FACULTY OF LAW, ECONOMICS AND FINANCE

CREA Discussion Paper 2014-07

Center for Research in Economic Analysis
University of Luxembourg

Specific Markov-switching behaviour for ARMA parameters

available online : http://wwwfr.uni.lu/recherche/fdef/crea/publications2/discussion_papers

Jean-François Carpantier, CREA, Université du Luxembourg
Arnaud Dufays, CREST, Ecole Nationale de la Statistique et de l'Administration Economique, Paris

March, 2014

For editorial correspondence, please contact: crea@uni.lu
University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faïencerie
L-1511 Luxembourg

uni.lu UNIVERSITÉ DU LUXEMBOURG

Specific Markov-switching behaviour for ARMA parameters

Jean-François Carpantier[1] and Arnaud Dufays[2]

March 24, 2014

Abstract

We propose an estimation method that circumvents the path dependence problem existing in Change-Point (CP) and Markov Switching (MS) ARMA models. Our model embeds a sticky infinite hidden Markov-switching structure (sticky IHMM), which makes possible a self-determination of the number of regimes as well as of the specification : CP or MS. Furthermore, CP and MS frameworks usually assume that all the model parameters vary from one regime to another. We relax this restrictive assumption. As illustrated by simulations on moderate samples (300 observations), the sticky IHMM-ARMA algorithm detects which model parameters change over time. Applications to the U.S. GDP growth and the DJIA realized volatility highlight the relevance of estimating different structural breaks for the mean and variance parameters.

Keywords: Bayesian inference, Markov-switching model, ARMA model, Infinite hidden Markov model, Dirichlet Process

JEL Classification: C11, C15, C22, C58.

[1]University of Luxembourg, CREA, Avenue de la faiencerie, 162A, L-1511 Luxembourg. Email: jean-francois.carpantier@uni.lu.

[2]Corresponding author. Ecole Nationale de la Statistique et de l'Administration Economique, CREST, 15, Boulevard Gabriel Péri, 92245 Malakoff, France. Email: arnaud.dufays@ensae.fr. Website: https://sites.google.com/site/websiteofarnauddufays/

This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

Econometricians have been proposing models with evolving parameters for more than twenty years. In the context of abrupt switches in the parameters it really started with the framework of Hamilton (1989). It consists in enriching a model with a discrete latent variable so that the parameters can abruptly switch from one state to another. By assuming Markovian transitions, a forward-backward algorithm has been designed for direct evaluation of the likelihood by integrating the latent variable (Rabiner (1989)). Hamilton developed the methodology for recurrent states that is now known as Markov-switching (MS) model. Some years afterwards, Chib (1998) designed the Change-point (CP) model. It arises when the transitions are not reversible, a feature that brings non-stationarity. During the last two decades these two frameworks have been intensively used due to their flexibility (e.g. Hamilton and Susmel (1994), Pesaran, Pettenuzzo, and Timmermann (2006)). However the methodology relies on an assumption problematic for models exhibiting path dependency such as the Auto-regressive Moving Average (ARMA, David and Whittle (1952)) and the Generalized Auto-regressive Conditional Heteroskedastic (GARCH, Bollerslev (1986)) models. The difficulty occurs because a hidden variable at time t (the lagged error term in the ARMA model) depends on the entire path of states that have been followed until time t. The computation thus exponentially grows with time (at least for MS framework) and becomes infeasible for relatively small time series. Recently a lot of efforts have been devoted to the CP- and MS-GARCH models resulting in important improvements in estimation methods (see among others Haas, Mittnik, and Paolella (2004), Francq and Zakoian (2008), Bauwens, Preminger, and Rombouts (2010), Bauwens, Dufays, and Rombouts (2011)). We rely on this literature to propose a new flexible estimation approach for CP- and MS-ARMA models.

Selecting the number of regimes has also been considered as a relevant question for many years. Current methods usually apply penalized-likelihood approaches. In Bayesian inference it leads to the computation of the marginal likelihood. Although many numerical tools have been developed for its estimation (for a recent review see Ardia, Hoogerheide, and van Dijk. (2009)), it still remains a tedious calculation for complex models (e.g. Bauwens, Dufays, and Rombouts (2011)). The sticky infinite hidden Markov model (sticky IHMM) (Fox, Sudderth, Jordan, and Willsky (2007)) allows bypassing the marginal likelihood computation by building a Markov-chain with a potentially infinite number of states. It therefore encompasses any number of regimes. Dirichlet process (Ferguson (1973) and Sethuraman (1994)) and hierarchical Dirichlet process (Teh, Jordan, Beal, and Blei (2006)) constitute the core of the IHMM. The framework has successfully been applied in various fields such as genetics (Beal and Krishnamurthy (2006)), visual recognition (Kivinen, Sudderth, and Jordan (2007)) and economics on Auto-regressive models (Jochmann (2010) and Song (2011)) and on GARCH models (Dufays (2012)).

Our contribution is twofold. Estimating CP- and MS-ARMA models is challenging. Relying on Dufays (2012), we estimate these models by combining the model of Klaassen (2002) with a Metropolis-Hastings step. By embedding the sticky IHMM framework, the model encompasses the CP and the MS settings by covering a potentially infinite number of regimes. We thus innovate by efficiently estimating the CP- and the MS-ARMA models and by incorporating the sticky IHMM into these models. Secondly, standard CP and MS models assume that all the parameters of the model vary from one regime to another. Another contribution of the paper lies in relaxing this assumption. We propose a model that allows for different break dates in the mean and in the variance. Given the flexibility of the sticky IHMM framework, this can be easily extended to any parameters of the model.

In Section 2, we present the model and Section 3 discusses how to relax the standard assumption stating that all the model parameters change from one regime to another. The Section 4 covers the estimation procedure. The prior elicitation and the label switching problem are addressed in Section 5. The sticky IHMM algorithm is explored on artificial data in Section 6. Two applications, namely on the U.S. quarterly GDP growth and on the DJIA daily realized volatility, are finally proposed in Section 7. A general conclusion ends the paper.

2 The model

We consider the sticky infinite hidden MS-ARMA(1,1) (sticky IHMM-ARMA) model defined as

$$\begin{aligned} y_t &= \mu_{s_t^a} + \beta_{s_t^a} y_{t-1} + \phi_{s_t^a} \epsilon_{t-1} + \epsilon_t \\ \epsilon_t &\sim i.i.d.\ N(0, \sigma^2_{s_t^b}) \end{aligned}$$

where s_t^a, s_t^b are integer random variables taking values in $\mathbb{N}^2$. We define $Y_{1:T} = \{y_1, ..., y_T\}'$, $S^a_{1:T} = \{s^a_1, ..., s^a_T\}'$ and $S^b_{1:T} = \{s^b_1, ..., s^b_T\}'$ where T denotes the sample size. The set $\Theta = \{\mu_1, \ldots, \mu_j, \ldots, \beta_1, \ldots, \beta_j, \ldots, \phi_1, \ldots, \phi_j, \ldots, \sigma_1, \ldots, \sigma_j, \ldots\}'$ gather all the ARMA parameters of the model and the set $\Theta_j = \{\mu_j, \beta_j, \phi_j, \sigma_j\}'$ contains all the continuous parameters of the regime j. The latent state processes $\{s_t^a, s_t^b\}$ are independent first order Markovian with transition matrices (see Section 3 for à priori dependent latent state vectors)

$$P^a = \begin{pmatrix} p^a_{11} & p^a_{12} & p^a_{13} & \cdots \\ p^a_{21} & p^a_{22} & p^a_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ p^a_{i1} & p^a_{i2} & p^a_{i3} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \quad P^b = \begin{pmatrix} p^b_{11} & p^b_{12} & p^b_{13} & \cdots \\ p^b_{21} & p^b_{22} & p^b_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ p^b_{i1} & p^b_{i2} & p^b_{i3} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

These transition matrices characterize a Markov-switching model (MS) with a potentially infinite number of breaks. Note that another distributional assumption than the normal one, a different number of latent state vectors or a higher order dimension than the ARMA(1,1) configuration can be handled.

Inferring on an infinite number of breaks can be dealt with a sticky infinite hidden Markov model (sticky IHMM, Teh, Jordan, Beal, and Blei (2006), Fox, Sudderth, Jordan, and Willsky (2008)). A complete estimation procedure is developed in Dufays (2012) for univariate and multivariate GARCH models. Dufays (2012) relies on the beam sampler (Van Gael, Saatci, Teh, and Ghahramani (2008)) to tackle the infinite dimension. This paper rests on the *degree L weak limit approximation* (Ishwaran and Zarepour (2002)) of the Dirichlet process that suggests estimating a truncated model given a large finite L number of regimes instead of an infinite one. If the number of regime L is large enough, the results are scarcely different

(Kurihara, Welling, and Teh (2007)). Under the degree L weak limit approximation, each row of the transition matrix is truncated by a finite Dirichlet distribution (instead of being driven by a Dirichlet process) given a hierarchical structure composed of univariate sticky IHMM parameters ($\forall j \in [a,b], \alpha^j \in \Re^+$ and $\kappa^j \in \Re^+$ and global weights $\pi^j = \{\pi^j_1, \ldots, \pi^j_L\} \in [0,1]^L$):

$$\begin{aligned} p^a_i = \{p^a_{i1}, p^a_{i2}, ..., p^a_{iL}\}|\pi^a, \alpha^a, \kappa^a &\sim Dir(\alpha^a\pi^a_1, \alpha^a\pi^a_2, ..., \alpha^a\pi^a_i + \kappa^a, ..., \alpha^a\pi^a_L) \\ p^b_i = \{p^b_{i1}, p^b_{i2}, ..., p^b_{iL}\}|\pi^b, \alpha^b, \kappa^b &\sim Dir(\alpha^b\pi^b_1, \alpha^b\pi^b_2, ..., \alpha^b\pi^b_i + \kappa^b, ..., \alpha^b\pi^b_L) \end{aligned}$$

The global weights which build the entire transition matrices follow a finite symmetric Dirichlet prior given the IHMM parameters ($\eta^a \in \Re^+$ and $\eta^b \in \Re^+$) :

$$\begin{aligned} \pi^a = \{\pi^a_1, \pi^a_2, ..., \pi^a_L\}|\eta^a &\sim Dir(\frac{\eta^a}{L}, \frac{\eta^a}{L}, ..., \frac{\eta^a}{L}) \\ \pi^b = \{\pi^b_1, \pi^b_2, ..., \pi^b_L\}|\eta^b &\sim Dir(\frac{\eta^b}{L}, \frac{\eta^b}{L}, ..., \frac{\eta^b}{L}) \end{aligned}$$

Throughout the paper the sticky IHMM parameters $\alpha = \{\alpha^a, \alpha^b\}, \kappa = \{\kappa^a, \kappa^b\}, \eta = \{\eta^a, \eta^b\}$ are brought together in the parameter set H_{Dir}.

The parameter set Θ also benefits from a hierarchical layer in order to take into account information from existing regimes. As advocated by Song (2011), the structure improves the birth of new regimes by drawing realistic parameters from a common distribution. Table 1 displays the hierarchical structure of the model parameters. The hyper-parameters $\underline{\mu}, \underline{\Sigma}, \underline{V}, \underline{v}, \underline{e_a}, \underline{f_a}$ and $\underline{f_b}$ are defined in Section 5.

Hierarchical Distributions of the ARMA parameters	
For each regime i : $\{\mu_i, \beta_i, \phi_i\} \sim \text{N}(\bar{\mu},\bar{\Sigma})\ \delta_{\{\lvert\beta_i\rvert<1, \lvert\phi_i\rvert<1\}}$	
Hierarchical parameter : $\bar{\mu} \sim N(\underline{\mu}, \underline{\Sigma})$	Hierarchical parameter : $\bar{\Sigma}^{-1} \sim W(\underline{V}, \underline{v})$
Hierarchical Distributions of the variances	
For each regime i : $\{\sigma_i^{-2}\} \sim \text{G}(\bar{e},\bar{f})$	
Hierarchical parameter : $\bar{e} \sim Exp(\underline{e_a})$	Hierarchical parameter : $\bar{f}^{-1} \sim G(\underline{f_a}, \underline{f_b})$

Table 1: Hierarchical Distributions of the model parameters. $\delta_{\{a>0\}}$ is the Dirac function taking the value one if the constraint $a > 0$ holds and zero otherwise.

3 Multiple state vectors

Standard CP and MS frameworks implicitly assume that all the model parameters vary from one regime to another. However it is highly likely that only a subset of parameters evolves when facing a structural break. Holding to this hypothesis can spuriously increase the number of parameters and lead to inefficient estimates. Clearly, relaxing this assumption would help to keep a more parsimonious model while taking structural breaks into account, but it also seriously hardens the estimation.

A straightforward solution to this estimation problem consists in increasing the number of state vectors and in linking them with the model parameters (see Eo (2012)). In classical CP and MS approaches, where the optimal number of regimes is computed by maximizing a criterion such as the marginal likelihood, assuming that L denotes the maximum number of regimes that we want to consider and q stands for the number of parameters, the strategy leads to estimate L^q different models. For instance 81 posterior distribution and marginal likelihood estimations are required for the MS-ARMA(1,1) model with a maximum of considered regimes equal to three. The method is too computationally demanding and not optimal.

By incorporating the sticky IHMM, as done in the previous Section, the same strategy can be used without having to consider more than one model. Indeed the sticky IHMM allows for self-determining the optimal number of regimes whatever the number of state vectors in presence. Before covering the model estimation by MCMC algorithm (Section 4), we first discuss different possible implementations of using a sticky IHMM structure with several latent state vectors.

3.1 Discussion on the priors

In order to limit the number of model parameters, the state vectors could share the same transition matrix. This configuration would rule out quick and slow switches for the same regime for different parameters and would strongly restrict the model. We therefore prefer to assume two different transition matrices for the state vectors.

Specifying that the transition matrices are à priori independent, as in Section 2, is not realistic.

In most cases, the state vectors should indeed be correlated in some sense. For example, we could expect that a switch in the variance is more likely to occur when a change in the mean has just happened (and vice versa). Incorporating dependence between the state vectors can therefore be relevant and could be captured by a specific parameter $\lambda = \{\lambda_1, ..., \lambda_L\}' \in [0,1]^L$ that denotes the probability of staying in the same regime when the other state vector switches to another one. The dependence of the variance state vector with respect to the mean state vector can be modelled as follows :

$$f(s_t^b | s_{t-1}^b = i, P^b, d_t, \lambda) \quad = \quad p_i^b \quad \text{if} \quad d_t \quad > \quad D \tag{1}$$

$$= \quad \{\frac{1-\lambda_i}{L-1}, ..., \lambda_i, ..., \frac{1-\lambda_i}{L-1}\} \text{ otherwise} \tag{2}$$

where d_t stands for the duration of the last observed regime at time t in the mean state vector $S_{1:t}^a$ and D is a user-defined distance parameter. The parameter D will typically be equal to 4 for quarterly data, 12 for monthly data and 25 for daily data. Under this framework, the parameter λ_j related to the regime j is driven by a Beta distribution :

$$\lambda_j | \alpha^b, \pi^b, \kappa^b \quad \sim \quad \text{Beta}(\alpha^b \pi_j^b + \kappa^b, \alpha^b(1 - \pi_j^b)) \tag{3}$$

The posterior distribution of λ_j given the other parameters of the model is conjugated with the prior distribution and turns out to be a Beta distribution. When dependence between state vectors is à priori assumed, the MCMC scheme is then just embedded with L additional Gibbs steps. The applications in Section 7 are performed with and without including the dependence parameter λ.

3.2 Discussion on the state vector updates

The state vectors can be jointly or independently sampled. The former approach theoretically leads to a better exploration of the support, an enjoyable feature for the MCMC algorithm. Sampling the state vectors jointly is unquestionably the best technique for models without any path dependency problems. However in the presence of path dependence models the state vector updates rest on a Metropolis-Hastings (M-H) step (as it will become clear in the next Section) that could stick the MCMC simulation if the proposal distribution (of the M-H step) turns out to be a poor approximation of the targeted conditional distribution. Jointly updating several state vectors increases the dimension problem and more rejections of the proposal draws are bound to happen. The sampling strategy of the state vectors is therefore model-specific. In our exercises, we disunite the drawing of the break dates of the mean with the one of the variance. As it is well-known in the MS literature, the mean state vector exhibits a path dependence problem and must be for example sampled with the M-H forward-backward algorithm of Section 4. On the contrary the variance state vector does not display such a dependency. We can directly apply the standard forward-backward algorithm conditional to the mean state vector without any additional difficulty. This observation made us choose to separately draw the two state vectors.

4 Estimation

Estimation is feasible by explicitly treating $S^a_{1:T}, S^b_{1:T}$ as parameters. To infer the model parameters we develop an MCMC sampler that iteratively draws from eight full conditional distributions (see Table 2).

1. $f(S^a_{1:T}\|\Theta, P^a, S^b_{1:T}, Y_{1:T})$	5. $f(\Theta\|\bar{\mu}, \bar{\Sigma}, \bar{e}, \bar{f}, H_{Dir}, S^a_{1:T}, S^b_{1:T}, Y_{1:T})$
2. $f(P^a\|\Theta, H_{Dir}, \pi^a, S^a_{1:T}, Y_{1:T})$	6. $f(\bar{\mu}, \bar{\Sigma}, \bar{e}, \bar{f}\|\Theta, H_{Dir}, S^a_{1:T}, S^b_{1:T}, Y_{1:T})$
3. $f(S^b_{1:T}\|\Theta, P^b, S^a_{1:T}, Y_{1:T})$	7. $f(\pi^a, \pi^b\|\Theta, P^a, P^b, H_{Dir}, S^a_{1:T}, S^b_{1:T}, Y_{1:T})$
4. $f(P^b\|\Theta, H_{Dir}, \pi^b, S^b_{1:T}, Y_{1:T})$	8. $f(H_{Dir}\|P^a, P^b, \pi^a, , \pi^b, S^a_{1:T}, S^b_{1:T}, Y_{1:T})$

Table 2: sticky IHMM-ARMA MCMC sampler.

Besides sampling the mean state vector from $f(S^a_{1:T}|\Theta, P^a, H_{Dir}, \pi^a, Y_{1:T})$, drawing from the other full conditional distributions (from 2 to 8) are standard. The entire MCMC scheme is detailed in Appendix A. We now concentrate on the most challenging item of the Gibbs sampler : $f(S^a_{1:T}|\Theta, P^a, S^b_{1:T}, Y_{1:T})$.

Updating the state vector is usually done by the forward-backward algorithm (Rabiner (1989), Chib (1998)). The algorithm perfectly works for Auto-regressive (AR) model but fails for moving average processes due to the path dependence exhibited by the lag of the error term. In an AR model, the likelihood only depends on the current state while in the MA model, the likelihood depends on the whole path that have been used so far. The forward-backward method is then completely intractable since the computation exponentially grows with time. Here we adapt the method of Dufays (2012) that combines a M-H approach with an approximate model. We sample a state vector from a modified MS-ARMA model that gets rid of the path dependence problem. The M-H step then ensures that the posterior distribution is not altered. GARCH and ARMA models are closely related when tackling the path dependence problem. Reliable approximations of the MS-GARCH have been proposed by Gray (1996), Klaassen (2002), Haas, Mittnik, and Paolella (2004). We focus on the Klaassen (2002) model[1] and accommodate it to the current specification :

$$y_t = \mu_{s_t^a} + \beta_{s_t^a} y_{t-1} + \phi_{s_t^a} \tilde{\epsilon}_{t-1,s_t^a} + \epsilon_t$$

where $\tilde{\epsilon}_{t-1,s_t^a} = E[\epsilon_{t-1}|Y_{1:t-1}, s_t^a, \Theta, P^a, S_{1:t-1}^b]$. We derive the computation of $\tilde{\epsilon}_{t-1,s_t^a}$ in Appendix B. The approximation rules out the path dependence problem since the vector $\tilde{\epsilon}_t = \{\tilde{\epsilon}_{t,1}, ..., \tilde{\epsilon}_{t,L}\}'$ is known at time t for any regime.

We therefore sample a new state vector $S_{1:T}^{a'}$ from the MS-ARMA approximation employing the forward-backward algorithm. The proposed parameter is then accepted according to the M-H ratio :

$$\begin{aligned} \alpha(S_{1:T}^{a'}, S_{1:T}^a|Y_{1:T}, \Theta, S_{1:T}^b, P^a) &= \min\{1, \frac{f(S_{1:T}^{a'}|Y_{1:T}, \Theta, S_{1:T}^b, P^a)q(S_{1:T}^a|Y_{1:T}, \Theta, S_{1:T}^b, P^a)}{f(S_{1:T}^a|Y_{1:T}, \Theta, S_{1:T}^b, P^a)q(S_{1:T}^{a'}|Y_{1:T}, \Theta, S_{1:T}^b, P^a)}\} \\ &= \min\{1, \frac{f(Y_{1:T}|S_{1:T}^{a'}, \Theta, S_{1:T}^b, P^a)f(S_{1:T}^{a'}|P^a)q(S_{1:T}^a|Y_{1:T}, \Theta, S_{1:T}^b, P^a)}{f(Y_{1:T}|S_{1:T}^{a'}, \Theta, S_{1:T}^b, P^a)f(S_{1:T}^{a'}|P^a)q(S_{1:T}^{a'}|Y_{1:T}, \Theta, S_{1:T}^b, P^a)}\} \end{aligned}$$

where $q(.|Y_{1:T}, \Theta, S_{1:T}^b, P^a)$ is the proposal distribution of $S_{1:T}^a$ derived from the forward-backward algorithm and is equal to

$$q(s_T^a|Y_{1:T}, \Theta, S_{1:T}^b, P^a)q(s_{T-1}^a|Y_T, s_T^a, \Theta, S_{1:T}^b, P^a)...q(s_1^a|Y_{1:T}, S_{2:T}^a, \Theta, S_{1:T}^b, P^a)$$

[1]The model of Haas, Mittnik, and Paolella enjoys interesting features such as a simple implementation and explicit statistic derivations but, as documented in Dufays (2012) for GARCH model, the Klaassen's model seems to provide a better approximation of the MS-GARCH model. In our case, estimations with an adapted model of Haas, Mittnik, and Paolella have also been carried out and lead to similar results to those obtained from the Klassen's model. However Klaassen's acceptance rates are always higher than those of Haas, Mittnik, and Paolella.

The proposal $S^{a'}_{1:T}$ will be hardly accepted if an entire state vector is drawn from the MS-ARMA approximation in one block. To ensure good MCMC mixing properties we update the state vector in small block of random size (see e.g. Chib and Ramamurthy (2010) in a DSGE context). It avoids some sticking situations where the MCMC algorithm always rejects the proposed parameter and also enhances the acceptance rate.

5 Prior elicitation and label switching problem

Table 3 reports the prior distributions and their related hyper-parameters. Regarding the sticky IHMM parameter set, the priors are chosen to get conjugate posterior distributions as advocated by Fox, Sudderth, Jordan, and Willsky (2008). A hierarchical Normal-Wishart distribution is assumed for the ARMA parameters. Gathering information of existing regimes, this structure facilitates births of new regimes whilst not complicating the MCMC simulation by providing conjugate distributions. A similar structure is applied to the variance (see e.g. Pesaran, Pettenuzzo, and Timmermann (2006)).

The posterior distribution is invariant to the label of the state vector. If a label switch occurs in the state vectors during the MCMC simulation, the usual summary statistics such as the posterior means and standard deviations are misleading. Indeed these statistics depend on the label of the state. Different solutions exist to solve the issue. The prior distributions can be chosen to rule out the label switching problem by constraining the support of the parameters given the regimes (see Bauwens, Preminger, and Rombouts (2010)). Nevertheless finding appropriate constraints to preclude all the possible switches without truncating the posterior distribution can be difficult. Otherwise as advocated by Geweke (2007), the label switching issue can be completely ignored in the MCMC simulation. In this case either the reported summary statistics are label invariant (see Song (2011), Dufays (2012)) or a loss function is used to sort the posterior draws in one specific label ordering (see e.g. Marin, Mengersen, and Robert (2005), Bauwens, Dufays, and Rombouts (2011)). We apply the latter approach but with a new sorting strategy.

In the empirical exercise, we are interested in the posterior distribution given a fixed number of regimes (say L_a for the mean and L_b for the variance). In order to sort the posterior sample, note that the posterior distribution $\Theta_t|Y_{1:T}, L_a, L_b$ is label invariant and constitutes

Prior Distributions of the Dirichlet parameters			
For the Mean:	$\eta \sim G(5,1)$	$\lambda+\kappa \sim G(125,\frac{1}{5})$	$\rho = \frac{\kappa}{\lambda+\kappa} \sim Beta(10,1)$
For the Variance:	$\eta \sim G(5,1)$	$\lambda+\kappa \sim G(1000,1)$	$\rho = \frac{\kappa}{\lambda+\kappa} \sim Beta(10000,6)$
Prior Distributions of the ARMA parameters			
For each regime i : $\{\mu_i,\beta_i,\phi_i\} \sim \mathrm{N}(\bar{\mu},\bar{\Sigma})\ \delta_{\{\vert\beta_i\vert<1,\vert\phi_i\vert<1\}}$			
Hierarchical parameter : $\bar{\mu}$		Hierarchical parameter : $\bar{\Sigma}$	
$\bar{\mu}$	$\sim \mathrm{N}(\underline{\mu},\underline{\Sigma})$	$\bar{\Sigma}^{-1}$	$\sim \mathrm{W}(\underline{V},\underline{v})$
$\underline{\mu}$	$= \{0,0,0\}$	$\underline{V}$	$= \frac{1}{5\underline{v}}I_3$
$\underline{\Sigma}$	$= 0.1I_3$	$\underline{v}$	$= 5$
Prior Distributions of the variances			
For each regime i : $\{\sigma_i^{-2}\} \sim \mathrm{G}(\bar{e},\bar{f})$			
Hierarchical parameter : $\bar{e}$		Hierarchical parameter : $\bar{f}$	
$\bar{e}$	$\sim \mathrm{Exp}(\underline{e_a})$	$\bar{f}^{-1}$	$\sim \mathrm{G}(\underline{f_a},\underline{f_b})$
$\underline{e_a}$	$= \frac{1}{2}$	$\underline{f_a}$	$= 8$
		$\underline{f_b}$	$= 4$

Table 3: Prior Distributions. I_d stands for the identity matrix of dimension d and $\delta_{\{a>0\}}$ is the Dirac function taking the value one if the constraint $a > 0$ holds and zero otherwise.

an interesting candidate for recovering the posterior distribution of each regime. We only need to find a period of time where the posterior distribution of each regime matches with the label invariant distribution. We develop the procedure for the mean parameters and apply the same algorithm for the variance parameters in the simulations and applications. The sorting strategy consists in three steps (where the set $\theta = \{\mu_1, ..., \mu_L, \beta_1, ..., \beta_L, \phi_1, ..., \phi_L\}'$ gathers all the ARMA parameters and $\theta_j = \{\mu_j, \beta_j, \phi_j\}'$ denotes the ARMA parameters of the regime j) :

1. Apply a K-means algorithm (e.g. MacQueen (1967)) on the posterior approximation of $E[\theta_t|L_a, L_b, Y_{1:T}]$ for all $t \in [1, T]$ with L_a number of clusters.[2] The output is the best repartition of the regimes (say $S^*_{1:T}$).

2. For each regime $j = 1, ..., L_a$ and for all t such that $s^*_t = j$, find the time t_j such that the squared Euclidean distance loss function on the empirical standardized standard deviations of $\theta_t|L_a, L_b, Y_{1:T}$ is minimized. This loss function is intuitively minimized for the time where the probability of being in regime j is maximized.

3. For each regime $j = 1, ..., L_a$, match the posterior distributions : $\theta_j|L_a, L_b, Y_{1:T} = \theta_{t_j}|L_a, L_b, Y_{1:T}$.

6 Simulation

The proposed algorithm is used on artificial data. We first expose our data generated processes (DGPs) and then report the simulation results.

Data generated processes

Four different scenarii allowing a structural break in each model parameters are specified. One artificial data of 300 observations is sampled from each DGP displayed in Table 4 and the sticky IHMM-ARMA model is then estimated on it. The DGPs are MS specifications with two switches: one occurring after the 99th observation and the second one after the 199th observation. The first regime therefore counts 200 observations.

Results

The MCMC simulations iterate 45000 times with a burn-in phase of 5000 draws. The maximum number of states is set to 10 for the two different Markov-chain specifications (i.e. mean

[2]The K-means algorithm is available in mathematical programs such as Matlab or R.

Name	Type	Regimes	Break point	μ	β	ϕ	σ^2
DGP_μ	MS	2	$\{99, 199\}$ obs.	$\{1.5; -0.5\}$	$\{0.8; 0.8\}$	$\{0.5; 0.5\}$	{1;1}
DGP_β	MS	2	$\{99, 199\}$ obs.	$\{0.5; 0.5\}$	$\{0.5; 0.9\}$	$\{0.2; 0.2\}$	{1;1}
DGP_ϕ	MS	2	$\{99, 199\}$ obs.	$\{0.5; 0.5\}$	$\{0.8; 0.8\}$	$\{0.5; -0.2\}$	{1;1}
DGP_σ	MS	2	$\{99, 199\}$ obs.	$\{0.5; 0.5\}$	$\{0.8; 0.8\}$	$\{0.2; 0.2\}$	{1;2.5}

Table 4: Data Generating Processes of the four simulated series.

and variance) and the starting point consists in the maximum likelihood estimates given no structural breaks. This specification will be used throughout the paper.

Table 5 shows the posterior probabilities of having a specific number of regimes. For each artificial data, the right specification has been identified.

				DGP_μ			
# Regimes (ARMA,Variance)	[1,1]	[2,1]	[3,1]	[2,2]	[4,1]	[3,2]	[5,1]
	0	**0.73**	0.19	0.04	0.02	0.01	0.00
				DGP_β			
# Regimes (ARMA,Variance)	[1,1]	[2,1]	[3,1]	[4,1]	[2,2]	[3,2]	[5,1]
	0	**0.46**	0.27	0.10	0.07	0.04	0.03
				DGP_ϕ			
# Regimes (ARMA,Variance)	[1,1]	[2,1]	[3,1]	[4,1]	[5,1]	[2,2]	[3,2]
	0	**0.53**	0.30	0.10	0.02	0.02	0.01
				DGP_σ			
# Regimes (ARMA,Variance)	[1,1]	[1,2]	[1,3]	[2,2]	[2,3]	[3,2]	[3,3]
	0	**0.36**	0.35	0.08	0.07	0.03	0.03

Table 5: Posterior probabilities of the number of regimes for four simulated series generated from DGP displayed in Table 4. The true number of regimes is in bold.

The posterior means and their standard deviations are reported in Table 6. We also display the maximum likelihood estimates given the true states. The posterior estimates and their

MLE counterparts are always very similar and thus confirm the estimations.

7 Applications

In this section, the sticky IHMM-ARMA model is estimated on two (low- and high-frequency) empirical series: the quarterly U.S. GDP growth and the daily Dow Jones Industrial Average realized volatility. Allowing for two different break vectors (one for the mean and a second for the variance) clearly enhances the flexibility of the model and sheds some lights on widely known structural breaks in the literature. For instance, focusing on the U.S. GDP growth, the sticky IHMM-ARMA model confirms that the beginning of the great moderation era is characterized by a drop in the variance while no change in the mean parameters is observable. Regarding the realized volatility, the sticky IHMM-ARMA model also turns out to be a promising alternative to volatility models estimated on returns series.

7.1 Breaks and cycles in U.S. GDP growth

In this section, we revisit the seminal paper of Hamilton (1989) on the business cycles by applying the sticky IHMM-ARMA model to U.S. quarterly GDP data from 1959Q1 to 2011Q3 (211 observations). More specifically, we complicate the Hamilton's approach since the Auto-regressive specification is augmented with an MA term. To begin with, Table 7 provides the summary statistics of the series. It exhibits a small but significant auto-correlation. The normality assumption is rejected for the entire sample but seems plausible for the first part of the series (until 1985). Finally, the tests assuming under the null the stationarity of the variance over the series clearly reject this assumption. It confirms the great moderation phenomenon and indicates that breaks should be taken into account.
The results, reported in Table 8, provide evidence of three regimes for the mean parameters and two regimes for the variance.

As reported in Table 9 and in Figure 1, we first observe that a break occurs in the variance in 1983, at the beginning of the Great Moderation, as originally detected by Kim and Nelson (1999), and Perez-Quiros and McConnell (2000), with a volatility divided by four after 1983 (see Table 10). Secondly, we note that the model identifies separate breaks for the variance

Regime	μ		β		ϕ		σ^2	
				DGP_μ				
	MCMC	MLE	MCMC	MLE	MCMC	MLE	MCMC	MLE
1	1.77	1.67	0.75	0.76	0.51	0.49	1.08	1.07
	(0.31)	(0.33)	(0.04)	(0.05)	(0.06)	(0.07)	(0.09)	(0.09)
2	-0.60	-0.58	0.76	0.75	0.36	0.37	—	—
	(0.15)	(0.16)	(0.04)	(0.05)	(0.09)	(0.10)	—	—
				DGP_β				
	MCMC	MLE	MCMC	MLE	MCMC	MLE	MCMC	MLE
1	0.39	0.39	0.64	0.61	0.05	0.06	0.92	0.92
	(0.12)	(0.12)	(0.09)	(0.08)	(0.12)	(0.11)	(0.08)	(0.08)
2	0.91	1.02	0.84	0.82	0.22	0.22	—	—
	(0.38)	(0.34)	(0.06)	(0.06)	(0.11)	(0.11)	—	—
				DGP_ϕ				
	MCMC	MLE	MCMC	MLE	MCMC	MLE	MCMC	MLE
1	0.57	0.58	0.73	0.73	0.60	0.57	1.02	1.02
	(0.18)	(0.17)	(0.06)	(0.06)	(0.08)	(0.07)	(0.09)	(0.08)
2	0.51	0.48	0.80	0.80	-0.12	-0.16	—	—
	(0.18)	(0.20)	(0.07)	(0.07)	(0.14)	(0.12)	—	—
				DGP_σ				
	MCMC	MLE	MCMC	MLE	MCMC	MLE	MCMC	MLE
1	0.55	0.59	0.78	0.76	0.19	0.22	0.99	0.95
	(0.11)	(0.13)	(0.03)	(0.04)	(0.05)	(0.07)	(0.11)	(0.09)
2	—	—	—	—	—	—	2.54	2.52
	—	—	—	—	—	—	(0.48)	(0.36)

Table 6: Posterior means and standard deviations related to the most observed number of regimes for the four different simulated data.

Table 7: Summary statistics of quarterly U.S. GDP data.

Average	0.75977
Std error	0.88397
Skewness	-0.31552
Excess kurtosis	1.4286
First auto correlation	0.25152
P-val. ACF test (LBQ) with 4 lags	5.5372e-008
P-val. normality test (JB)	0.0020581
P-val. normality test (1st half sample)	0.37691
P-val. normality test (2nd half sample)	0.001
F-test : same variances (half-sample)	0.037655
Bartlett test : same variances (5 batches of equal size)	0.0012432
Nb. Obs	210

# Regimes (ARMA,Variance)	[3,2]	[4,2]	[2,2]	[5,2]	[6,2]	[3,3]
Prob.	0.33	0.23	0.21	0.11	0.04	0.02

Table 8: U.S. GDP growth : posterior means of the probability of having a specific number of regime

and the mean, since the break in the variance in 1983 has no simultaneous counterpart in the mean equation. As regards the mean equation, the model switches between bust and boom periods (second and third regimes), known as business cycles, as captured by the Hamilton (1989)'s approach. Figure 1 emphasizes that most NBER breaks have been detected by the model. We finally note that a first regime in the mean equation captures the specific 2008-2009 recession.

INSERT FIGURE 1

Specifying dependence between state vectors, as documented in Equations (2) and (3) of Section 3, does not change the results. The very same structural break for the variance is detected and no additional change is observed. Table 11 reports the probabilities of staying in a regime when a break has just occurred in the mean parameters ($D = 4$, see discussion in Section 3) compared to the same probabilities when no switch has happened in the mean parameter. These probabilities are extremely high and the differences are not statistically

Mean dynamic		Variance dynamic	
Date	Regime	Date	Regime
1959Q3	2 into 3	1983Q4	1 into 2
1960Q3	3 into 2	—	—
1960Q4	2 into 3	—	—
1973Q4	3 into 2	—	—
1975Q1	2 into 3	—	—
1979Q4	3 into 2	—	—
1980Q3	2 into 3	—	—
1981Q1	3 into 2	—	—
1982Q3	2 into 3	—	—
1990Q1	3 into 2	—	—
1991Q1	2 into 3	—	—
2000Q4	3 into 2	—	—
2001Q1	2 into 3	—	—
2001Q2	3 into 2	—	—
2001Q3	2 into 3	—	—
2007Q3	3 into 2	—	—
2008Q2	2 into 1	—	—
2009Q1	1 into 2	—	—
2009Q3	2 into 3	—	—

Table 9: U.S. GDP growth : break dates for the mean and the variance.

different from zero.

Regime	μ	β	ϕ	σ^2
1	-0.70	0.29	0.05	1.01
	(0.74)	(0.42)	(0.54)	(0.16)
2	0.41	0.38	-0.04	0.28
	(0.31)	(0.24)	(0.33)	(0.05)
3	0.52	0.37	-0.13	—
	(0.23)	(0.25)	(0.23)	—

Table 10: U.S. GDP growth : posterior means and standard deviations given the most observed number of regimes (two for the mean and two for the variance).

Regime	1	2
p_{ii}	0.9986	0.9994
	0.001	0.0007
λ_i	0.9993	0.9994
	0.001	0.0008
$\lambda_i - p_{ii}$	0.0006	-0.0004
	0.0015	0.0011

Table 11: U.S. GDP growth : posterior means and standard deviations given the most observed number of regimes of the probabilities of staying in the same regime as well as their differences.

7.2 DJIA realized volatility

Volatility models are usually required to estimate the time-varying variance of financial time series. Since it is a well-known stylized fact that the volatility of financial returns is highly persistent, the GARCH and the stochastic volatility models use lags of the (unobserved) conditional variance to predict the current one. On the contrary, if the volatility is directly observable (examples are the realized volatility or the daily log price range), estimating an ARMA model is a natural way to fit the data.[3]

A good ARMA model for realized volatility must take into account two features. First, since many papers document the presence of structural breaks in the volatility (Diebold (1986), Lamoureux and Lastrapes (1990), Hilebrand (2005)), the model should allow for the presence of breaks in the mean dynamic. Second, it has been empirically demonstrated that the variance of the realized volatility is not constant over time.[4] The model should therefore also allow for variance changes in the volatility. These features make the sticky IHMM-ARMA model an interesting candidate to fit the realized volatility.

We here apply the sticky IHMM-ARMA model on log daily Dow Jones realized volatility data, as provided by the Univariate HEAVY paper of Shephard and Sheppard (2010), over the period from 02 January 2004 to 27 February 2009 (1295 observations). Table 12 documents the summary statistics of the series. As expected, the financial series exhibits a strong and significant persistence. The first auto-correlation is equal to 0.8182. The skewness highly deviates from zero and the kurtosis is above 3 indicating that the marginal distribution is asymmetric and exhibits fatter tails than the normal one. The Jarque-Bera test confirms that the observations are far from realizations of a normal distribution but it seems to be due to the second half of the sample. Finally, the tests on variances suggest investigating the presence of structural breaks in the second moment.

[3] Furthermore, in the context of GARCH model, a direct equivalence between the GARCH model and the ARMA model on the squared returns can easily be put forward.

[4] Corsi, Mittnik, Pigorsch, and Pigorsch (2008) show that the residuals of the commonly used ARMA models for realized volatility exhibit volatility clustering. They therefore use a specification which allows the variance of the realized volatility to vary over time.

Table 12: Summary statistics of daily DJIA realized volatility.

Average	-0.94122
Std error	1.0826
Skewness	1.1805
Excess kurtosis	1.8496
First auto correlation	0.81824
P-val. ACF test (LBQ) with 50 lags	0
P-val. normality test (JB)	0.001
P-val. normality test (1st half sample)	0.46932
P-val. normality test (2nd half sample)	0.001
F-test : same variances (half-sample)	1.4339e-072
Bartlett's test : same variances (5 batches of equal size)	1.2598e-050
Nb. Obs	1295

The posterior probability of having a specific number of regimes is given in Table 13. The highest likely number of regimes are 3 for the mean and 2 for the variance.

# Regimes (ARMA,Variance)	[3,2]	[3,3]	[4,2]	[4,3]	[3,4]	[2,2]
	0.21	0.19	0.09	0.09	0.08	0.07

Table 13: DJIA realized volatility : posterior means of the probability of having a specific number of regimes.

The breaks in the mean and the variance are reported in Table 14 and illustrated in Figure 2. Regarding the mean, the regime 1, spanning from 26 February 2007 to 01 March 2007, is a one-week transition period, corresponding to a period of stress on the markets[5], between the 2004-2007 regime 2 and 2007-2009 regime 3. The end of February 2007 break confirms the results found by Dufays (2012) with a sticky IHMM-GARCH approach applied on the S&P500 daily returns. The six breaks (corresponding to two regimes) captured in the variance argue in favor of the relevance of considering "the volatility of the realized volatility", as is titled the paper of Corsi, Mittnik, Pigorsch, and Pigorsch (2008). Assigning the regime changes in the variance to specific events is quite difficult. We first note that most breaks occur in the second half of the sample (no break before April 2006), showing that the kurtosis of the

[5]The 27th February is the day that the Federal Home Loan Mortgage Corporation (Freddie Mac) announces that it will no longer buy the most risky subprime mortgages and mortgage-related securities. This is a major landmark in the emergence of the 2007-2008 financial crisis.

returns increased during the financial crisis.[6] We then observe that a break captures the mid-September impact of the Lehman Brother collapse.

INSERT FIGURE 2

Table 14: DJIA realized volatility : break dates obtained by the K-means algorithm (see Section 5).

Mean dynamic		Variance dynamic	
Date	Regime	Date	Regime
26/02/2007	2 into 1	25/04/2006	1 into 2
1/03/2007	1 into 2	11/07/2006	2 into 1
18/07/2007	2 into 3	20/11/2006	1 into 2
		7/02/2008	2 into 1
		10/09/2008	1 into 2
		5/01/2009	2 into 1

The impulse response functions are displayed in Figure 3. The third regime (starting in July 2007) is clearly more persistent than the second one and reflects the larger market instability. This higher persistence is also apparent in the larger auto-regressive parameter estimate (β) of regime 3, compared to regime 2, as given in Table 15.

INSERT FIGURE 3

As discussed in Section 3, we also specify dependence between state vectors. More precisely, we assume that when a break in the mean has just occurred ($D = 25$), a change in the variance is subject to another probability than the à priori independent transition matrix (see equations (1),(2),(3)). Assuming this dependence does not change the results (not reported for saving space).

8 Conclusion

CP and MS frameworks offer a lot of flexibility by allowing abrupt switches in the parameters. Although efficient algorithms have been developed to handle these models during the last two

[6]See Jondeau and Rockinger (2003) for a study of the conditional kurtosis of the S&P500 over the period 1971-1999.

Regime	μ	β	ϕ	σ^2
1	-0.06	0.35	-0.19	0.18
	(0.40)	(0.65)	(0.44)	(0.01)
2	-0.16	0.88	-0.48	0.37
	(0.15)	(0.10)	(0.10)	(0.03)
3	-0.05	0.94	-0.41	—
	(0.13)	(0.08)	(0.13)	—

Table 15: DJIA realized volatility: posterior means and standard deviations related to the most observed number of regimes (three for the mean and two for the variance).

decades, there are still cases where the solutions are not fully satisfactory, namely for CP- and MS-ARMA models due to their path dependence nature. We propose a new approach to estimate these CP- and MS-ARMA models by combining the model of Klaassen (2002) with a Metropolis-Hastings step. Our approach relies on a sticky infinite hidden Markov framework, which encompasses the CP and MS setting by covering a potentially infinite number of regimes. This structure has two clear advantages (and can be applied to models without path dependence). First it allows for a self-determination of the specification (CP and/or MS) and of the number of regimes. Second, it provides a way of relaxing the implicit assumption that all parameters change together when a structural break occurs. In other words, our model can have a different number of breaks for the mean and variance parameters. In addition, since some correlation between the breaks in the mean and in the variance is quite likely, we also discuss how to specify the dependence between these two dynamics.

Results on simulated series document the ability of the sticky IHMM-ARMA model to detect the right specifications and the true number of regimes. Empirical examples on the quarterly U.S. GDP growth and the log daily DJIA realized volatility highlight the relevance of allowing for different structural breaks in the mean and in the variance. In particular, for the US GDP growth we find a structural break in the variance at the beginning of the great moderation era with no simultaneous break in the mean parameters. Our model also captures quite precisely the breaks reported by the NBER's Business Cycle Dating Committee. Regarding the DJIA volatility, we show that the sticky IHMM-ARMA model, combined with realized volatility data, is a relevant alternative to MS/CP GARCH models applied on return series by capturing similar breaks on the mean equation. Our results also confirm those of Corsi, Mittnik, Pigorsch, and Pigorsch (2008) where the volatility of the realized volatility was shown to change over time.

A sticky IHMM-ARMA 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 L 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 :

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

Remind that the set $\Theta = \{\mu_1, \ldots, \mu_L, \beta_1, \ldots, \beta_L, \phi_1, \ldots, \phi_L, \sigma_1, \ldots, \sigma_L\}'$ contain all the continuous parameters of the model, the set $\Theta_j = \{\mu_j, \beta_j, \phi_j, \sigma_j\}'$ denotes the continuous parameters of regime j, the set $\theta = \{\mu_1, ..., \mu_L, \beta_1, ..., \beta_L, \phi_1, ..., \phi_L\}'$ gathers all the ARMA parameters and $\theta_j = \{\mu_j, \beta_j, \phi_j\}'$ denotes the ARMA parameters of the regime j.
The MCMC simulation iterates between the following steps.

1. Sampling S^a_T from $f(S^a_{1:T}|\Theta, P^a, S^b_{1:T}, Y_{1:T})$: see Section 4.

2. Sampling P^a from $f(P^a|H_{Dir}, \pi^a, S^a_{1:T}, Y_{1:T})$: for j=1,...,L, sample $p^a_{j,1:L} \sim \text{Dir}(\alpha^a\pi^a_1 + n^a_{j,1}, ..., \alpha^a\pi^a_j + \kappa^a + n^a_{j,j}, ..., \alpha^a\pi^a_L + n^a_{j,L})$ where $n^a_{j,k}$ denotes the number of transition from state j to k observed in the state vector $S^a_{1:T}$.

3. Sampling S^b_T from $f(S^b_{1:T}|\Theta, P^b, S^a_{1:T}, Y_{1:T})$ by forward-backward algorithm (see Rabiner (1989)).

4. Sampling P^b from $f(P^b|H_{Dir}, \pi^b, S^b_{1:T}, Y_{1:T})$: for j=1,...,L, sample $p^b_{j,1:L} \sim \text{Dir}(\alpha^b\pi^b_1 + n^b_{j,1}, ..., \alpha^b\pi^b_j + \kappa^b + n^b_{j,j}, ..., \alpha^b\pi^b_L + n^b_{j,L})$ where $n^b_{j,k}$ denotes the number of transition from state j to k observed in the state vector $S^b_{1:T}$.

5. Sampling $\{\alpha^a, \kappa^a, \eta^a\}$ from $f(\alpha^a, \kappa^a, \eta^a|P^a, \pi^a, S^a_{1:T}, Y_{1:T})$:

   (a) Introduce auxiliary variables :

   - Sampling m : For j=1,...,L, and k=1,...,L. Set $m_{j,k} = 0$. For i=1,...,$n_{j,k}$ sample $x_i \sim \text{Bernoulli}(\frac{\alpha^a\pi^a_k + \kappa^a 1_{\{j=k\}}}{i-1+\alpha^a\pi^a_k+\kappa^a 1_{\{j=k\}}})$ and increment $m_{j,k} = 0$ if $x_i = 1$.
   - Sampling r : For j=1,...,L. $r_j \sim \text{Binomial}(m_{j,j}, \frac{\rho}{(1-\rho)\pi^a_j+\rho})$ where $\rho = \frac{\alpha^a}{\alpha^a+\kappa^a}$

- 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,...,L, if $\bar{m}_{.,k} > 0$ then increment $\bar{K}$

(b) Sampling α^a and κ^a

- Sample auxiliary variables : for i=1,...,L, $q_i \sim \text{Beta}(\alpha^a + \kappa^a + 1, n_{i,.})$ and $s_i \sim \text{Bernoulli}(\frac{n_{i,.}}{n_{i,.}+\alpha^a+\kappa^a})$
- Sample $\rho = \frac{\kappa^a}{\alpha^a+\kappa^a} \sim \text{Beta}(\rho_{\text{hyp1}} + r_., \rho_{\text{hyp2}} + m_{.,.} - r_.)$ where ρ_{hyp1} and ρ_{hyp2} denotes the hyper-parameters of ρ (see Table 3)
- Sample $\alpha^a + \kappa^a \sim \text{G}(a_{\text{hyp}} + m_{.,.} - s_., (\frac{1}{b_{\text{hyp}}} - \log q_.)^{-1})$ where a_{hyp} and b_{hyp} denotes the hyper-parameters of $\alpha^a + \kappa^a$ (see Table 3)
- set $\alpha^a = (1-\rho)(\alpha^a + \kappa^a)$ and $\kappa^a = \rho(\alpha^a + \kappa^a)$

(c) Sampling η^a

- Sample auxiliary variables : $\tilde{q} \sim \text{Beta}(\eta^a + 1, \bar{m}_{.,.})$ and $\tilde{s} \sim \text{Bernoulli}(\frac{\bar{m}_{.,.}}{\bar{m}_{.,.}+\eta^a})$
- Sample $\eta^a \sim \text{G}(\eta^a_{\text{hyp1}} + \bar{K} - \tilde{s}, \{\frac{1}{\eta^a_{\text{hyp2}}} - \log \tilde{q}\}^{-1})$ where η^a_{hyp1} and η^a_{hyp2} denotes the hyper-parameters of η^a (see Table 3)

6. Sampling $\{\alpha^b, \kappa^b, \eta^b\}$ from $f(\alpha^b, \kappa^b, \eta^b | P^b, \pi^b, S^b_{1:T}, Y_{1:T})$: similar to previous item.

7. Sampling π^a from $f(\pi^a | P^a, H_{Dir}, S^a_{1:T}, Y_{1:T}) \sim \text{Dir}(\frac{\eta^a}{L} + \bar{m}_{.,1}, ..., \frac{\eta^a}{L} + \bar{m}_{.,L})$.

8. Sampling π^b from $f(\pi^b | P^b, H_{Dir}, S^b_{1:T}, Y_{1:T})$: similar to previous item.

9. For $j = 1, ..., L$, sampling μ_j, β_j, ϕ_j from $f(\mu_j, \beta_j, \phi_j | \bar{\mu}, \bar{\Sigma}, S^a_{1:T}, S^b_{1:T}, \{\sigma^2_1, ..., \sigma^2_L\}, Y_{1:T})$ by M-H exactly as in Henneke, Rachev, Fabozzi, and Nikolov (2011).

10. For $j = 1, ..., L$, sampling σ_j^{-2} from $f(\sigma_j^{-2} | \bar{e}, \bar{f}, S^a_{1:T}, S^b_{1:T}, Y_{1:T}) \sim G(0.5 n^b_{.,j} + \bar{e}, (0.5 \sum_{t=1}^T \epsilon_t^2 \delta_{\{s^b_t = j\}} + \frac{1}{\bar{f}})^{-1})$ where $\delta_{\{s^b_t=j\}}$ is the Dirac function equal to one if $s^b_t = j$ and zero otherwise.

11. Sampling $\bar{\mu}, \bar{\Sigma}$ from $f(\bar{\mu}, \bar{\Sigma} | \Theta)$:
    - Drawing $\bar{\mu}$ from $f(\bar{\mu} | \bar{\Sigma}, \Theta) \sim N(\mu_{\text{post}}, \Sigma_{\text{post}})$ where $\Sigma_{\text{post}} = (\underline{\Sigma}^{-1} + L\bar{\Sigma}^{-1})^{-1}$ and $\mu_{\text{post}} = \Sigma_{\text{post}}(\underline{\Sigma}^{-1}\underline{\mu} + \sum_{j=1}^L \bar{\Sigma}^{-1}\theta_j)$.
    - Drawing $\bar{\Sigma}^{-1}$ from $f(\bar{\Sigma}^{-1} | \bar{\mu}, \Theta) \sim \text{Wishart}(\underline{v} + L, (\underline{V}^{-1} + \sum_{j=1}^L (\theta_j - \bar{\mu})(\theta_j - \bar{\mu})')^{-1})$.

12. Sampling $\bar{e}, \bar{f}^{-1}$ from $f(\bar{e}, \bar{f}^{-1} | \{\sigma^2_1, ...\sigma^2_L\}, Y_{1:T})$:

- Drawing $\bar{e}$ from $f(\bar{e}|\bar{f}, \{\sigma_1^2, ...\sigma_L^2\}, Y_{1:T})$ by Metropolis. The proposal distribution is Normal with variance equal to 0.5.
- Drawing $\bar{f}^{-1}$ from $f(\bar{f}^{-1}|\bar{e}, \{\sigma_1^2, ...\sigma_L^2\}, Y_{1:T}) \sim G(L\bar{e} + \underline{f_a}, (\frac{1}{\underline{f_b}} + \sum_{j=1}^{L} \sigma_j^{-2})^{-1})$.

B Adapted MS-ARMA model

In this section, we detail the computation of $\tilde{\epsilon}_{t-1,s_t^a} = E[\epsilon_{t-1}|Y_{1:t-1}, s_t^a, \Theta, P^a, S_{1:t-1}^b]$ that is relevant for estimating the model of Klaassen (2002). By the definition of the expectation, we have that :

$$E[\epsilon_{t-1}|Y_{1:t-1}, s_t^a, \Theta, P^a, S_{1:t-1}^b] = \sum_{i=1}^{L} \epsilon_{\{t-1,s_{t-1}^a=i\}} f(s_{t-1}^a = i|Y_{1:t-1}, s_t^a, \Theta, P^a, S_{1:t-1}^b)$$

where $\epsilon_{\{t-1,s_{t-1}^a=i\}} = y_{t-1} - \mu_{\{s_{t-1}^a=i\}} - \beta_{\{s_{t-1}^a=i\}} y_{t-2} - \phi_{\{s_{t-1}^a=i\}} \tilde{\epsilon}_{\{t-2,s_{t-1}^a=i\}}$ We now detail how to compute the conditional distribution $f(s_{t-1}^a|Y_{1:t-1}, s_t^a, \Theta, P^a, S_{1:t-1}^b)$.

$$\begin{aligned}
f(s_{t-1}^a|Y_{1:t-1}, s_t^a, \Theta, P^a, S_{1:t-1}^b) &= \frac{f(s_t^a, s_{t-1}^a|Y_{1:t-1}, \Theta, P^a, S_{1:t-1}^b)}{f(s_t^a|Y_{1:t-1}, \Theta, P^a, S_{1:t-1}^b)} \\
&= \frac{f(s_{t-1}^a|Y_{1:t-1}, \Theta, P^a, S_{1:t-1}^b) f(s_t^a|s_{t-1}^a, P^a)}{f(s_t^a|Y_{1:t-1}, \Theta, P^a, S_{1:t-1}^b)} \\
&= \frac{f(s_{t-1}^a|Y_{1:t-1}, \Theta, P^a, S_{1:t-1}^b) f(s_t^a|s_{t-1}^a, P^a)}{\sum_{i=1}^{L} f(s_{t-1}^a = i|Y_{1:t-1}, \Theta, P^a, S_{1:t-1}^b) f(s_t^a|s_{t-1}^a = i, P^a)}
\end{aligned}$$

In the light of the previous calculation, the forward step of the forward-backward algorithm provides all the quantities to compute at each time the conditional distribution $f(s_{t-1}^a|Y_{1:t-1}, s_t^a, \Theta, P^a, S_{1:t-1}^b)$.

References

ARDIA, D., L. HOOGERHEIDE, AND H. VAN DIJK. (2009): "To Bridge, to Warp or to Wrap? A Comparative Study of Monte Carlo Methods for Efficient Evaluation of Marginal Likelihoods," *Tinbergen Institute Discussion paper TI*, pp. 2009–017/4.

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.*

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

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

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

CORSI, F., S. MITTNIK, C. PIGORSCH, AND U. PIGORSCH (2008): "The Volatility of Realized Volatility," *Econometric Reviews*, 27(1-3), 46–78.

DAVID, F. N., AND P. WHITTLE (1952): "Hypothesis-Testing in Time Series Analysis," *Biometrika*, 39.

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.*

Eo, Y. (2012): "Bayesian Inference about the Types of Structural Breaks When There Are Different Breaks in Many Parameters," Available at SSRN: http://ssrn.com/abstract=2011825 or http://dx.doi.org/10.2139/ssrn.2011825.

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.

——— (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.

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

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

Haas, M., S. Mittnik, and M. Paolella (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.

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.

Hilebrand, E. (2005): "Neglecting parameter changes in GARCH models," *Journal of Econometrics*, 129, 121–138.

Ishwaran, H., and M. Zarepour (2002): "Exact and Approximate Sum Representations for the Dirichlet Process," *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 30(2), pp. 269–283.

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

Jondeau, E., and M. Rockinger (2003): "Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements," *Journal of Economic Dynamics and Control*, 27(10), 1699–1737.

Kim, C.-J., and C. R. Nelson (1999): "Has The U.S. Economy Become More Stable? A Bayesian Approach Based On A Markov-Switching Model Of The Business Cycle," *The Review of Economics and Statistics*, 81(4), 608–616.

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.

Kurihara, K., M. Welling, and Y. Teh (2007): "Collapsed variational Dirichlet process mixture models," in *In proc. Twentieth International Joint Conference on Artificial Intelligence (IJCAI07).*

Lamoureux, C., and W. Lastrapes (1990): "Persistence in variance, structural change, and the GARCH model," *Journal of Business and Economic Statistics*, 8, 225–234.

MacQueen, J. B. (1967): "Some Methods for Classification and Analysis of MultiVariate Observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. by L. M. L. Cam, and J. Neyman, vol. 1, pp. 281–297. University of California Press.

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.

Perez-Quiros, G., and M. M. McConnell (2000): "Output Fluctuations in the United States: What Has Changed since the Early 1980's?," *American Economic Review*, 90(5), 1464–1476.

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

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

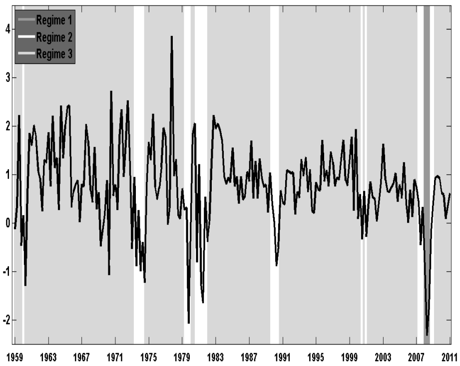
Sethuraman, J. (1994): "A constructive definition of dirichlet priors," *Statistica Sinica*, 4, 639–650.

Shephard, N., and K. Sheppard (2010): "Realising the future: forecasting with high-frequency-based volatility (HEAVY) models," *Journal of Applied Econometrics*, 25(2), 197–231.

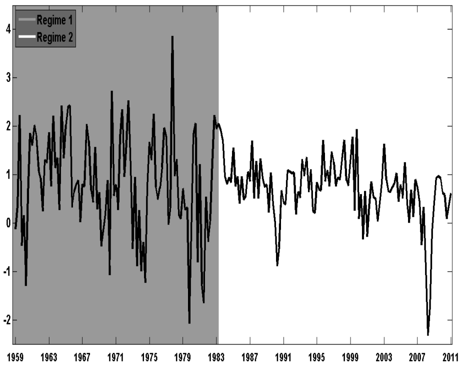
Song, Y. (2011): "Modelling Regime Switching and Structural Breaks with an Infinite Dimension Markov Switching Model," *Working Paper Series*.

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

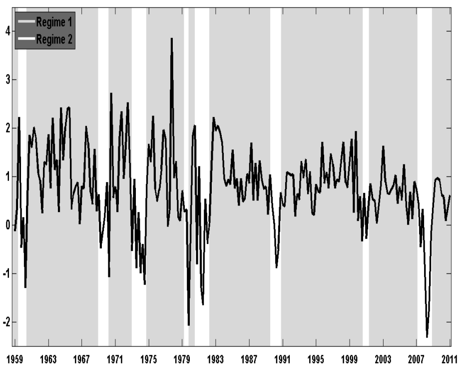
Van Gael, J., Y. Saatci, Y. Teh, and Z. Ghahramani (2008): "Beam sampling for the infinite hidden Markov model," *Proceedings of the 25th International Conference on Machine learning*.



Breaks in the mean

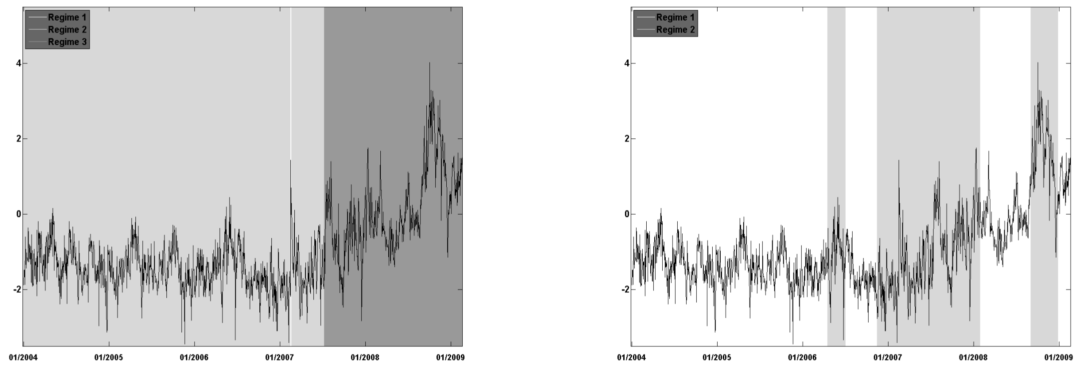


Breaks in the variance



NBER breaks

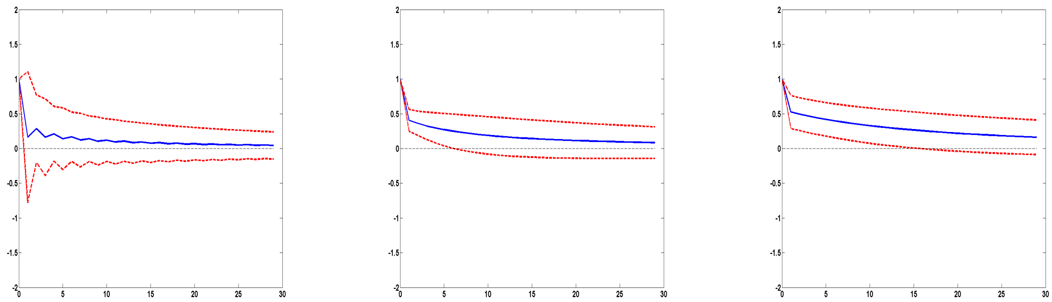
Figure 1: U.S. GDP growth: breaks in the mean and the variance obtained by the K-means algorithm (see Section 5).



Breaks in the mean

Breaks in the variance

Figure 2: DJIA realized volatility: breaks in the mean and the variance.



Regime 1

Regime 2

Regime 3

Figure 3: DJIA realized volatility: impulse response functions of the three observed regimes.