Factor Analysed Hidden Markov Models for Conditionally Heteroskedastic Financial Time Series

Mohamed SAIDANE [1] & Christian LAVERGNE [2]

[1] INRIA Rhône-Alpes - Projet MISTIS
Zirst - 655 Avenue de l'Europe - Montbonnot - 38334 Saint Martin Cedex - France
Email : Mohamed.Saidane@inrialpes.fr

[2] Unité Mixte de Recherche CNRS / Université Montpellier II
Institut de Mathématiques et de Modélisation de Montpellier
Place Eugène Bataillon 34095 Montpellier Cedex 5, France
Email : Christian.Lavergne@math.univ-montp2.fr

Abstract

In this paper, we investigate a class of switching latent dynamic factor models for conditionally heteroskedastic financial time series. The proposed innovation in this paper can be viewed in two different ways. It can be presented as a generalization of the standard factor analysed hidden Markov model (Rosti and Gales 2002) since we will allow the common or specific variance of the model to be a stochastic function of time. Alternatively, we can present the contribution as an extension to conditionally heteroskedastic factor analysis (Demos and Sentana [1998]) where the low dimensional subspace is modelled with a Gaussian hidden Markov model (HMM) and the observation noise by a Gaussian model. More specifically, we concentrate on situations where the factor variances are modelled by univariate GQARCH processes. Hence, our main contribution relies upon the use of a piece-wise multivariate and linear process - which we can also regard as a linear and dynamic system with mixture of states - for modelling the regime switches. In particular, we supposed that the observed series can be modelled using a time varying parameter model with the assumption that the evolution of these parameters is governed by a first-order hidden Markov process. An EM algorithm for the parameter optimisation is presented along with a number of methods to increase the efficiency of training. The various regimes, the common factors and their volatilities are supposed unobservable and the inference must be carried out from the observable process. Finally, a Monte Carlo simulation study is carried out to investigate the properties of the latente structure inference and estimation procedures.

Key words : Dynamic Factor Models, EM Algorithm, Extended Kalman filter, Generalized Pseudo-Bayesian Method, GQARCH Processes, HMM, Model Selection, Monte Carlo Simulations, Switching State-Space Models, Viterbi Approximation.

1 Introduction

Many issues in finance require the analysis of the variances and covariances of a large number of assets. For instance, asset pricing theories often derive restrictions on mean returns or risk premia from restrictions on cross sectional correlations of a large collection of assets. Similarly, portfolio allocation models exploit the imperfect correlation of different asset returns in order to reduce risk by means of well diversified portfolios consisting of a large number of assets.

Traditionally, these issues were considered in a static framework, but recently the emphasis has shifted to intertemporal models, in which agents actions are based on the distribution of returns conditional on their time-varying information set. At least partly, this shift in interest has been motivated by the fact that time variation in the volatility of asset prices is nowadays well documented and widely recognized as one of the characteristics of financial markets. In parallel with these theoretical developments, a large family of statistical models for the time variation in conditional variances has grown up, mostly but not only, around the Atutoregressive Conditional Heteroskedasticity (ARCH) model of Engle [1982], and numerous applications have already appeared (see Bollerslev, Chou and Kroner [1992] for a recent survey). By and large, though, most applied work has been on univariate financial time series, as the application of these models in a multivariate context has been hampered by the large number of parameters involved. To avoid this problem, Demos and Sentana [1998] proposed a multivariate approach based on the same idea as traditional (i.e. conditionally homoskedastic) factor analysis. That is, it is assumed that each of q observed variables is a linear combination of k ($k < q$) common factors plus an idiosyncratic term, but allowing for ARCH type effects in the underlying factors.

In spite of the interest carried by this type of models, certain financial return series are still not well characterized by such a specification. Indeed, financial time series are often characterized by a non constant - conditional or unconditional - volatility. The volatility shocks tend to persist through time and can, thus, affect certain fundamental behaviors of the financial returns, such as the leptokurtic aspect and the mean reverting phenomenon. In general, the variation in the exchange rate regime, deregulation, financial opening, the financial markets debacles, shocks of policy such as tax and trade reforms, can be modelled as structural changes of the parameters as well as variance changes of one of the shocks.

For modelling nonconstancy, we proposed a new approach based on the combination of conditionally heteroskedastic factor models with hidden Markov models. The original idea of this work is the use of a piece-wise multivariate and linear process - which we can also regard as a linear and dynamic system with mixture of states - for modelling the regime switches. In particular, we supposed that the observed series can be approximated using a time varying parameter model with the assumption that the evolution of these parameters is governed by a first-order hidden Markov process with m states.

Paper is organized as follows. In section 2, we will introduce the general form of the model in its simplest version. It is in fact a standard factor analysis model combined with a first-order hidden Markov process. We will study, thereafter, its likelihood function for finally estimating its parameters by using an exact EM algorithm inspired by the Baum and Welch algorithm for HMMs. In the third section we will extend the standard model for studying the co-movements of financial time series characterized by a dynamic heteroskedasticity in the variances. We will study it thereafter in a multi-phasic space-state structure, in order to estimate common factors by using an extended version of the Kalman filter based on the "moment matching" technique. The complete likelihood function and the EM algorithm will be presented in the fourth section, where we will discuss with much more details the estimate of the parameters of the conditionally heteroskedastic component based on the restoration of the discrete and continuous hidden states by using, either posterior probabilities already provided by the generalized pseudo-bayesian algorithm (GPB1), or an approximated version of the Viterbi algorithm. In this section, we will also present another alternative approach for the inference of the latent structures and the estimate of the parameters of these models, based on the Viterbi approximation. In this case, we consider the model as a dynamic bayesian network with mixture of states. In the section 5, the model selection problem is considered and we derived possible penalized criteria for choosing among several specific models, focusing essentially on the number of hidden states. Finally, in the last section we will study and test the quality of the estimates through Monte Carlo simulations.

2 Factor Analysed Hidden Markov Models

The factor analysed hidden Markov model (FAHMM) is a dynamic state space generalisation of a multiple component factor analysis system. The k-dimensional state vectors are generated by a standard diagonal covariance Gaussian mixture HMM. The q-dimensional observation vectors are generated by a multiple noise component factor analysis observation process. A generative model for FAHMM can be described as follows [1] :

$$S_t \sim P(S_t = j / S_{t-1} = i)$$

$$\mathbf{f}_t = \mathbf{w}_{s_t}, \quad \text{où} \quad \mathbf{w}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_j)$$

$$\mathbf{y}_t = \mathbf{X}_{s_t}\mathbf{f}_t + \varepsilon_{s_t} \quad \text{avec} \quad \varepsilon_j \sim \mathcal{N}(\theta_j, \mathbf{\Psi}_j)$$

The HMM state transition probabilities from state i to state j are represented by p_{ij} and the parameters $\{\mathbf{0}, \theta_j\}$ are, respectively, the means of the $(k \times 1)$ vectors of latent factors and the $(q \times 1)$ observation vectors for all $j = 1..., m$ and $t = 1..., n$. The factor loading matrices, the conditionally heteroskedastic factor diagonal variance-covariance matrices and the idiosyncratic factors ε_t and the common factors $\mathbf{f}_t$ for each state, respectively, are indicated by $\mathbf{X}_j$, $\mathbf{\Psi}_j$ and $\mathbf{H}_j$, of dimensions $(q \times k)$, $(q \times q)$ and $(k \times k)$ with $k \ll q$.

An important aspect of any generative model is the complexity of the likelihood calculations. The generative model above can be expressed by the two following Gaussian distributions :

$$\begin{aligned} p(\mathbf{f}_t / S_t = j) &= \mathcal{N}(\mathbf{0}, \mathbf{H}_j) \quad (1) \\ p(\mathbf{y}_t / \mathbf{f}_t, S_t = j) &= \mathcal{N}(\theta_j + \mathbf{X}_j \mathbf{f}_t, \mathbf{\Psi}_j) \quad (2) \end{aligned}$$

The likelihood of an observation $\mathbf{y}_t$ given the state $S_t = j$ can be obtained by integrating the state vector $\mathbf{f}_t$ out of the product of the above Gaussians. The resulting likelihood is also a Gaussian and can be written as :

$$b_j(\mathbf{y}_t) = p(\mathbf{y}_t / S_t = j) = \mathcal{N}(\mathbf{y}_t \,/\, \theta_j, \mathbf{\Sigma}_j) \quad (3)$$

[1] The $\sim$ symbol in $S_t \sim P(S_t/S_{t-1})$ is used to represent a discrete Markov chain. Normally it means the variable on the left hand side is distributed according to the probability density function on the right hand side as in $\mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_t)$.

where $\mathbf{\Sigma}_j = \mathbf{X}_j \mathbf{H}_j \mathbf{X}_j' + \mathbf{\Psi}_j$. The likelihood calculation requires inverting m full $q \times q$ covariance matrices. If the amount of memory is not an issue, the inverses and the corresponding determinants for all the discrete states in the system can be computed prior to starting off with the training and recognition. However, this can rapidly become impractical for a large system. A more memory efficient implementation requires the computation of the inverses and determinants for each time instant. These can be efficiently obtained using the following equality for matrix inverses :

$$\left[\mathbf{X}_j \mathbf{H}_j \mathbf{X}_j' + \mathbf{\Psi}_j\right]^{-1} = \mathbf{\Psi}_j^{-1} - \mathbf{\Psi}_j^{-1} \mathbf{X}_j \left[\mathbf{X}_j' \mathbf{\Psi}_j^{-1} \mathbf{X}_j + \mathbf{H}_j^{-1}\right]^{-1} \mathbf{X}_j' \mathbf{\Psi}_j^{-1} \tag{4}$$

2.1 Optimising FAHMM Parameters

The maximum likelihood (ML) criterion may be used to optimise the FAHMM parameters. It is also possible to find a discriminative training scheme such as minimum classification error (L. Saul and M. Rahim [1999]). However, in this work only ML training is considered. In common with standard HMM training described in L.R. Rabiner [1989], the expectation maximisation (EM) algorithm is used. For a sequence of observation vectors $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n\}$, a sequence of continuous state vectors $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n\}$ and a sequence of discrete HMM states $\mathcal{S} = \{S_1, S_2, ..., S_n\}$, the auxiliary function for FAHMMs can be written as :

$$p(\mathcal{Y}, \mathcal{F}, \mathcal{S}) = p(S_1) \prod_{t=2}^{n} p(S_t / S_{t-1}) \prod_{t=1}^{n} p(\mathbf{f}_t / S_t) p(\mathbf{y}_t / \mathbf{f}_t, S_t) \tag{5}$$

where $p(S_1) = \pi_{s_1}$ and $p(S_t/S_{t-1}) = p_{S_{t-1} S_t}$ are the initial state and the discrete state transition probabilities.

All the sufficient statistics, necessary for the implementation of the E step, are evaluated using the parameters from the previous iteration. This derivation assumes that the first discrete state is always the initial state and all states are emitting. It is easy to extend the derivation for use with explicit initial discrete state probabilities and to include non-emitting states. A set of parameters, $\widehat{\Theta}$, that maximise the auxiliary function is found during the maximisation step

$$\widehat{\Theta} = \arg\max_{\Theta} Q(\Theta, \widehat{\Theta})$$

These parameters will be used as the set of old parameters in the following iteration, $\widehat{\Theta} \longrightarrow \Theta^{(i+1)}$. The auxiliary function for FAHMMs can be written as :

$$\begin{aligned} Q(\Theta, \Theta^{(i)}) &= \mathbb{E}\left[\log p(\mathcal{Y}, \mathcal{F}, \mathcal{S} / \Theta^{(i)}) / \mathcal{Y}, \Theta\right] \\ &= \sum_{\forall \mathcal{S}} \int p(\mathcal{F} / \mathcal{Y}, \mathcal{S}, \Theta) p(\mathcal{S} / \mathcal{Y}, \Theta) \log p(\mathcal{Y}, \mathcal{F}, \mathcal{S} / \Theta^{(i)}) \, df \end{aligned} \tag{6}$$

where all the possible discrete state and continuous state sequences of length n are included in the sum and the integral. The set of current model parameters is represented by $\Theta^{(i)}$.

2.1.1 Forward-Backward Algorithm

The likelihood of being in discrete state j and the observations up to time instant t is represented by the forward variable, $\alpha_j(t) = p(S_t = j,\ \mathcal{Y}_{1:t})$. Assuming that the first observation is generated by the first discrete state, the forward variable is initialised as :

$$\begin{cases} b_1(\mathbf{y}_1) & , \quad j = 1 \\ 0 & , \quad j \neq 1 \end{cases}$$

Using the conditional independence assumption in FAHMM, the forward variable at time instant t is defined by the following recursion :

$$\alpha_j(t) = p(S_t = j, \mathcal{Y}_{1:t}) = b_j(\mathbf{y}_t) \sum_{i=1}^{m} p_{ij} \alpha_i(t-1) \tag{7}$$

The likelihood of the observations from $t+1$ to n given being in state i at time instant t is represented by the backward variable, $\beta_i(t) = p(\mathcal{Y}_{t+1:n}/S_t = i)$. The backward variable is initialised as $\beta_i(n) = 1$ for all $i \in [1, m]$. Using the conditional independence assumption in FAHMM, the backward variable at time instant $t-1$ is defined by the following recursion :

$$\beta_i(t-1) = p(\mathcal{Y}_{t:n}/S_{t-1} = i) = \sum_{j=1}^{m} p_{ij} b_j(\mathbf{y}_t) \beta_j(t) \tag{8}$$

The likelihood of the observation sequence, $\mathcal{Y}$, can be represented in terms of the forward and backward variables as follows :

$$p(\mathcal{Y}) = \sum_{i=1}^{m} p(S_t = i, \mathcal{Y}_{1:t}) p(\mathcal{Y}_{t+1:n}/S_t = i) = \sum_{i=1}^{m} \alpha_i(t) \beta_i(t) \tag{9}$$

The probability of being in state j at time t given the observation sequence is needed in the parameter update formulae. This likelihood can be expressed in terms of the forward and back- ward variables as follows :

$$\gamma_j(t) = p(S_t = j/\mathcal{Y}) = \frac{\alpha_j(t)\beta_j(t)}{\sum_{i=1}^{m} \alpha_i(t)\beta_i(t)}$$

The joint probability of being in state i at time instant $t-1$ and in state j at time instant t given the observation sequence is needed in the transition parameter update formulae. This likelihood can be expressed in terms of the forward and backward variables as follows :

$$\xi_{ij}(t) = p(S_{t-1} = i, S_t = j/\mathcal{Y}) = \frac{\alpha_i(t-1) p_{ij} b_j(\mathbf{y}_t) \beta_j(t)}{\sum_{i=1}^{m} \alpha_i(t)\beta_i(t)} \tag{10}$$

2.1.2 Continuous State Posterior Statistics

Given the current discrete state, $S_t = j$, the joint likelihood of the current observation and continuous state vector is a Gaussian,

$$p(\mathbf{y}_t, \mathbf{f}_t / S_t = j) = \mathcal{N}\left[\begin{pmatrix} \theta_j \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{X}_j\mathbf{H}_j\mathbf{X}_j' + \mathbf{\Psi}_j & \mathbf{X}_j\mathbf{H}_j \\ \mathbf{H}_j\mathbf{X}_j' & \mathbf{H}_j \end{pmatrix}\right] \tag{11}$$

The posterior distribution is also a Gaussian and can be written as :

$$p(\mathbf{f}_t / \mathbf{y}_t, S_t = j) = \mathcal{N}\left[\mathbf{K}_j(\mathbf{y}_t - \theta_j), \mathbf{H}_j - \mathbf{K}_j\mathbf{X}_j\mathbf{H}_j\right] \tag{12}$$

where $\mathbf{K}_j = \mathbf{H}_j\mathbf{X}_j'\left[\mathbf{X}_j\mathbf{H}_j\mathbf{X}_j' + \mathbf{\Psi}_j\right]^{-1}$. We denote

$$\begin{aligned} \widehat{\mathbf{f}}_j(t) &= \mathbf{K}_j(\mathbf{y}_t - \theta_j) \\ \widehat{\mathbf{R}}_j(t) &= \mathbf{H}_j - \mathbf{K}_j\mathbf{X}_j\mathbf{H}_j + \widehat{\mathbf{f}}_j(t)\widehat{\mathbf{f}}_j(t)' \end{aligned}$$

These statistics are used, also, in the parameter update formulae below.

2.1.3 Parameter Update Formulae

Given the two sets of sufficient statistics above the model parameters can be optimised by solving a standard maximisation problem. Maximising the auxiliary function in Equation (6) with respect to the initial state probabilities π_j (respectively the discrete state transition probabilities, p_{ij}), can be carried out using the Lagrange multiplier together with the sum to unity constraint $\sum_{j=1}^{m} \pi_j = 1$ (respectively $1 - \sum_{j=1}^{m} p_{ij} = 0$), where $p(S_1) = \pi = [\pi_1, \pi_2, ..., \pi_m]'$ and $\pi_j = p(S_1 = j)$. Solving for π_j and p_{ij}, the new initial state and discrete state transition probabilities can be written as :

$$\widehat{\pi}_j = \frac{\gamma_j(1)}{\sum_{i=1}^{m} \gamma_i(1)} \qquad \text{and} \qquad \widehat{p}_{ij} = \frac{\sum_{t=2}^{n} \xi_{ij}(t)}{\sum_{t=2}^{n} \gamma_i(t-1)}$$

Thereafter if we denote by $\mathbf{x}_{jl}$ the l-th row vector of $\mathbf{X}_j$, the maximization of (6) with respect to $\mathbf{X}_j$ gives us :

$$\widehat{\mathbf{x}}_{jl} = \mathbf{k}_{jl}'\mathbf{G}_{jl}^{-1}$$

with

$$\begin{aligned} \mathbf{G}_{jl} &= \frac{1}{\psi_{jl}} \sum_{t=1}^{n} \gamma_j(t)\widehat{\mathbf{R}}_j(t) \\ \mathbf{k}_{jl} &= \frac{1}{\psi_{jl}} \sum_{t=1}^{n} \gamma_j(t)(y_{tl} - \theta_{jl})\widehat{\mathbf{f}}_j(t) \end{aligned}$$

where ψ_{jl} is the l-th diagonal element of the idiosyncratic covariance matrix $\mathbf{\Psi}_j$, y_{tl} and θ_{jl} are the l-th elements of the current observation and the observation noise mean vectors, respectively.

Maximising the auxiliary function with respect to the observation noise mean vector, θ_j, yields

$$\widehat{\theta}_j = \frac{\sum_{t=1}^{n} \gamma_j(t)(\mathbf{y}_t - \mathbf{X}_j \widehat{\mathbf{f}}_j(t))}{\sum_{t=1}^{n} \gamma_j(t)}$$

In a similar way, the idiosyncratic variances will be given by :

$$\widehat{\mathbf{\Psi}}_j = \frac{1}{\sum_{t=1}^{n} \gamma_j(t)} \sum_{t=1}^{n} \gamma_j(t) diag \left[\left(\mathbf{y}_t - \mathbf{X}_j \widehat{\mathbf{f}}_j(t) - \theta_j \right) \left(\mathbf{y}_t - \mathbf{X}_j \widehat{\mathbf{f}}_j(t) - \theta_j \right)' + \mathbf{X}_j \mathbf{H}_j \mathbf{X}_j' \right]$$

and the common factor variances by :

$$\widehat{\mathbf{H}}_j = \frac{1}{\sum_{t=1}^{n} \gamma_j(t)} diag \left\{ \sum_{t=1}^{n} \gamma_j(t) \widehat{\mathbf{R}}_j(t) \right\}$$

3 Conditionally Heteroskedastic FAHMM

the model which we will present now is a generalization of the models suggested by Diebold and Nerlove [1988], Engle, Ng and Rothschild [1990] ; Harvey, Ruiz and Sentana [1992] ; Ng, Engle and Rothschild [1992] ; King, Sentana and Wadhwani [1991] ; Kroner [1987] ; Sentana [1998], Sentana and Shah [1994], Demos and Sentana [1998], Aguilar and West [2000] and Rosti and Gales [2002]. It is a combination of the GQARCH(1,1) factor models and HMM models. This new specification called conditionally heteroskedastic factor analysed hidden Markov model (CHFAHMM) is defined by :

$$\mathbf{y}_t = \theta_{S_t} + \mathbf{X}_{S_t} \mathbf{f}_{S_t} + \epsilon_{S_t} \quad \text{where} \quad \mathbf{f}_{S_t} = \mathbf{H}_{S_t}^{1/2} \mathbf{f}_t^*, \quad \text{and} \tag{13}$$

$$\begin{pmatrix} \mathbf{f}_t^* \\ \epsilon_{S_t} \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}_{S_t} \end{pmatrix} \right]$$

where $S_t \sim P(S_t = j / S_{t-1} = i)$ is a hidden Markov chain indicating the state or the regime at the date t, $\mathbf{y}_t$ is a $q \times 1$ random vector of observable variables (financial returns). In an unspecified state $S_t = j$ $(j = 1, ..., m)$, θ_j are the $q \times 1$ mean vectors, $\mathbf{f}_{jt}$ the $k \times 1$ vectors of unobserved common factors [2], ϵ_{jt} the $q \times 1$ vectors of idiosyncratic noises, $\mathbf{X}_j$ the $q \times k$ factor loadings matrices, with $q \geq k$ and $rank(\mathbf{X}_j) = k$, $\mathbf{\Psi}_j$ are $q \times q$ positive semi-definite matrices

[2] the number of factors k can change with the regime but in this work, we will consider the case where there is only one factor

of constant idiosyncratic variances, and $\mathbf{H}_{jt}$ the $k \times k$ diagonal and definite positive matrices (uncorrelated factors) whose elements are the variances of the common factors presumedly dynamic in variable time and their parameters according to the mode. In particular, we suppose that the variances of the common factors follow switching GQARCH(1,1) processes, the l *th* element of this matrix $\mathbf{h}_t^{(j)}$ under an unspecified regime $S_t = j$ since $S_{t-1} = i$ is $h_{lt}^{(j)} = 1 + \gamma_j^l f_{t-1}^{(i)} + \alpha_j^l f_{t-1}^{(i)2} + \delta_j^l h_{t-1}^{(i)}$.

The use of this model to analyse and forecast the financial returns will enable us to better characterize the stock price dynamics and to solve the various problems related to the changes of the internal and unobservable structure of financial data. They are problems of the type :

1. Can we distinguish distinct regimes in stock market returns ?
2. How do the regimes differ ?
3. How frequent are regime switches and when do they occur ?
4. Which the impact of a particular regime on volatility ?
5. Are regime switches predictable ?

3.1 The Likelihood Function

the log-likelihood function in the simplest case where the transition probabilities are constant, is given by :

$$\mathcal{L}(\Theta/\mathbf{y}) = c - \frac{1}{2}\sum_{t=1}^{n}\sum_{j=1}^{m} p(S_t = j/D_{t-1};\Theta)\left[\log|\mathbf{\Sigma}_{t/t-1}^{(j)}| + (\mathbf{y}_t - \theta_j)'\mathbf{\Sigma}_{t/t-1}^{(j)-1}(\mathbf{y}_t - \theta_j)\right] \quad (14)$$

where $\mathbf{\Sigma}_{t/t-1}^{(j)} = \mathbf{X}_j \mathbf{H}_{t/t-1}^{(j)} \mathbf{X}_j' + \mathbf{\Psi}_j$ and $h_{lt/t-1}^{j} = 1 + \gamma_j^l f_{lt-1/t-1} + \alpha_j^l (f_{lt-1/t-1}^2 + h_{lt-1/t-1}) + \delta_j^l h_{lt-1/t-2}$. D_{t-1} is the information set that we would have at time $t-1$ (all the past of $\mathbf{y}$: $\mathcal{Y}_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}....\}$; all the past of $\mathbf{f}$: $\mathcal{F}_{t-1} = \{\mathbf{f}_{t-1}, \mathbf{f}_{t-2}....\}$ as well as the complete sequence of the discrete state variable $\mathcal{S}$ until the date $t-1$).

The above likelihood function can be maximised with respect to the parameter vector $\Theta_j' = \left(vec(\theta_j)', vec(\mathbf{X}_j)', vech(\mathbf{\Psi}_j)', \gamma_j', \alpha_j', \delta_j'\right)$, for $j = 1, ..., m$ as well as the initial state and transition probabilities, π_j and p_{ij} (respectively), for $i, j = 1..., m$ by solving the associated first order conditions. But since they are particulary complicated in this case, a numerical approach is usually required. However, this leads to a number of practical problems. Using a first derivative ML approach, the Kalman filter and smoother must be used to produce estimates of the unobservable $\mathbf{f}_t$'s and their time varying variances in each regime once for every parameter per iteration. Using ML we must, also, calculate the prediction, filtering and smoothing probabilities of $S_t = j$ $(j = 1, ..., m)$ given a particular sequence of observations. This results in a very time consuming procedure wich is disproportionately more so as the number of series considered increases. Not surprisingly, empirical applications of this model have been limited to cases where q is relatively small.

3.2 A Switching State-Space Representation : The GPB1 method

The conditionally heteroskedastic factor model in (13) can be regarded as a random field with indices $i = 1, ..., q$, $t = 1, ..., n$ and $j = 1, ..., m$. Therefore, it is not surprising that it has a switching time-series state-space representation, with $\mathbf{f}_{jt}$ as the continuous state variables. The transition and measurement equations are given by :

$$\mathbf{y}_t = \theta_{S_t} + \mathbf{X}_{S_t}\mathbf{f}_t + \epsilon_{S_t}$$

$$S_t \sim P(S_t = j/S_{t-1} = i)$$

$$\mathbf{f}_t \sim N(\mathbf{0}, \mathbf{H}_{S_t}) \qquad S_t \neq S_{t-1}$$

$$\mathbf{f}_t = \mathbf{0}.\mathbf{f}_{t-1} + \mathbf{f}_{S_t} \qquad S_t = S_{t-1}$$

For the derivation of the filtering and smoothing equations we will use the generalized pseudo-bayesian method GPB1 based on the moment matching technique. These statistics, will thereafter be introduced into a conditional EM algorithm in order to estimate all the parameters of the model.

For the implementation of the filtering and smoothing algorithms, let us start by defining some notation.

$$\begin{aligned}
\mathbf{f}^{i(j)}_{t/\tau} &= \mathbb{E}\left[\mathbf{f}_t/\mathcal{Y}_{1:\tau}, S_{t-1} = i, S_t = j\right] \\
\mathbf{f}^{(j)k}_{t/\tau} &= \mathbb{E}\left[\mathbf{f}_t/\mathcal{Y}_{1:\tau}, S_t = j, S_{t+1} = k\right] \\
\mathbf{f}^{j}_{t/\tau} &= \mathbb{E}\left[\mathbf{f}_t/\mathcal{Y}_{1:\tau}, S_t = j\right]
\end{aligned}$$

If $\tau = t$, these are called filtered statistics ; if $\tau > t$, they are called smoothed statistics ; and if $\tau < t$, they are called predicted statistics. Notice how the superscript inside the brackets is the value of the switch node at time t (the subscript value) ; the superscript to the left is the value of S_{t-1}, and to the right, S_{t+1}. We need these subtle distinctions to handle the cross-variance terms correctly. We also define the following :

$$\begin{aligned}
h^{j}_{t/\tau} &= Cov\left(\mathbf{f}_t/\mathcal{Y}_{1:\tau}, S_t = j\right) \\
h^{j}_{t,t-1/\tau} &= Cov\left(\mathbf{f}_t, \mathbf{f}_{t-1}/\mathcal{Y}_{1:\tau}, S_t = j\right) \\
h^{i(j)}_{t,t-1/\tau} &= Cov\left(\mathbf{f}_t, \mathbf{f}_{t-1}/\mathcal{Y}_{1:\tau}, S_{t-1} = i, S_t = j\right) \\
M_{t-1,t/\tau}(i,j) &= p\left(S_{t-1} = i, S_t = j/\mathcal{Y}_{1:\tau}\right) \\
M_{t/\tau}(j) &= p(S_t = j/\mathcal{Y}_{1:\tau}) \\
L^{j}_{t} &= p(\mathbf{y}_t/\mathcal{Y}_{1:t-1}, S_t = j)
\end{aligned}$$

where L^j_t is the likelihood of the innovation at time t, given that the current model is j.

3.2.1 Filtering Algorithm

We perform the following steps in sequence.

$$\mathbf{f}_{t/t-1}^{i(j)} = \mathbf{0}.\mathbf{f}_{t-1/t-1}^{(i)} = \mathbf{0} \qquad \forall\, i, j = 1, ..., m \quad \text{et} \tag{15}$$

$$h_{t/t-1}^{i(j)} = 1 + \gamma_j f_{t-1/t-1}^{(i)} + \alpha_j \left[f_{t-1/t-1}^{(i)\,2} + h_{t-1/t-1}^{(i)} \right] + \delta_j h_{t-1/t-2}^{(i)} \tag{16}$$

Then we compute the error in our prediction (the innovation), the variance of the error, the Kalman gain matrix, and the likelihood of this observation.

$$\mathbf{e}_t = \mathbf{y}_t - \theta_j - \mathbf{X}_j \mathbf{f}_{t/t-1}^{i(j)}$$

$$\mathbf{\Sigma}_t = \mathbf{X}_j \mathbf{H}_{t/t-1}^{i(j)} \mathbf{X}_j' + \mathbf{\Psi}_j$$

$$K_t(i,j) = \mathbf{H}_{t/t-1}^{i(j)} \mathbf{X}_j' \mathbf{\Sigma}_t^{-1}$$

$$L_t(i,j) = \mathcal{N}\left(\mathbf{e}_t \,/\, \mathbf{0}, \mathbf{\Sigma}_t\right)$$

Finally, we update our estimates of the mean and the variance.

$$\mathbf{f}_{t/t}^{i(j)} = \mathbf{f}_{t/t-1}^{i(j)} + K_t(i,j)\mathbf{e}_t \tag{17}$$

$$\mathbf{H}_{t/t}^{i(j)} = (\mathbf{I}_k - K_t(i,j)\mathbf{X}_j)\mathbf{H}_{t/t-1}^{i(j)} = \mathbf{H}_{t/t-1}^{i(j)} - K_t(i,j)\mathbf{\Sigma}_t K_t(i,j)' \tag{18}$$

and we calculate the probabilities :

$$M_{t-1,t/t}(i,j) = p(S_{t-1} = i, S_t = j/\mathcal{Y}_{1:t}) = \frac{L_t(i,j) p_{ij} M_{t-1/t-1}(i)}{\sum_{i=1}^{m}\sum_{j=1}^{m} L_t(i,j) p_{i,j} M_{t-1/t-1}(i)}$$

$$M_{t/t}(j) = \sum_{i=1}^{m} M_{t-1,t/t}(i,j)$$

and

$$W_{i/j}(t) = p(S_{t-1} = i/S_t = j, \mathcal{Y}_{1:t}) = M_{t-1,t/t}(i,j)/M_{t/t}(j)$$

and the statistics

$$\mathbf{f}_{t/t}^{j} = \sum_{i=1}^{m} W_{i/j}(t) \mathbf{f}_{t/t}^{i(j)}$$

$$h_{t/t}^{j} = \sum_{i=1}^{m} W_{i/j}(t) h_{t/t}^{i(j)} + \sum_{i=1}^{m} W_{i/j}(t) \left[\mathbf{f}_{t/t}^{i(j)} - \mathbf{f}_{t/t}^{j}\right]\left[\mathbf{f}_{t/t}^{i(j)} - \mathbf{f}_{t/t}^{j}\right]'$$

$$h_{t/t-1}^{(j)} = \sum_{i=1}^{m} W_{i/j}(t) h_{t/t-1}^{i(j)} + \sum_{i=1}^{m} W_{i/j}(t) \left[\mathbf{f}_{t/t-1}^{i(j)} - \mathbf{f}_{t/t-1}^{(j)}\right]\left[\mathbf{f}_{t/t-1}^{i(j)} - \mathbf{f}_{t/t-1}^{(j)}\right]'$$

3.2.2 Smoothing Algorithm

Given the degenerate nature of the (time-series) transition equation, the smoother gain matrix is always null, and smoothing is unnecessary in this case,

$$\begin{aligned} \mathbf{f}_{t/n}^{(j)k} &= \mathbf{f}_{t/t}^{j} + J_t^{(j)k}\left[\mathbf{f}_{t+1/n}^{k} - \mathbf{f}_{t+1/t}^{j(k)}\right] = \mathbf{f}_{t/t}^{j} \\ \mathbf{H}_{t/n}^{(j)k} &= \mathbf{H}_{t/t}^{j} + J_t^{(j)k}\left[\mathbf{H}_{t+1/n}^{k} - \mathbf{H}_{t+1/t}^{j(k)}\right] J_t^{(j)k'} = \mathbf{H}_{t/t}^{j} \end{aligned}$$

Thereafter we calculate the probabilities,

$$U_t^{j/k} = p(S_t = j/S_{t+1} = k, \mathcal{Y}_{1:n}) \simeq \frac{M_{t/t}(j)p_{jk}}{\sum_{j'=1}^{m} M_{t/t}(j')p_{j'k}}$$

where the approximation arised because S_t is not conditionally independente of the future evidence $\mathbf{y}_{t+1}, ..., \mathbf{y}_n$, given S_{t+1}. This approximation will not be too bad provided future evidence does not contain much information about S_t beyond than that contained in S_{t+1}.

Pour la mise à jour des paramètres, nous avons besoin aussi des probabilités

$$\begin{aligned} M_{t,t+1/n}(j,k) &= U_t^{j/k} M_{t+1/n}(k) \\ M_{t/n}(j) &= \sum_{k=1}^{m} M_{t,t+1/n}(j,k) \end{aligned}$$

4 The EM Algorithm

The application of a conditional version of the EM algorithm (Dempster, Laird and Rubin [1977]) to estimate conditionally heteroskedastic factor models without regime switching requires the division of the entire set parameters in two groups (Demos and Sentana [1998]). The parameters of the first group will be estimated by maximizing the conditional expectation of the complete log-likelihood function ; and parameters of the second group, i.e. the parameters of the conditionally heteroscedastic component ($\Phi = \{\gamma, \alpha, \delta\}$), will be updated by maximizing the observed log-likelihood function through the BHHH algorithm (Berndt-Hall-Hall-Hausman). This algorithm could still be applied to our switching model, but in this case it is necessary to introduce a Viterbi approximation step before the maximization of the observed log-likelihood to detect the switching regime date.

The joint likelihood of an utterance $\mathcal{Y}$, continuous state vector sequence $\mathcal{F}$ (common factors), and HMM state sequence $\mathcal{S}$ is given by :

$$p(\mathcal{Y}, \mathcal{F}, \mathcal{S}) = p(S_1) \prod_{t=2}^{n} p(S_t/S_{t-1}) \prod_{t=1}^{n} p(\mathbf{f}_t/S_t, D_{t-1}) p(\mathbf{y}_t/\mathbf{f}_t, S_t, D_{t-1}) \tag{19}$$

where $D_{t-1} = \{\mathcal{Y}_{1:t-1}, \mathcal{F}_{1:t-1}\}$, is the information set that we would have at time $t-1$, $p(S_1) = \pi_{s_1}$ the initial state probability and $p(S_t/S_{t-1}) = p_{S_{t-1}S_t}$ the transition probabilities.

E Step :

If one poses $D_{ni} = \left\{\mathcal{Y}_{1:n}, \Theta^{(i)}\right\}$, the conditional expectation of the complete log-likelihood will be given by :

$$\begin{aligned} Q(\Theta/\Theta^{(i)}) &\simeq \sum_{j=1}^{m} M_{1/n}(j)\log(p(S_1)) - \sum_{t=2}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} M_{t-1,t/n}(i,j)\log(p_{ij}) \\ &- \frac{1}{2}\sum_{j=1}^{m}\sum_{t=1}^{n} M_{t/n}(j)\left[\log|\mathbf{\Psi}_j| + \mathbb{E}\left\{(\mathbf{y}_t - \theta_j - \mathbf{X}_j\mathbf{f}_{jt})'\mathbf{\Psi}_j^{-1}(\mathbf{y}_t - \theta_j - \mathbf{X}_j\mathbf{f}_{jt})/D_{ni}\right\}\right] \\ &- \frac{1}{2}\sum_{j=1}^{m}\sum_{l=1}^{k}\sum_{t=1}^{n} M_{t/n}(j)\mathbb{E}\left[\log(h_{lt}^j) + f_{lt}^2/h_{lt}^j/D_{ni}\right] \end{aligned} \tag{20}$$

CM1 Step :

As in the case of CHFAHMM models, the initial state probabilities and the transition probabilities will be obtained by replacing in the formulas of $\widehat{\pi}_j$ and $\widehat{p}_{ij}$ the posterior probabilities $\gamma_j(t)$ and $\xi_{ij}(t)$ by the estimates, $M_{t/n}(j)$ and $M_{t-1,t/n}(i,j)$, provided by the extended Kalman filter. Equivalently, the other parameters i.e. the factor loadings matrices $\mathbf{X}_j$, the mean vectors θ_j and the idiosyncratic variances $\mathbf{\Psi}_j$ will be obtained by quite simply replacing the values of $\widehat{\mathbf{f}}_j(t)$, $\mathbf{H}_j$ and $\widehat{\mathbf{R}}_j(t)$ in the case of FAHMM by the outputs of the Kalman filter algorithm $\mathbf{f}_{t/n}^j$, $\mathbf{H}_{t/n}^j$ and $\left[\mathbf{H}_{t/n}^j + \mathbf{f}_{t/n}^j\mathbf{f}_{t/n}^{j'}\right]$ (respectively).

CM2 Step : Conditional Variance Parameters Estimation

In the previous step we maximized the conditional expectation of the complete log-likelihood function (equation (20)) with respect to the parameters, π_j, p_{ij}, θ_j, $\mathbf{X}_j$ and $\mathbf{\Psi}_j$ by an EM algorithm based on the generalized pseudo-bayesian approximation GPB1. For the implementation of this algorithm, we use each time the parameters of the conditional variance which one already found in the preceding iteration. Now, and being given the new values given by the CM1 step, we can update the values of γ_j, α_j and δ_j by maximizing the observed log-likelihood function given by the equation (14) with respect to the parameters of the conditionally heteroskedastic component, and we repeat these iterations until convergence. However, for the implementation of the optimization algorithm it is necessary to identify the switching regime dates. The identification of these points can be carried out, either by using the smoothing probabilities $M_{t/n}(j)$ given by the Kalman smoothing algorithm, or also by an approximated version of the Viterbi algorithm. Once these points will be known, on each segment of data we maximize the observed log-likelihood $\mathcal{L}^{(i)}$ by using a BHHH algorithm :

$$\Phi_j^{(i+1)} = \Phi_j^{(i)} + \left[H_{(i)}\left(\Phi_j^{(i)}\right)\right]^{-1} g_{(i)}\left(\Phi_j^{(i)}\right) \tag{21}$$

where $\Phi_j^{(i)}$ is the vector of conditional variance parameters of the i-th iteration ; $H_{(i)}\left(\Phi_j^{(i)}\right)$ the Hessian matrix (the matrix of the second derivatives of $\mathcal{L}^{(i)}$ with respect to parameters evaluated at $\Phi_j^{(i)}$) ; and $g_{(i)}\left(\Phi_j^{(i)}\right)$ the negatif gradient of $\mathcal{L}^{(i)}$ evaluated at $\Phi_j^{(i)}$.

4.1 Latent Struture Inference Using Viterbi Approximation

The task of Viterbi approximation approach is to find the best sequence of switching states S_t and common factors $\mathbf{f}_t$ that minimizes the Hamiltonian cost in equation (22) for a given observation sequence $\mathcal{Y}_{1:n}$. It is well known how to apply Viterbi inference to discrete state hidden Markov models (L. R. Rabiner and B. Juang [1993]) and continuous state Gauss-Markov models (R. E. Kalman and R. S. Bucy [1961]). Here we develop an algorithm for Viterbi inference in CHFAHMMs.

$$\begin{aligned}
H(\mathcal{F},\mathcal{S},\mathcal{Y}) &= \frac{1}{2}\sum_{t=1}^{n}\sum_{j=1}^{m}\left[(\mathbf{y}_t-\mathbf{X}_j\mathbf{f}_{jt}-\theta_j)'\mathbf{\Psi}_j^{-1}(\mathbf{y}_t-\mathbf{X}_j\mathbf{f}_{jt}-\theta_j)+\log|\mathbf{\Psi}_j|\right]S_t(j)\\
&+ \frac{1}{2}\sum_{t=2}^{n}\sum_{j=1}^{m}\left[\mathbf{f}_{jt}'\mathbf{H}_{jt}^{-1}\mathbf{f}_{jt}+\log|\mathbf{H}_{jt}|\right]S_t(j)+\frac{n}{2}(m+q)\log 2\pi\\
&+ \frac{1}{2}\sum_{j=1}^{m}\left[(\mathbf{f}_{j1})'\mathbf{H}_{j1}^{-1}(\mathbf{f}_{j1})+\log|\mathbf{H}_{j1}|\right]S_1(j)\\
&+ \sum_{t=2}^{n}S_t'(-\log\mathbf{P})S_{t-1}+S_1'(-\log\pi_1)
\end{aligned} \tag{22}$$

If the best sequence of switching states is denoted $\mathcal{S}_n^*$ we can then approximate the desired posterior $p(\mathcal{F},\mathcal{S}/\mathcal{Y})$ as [3] :

$$p(\mathcal{F},\mathcal{S}/\mathcal{Y}) = p(\mathcal{F}/\mathcal{S},\mathcal{Y})p(\mathcal{S}/\mathcal{Y}) \simeq p(\mathcal{F}/\mathcal{S},\mathcal{Y})\mu(\mathcal{S}_n-\mathcal{S}_n^*)$$

i.e. the switching sequence posterior $p(\mathcal{S}/\mathcal{Y}_{1:n})$ was approximated by its mode. More formally, we are looking for the switching sequence $\mathcal{S}_n^*$ such that

$$\mathcal{S}_n^* = \arg\max_{\mathcal{S}_{1:n}} p(\mathcal{S}_{1:n}/\mathcal{Y}_{1:n})$$

It is easily to shown that a (suboptimal) solution to this problem can be obtain by recursive optimization of the probability of the best sequence at time t.

$$\begin{aligned}
J_{t,j} &= \max_{\mathcal{S}_{t-1}} p(\mathcal{S}_{t-1},S_t=j,\mathcal{Y}_t)\\
&\simeq \max_i \Big\{p(\mathbf{y}_t/S_t=j,S_{t-1}=i,\mathcal{S}_{t-2}^*(i),\mathcal{Y}_{t-1})p(S_t=j/S_{t-1}=i)\\
&\quad \max_{\mathcal{S}_{t-2}} p(\mathcal{S}_{t-2},S_{t-1}=i,\mathcal{Y}_{t-1})\Big\}
\end{aligned} \tag{23}$$

[3] $\mu(x)=1$ for $x=\emptyset$ et zero otherwise.

Here $\mathcal{S}_{t-2}^{*}(i)$ is the "best" switching sequence up to time $t-1$ when CHFAHMM is in state i at time $t-1$, $\mathcal{S}_{t-2}^{*}(i) = \arg\max_{\mathcal{S}_{t-2}} J_{t-1,i}$.

Define first the "best" partial cost up to time t of the measurement sequence $\mathcal{Y}_t$ when the switch is in state j at time t :

$$J_{t,j} = \min_{\mathcal{S}_{t-1},\mathcal{F}_t} H\left[\{\mathcal{S}_{t-1}, S_t = j\}, \mathcal{F}_t, \mathcal{Y}_t\right] \tag{24}$$

Namely, this cost is the least cost over all possible sequences of switching states $\mathcal{S}_{t-1}$ and corresponding factor model states $\mathcal{F}_t$. This partial cost is essential in Viterbi like total cost minimisation. For a given switch state transition $i \rightarrow j$ it is now easy to establish relationship between the filtered and the predicted estimates (equations (15) and (16)). From the theory of Kalman estimation (B.O. Anderson et J.B. Moore 1979) and given a new observation $\mathbf{y}_t$ at time t each of these predicted estimates can now be filtered using Kalman measurement update framework (equation (17) and (18)). hence, each of these $i \rightarrow j$ transitions has a certain innovation cost $J_{t,t-1,i,j}$ associated with it, as defined in Equation (25).

$$\begin{aligned} J_{t,t-1,i,j} &= \frac{1}{2}\left(\mathbf{y}_t - \mathbf{X}_j\mathbf{f}_{t/t-1}^{i(j)} - \theta_j\right)'\left[\mathbf{X}_j\mathbf{H}_{t/t-1}^{i(j)}\mathbf{X}_j' + \mathbf{\Psi}_j\right]^{-1}\left(\mathbf{y}_t - \mathbf{X}_j\mathbf{f}_{t/t-1}^{i(j)} - \theta_j\right) \\ &+ \frac{1}{2}\log\left|\mathbf{X}_j\mathbf{H}_{t/t-1}^{i(j)}\mathbf{X}_j' + \mathbf{\Psi}_j\right| - \log p_{ij} \end{aligned} \tag{25}$$

One portion of the innovation cost reflects the continuous state transition (the common factors), as indicated by the innovation terms in equation (25). The remaining cost is due to switching from state i to state j, $-\log p_{ij}$.

Obviously, for every current switching state j there are m possible previous switching states where the system could have originated from. To minimize the overall cost at every time step t and for every switching state j one "best" previous state i is selected :

$$J_{t,j} = \min_i\{J_{t,t-1,i,j} + J_{t-1,i}\} \tag{26}$$

$$\delta_{t-1,j} = \arg\min_i\{J_{t,t-1,i,j} + J_{t-1,i}\} \tag{27}$$

The index of this state is kept in the state transition record $\delta_{t-1,j}$. Consequently, we now obtain a set of m best filtered continuous states and their variances at time t :

$$\mathbf{f}_{t/t}^{j} = \mathbf{f}_{t/t}^{\delta_{t-1,j}\,(j)} \quad \text{et} \quad \mathbf{H}_{t/t}^{j} = \mathbf{H}_{t/t}^{\delta_{t-1,j}\,(j)} \tag{28}$$

Once all n observations $\mathcal{Y}_{1:n}$ have been fused the best overall cost is obtained as :

$$J_n^* = \min_j J_{n,j} \tag{29}$$

To decode the "best" switching state sequence one uses the index of the best final state, $j_n^* = \arg\min_j J_{n,j}$ then traces back through the state transition record $\delta_{t-1,j}$,

$$j_t^* = \delta_{t,j_{t+1}^*} \tag{30}$$

Switching model's sufficient statistics are now simply $\mathrm{E}(S_t/.) = S_t(j^*)$ and $\mathrm{E}(S_t S_{t-1}'/.) = S_t(j^*)S_t(j^*)'$.[4] Given the "best" switching state sequence the sufficient factor model statistics can be easily obtained using the Rauch-Tung-Streiber smoothing [1963]. For example,

$$\mathrm{E}(\mathbf{f}_t, S_t(j)/.) = \begin{cases} \mathbf{f}_{t/n}^{j_t^*} & j = j_t^* \\ \mathbf{0} & \text{otherwise} \end{cases}$$

for $j = 1, ..., m$.

4.2 Initialization

In practice, the key point for initialization is to start with a good segmentation of the data set, where by segmentation we mean a partition of the data, with each part modelled by a particular dynamic factor model. In the case of our model, the best strategy of initialization consists in implementing a standard EM algorithm for the data by supposing that the factors are homoskedastic[5]. Thereafter and given the output of the EM algorithm, we can implement a Viterbi algorithm or we use directly the posterior probabilities $p(S_t = j/\mathcal{Y})$ in order to obtain the optimal sequence of states. In a second step, a particular simple conditionally heteroskedastic factor model will be initialized for each segment. With this intention, one uses the empirical covariance matrices like estimates of the idiosyncratic variance matrices $\mathbf{\Psi}_j$ and the empirical means like estimates of the means θ_j. The parameters of the conditionally heteroskedastic variances will be initialized by applying a GQARCH(1,1) model to each segment of data. Finally, the elements of the transition matrix $\mathbf{P}$, can also be initialized by counting the number of transitions from state i to state j, $(i, j = 1..., m)$, and dividing it by the number of transitions from state i to any other state.

5 Model Selection

In the context of our work, a particular type of idiosyncratic covariance matrices (a diagonal form) $\mathbf{\Psi}_j$, $j = 1, ..., m$, was chosen to reduce the number of free parameters and thus guarantee the identification of the model. We considered, also, the existence of only one common factor. More important is to determine a reliable number m of hidden states. A classical way to deal with this dimension selection problem is to minimize a penalized log-likelihood criteria, of the form "Deviance+Penality", where

$$\text{Deviation} = -2\,\mathcal{L}(\widehat{\Theta}/\mathcal{Y})$$

[4] The operator $\mathrm{E}(/.)$ denotes conditional expectation with respect to the posterior distribution, e.g. $\mathrm{E}(\mathbf{f}_t/.) = \sum_{\mathcal{S}} \int_{\mathcal{F}} \mathbf{f}_t p(\mathcal{F}, \mathcal{S}/\mathcal{Y})$.

[5] In practice, 20 iterations of the EM algorithm are largely sufficient.

$\mathcal{L}(\widehat{\Theta}/\mathcal{Y})$ being the observed log-likelihood evaluated at the parameter value maximizing it. The deviance is a measure of the fit of the model to the learning sample $\mathcal{Y}$. The Penalty term is an increasing function of the number of free parameters in the model (see for instance Burnham and Anderson 1998). Thus the art of model selection lies on choosing a good penalty function.

In the existing literature, traditional model selection criteria based on likelihood include variants of AIC Akaike [1974], the Schwartz or Bayesian criteria, or BIC, and related information criteria such as the ICOMP methods of Bozdogan and Ramirez [1987] and Bozdogan and Shigemasu [1998]. These various criteria respect the fundamental principles of the choice of a model : good adjustment, parsimony and objectivity.

Many adaptations of the AIC are available. Schwarz [1978] suggested the BIC which increases information on the number of parameters with the number of observations. This criterion is as follows for a model $\mathcal{M}$:

$$\text{BIC} = -2\,\mathcal{L}(\widehat{\Theta}/\mathcal{Y}) + v_{\mathcal{M}} \log n \tag{31}$$

$v_{\mathcal{M}}$ denoting the number of free parameters of the model $\mathcal{M}$. It has been proved to be reliable on an empirical ground for mixture models (see for instance Roeder and Wasserman, 1997) and, on a theoretical ground, it has been proved to provide a consistent estimate of the dimension of a hidden Markov model under reasonable assumptions (Gassiat, 2002). The AIC criterion makes use of the less stringent penalty term $2\,v_{\mathcal{M}}$ and, in many circumstances, is expected to select too complex models (see Burnham and Anderson 1998).

These various criteria implicitly assume that the sampling distribution is belonging to, at least, one of the models in competition. This assumption is most often unrealistic and can lead to underpenalize complex models. Taking into account the modelling purpose can counter efficiently this tendancy. This point of view is much sensible for hidden structure models. In this setting, discovering the hidden structure is often of primary interest for the user to derive a reliable clustering of its data set. Thus, we propose a model selection criterion favoring minimal missing information models. This criterion, the so called ICL criterion previously proposed in (Biernacki et al. 2000) for the mixture model, can be generalized to any hidden structure model. In this case, it seems preferable to us to base the selection of a model on the maximization of the integrated complete log-likelihood function defined by :

$$p(y, z|\mathcal{M}) = \int p(y, z|\mathcal{M}, \Theta_{\mathcal{M}})\pi(\Theta_{\mathcal{M}})d\,\Theta_{\mathcal{M}} \tag{32}$$

where z indicates the hidden variables : the latent common factors and the discrete hidden state variables. An equivalent approximation for the integrated complete log-likelihood function is given by

$$\text{ICL}(\mathcal{M}) = \text{BIC}(\mathcal{M}) + \log\, p(z|y, \widehat{\Theta}_{\mathcal{M}}) \tag{33}$$

$\log\, p(z|y, \widehat{\Theta}_{\mathcal{M}})$ being a measurement of the missing information carried by the model $\mathcal{M}$. This writing highlights that the ICL criterion overpenalize the models at significant missing information compared to BIC.

Such as it is defined, the ICL criterion is not calculable since the states z are not observed. A natural approximation for $\log\, p(z|y, \widehat{\Theta}_{\mathcal{M}})$ is given by :

$$\log\, p(z|y,\widehat{\Theta}_{\mathcal{M}}) = \max_z \left[\log\, p(z|y,\widehat{\Theta}_M)\right] \tag{34}$$

In the case of CHFAHMM models, (34) can be resolved by the Viterbi algorithm.

6 Monte Carlo Simulations

We conducted a series of experiments to explore some importante properties of the algorithms and the resulting CHFAHMM estimator. We report on a three experiments that were designed to adress the following issues :

(A) The most important question we ask about the CHFAHMM estimator is whether it is a consistent estimator for Θ. Once we are assured of consistency, the natural question to ask is what are reasonable sequence sizes required to obtain accurate and stable estimates ?

(B) The most classical question is to ask if the estimates are asymptotically normally distributed, and what sequence size is needed for such an approximation.

(C) An important question when using a multi-class model, is to choose a reliable model, containing enough parameters to ensure a realistic fit to the learning data set, but not too much to avoid overfitting and poor performances for future use of the model. To answer this question, we will use four selection criteria : AIC, BIC, ICOMP and the ICL criterion favoring minimal missing information model.

6.1 Simulation A : Accuracy and Stability of the CHFAHMM Estimates

6.1.1 The Kullback-Leibler Divergence

simulations which we will present now are based on models with 6 series of observations, two hidden Markovian states and only one GQARCH(1,1) common factor. Our objective is to study the behaviour of the estimates of the CHFAHMM when the size of the sequence n increases by 1500 to 3000. With this intention, we generated sequences of observations of sizes n = 1500, 2000 and 3000 (with a hundred replications for each simulation). The parameters of this simulation, as well as the initialization values of the EM algorithm ($\Theta^{(0)}$), are given in table 1. The iterations of this algorithm stop when the relative change in each component of the values of the estimated parameters, are all smaller than a threshold value selected for example equal to 0.0001 (the initial probabilities are excluded). For the inference of the latent structures and the parameters estimation we used the generalized pseudo-bayesian approximation GPB1.

A natural metric to mesure the distance of estimators from the true model parameters is the Kullback-Leibler divergence,

$$K(\Theta_0,\Theta) \stackrel{def}{=} \lim_n \frac{1}{n}\left\{\log\, \mathcal{L}(\mathbf{y}_1,....,\mathbf{y}_n;\Theta_0) - \log\, \mathcal{L}(\mathbf{y}_1,....,\mathbf{y}_n;\Theta)\right\}$$

where Θ_0 is the true parameter. For a finite sequence of length n, we define the sample Kullback-Leibler divergence between two parameter points as :

Parameter values for true model

État 1				État 2			
θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ	θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ
1.0	2.0	1.0		2.0	3.0	2.0	
2.0	2.0	1.0	0.1	3.0	3.0	2.0	0.2
1.0	2.0	1.0	0.3	2.0	3.0	2.0	0.2
2.0	2.0	1.0	0.4	3.0	3.0	2.0	0.6
1.0	2.0	1.0		2.0	3.0	2.0	
2.0	2.0	1.0		3.0	3.0	2.0	

Initial values for the EM algorithm

État 1				État 2			
θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ	θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ
0.0	1.0	0.5		1.0	1.0	0.5	
1.0	1.0	0.5	0.1	1.0	0.5	0.5	0.1
0.5	0.5	0.5	0.1	1.0	0.5	0.5	0.1
1.0	3.5	0.5	0.1	1.0	1.0	0.5	0.1
0.0	4.0	0.5		1.0	0.5	0.5	
0.5	1.0	0.5		1.0	0.5	0.5	

TAB. 1 – Simulation parameters.

$$K_n(\Theta_0,\Theta) \overset{def}{=} \frac{1}{n}\left\{\log \mathcal{L}(\mathbf{y}_1,....,\mathbf{y}_n;\Theta_0) - \log \mathcal{L}(\mathbf{y}_1,....,\mathbf{y}_n;\Theta)\right\}$$

We shall use the stochastic distance function $\widetilde{K}_n(\Theta_0,\widetilde{\Theta})$ to measure the distance between CHFAHMM estimate $\widetilde{\Theta}$ and Θ_0. This distance measure was effectively used in earlier studies (see Juang and Rabiner [1985]).

For each value of n, the estimation procedure was carried out a hundred times, and the distances $\widetilde{K}_n(\Theta_0,\widetilde{\Theta}_n)$ between each of the hundred estimators and the true parameter Θ_0 were evaluated on a new sequence, independent of the first hundred sequences used to obtain the estimators. This way we prevent the potential underestimation of the distance as a result of estimating the parameter and evaluating its performance on the same sequence.

Box plots of the sets of distances for the various values of n are presented under a unified scale in Figure 1. The plots clearly show a general decrease in average and spread of the distances with increasing n. Given that small values of $\widetilde{K}_n$ imply similarity between Θ_0 and $\widetilde{\Theta}_n$, the results of this experiment suggest increasing accuracy and stability of the sequence of CHFAHMM estimators as n increases.

Table 2 show the mean, median and standard deviation of the parameter estimates for simulations with different values of n.

6.1.2 Comparison of the two Approaches

In this simulation we applied the 2 training methods which we have developed. The first approach is based on the generalized pseudo-bayesian method GPB1. The second on the Viterbi

Simulation with $n = 1500$

. mean, (.) standard deviation, ⟨.⟩ median

State 1				State 2			
θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ	θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ
0.8820	2.0277	0.9978		1.8600	3.0602	2.0150	
(0.11)	(0.05)	(0.06)		(0.16)	(0.04)	(0.13)	
⟨0.87⟩	⟨2.02⟩	⟨0.99⟩		⟨1.88⟩	⟨3.06⟩	⟨2.02⟩	
1.8938	2.0264	0.9975		2.8530	3.0619	1.9775	
(0.11)	(0.04)	(0.06)		(0.17)	(0.04)	(0.11)	
⟨1.88⟩	⟨2.02⟩	⟨0.99⟩	0.0664	⟨2.85⟩	⟨3.07⟩	⟨1.96⟩	0.1914
0.8817	2.0255	0.9994	(0.09)	1.8570	3.0617	1.9956	(0.11)
(0.12)	(0.04)	(0.06)	⟨0.07⟩	(0.16)	(0.04)	(0.10)	⟨0.18⟩
⟨0.87⟩	⟨2.02⟩	⟨0.99⟩	0.2994	⟨1.87⟩	⟨3.06⟩	⟨2.00⟩	0.2063
1.8866	2.0257	1.0014	(0.06)	2.8587	3.0589	1.9878	(0.04)
(0.11)	(0.04)	(0.07)	⟨0.29⟩	(0.17)	(0.03)	(0.212)	⟨0.20⟩
⟨1.88⟩	⟨2.02⟩	⟨0.99⟩	0.3936	⟨2.85⟩	⟨3.06⟩	⟨2.00⟩	0.5817
0.8865	2.0313	1.0076	(0.05)	1.8537	3.0608	1.9818	(0.04)
(0.10)	(0.04)	(0.06)	⟨0.39⟩	(0.17)	(0.04)	(0.14)	⟨0.58⟩
⟨0.88⟩	⟨2.03⟩	⟨1.01⟩		⟨1.85⟩	⟨3.06⟩	⟨1.97⟩	
1.8894	2.0269	1.0026		2.8517	3.0611	1.9893	
(0.11)	(0.05)	(0.07)		(0.16)	(0.04)	(0.12)	
⟨1.88⟩	⟨2.03⟩	⟨1.00⟩		⟨2.86⟩	⟨3.06⟩	⟨1.99⟩	

Simulation with $n = 2000$

0.8832	2.0254	0.9993		1.8581	3.0569	1.9980	
(0.09)	(0.04)	(0.05)		(0.15)	(0.03)	(0.10)	
⟨0.87⟩	⟨2.02⟩	⟨0.99⟩		⟨1.86⟩	⟨3.05⟩	⟨1.99⟩	
1.8905	2.0249	1.0019		2.8617	3.0532	2.0040	
(0.09)	(0.04)	(0.05)		(0.15)	(0.03)	(0.09)	
⟨1.88⟩	⟨2.02⟩	⟨1.00⟩	0.0695	⟨2.86⟩	⟨3.05⟩	⟨1.99⟩	0.1884
0.8828	2.0227	0.9882	(0.09)	1.8539	3.0534	1.9788	(0.08)
(0.08)	(0.03)	(0.06)	⟨0.07⟩	(0.15)	(0.03)	(0.10)	⟨0.20⟩
⟨0.87⟩	⟨2.02⟩	⟨0.99⟩	0.3039	⟨1.86⟩	⟨3.05⟩	⟨1.98⟩	0.2013
1.8900	2.0216	0.9960	(0.05)	2.8584	3.0527	1.9776	(0.04)
(0.308)	(0.04)	(0.06)	⟨0.31⟩	(0.15)	(0.03)	(0.11)	⟨0.20⟩
⟨1.88⟩	⟨2.01⟩	⟨0.99⟩	0.3872	⟨2.86⟩	⟨3.05⟩	⟨1.98⟩	0.5890
0.8839	2.0219	0.9983	(0.04)	1.8573	3.0590	2.0039	(0.03)
(0.09)	(0.03)	(0.06)	⟨0.38⟩	(0.15)	(0.03)	(0.11)	⟨0.59⟩
⟨0.87⟩	⟨2.01⟩	⟨0.99⟩		⟨1.86⟩	⟨3.06⟩	⟨2.00⟩	
1.8851	2.0227	1.0007		2.8547	3.0540	1.9986	
(0.08)	(0.04)	(0.05)		(0.14)	(0.03)	(0.11)	
⟨1.87⟩	⟨2.02⟩	⟨1.00⟩		⟨2.85⟩	⟨3.05⟩	⟨1.99⟩	

Simulation with $n = 3000$

0.8969	2.0232	0.9955		1.8315	3.0579	1.9949	
(0.08)	(0.03)	(0.05)		(0.14)	(0.02)	(0.08)	
⟨0.90⟩	⟨2.02⟩	⟨0.99⟩		⟨1.84⟩	⟨3.06⟩	⟨2.00⟩	
1.8977	2.0223	0.9988		2.8359	3.0568	1.9878	
(0.08)	(0.03)	(0.04)		(0.13)	(0.03)	(0.08)	
⟨1.90⟩	⟨2.02⟩	⟨0.99⟩	0.1030	⟨2.85⟩	⟨3.06⟩	⟨2.00⟩	0.1826
0.8966	2.0219	0.9948	(0.06)	1.8309	3.0554	1.9905	(0.08)
(0.08)	(0.03)	(0.05)	⟨0.09⟩	(0.13)	(0.03)	(0.10)	⟨0.18⟩
⟨0.91⟩	⟨2.02⟩	⟨0.99⟩	0.2986	⟨1.84⟩	⟨3.06⟩	⟨1.99⟩	0.1952
1.8999	2.0221	0.9968	(0.05)	2.8349	3.0592	1.9847	(0.03)
(0.09)	(0.03)	(0.04)	⟨0.30⟩	(0.12)	(0.03)	(0.08)	⟨0.20⟩
⟨1.91⟩	⟨2.02⟩	⟨1.00⟩	0.3917	⟨2.84⟩	⟨3.06⟩	⟨1.98⟩	0.5921
0.8981	2.0214	1.0020	(0.04)	1.8296	3.0581	1.9872	(0.02)
(0.09)	(0.03)	(0.04)	⟨0.39⟩	(0.12)	(0.02)	(0.09)	⟨0.59⟩
⟨0.90⟩	⟨2.02⟩	⟨1.00⟩		⟨1.85⟩	⟨3.06⟩	⟨1.98⟩	
1.8962	2.0219	0.9975		2.8316	3.0596	2.0022	
(0.08)	(0.03)	(0.04)		(0.13)	(0.03)	(0.10)	
⟨1.90⟩	⟨2.03⟩	⟨1.00⟩		⟨2.84⟩	⟨3.06⟩	⟨2.01⟩	

TAB. 2 – Results of the GPB1 method

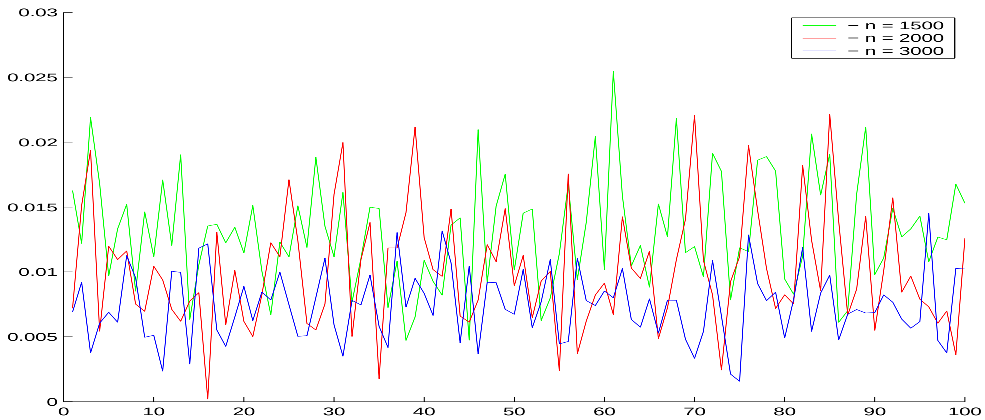

FIG. 1 – Box plots of $\widetilde{K}(\Theta_0, \widetilde{\Theta}_n)$.

approximation by regarding the factor model as a dynamic bayesian network with mixture of states. In this case, we have used the same sample of data simulated from a CHFAHMM model with two hidden markovian states, $n = 2000$, $q = 6$, $k = 1$ and a GQARCH(1,1) structure characterizing the two regimes (table 1). The results are given in the table 3 which shows a clear improvement in the conditional variance parameters estimation carried by the method based on the Viterbi approximation.

We have applied, also, the algorithms of identification of the optimal sequence for different numbers of observations and a hundred of replications [6]. All the results are given in table 4, where $t + 1$ is the true date of regime switching. This table shows us that only the Viterbi algorithm and the smoothing algorithm are able to detect the points of change. The other algorithms give in the majority of the cases results shifted of one or two periods. Such a result is related to the quite particular structure of the transition matrix (the existence of non emitting states). We notice, also, that the number of observations does not have any effect on the performance of the algorithms and that the filtering and the prediction algorithms for a horizon of 1 and 2 periods give the same results exactly, and for this reason we gave the results of the filtering algorithm only in table 4.

6.2 Simulation B : Asymptotic Normality of the CHFAHMM Estimates

This experiment investigates the asymptotic distribution of $\widetilde{\Theta}_n$, the CHFAHMM estimate. We generated 50 sequences of size n following the same procedure as described in simulation A.

[6] We have used the Viterbi algorithm, the Smoothing algorithm, the Filtering algorithm and the Prediction algorithm for one and two periods.

	State 1				State 2		
The Generalized Pseudo-Bayesian Method							
θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ	θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ
0.8832	2.0254	0.9993	0.0695	1.8581	3.0569	1.9980	0.1884
(0.09)	(0.04)	(0.05)	(0.09)	(0.15)	(0.03)	(0.10)	(0.08)
1.8905	2.0249	1.0019	0.3039	2.8617	3.0532	2.0040	0.2013
(0.09)	(0.04)	(0.05)	(0.05)	(0.15)	(0.03)	(0.10)	(0.04)
0.8828	2.0227	0.9882	0.3872	1.8539	3.0534	1.9788	0.5890
(0.08)	(0.04)	(0.06)	(0.04)	(0.15)	(0.03)	(0.10)	(0.03)
1.8900	2.0216	0.9960		2.8584	3.0527	1.9776	
(0.08)	(0.04)	(0.06)		(0.15)	(0.03)	(0.10)	
0.8839	2.0219	0.9983		1.8573	3.0590	2.0039	
(0.09)	(0.03)	(0.06)		(0.15)	(0.03)	(0.10)	
1.8851	2.0227	1.0007		2.8547	3.0540	1.9986	
(0.09)	(0.04)	(0.05)		(0.15)	(0.03)	(0.11)	
The Viterbi Approximation							
0.9839	1.9980	0.9989	0.1017	1.9726	3.0134	1.9988	0.2046
(0.10)	(0.07)	(0.05)	(0.09)	(0.19)	(0.05)	(0.10)	(0.09)
1.9912	1.9973	1.0016	0.3022	2.9759	3.0097	2.0047	0.1992
(0.10)	(0.07)	(0.05)	(0.05)	(0.19)	(0.05)	(0.10)	(0.04)
0.8934	1.9952	0.9878	0.3977	1.9681	3.0099	1.9797	0.5982
(0.09)	(0.06)	(0.06)	(0.04)	(0.19)	(0.05)	(0.11)	(0.03)
1.9906	1.9940	0.9958		2.9726	3.0092	1.9783	
(0.09)	(0.06)	(0.06)		(0.20)	(0.05)	(0.11)	
0.9845	1.9945	0.9980		1.9718	3.0154	2.0046	
(0.10)	(0.06)	(0.06)		(0.20)	(0.05)	(0.11)	
1.9856	1.9952	1.0006		2.9690	3.0105	1.9992	
(0.09)	(0.06)	(0.05)		(0.19)	(0.05)	(0.11)	

TAB. 3 – The two approaches results.

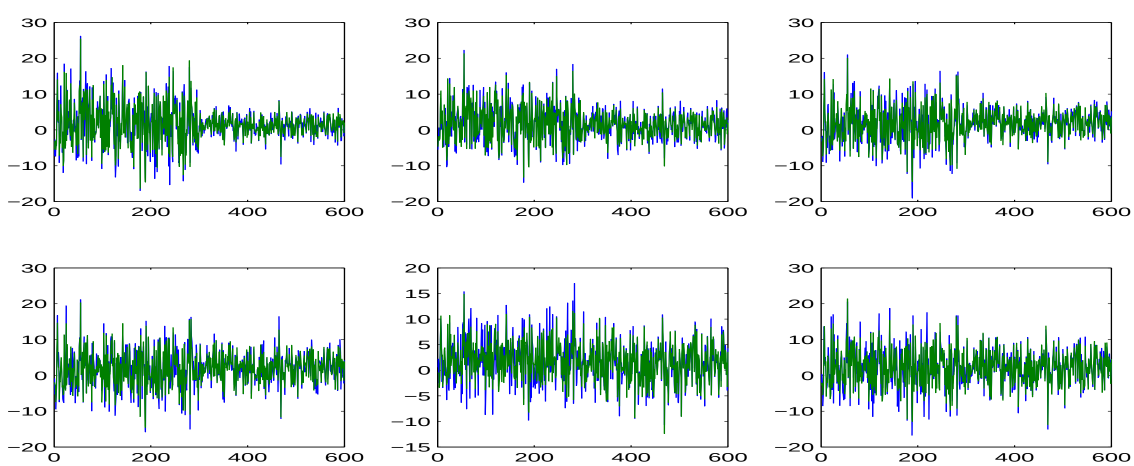

FIG. 2 – Simulated and estimated trajectories of a CHFAHMM model with 2 hidden states, 1 GQARCH(1,1) factor and 6 observable variables.

The Viterbi Algorithm

n	$t-3$	$t-2$	$t-1$	$t=n/2$	$t+1$	$t+2$	$t+3$
600	0	0	0	9	87	4	0
800	1	0	0	8	83	7	1
1000	0	1	1	7	88	1	2
1200	0	2	1	7	89	1	0
1500	0	0	1	14	80	4	1

The Smoothing Algorithm

n	$t-3$	$t-2$	$t-1$	$t=n/2$	$t+1$	$t+2$	$t+3$
600	0	0	0	9	87	4	0
800	1	0	0	8	82	8	1
1000	0	1	1	7	88	1	2
1200	0	1	3	7	88	1	0
1500	0	0	1	13	81	4	1

The Filtering Algorithm

n	$t-3$	$t-2$	$t-1$	$t=n/2$	$t+1$	$t+2$	$t+3$
600	0	0	0	0	5	42	41
800	1	0	0	0	0	45	42
1000	0	1	1	0	1	41	47
1200	0	0	0	0	2	50	38
1500	0	0	0	0	1	51	38

TAB. 4 – Identification of the optimal sequence

We used the Shapiro-Wilk statistic to test the univariate normality of each component of $\widetilde{\Theta}_n$. The sequence size n was increased until the normality null hypotheses [7] could not be rejected at the 0.02 level of significance. The steps in making the $\mathcal{W}$ test for normality, are :

- Calculate the statistic $\mathcal{W}$,

$$\mathcal{W} = \frac{\left[\sum_{i=1}^{r} a_{v+1-i}\left(\Theta_{(v+1-i)} - \Theta_{(i)}\right)\right]^2}{\sum_{i=1}^{v}\left(\Theta_{(i)} - \overline{\Theta}\right)^2}$$

where $r = (v-1)/2$, if v (the number of replications) is odd, and $r = v/2$ if v is even. The exact distribution of $\mathcal{W}$ under the null hypotheses will depend on v, but not on the true values of m and σ. This distribution is not known, and Shapiro and Wilk provided Monte Carlo percentage points for use with the test, for sample sizes $v \leq 50$. Shapiro and Wilk [1968] gave an approximation to the null distribution of $\mathcal{W}$. [8].

- If $\mathcal{W}$ is less than the value 0.938 given in the lower tail (in table 5.5, p. 211 Agostino and Stephens [1986]) for appropriate values of $v = 50$ and $\alpha = 2\%$, we reject H_0 at level α.

6.3 Simulation C : Model Selection

to study the aptitude of the various criteria of selection to choose the suitable dimension of the model, we put in competition models at factors which differ by their hidden structures.

[7] H_0 : the Θ_i are a random sample from $\mathcal{N}(m, \sigma)$, with m and σ unknown.

[8] The simulated values of a_i as well as the thresholds of significance of this test for the samples of size to most equal to 50 was presented in Agostino and Stephens [1986].

Simulations with $n = 1000$

State 1				State 2			
θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ	θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$	Φ
0.9792	0.9676	0.9797		0.9738	0.9519	0.9802	
0.9725	0.9643	0.9700	0.9734	0.9759	0.9823	0.9714	0.9809
0.9720	0.9831	0.9768	0.9526	0.9912	0.9481	0.9768	0.9832
0.9713	0.9763	0.9586	0.9798	0.9749	0.9686	0.9605	0.9845
0.9842	0.9825	0.9728		0.9758	0.9588	0.9769	
0.9805	0.9693	0.9777		0.9842	0.9837	0.9705	

TAB. 5 – The Shapiro-Wilk $\mathcal{W}$ Statistics

EM Algorithm + GPB1

Critère	CHFAHMM	FAHMMNS	FAHMM1	FAHMM2	CHFA	FA1	FA2	FA3
AIC	90	10	0	0	0	0	0	0
BIC	100	0	0	0	0	0	0	0
ICL	100	0	0	0	0	0	0	0
ICOMP	7	93	0	0	0	0	0	0

EM Algorithm + Viterbi

Critère	CHFAHMM	FAHMMNS	FAHMM1	FAHMM2	CHFA	FA1	FA2	FA3
AIC	92	8	0	0	0	0	0	0
BIC	100	0	0	0	0	0	0	0
ICL	100	0	0	0	0	0	0	0
ICOMP	17	83	0	0	0	0	0	0

TAB. 6 – Model Selection.

The three first models are standard factor analysis models (FA) with 1, 2 and 3 common factors respectively (without regime switching). The fourth is a GQARCH(1,1) conditionally heteroskedastic factor model without regime switching (CHFA). Models 5 and 6 are FAHMM models with 1 and 2 common factors and 2 hidden markovian states. The seventh is a FAHMM model with $k = 1$, and where the variances of the common factors are not subjected to the constraint of orthogonality (FAHMMNS). The last and eighth model is a CHFAHMM (the true model) with only one GQARCH(1,1) common factor and 2 hidden markovian states. In this simulation we used $q = 6$ observable variables and $n = 1000$ observations, while respecting the constraint of parsimony on the maximum number of common factors (this number should not exceed 3 in our case). Truths parameters of this model are given in table 1. In each replications, we estimated the parameters of the models in competition for finally calculating their corresponding likelihoods and AIC, BIC, ICL and ICOMP values. The results for 100 replications are given in table 6. This table gives the number of times that each k-factors model reaches the highest log-likelihood. For example by applying an EM algorithm based on the pseudo-bayesian generalized approximation for estimating the models, the AIC criterion selects CHFAHMM model 90 times among 100 and FAHMMNS model with $k = 1$ and $m = 2$, 10 times. Good results were also found by using the BIC and ICL criteria. On the other hand, relatively weak results were completed with the criterion ICOMP which tends to prefer a great number of factors in most of the time (choice of the FAHMM with $k = 2$ and $m = 2$).

We also applied this method GPB1 to estimate the parameters of a CHFAHMM model with two hidden Markovian states and only one factor. On the same sample of data we considered standard and nonstandard FAHMM models, a CHFA model and a FA model.

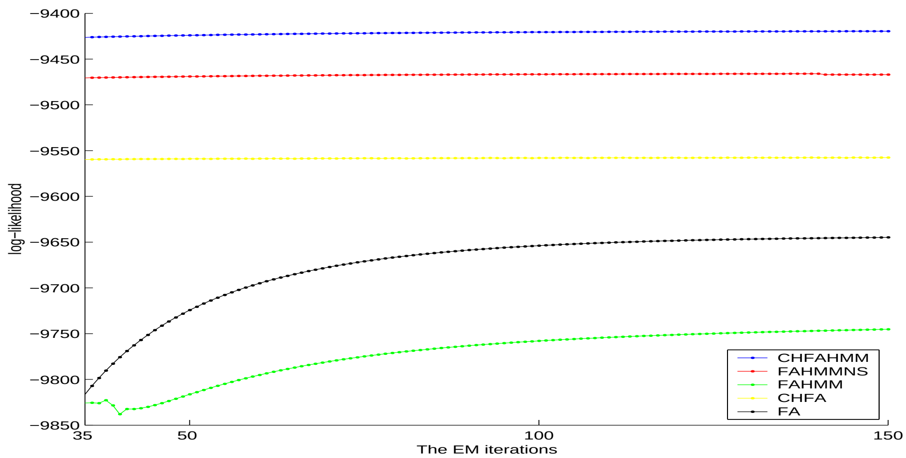

FIG. 3 – Log-likelihoods of the different models

parameters of this simulation, as well as the results for the various models are given in Figures 3, 4, 5 and 6. These Figures show that the FAHMM does not detect the switching regime point and the box-plots of the discrete state posterior probabilities shows that the hidden structure of the data is characterized by the presence of two almost identical regimes and who have a no persistent effect on the observations dynamics. The box-plots of the estimates of the common factors and their hidden volatility show that, except the CHFAHMM model, all the other models lead either to an over-estimate or an under-estimate.

Bibliographie

[1] Aguilar O., and West M. (2002), "Bayesian Dynamic Factor Model for Portfolio Allocation", Journal of Business & Economic Statistics, V 18, N°3, 338-357.

[2] Baum, L., Petrie, T., Soules, G., and Weiss, N. (1970), "A Maximisation technique occuring in the statistical analysis of probabilistic functions of Markov chains", The Annals of Mathematical Statistics, V 41, 164-171.

[3] Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum Likelihood from incomplete data via the EM algorithm", J. Royal Statistical Society Series B, V 39, 1-38.

[4] Demos A., and Sentana E. (1998), "An EM Algorithm for Conditionally Heteroscedastic Factor Models", Journal of Business & Economic Statistics, V 16, N° 3, 357-361.

[5] Diebold F., and Nerlove M. (1989), "The Dynamics of Exchange Rate Volatility : A Multivariate Latente Factor ARCH Model", Journal of Applied Econometrics, V 4, 1-21.

[6] Engle R. (1987), "Multivariate ARCH with Factor Structures-Cointegration in Variance", mimeo, University of California, San Diego, Dept. of Economics.

[7] Engle R., Ng V. K., and Rothschild M. (1990), "Asset Pricing with a Factor-ARCH Structure : Empirical Estimates for Treasury Bills", Journal of Econometrics, V 45, 213-237.

[8] Ghahramani Z. and Geoffrey E. H. (1999), "Variational Learning for Switching State-Space Models", Neural Computation.

[9] Celeux G., Marques J. S. and Nascimento J. (2003), "Learning Switching Dynamic Models for Objects Tracking", Rapport de recherche N° 4863, Unité de recherche INRIA Rhône-Alpes.

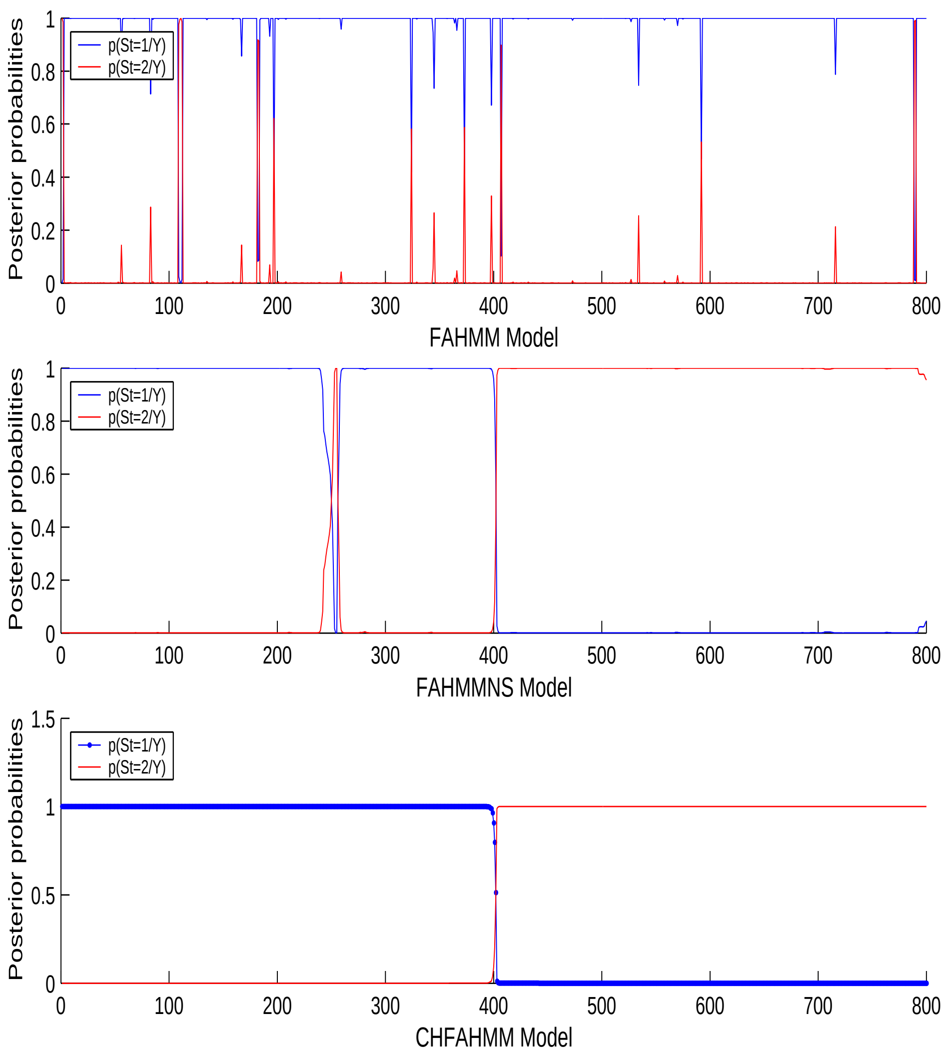

FIG. 4 – Posterior probabilities of the discrete hidden state variable

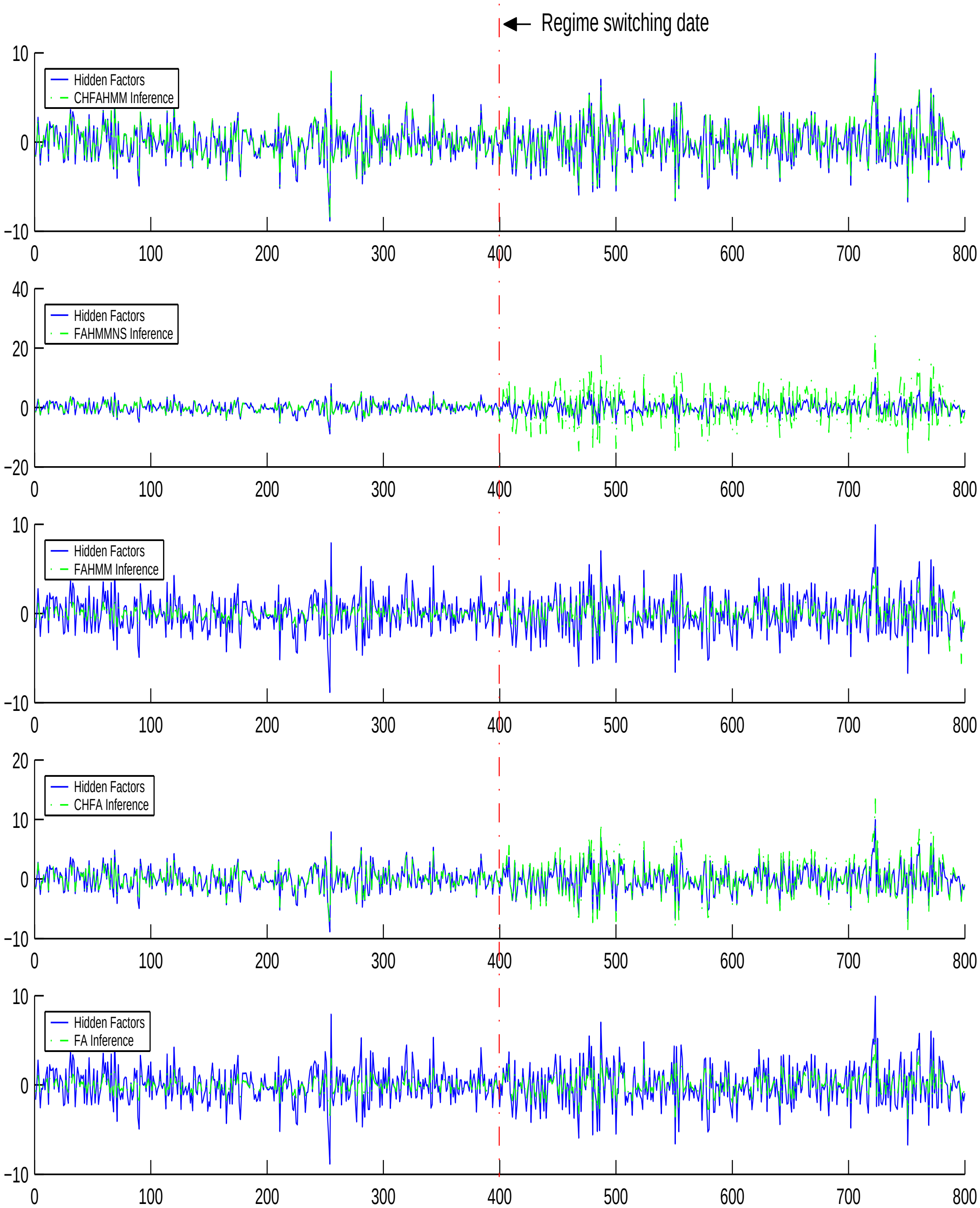

FIG. 5 – Latente factor estimation

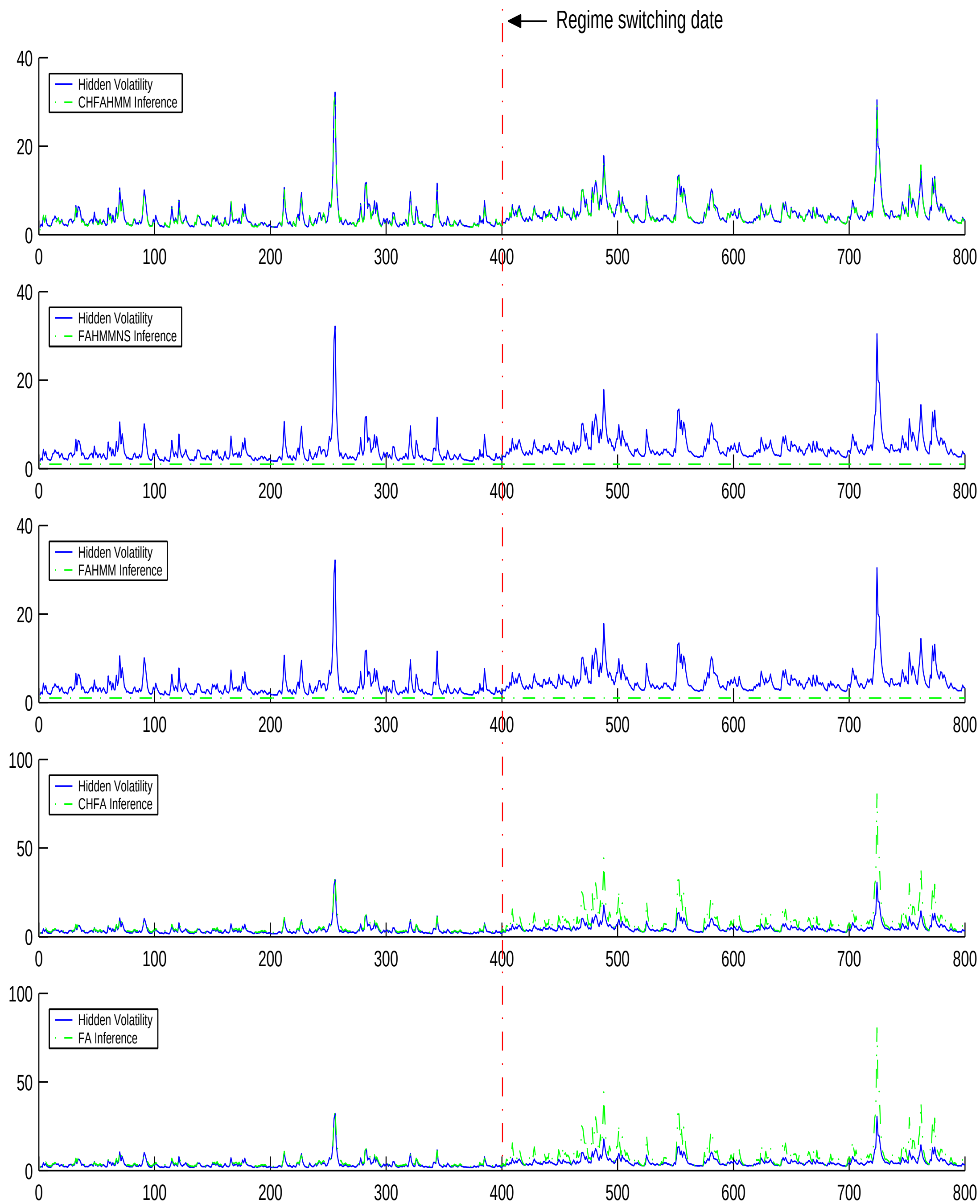

Fig. 6 – Unobservable volatility estimation

[10] Harvey A., Ruiz E. and Sentana E. (1992), "Unobserved component time series models with ARCH disturbances", Journal of Econometrics, V 52, 129-157.
[11] King M., Sentana E., and Wadhwani S. (1994) Volatility and Links Between National Stock Markets", Econometrica, V 62, 901-933.
[12] Kalman, R. E and Bucy, R. S. (1961), "New results in linear filtering and prediction", Journal of Basic Engineering (ASME), V 83, 95-108.
[13] Kroner K. F. (1987), "Estimating and testing for factor GARCH", University of California at San Diego, mimeo.
[14] Lin W. (1992), "Alternative estimators for factor GARCH models - a Monte Carlo comparaison", Journal of Applied Econometrics V 7, 259-279.
[15] Ng V., Engle R. and Rothschild M. (1992), "A multi-dynamic-factor model for stock returns", Journal of Econometrics, V 52, 245-266.
[16] Rabiner L. R (1989), "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE, V 77, N° 2, 257-286.
[17] Rosti A., and Gales M. J. F. (2002), "Factor Analysed Hidden Markov Models", ICASSP, V 1, 949-952.
[18] Roweis S. and Ghahramani Z. [1999], "A Unifying Review of Linear Gaussian Models", Neural Computation, V 11, N° 2, 305-345.
[19] Rubin D., and Thayer D. (1982), "EM Algorithms for ML Factor Analysis", Psychometrika, V 47, 69-76.
[20] Ruud A. (1991), "Extensions of estimation methods using the EM algorithm", Journal of Econometrics, V 49, 305-341.
[21] Sentana E. (2000), "The likelihood function of conditionally heteroscedastic factor models", Annales d'Economie et de Statistique, N°. 58, 1-19.
[22] Sentana E. (1995), "Quadratic ARCH Models", Review of Economic Studies, V 62, 639-661.
[23] Sentana E. (1992), "Identification of Multivariate Conditionally Heteroskedastic Factor Models", Financial Market Group Discussion Paper 139, London School of Economics.
[24] Sentana E., and Shah M. (1994), "An Index of Comovements in Financial Time Series", CEMFI Working Paper 9415.
[25] Shumway, R. H., and Stoffer, D. S. (1991), "Dynamic linear models with switching", Journal of the American Statistical Association V 86, 763-769.