2012/43

Infinite-state Markov-switching for dynamic volatility and correlation models

Arnaud Dufays

CORE

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium
http://www.uclouvain.be/core

CORE DISCUSSION PAPER
2012/43

Infinite-state Markov-switching for dynamic volatility and correlation models

Arnaud DUFAYS[1]

November 2012

Abstract

Dynamic volatility and correlation models with fixed parameters are restrictive for time series subject to breaks. GARCH and DCC models with changing parameters are specified using the sticky infinite hidden Markov-chain framework. Estimation by Bayesian inference determines the adequate number of regimes as well as the optimal specification (Markov-switching or change-point). The new estimation algorithm is studied in terms of mixing properties and computational time. Applications highlight the flexibility of the model.

Keywords: Bayesian inference, Markov-switching, GARCH, DCC, infinite hidden Markov model, Dirichlet process

JEL Classification: C11, C15, C22, C58

[1] Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: arnaud.dufays@uclouvain.be. Website: https://sites.google.com/site/websiteofarnauddufays/

The author warmly thanks Luc Bauwens for his precious advices and for his support. He is also grateful to Jeroen Rombouts for useful comments that increased the quality of the manuscript. Research supported by the contract "Projet d'Actions de Recherches Concertées" 12/17-045 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain". The author is a F.S.R. Research Fellow and assumes the scientific responsibility.

1 Introduction

Univariate and multivariate GARCH (generalized auto-regressive conditional heteroskedastic models) are widely used to forecast volatilities and correlations. However GARCH models with fixed parameters are restrictive since financial time series are prone to exhibit breaks, especially over long periods. From time to time, the level of volatilities increases sharply, due to financial tensions, and decreases when they end, and these changes also affect the levels of correlations. Ignoring these breaks by assuming constant parameters in econometric models typically leads to forecasts that are far from realizations (see for example Stock and Watson (1996)) and often gives the spurious impression of a nearly integrated property of the time series (see, e.g., Diebold (1986), Lamoureux and Lastrapes (1990), and Hillebrand (2005)). An important econometric challenge is therefore to detect a structural break as soon as possible. An interesting way to introduce breaks in GARCH models is enriching them with a dynamic discrete latent state Markov process (say S_T), in such a way that the parameters can abruptly switch from one value to another. These models are called Markov-switching (MS) GARCH models when the Markov chain is recurrent (see e.g. Francq and Zakoian (2008) and Bauwens, Preminger, and Rombouts (2010)) and change-point (CP) GARCH models, see e.g. He and Maheu (2010) and Bauwens, Dufays, and De Backer (2011), when the states are not recurrent. Estimation of these models by the method of maximum likelihood is numerically infeasible, due to the path dependence problem. This occurs because the conditional variance at time t depends on the entire sequence of regimes visited up to time t. Bayesian estimation by a MCMC algorithm is practicable, by embedding the vector of states in the parameter space, and therefore simulating them, as done by Bauwens, Preminger, and Rombouts (2010)) and Bauwens, Dufays, and Rombouts (2011)).

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 (see Bauwens, Dufays, and Rombouts (2011)). 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. Our main contribution is to develop Bayesian inference for sticky infinite hidden Markov-GARCH and DCC models, thus for models *with* path dependence.

More precisely, our contribution is twofold. Though one can apply the forward-backward algorithm (Rabiner (1989), Hamilton (1989) and Chib (1996)) within a Gibbs sampler for inferring on a model with structural breaks[1], this does not apply to models subject to path dependence. We circumvent the issue by employing a Metropolis-Hastings algorithm, where the proposal density is based on the model of Klaassen (2002), which is used as an approximation to the MS-GARCH model[2]. We show that this preserves the invariance of the posterior distribution. We additionally illustrate that the new algorithm outperforms the existing MCMC alternatives in term of computational time, while the mixing properties remain competitive with respect to the Particle MCMC algorithm (PMCMC) of Bauwens, Dufays, and Rombouts (2011). 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 algorithm 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

[1] Hamilton (1989) used a forward algorithm to integrate the latent variables S_T and by doing so, was able to compute the likelihood at any parameter value. Chib (1996) embedded a backward step in the algorithm in order to ease the sampling of the latent variables. These methods rely on the assumption that the likelihood at time t only depends on the current state (i.e. no path dependence).

[2] MS-GARCH models that circumvent the path dependence problem were proposed by Gray (1996), Klaassen (2002) and Haas, Mittnik, and Paolella (2004).

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).

In the next section, we present the infinite hidden Markov-switching GARCH model (IHMS-GARCH) and the Bayesian estimation algorithm of its parameters. In Section 4, we compare the new algorithm with existing methods and illustrate this on simulated data. We highlight that the IHMS-GARCH model accurately estimates CP- and MS-GARCH models. Section 5.1 provides detailed results on the S&P500 index daily series and we compare our results with those of Bauwens, Dufays, and Rombouts (2011) for the MS-GARCH model. We finally apply the new method to the IHMS-DCC model in Section 6. Conclusions are presented in the last section. A brief review of the IHMM modeling framework is provided in Appendix C.

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 C. Before stating the model in subsection 2.2, we briefly define the Dirichlet process as well as its useful stick-breaking representation since they are directly used in our specification.

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 $\Theta \in \Re^d$), 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 $\{\Theta_k\}_{k=1}^{\infty}$ and is constructed by the following formulas (δ_{Θ_k} is the probability measure concentrated at Θ_k) :

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

which ensures that $G \sim DP(\eta, G_0)$.
The distribution over π is sometimes written $\pi \sim GEM(\eta)$[1] or $\pi \sim Stick(\eta)$.

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 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 :

$$y_t = \sigma_t \epsilon_t \tag{1}$$
$$\sigma_t^2 = \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t}\sigma_{t-1}^2 \tag{2}$$
$$\epsilon_t \sim N(0,1) \tag{3}$$
$$s_t | s_{t-1} = i, p_i \sim p_i = \{p_{i1}, p_{i2}, p_{i3}, ..., ..., ...\} \tag{4}$$
$$p_i | \pi, \lambda, \kappa \sim DP(\lambda + \kappa, \frac{\lambda\pi + \kappa\delta_i}{\lambda + \kappa}) \tag{5}$$
$$\pi | \eta \sim \text{Stick}(\eta) \tag{6}$$
$$\{\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}\} | \mu, \Sigma \sim N(\mu, \Sigma) \tag{7}$$
$$\mu \sim N(\underline{\mu}, \underline{\Sigma}) \tag{8}$$
$$\Sigma^{-1} \sim \text{Wishart}(\underline{V}, \underline{v}) \tag{9}$$

[1]GEM refers to Griffiths, Engen and McCloskey

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 $S_T = \{s_1, ..., s_T\}'$.

- **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

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

  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.

- **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

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

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.

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.

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 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 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 paper, 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 A. We now concentrate on the sampling of a complete state vector.

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. It is worth noticing that we could also use the Particle MCMC algorithm of Bauwens, Dufays, and Rombouts (2011) but our approach is less time consuming and easier to implement. 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 using the model of Klaassen (2002) as an approximation, conjugated with a Metropolis-Hastings step.

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 lets invariant the full posterior distribution. In the initial paper the distribution of $u_t|s_t, s_{t-1}, P$ is uniform : $U[0, p_{s_{t-1},s_t}]$[1]. To sample an entire vector U_T, we use the decomposition :

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

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 $\int_0^{p_{s_{t-1},s_t}} f(u_t|s_t, s_{t-1}, P)du_t = 1$.

Klaassen's 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. Relying on earlier studies of Gray (1996) and Klaassen (2002), we consider the following approximation of the MS-GARCH model :

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \tilde{\sigma}_{t-1,s_t}^2 \end{aligned} \tag{10}$$

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 B. The approximation gets rid of the path dependence problem since the likelihood $f(y_t|Y_{t-1}, \Theta, s_t)$ of the approximation only depends on the current state.

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 and P) :

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

Since there is no path dependence any more, each conditional density $q(s_t|Y_T, U_T, S^{t+1})$ is proportional to

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

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

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

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 $\{0 \le u_t \le p_{s_{t-1}, s_t}\}$.

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

$$\begin{aligned} \alpha(S_T, S'_T|Y_T, \Theta, P, U_T) &= \min\{1, \frac{f(S'_T|Y_T, U_T, \Theta, P) q(S_T|Y_T, U_T, \Theta, P)}{f(S_T|Y_T, U_T, \Theta, P) q(S'_T|Y_T, U_T, \Theta, P)}\} \\ &= \min\{1, \frac{f(Y_T|S'_T, U_T, \Theta, P) f(U_T|S'_T, P) f(S'_T|P) q(S_T|Y_T, U_T, \Theta, P)}{f(Y_T|S_T, U_T, \Theta, P) f(U_T|S_T, P) f(S_T|P) q(S'_T|Y_T, U_T, \Theta, P)}\}. \end{aligned}$$

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 algorithm. The number of regimes could then be determined using the marginal likelihood computed by the Chib's formula (Chib and Jeliazkov (2001)).

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 second subsection is dedicated to the mixing properties of the algorithm. These are compared with mixing properties of existing alternatives. The last subsection exposes summary statistics of the posterior distributions of several simulated data.

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

As it is shown in Table 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 really close to the one set by Bauwens, Dufays, and Rombouts (2011). GARCH parameters of each regime are independently driven by Normal distributions.

Prior Distributions of the Dirichlet processes		
$\eta \sim G(10, \frac{1}{2})$	$\lambda + \kappa \sim G(1000, 1)$	$\rho = \frac{\kappa}{\lambda+\kappa} \sim Beta(10000, 6)$
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
$\mu \sim \mathrm{N}(\underline{\mu},\underline{\Sigma})$		$\Sigma^{-1} \sim \mathrm{W}(\underline{V},\underline{v})$
$\underline{\mu} = \{0, \log(\frac{0.2}{0.8}), \log(\frac{0.8}{0.2})\}$		$\underline{V} = \frac{1}{5\underline{v}} I_3$
$\underline{\Sigma} = I_3$		$\underline{v} = 5$

Table 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 paper 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.

4.2 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) 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 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 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).

4.3 Comparison with other algorithms

We first compare our sampling method (called Kl-MH for Klaassen-Metropolis-Hastings) of the state vector with two other MCMC samplers (BPR for Bauwens, Preminger, and Rombouts (2010) and PMCMC for Bauwens, Dufays, and Rombouts (2011)). 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 $O(NT)$ where N stands for the number of particles. They choose $N = 250$ 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 ($O(KT)$ 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 displays the maximum autocorrelation time computed by batch means (see Geyer (1992)) and 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. The Kl-MH displays much better autocorrelation times than the BPR approach for all the simulated data. The comparison between the PMCMC and the Kl-MH is more delicate. The PMCMC method always exhibits better autocorrelation times than the Kl-MH algorithm but the differences are very small.

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

Table 4: Autocorrelation time of posterior structural break draws. Kl-MH stands for 'Klaassen Metropolis-Hastings' and denotes the method documented in the paper. The minimum autocorrelation times are in bold. For MS model, the autocorrelation time is computed on the number of observations in each regime.

Table 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 the Kl-MH is competitive with BPR for CP models, it clearly becomes the fastest method for MS models.

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

Table 5: Elapsed time in minutes for a MCMC simulation of 10000 draws

4.4 Results on simulated data

We present results of the IHMS-GARCH model on the different simulated data generated from the DGP of Table 3. Table 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.

	Regime 1	Regime 2	Regime 3	Regime 4	Regime 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 6: Posterior probabilities of the number of regimes for five simulated series generated from DGP displayed in Table 3. The true number of regimes is bolded.

Figure 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.

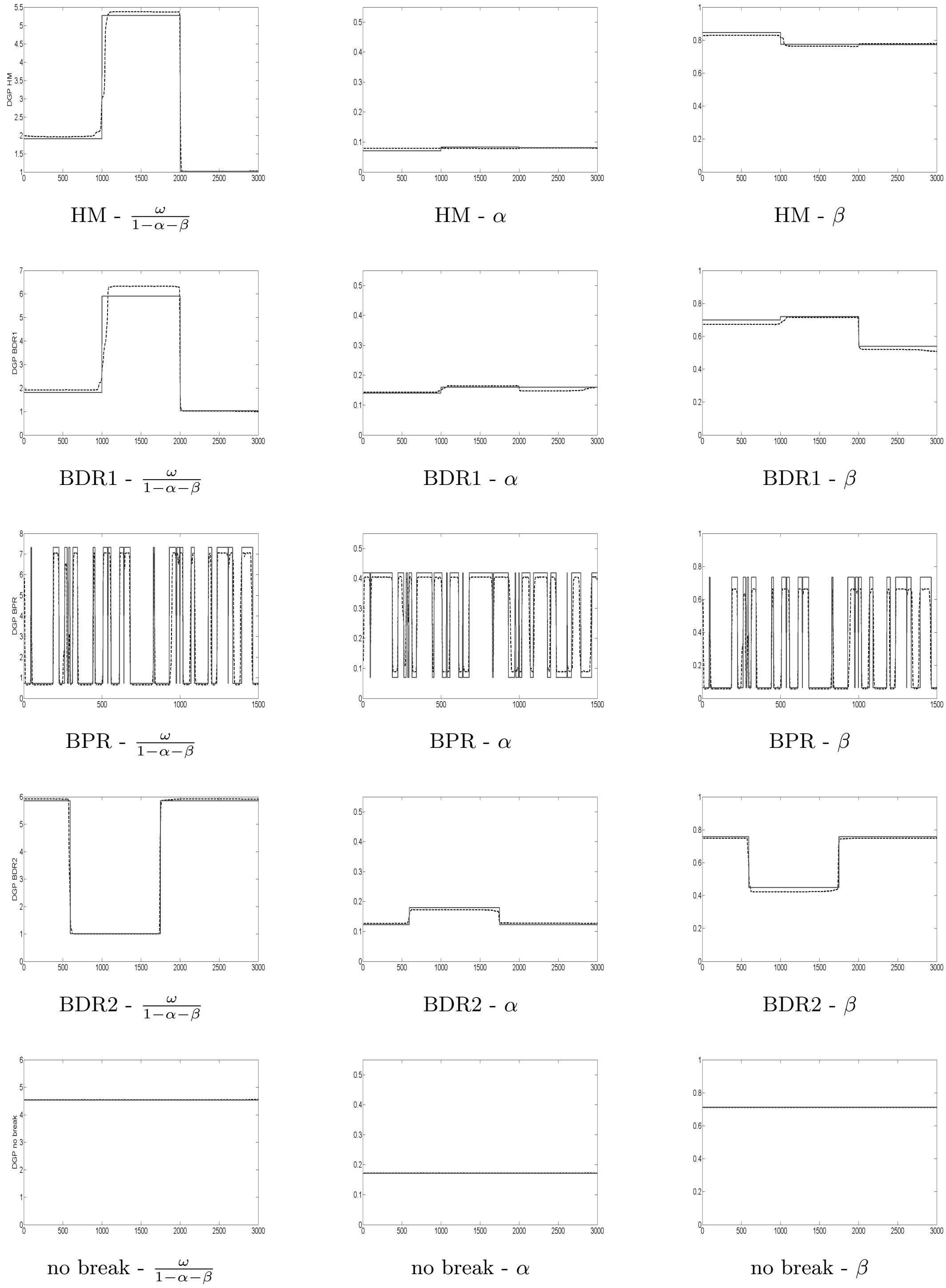


Figure 1: ML estimates given the true states (solid grey lines) compared to the posterior means (dashed black lines). Results for each simulated data are presented in row in the same order as in Table 3 (i.e. HM, BDR1, BPR, BDR2 and no break). The first column of graphics displays the (local) unconditional variance ($\frac{\omega_t}{1-\alpha_t-\beta_t}|Y_T$) whereas the two others respectively show the parameters $\alpha_t|Y_T$ and $\beta_t|Y_T$

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 $S\&P500$ daily index are documented in subsection 5.1. We next shortly provide posterior results for three commodities, namely Brent, Gold and Silver. This will help to devolatilize the returns in order to apply the DCC model in Section 6.

5.1 $S\&P500$ daily index

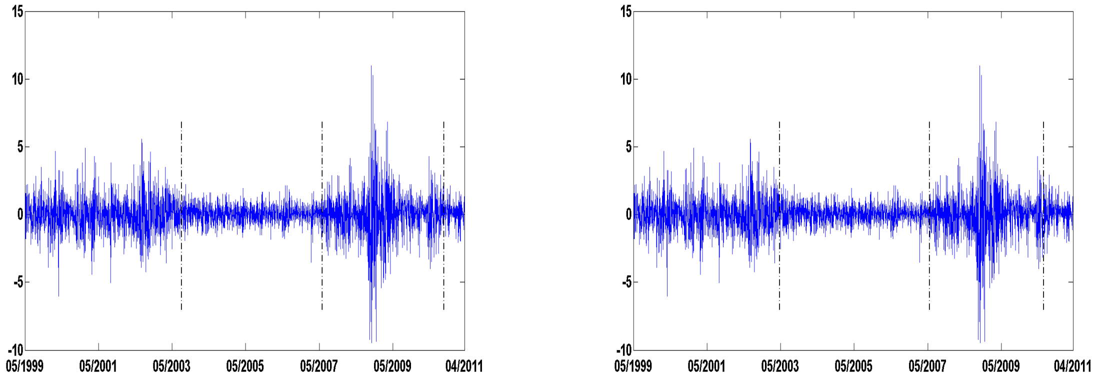
In this section we revisit the empirical exercise of Bauwens, Dufays, and Rombouts (2011). They use a MS and a CP model on the $S\&P500$ daily percentage returns from May 20, 1999 to April 25, 2011 (3000 observations). They further choose the optimal model and the number of regimes with the marginal likelihood. They find 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 7 displays the posterior distribution of the number of regimes.

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

Table 7: Posterior probabilities of the number of regimes for the $S\&P500$ daily index

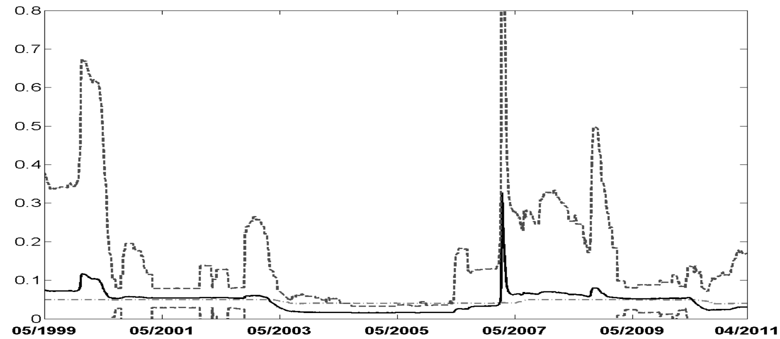
Figure 2 displays the time series with the estimated mode of the state vector at the most likely number of regimes. The regime switches occur at the same period as the best MS-GARCH model of Bauwens, Dufays, and Rombouts (2011).



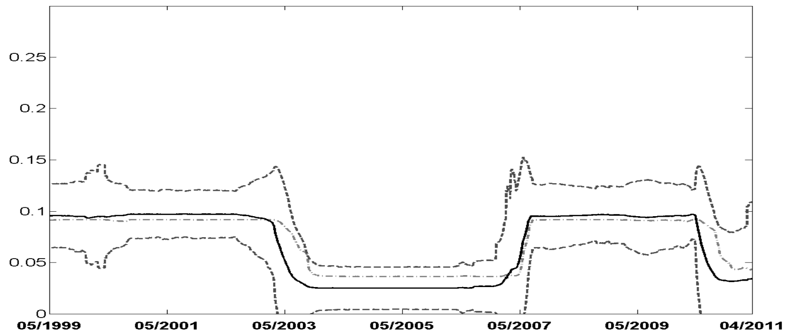
BDR - PMCMC   IHMS-GARCH

Figure 2: The left graphic shows the $S\&P500$ 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 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.



$S\&P500$ - $\frac{\omega}{1-\alpha-\beta}$



$S\&P500$ - α

$S\&P500$ - β

Figure 3: Posterior means of IHMS-GARCH parameters (black lines) with 95% percent confidence interval (dashed lines) for the $S\&P500$ 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 $\beta_t|Y_T$

5.2 Other financial time series

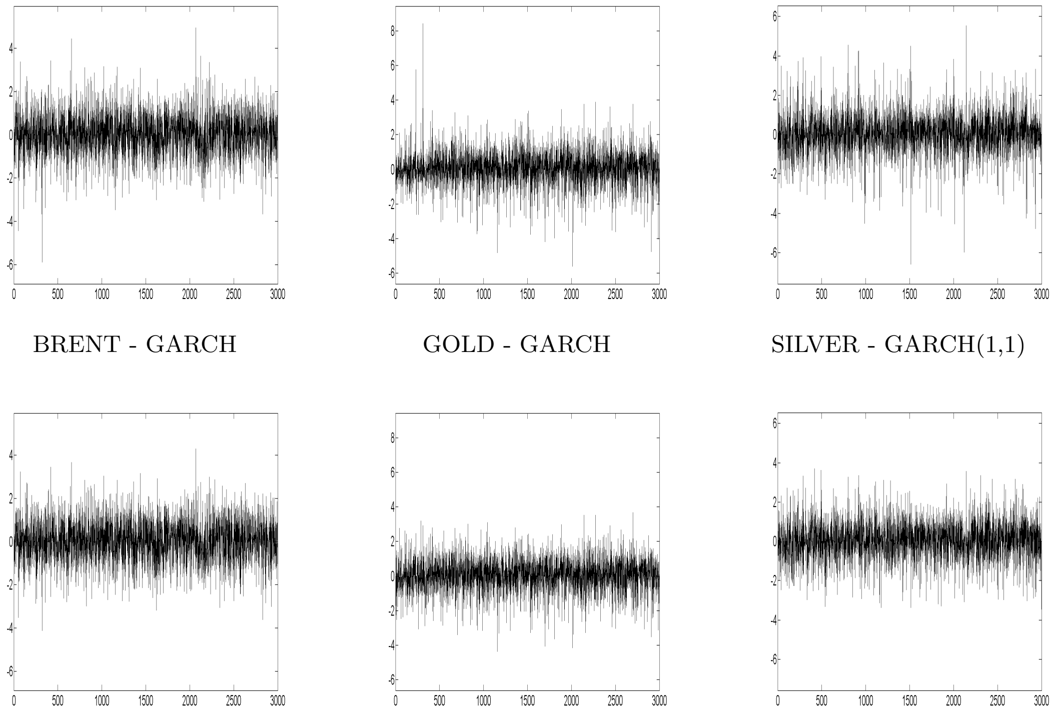
We briefly provide results on three other financial time series. These estimations are helpful for the multivariate correlation model of the next section that requires devolatilized time series as inputs. We shortly summarize in Table 8 the posterior distributions of three commodity percentage returns from July 3, 2000 to December 30, 2011 on a daily basis (3000 observations). Detailed results are available on request.

Table 8: Summaries of posterior distributions

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)

Figure 4 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.



BRENT - GARCH　GOLD - GARCH　SILVER - GARCH(1,1)

BRENT - IHMS-GARCH　GOLD - IHMS-GARCH　SILVER - IHMS-GARCH

Figure 4: Devolatilized financial time series with respect to the standard GARCH(1,1) (top) and to the IHMS-GARCH(1,1) (bottom) models.

6 Multivariate extension

The presented methodology is not limited to the univariate GARCH model. As an example we incorporate a infinite hidden MS framework in the Dynamic Conditional Correlation model (IHMS-DCC).

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 2:

$$\mathbf{y}_t = \epsilon_t \quad \text{where} \quad \epsilon_t \sim N(0, \Sigma_t) \tag{11}$$

$$\Sigma_t = \text{diag}\{Q_t\}^{-\frac{1}{2}} Q_t \text{diag}\{Q_t\}^{-\frac{1}{2}} \tag{12}$$

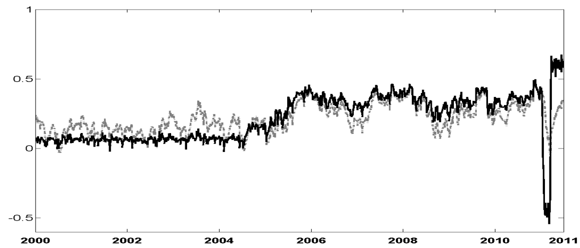
$$Q_t = Q^*_{s_t}(1 - \alpha_{s_t} - \beta_{s_t}) + \alpha_{s_t}\epsilon_{t-1}\epsilon'_{t-1} + \beta_{s_t} Q_{t-1} \tag{13}$$

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 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 also need to derive the appropriate approximation for the forward-backward Metropolis-Hastings step. The calculations, closely related to the Klaassen's approximation, are detailed in Appendix B.
We estimate the multivariate model on the three commodity devolatilized percentage returns (see subsection 5.2). The acceptance rate for the state vector amounts to 90.7 percent. Table 9 provides the posterior distribution of the number of regimes.

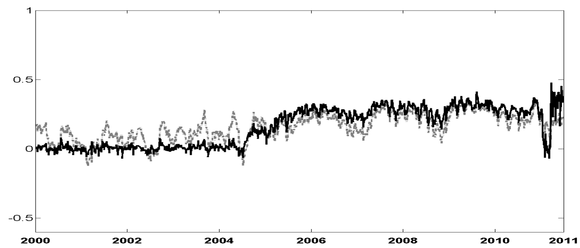
	Regime 1	Regime 2	Regime 3	Regime 4	Regime 5	Regime 6	Regime 7	Regime 8
Prob.	0	0.0149	**0.2984**	0.2592	0.2419	0.0920	0.0860	0.0076

Table 9: Posterior probabilities of the number of regimes for the devolatilized daily commodities (IHMS-DCC model)

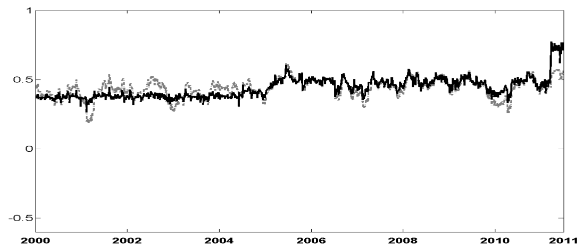
We display on figure 5 the posterior mean of each correlation over time. For the sake of comparison, we also expose the standard DCC correlations also based on the IHMS-GARCH devolatilized returns. The two dynamics are very similar although the standard DCC exhibits more volatile correlations during 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 5: IHMS-DCC(1,1) Correlation (in black) compared to DCC(1,1) correlations (in grey) for different commodities.

The graphic of the IHMS-DCC parameters lies in Appendix D. 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.

7 Conclusion

The GARCH model with fixed parameters has drained all its potential and more flexible models are now required. 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 paper 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 Bauwens, Dufays, and Rombouts (2011). 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. The algorithm could also handle other distributional assumptions than the Normal distribution. Further research will investigate this feature as well as its use and potential benefit to forecast financial time series.

Appendix

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 paper 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 4.1), 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 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 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 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 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 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 $f(\Sigma^{-1}|\mu, \Theta) \sim \text{Wishart}(\{\sum_{i=1}^{K}(\Theta_i - \mu)(\Theta_i - \mu)' + \underline{V}^{-1}\}^{-1}, \underline{v} + K)$

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 $f(u_t|s_t, s_{t-1}, P) \sim U[0, p_{s_{t-1}, s_t}]$.

B.1 Approximate GARCH : computation of the modified variance

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

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

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.

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

C The sticky infinite hidden Markov model

The sticky infinite hidden Markov model is based on Dirichlet processes and hierarchical Dirichlet processes. The Section 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.

C.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

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

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 :

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

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) :

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 : $G_j|G = \sum_{k=1}^{\infty} p_{jk} \delta_{\Theta_k} \sim DP(\lambda, G)$

$p_j = \{p_{jk}\}_{k=1}^{\infty} \sim \text{DP}(\lambda, \pi)$	Each row of the transition matrix is driven by a DP

3. Markov-switching model

$s_t|s_{t-1}, \{p_j\}_{j=1}^{\infty} \sim p_{s_{t-1}}$	First order Markovian with transition matrix $\{p_j\}_{j=1}^{\infty}$
$y_t|s_t, \{\Theta_k\}_{k=1}^{\infty} \sim F(\Theta_{s_t})$	Each state shares the same support (of G_0)

Table 10: Infinite hidden Markov Model (IHMM)

C.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. $E[p_{jk}|\lambda, \pi] = \pi_k \ \ \forall j$ (see Table 10)). The IHMM transition actually does not differ between a self-transition and a transition to another state, an unrealistic feature for time series.

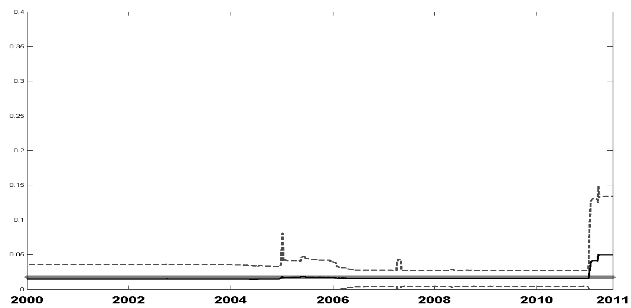
Fox, Sudderth, Jordan, and Willsky (2007) have developed a IHMM framework which excludes a high probability posterior with rapid switching. They called it 'the sticky HDP-HMM' or 'the sticky IHMM'. They specify a new parameter κ for self-transition bias and set a separate prior on this parameter. Their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

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

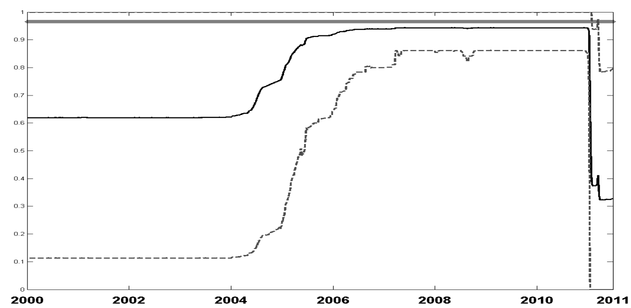
An amount $\kappa > 0$ is added to the j^{th} component of the (infinite) vector $\lambda\pi$. The new parameter implies a higher probability of staying in the same state in the next period than the original model (i.e. $E[p_{kk}|\lambda, \kappa, \pi] = \frac{\lambda \pi_k + \kappa}{\lambda + \kappa}$). Note that if $\kappa = 0$, we come back to the former specification (i.e the original IHMM of Table 10).

D Multivariate extension

The graphic 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 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 $\beta_t|Y_T$

References

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., 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.*

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

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

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

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

Ferguson, T. S. (1973): "A bayesian analysis of some nonparametric problem," *The annals of Statistics*, 1, No 2, 209–230.

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.

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

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(4), 473–511.

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

Haario, H., E. Saksman, and J. Tamminen (2001): "An Adaptive Metropolis Algorithm," *Bernoulli*, 7, No 2, 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.

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.*

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.

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

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

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

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

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

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

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.

TEH, Y., M. JORDAN, M. BEAL, AND D. M. BLEI (2006): "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, 101, 1566–1581.

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.*

Recent titles

CORE Discussion Papers

2012/4. Aitor CALO-BLANCO. Responsibility, freedom, and forgiveness in health care.
2012/5. Pierre PESTIEAU and Grégory PONTHIERE. The public economics of increasing longevity.
2012/6. Thierry BRECHET and Guy MEUNIER. Are clean technology and environmental quality conflicting policy goals?
2012/7. Jens L. HOUGAARD, Juan D. MORENO-TERNERO and Lars P. ØSTERDAL. A new axiomatic approach to the evaluation of population health.
2012/8. Kirill BORISSOV, Thierry BRECHET and Stéphane LAMBRECHT. Environmental maintenance in a dynamic model with heterogenous agents.
2012/9. Ken-Ichi SHIMOMURA and Jacques-François THISSE. Competition among the big and the small.
2012/10. Pierre PESTIEAU and Grégory PONTHIERE. Optimal lifecycle fertility in a Barro-Becker economy.
2012/11. Catherine KRIER, Michel MOUCHART and Abderrahim OULHAJ. Neural modelling of ranking data with an application to stated preference data.
2012/12. Matthew O. JACKSON and Dunia LOPEZ-PINTADO. Diffusion and contagion in networks with heterogeneous agents and homophily.
2012/13. Claude D'ASPREMONT, Rodolphe DOS SANTOS FERREIRA and Jacques THEPOT. Hawks and doves in segmented markets: A formal approach to competitive aggressiveness.
2012/14. Claude D'ASPREMONT and Rodolphe DOS SANTOS FERREIRA. Household behavior and individual autonomy: An extended Lindahl mechanism.
2012/15. Dirk VAN DE GAER, Joost VANDENBOSSCHE and José Luis FIGUEROA. Children's health opportunities and project evaluation: Mexico's *Oportunidades* program.
2012/16. Giacomo VALLETTA. Health, fairness and taxation.
2012/17. Chiara CANTA and Pierre PESTIEAU. Long term care insurance and family norms.
2012/18. David DE LA CROIX and Fabio MARIANI. From polygyny to serial monogamy: a unified theory of marriage institutions.
2012/19. Carl GAIGNE, Stéphane RIOU and Jacques-François THISSE. Are compact cities environmentally friendly?
2012/20. Jean-François CARPANTIER and Besik SAMKHARADZE. The asymmetric commodity inventory effect on the optimal hedge ratio.
2012/21. Concetta MENDOLICCHIO, Dimitri PAOLINI and Tito PIETRA. Asymmetric information and overeducation.
2012/22. Tom TRUYTS. Stochastic signaling: Information substitutes and complements.
2012/23. Pierre DEHEZ and Samuel FEREY. How to share joint liability: A cooperative game approach.
2012/24. Pilar GARCIA-GOMEZ, Erik SCHOKKAERT, Tom VAN OURTI and Teresa BAGO D'UVA. Inequity in the face of death.
2012/25. Christian HAEDO and Michel MOUCHART. A stochastic independence approach for different measures of concentration and specialization.
2012/26. Xavier RAMOS and Dirk VAN DE GAER. Empirical approaches to inequality of opportunity: principles, measures, and evidence.
2012/27. Jacques H. DRÈZE and Erik SCHOKKAERT. Arrow's theorem of the deductible : moral hazard and stop-loss in health insurance.
2012/28 Luc BAUWENS and Giuseppe STORTI. Computationally efficient inference procedures for vast dimensional realized covariance models.
2012/29. Pierre DEHEZ. Incomplete-markets economies: The seminar work of Diamond, Drèze and Radner.
2012/30 Helmuth CREMER, Pierre PESTIEAU and Grégory PONTHIÈRE. The economics of long-term care: a survey.
2012/31 Mathieu VAN VYVE, Laurence A. WOLSEY and Hande YAMAN. Relaxations for two-level multi-item lot-sizing problem.

Recent titles

CORE Discussion Papers - continued

2012/32 Jean-François MERTENS and Anna RUBINCHIK. Equilibria in an overlapping generations model with transfer policies and exogenous growth.
2012/33 Jean-François MERTENS and Anna RUBINCHIK. Pareto optimality of the golden rule equilibrium in an overlapping generations model with production and transfers.
2012/34 Hélène LATZER and Florian MAYNERIS. Income distribution and vertical comparative advantage. Theory and evidence.
2012/35 Pascal MOSSAY and Takatoshi TABUCHI. Preferential trade agreements harm third countries.
2012/36 Aitor CALO-BLANCO. The compensation problem with fresh starts.
2012/37 Jean-François CARPANTIER and Arnaud DUFAYS. Commodities volatility and the theory of storage.
2012/38 Jean-François CARPANTIER and Christelle SAPATA. Unfair inequalities in France: A regional comparison.
2012/39 Shui-Ki WAN, Shin-Huei WANG and Chi-Keung WOO. Total tourist arrival forecast: aggregation vs. disaggregation.
2012/40 Thierry BRECHET, Yann MENIERE and Pierre M. PICARD. The clean development mechanism in a global carbon market.
2012/41 Nikolaos GEORGANTZIS, Rafael MONER-COLONQUES, Vicente ORTS and José J. SEMPERE-MONERRIS. Theoretical and experimental insights on firms' internationalization decisions under uncertainty.
2012/42 Berardino CESI and Dimitri PAOLINI. Peer group and distance: when widening university participation is better.
2012/43 Arnaud DUFAYS. Infinite-state Markov-switching for dynamic volatility and correlation models.

Books

G. DURANTON, Ph. MARTIN, Th. MAYER and F. MAYNERIS (2010), *The economics of clusters – Lessons from the French experience*. Oxford University Press.

J. HINDRIKS and I. VAN DE CLOOT (2011), *Notre pension en heritage*. Itinera Institute.

M. FLEURBAEY and F. MANIQUET (2011), *A theory of fairness and social welfare*. Cambridge University Press.

V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.

I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.

W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.

L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.

J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.

M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.

J. HINDRIKS (2012), *Gestion publique*. De Boeck.

CORE Lecture Series

R. AMIR (2002), Supermodularity and complementarity in economics.

R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.

A. SHAPIRO (2010), Stochastic programming: modeling and theory.