARCH model, these algorithms could be extended to other path dependence processes such as Auto-Regressive Moving Average models.

Evidences of structural breaks in volatility processes of financial time series are empirically observed in each Chapter. Incorporating abrupt changes in parameters highly improves the in-sample fit for many financial time series. However contributions of structural break models are more disappointing in forecast exercises as detailed in Chapter 3 and 4. Given the results in the thesis, whether structural break models really help to improve the predictions remains an open question.

Further research could be devoted to this question. CP and MS models exhibit a large parameter set and uncertainties on parameters can be large especially for short regimes. Encouraging sparsity by using a double exponential or a Normal-gamma prior (Griffin and Brown (2010)) could improve the forecast abilities of these models. Forecast comparisons with competing models, such as the very general infinite mixture model of Griffin and Steel (2011), could also be carried out.
Furthermore, other empirically relevant issues are first the effect of relaxing the Gaussian innovation assumption, and second the impact of more complex volatility functions than the standard or the GJR-GARCH specifications.

Bibliography

Amado, C., and T. Teräsvirta (2008): "Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure," CREATES Research Papers 2008-08, University of Aarhus.

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.

Ardia, D., L. Hoogerheide, and H. van Dijk (2009): "AdMit : Adaptive Mixtures of Student-t Distributions," *The R Journal*, 1, 25–30.

Augustyniak, M. (2013): "Maximum Likelihood Estimation of the Markov-switching GARCH Model," *Computational Statistics and Data Analysis, forthcoming.*

Bachelier, L. (1900): "Théorie de la spéculation," *Thèse de doctorat.*

Bai, J., and P. Perron (1998): "Testing for and Estimation of Multiple Structural Changes," *Econometrica*, 66, 47–79.

——— (2003): "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*, 18, 1–22.

Balakrishnan, N., and V. B. Nevzorov (2005): "Dirichlet Distribution, in A Primer on Statistical Distributions," *John Wiley and Sons, Inc., Hoboken, NJ, USA., chapter 27.*

Bauwens, L., A. Dufays, and B. De Backer (2011): "Estimating and forecasting structural breaks in financial time series," *CORE discussion paper, 2011/55*.

Bauwens, L., A. Dufays, and J. V. K. Rombouts (2011): "Marginal Likelihood for Markov-switching and Change-Point GARCH Models," *CORE discussion paper, 2011/13*.

Bauwens, L., G. Koop, D. Korobilis, and J. V. K. Rombouts (2011): "The Contribution of Structural Break Models to Forecasting Macroeconomic Series," *The Rimini Centre for Economic Analysis, Working Paper 11-38*.

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

Beal, M. J., and P. Krishnamurthy (2006): "Gene expression time course clustering with countably infinite hidden Markov models," *Proceedings of the Conference on Uncertainty in Artificial Intelligence*.

Black, F., and M. Scholes (1973): "The pricing of options and corporate liabilities," *Journal of Political Economy*, 81, 637–654.

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.

Carlin, B. P., and S. Chib (1995): "Bayesian Model Choice via Markov chain Monte Carlo Methods," *Journal of the Royal Statistical Society*, B 57, 473–484.

Carlin, B. P., A. Gelfand, and A. Smith (1992a): "Hierarchical Bayesian analysis of change-point problems," *Applied Statistics*, 41, 389–405.

Carlin, B. P., A. E. Gelfand, and A. F. M. Smith (1992b): "Hierarchical Bayesian Analysis of Changepoint Problems," *Journal of the Royal Statistical Society*, C 41, 389–405.

Carpantier, J.-F., and A. Dufays (2012): "Commodities Volatility and the Theory of Storage," *CORE discussion paper, 2012/37*.

Carter, C. K., and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Chernoff, H., and S. Zacks (1964): "Estimating the current mean of a Normal distribution which is subject to changes in time," *Annals of Mathematical Statistics*, 35, 999–1018.

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

——— (1995b): "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, 327–335.

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.

Chib, S., and S. Ramamurthy (2010): "Tailored randomized block MCMC methods with application to DSGE models," *Journal of Econometrics*, 155(1), 19–38.

Del Negro, M., and F. Schorfheide (2011): "Bayesian Macroeconometrics," in *The Oxford Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. van Dijk, chap. 7, pp. 293–389. OUP, Oxford.

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

Dufays, A. (2012): "Infinite-State Markov-switching for Dynamic Volatility and Correlation Models," *CORE discussion paper, 2012/43*.

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

——— (2002): "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

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.

Ferguson, T. S. (1973): "A bayesian analysis of some nonparametric problem," *The annals of Statistics*, 1, 209–230.

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.

Fox, E., E. Sudderth, M. Jordan, and A. Willsky (2007): "The Sticky HDP-HMM : Bayesian Nonparametric Hidden Markov Models with Persistent States," *MIT Laboratory for Information and Decision Systems*, Tech.Rep.2777.

——— (2008): "An HDP-HMM for systems with state persistence," in *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5-9, 2008*, ed. by W. W. Cohen, A. McCallum, and S. T. Roweis, vol. 307 of *ACM International Conference Proceeding Series*, pp. 312–319. ACM.

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

Gelfand, A. E., and A. Smith (1990): "Sampling based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398–409.

Gelman, A., and D. Rubin (1992): "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472.

Geman, S., and D. Geman (1984): "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Patterni Analysis and Machine Intelligence*, 6, 721–741.

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.

Geyer, C. J. (1992): "Practical Markov Chain Monte Carlo," *Statistical Science*, 7, 473–511.

Giordani, P., and R. Kohn (2008): "Efficient Bayesian inference for multiple change-point and mixture innovation models," *Journal of Business and Economic Statistics*, 26, 66–77.

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

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

Green, P. J. (1995): "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

Griffin, J., and P. Brown (2010): "Inference with normal-gamma prior distributions in regression problems," *Bayesian Analysis*, 5, 171–188.

Griffin, J., and M. Steel (2011): "Stick-breaking autoregressive processes," *Journal of Econometrics*, 162(2), 383–396.

Haario, H., E. Saksman, and J. Tamminen (2001): "An Adaptive Metropolis Algorithm," *Bernoulli*, 7, 223–242.

Haas, M., S. Mittnik, and M. Paolella (2004): "A New Approach to Markov-Switching GARCH Models," *Journal of Financial Econometrics*, 2, 493–530.

Hamilton, J. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

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.

Henneke, J., S. Rachev, F. J. Fabozzi, and M. Nikolov (2011): "MCMC-based estimation of Markov Switching ARMA-GARCH models," *Applied Economics*, 43(3), 259–271.

Hillebrand, E. (2005): "Neglecting Parameter Changes in GARCH Models," *Journal of Econometrics*, 129, 121–138.

Jochmann, M. (2010): "Modeling U.S. Inflation Dynamics : A Bayesian Nonparametric Approach," *Working Paper Series*.

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.

Kivinen, J., E. Sudderth, and M. Jordan (2007): "Learning multiscale representations of natural scenes using Dirichlet processes," *Proceedings of the IEEE International Conference on Computer Vision*.

Klaassen, F. (2002): "Improving GARCH Volatility Forecasts with Regime-Switching GARCH," *Empirical Economics*, 27, 363–394, Universiteit Tubingen, Working Paper.

Koop, G., and S. Potter (2007): "Estimation and Forecasting with Multiple Breaks," *Review of Economic Studies*, 74, 763–789.

Lamoureux, C., and W. Lastrapes (1990): "Persistence in Variance, Structural Change, and the GARCH Model," *Journal of Business and Economic Statistics*, 8, 225–234.

Liu, C., and J. M. Maheu (2008): "Are there structural breaks in realized volatility?," *Journal of Financial Econometrics*, 6(3), 326–360.

Liu, J. S., W. H. Wong, and A. Kong (1994): "Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of the Estimators and Data Augmentation Schemes," *Biometrika*, 81, 27–40.

Malkiel, B. G., and E. F. Fama (1970): "Efficient Capital Markets: A review of theory and empirical work," *The journal of Finance*, 25, 383–417.

Mandelbrot, B. (1963): "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.

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.

Markowitz, H. M. (1959): "Portfolio Selection : Efficient Diversification of Investments," *Yale University Press*.

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.

Mincer, J., and V. Zarnowitz (1969): *The Evaluation of Economic Forecasts*Economic Forecasts and Expectations. in J.Mincer, New York: National Bureau of Economic Research.

Mira, A. (2001): "On Metropolis-Hastings algorithms with delayed rejection," *Metron*, 59, 3–4.

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

Neal, R. (2003): "Slice sampling," *The Annals of Statistics*, 31, 705–741.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057–1084.

Pesaran, M. H., and A. Timmermann (2004): "How costly is it to ignore breaks when forecasting the direction of a time series?," *International Journal of Forecasting*, 20, 411–425.

Pitman, J. (2006): *Combinatorial stochastic processes*, vol. 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.

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.

Rabiner, L. R. (1989): "A tutorial on hidden Markov models and selected applications in speech recognition," pp. 257–286.

Roberts, G. O., and J. S. Rosenthal (2001): "Optimal scaling for various Metropolis-Hastings algorithms," *Statistical Science*, 16, 351–367.

Sethuraman, J. (1994): "A constructive definition of Dirichlet Priors," 4, 639–650.

Shephard, N. (1994): "Partial Non-Gaussian State Space," *Biometrika*, 81, 115–131.

Sinharay, S., and H. S. Stern (2002): "On the Sensitivity of Bayes Factors to the Prior Distributions," *The American Statistician*, 56, 196–201.

Song, Y. (2011): "Modelling Regime Switching and Structural Breaks with an Infinite Dimension Markov Switching Model," *Working Paper Series*.

Stephens, D. (1994a): "Bayesian retrospective multiple-changepoint identification," *Applied Statistics*, 43, 159–178.

Stephens, D. A. (1994b): "Bayesian Retrospective Multiple-changepoint Identification," *Journal of the Royal Statistical Society*, C 43, 159–178.

Stock, J. H., and M. W. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14, 11–30.

Tanner, M., and W. Wong (1987): "The Calculation of the Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Teh, Y., M. Jordan, M. Beal, and D. M. Blei (2006): "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, 101, 1566–1581.

ter Braak, C. J. F. (2006): "A Markov Chain Monte Carlo Version of the Genetic Algorithm Differential Evolution: Easy Bayesian Computing For Real Parameter Spaces," vol. 16, pp. 239–249.

Van Gael, J., Y. Saatci, Y. Teh, and Z. Ghahramani (2008): "Beam sampling for the infinite hidden Markov model," *Proceedings of the 25th International Conference on Machine learning*.

Vrugt, J. A., and C. J. F. ter Braak (2011): "$DREAM_{(D)}$: an Adaptive Markov Chain Monte Carlo Simulation Algorithm to Solve Discrete, Noncontinuous, and Combinatorial Posterior Parameter Estimation Problems," *Hydrology and Earth System Sciences*, 15, 3701–3713.

Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009): "Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptative Randomized Subspace Sampling," *International Journal of Nonlinear Sciences and Numerical Simulations*, 10, 271–288.

Whiteley, N., C. Andrieu, and A. Doucet (2011): "Bayesian Computational Methods for Inference in Multiple Change-points Models," *Discussion paper, University of Bristol*.