Bayesian inference and state number determination for hidden Markov models: an application to the information content of the yield curve about inflation

N. CHOPIN[a]† , F. PELGRIN[b,c]
[a] CREST, INSEE, Timbre J120, 75675 Paris, cedex 14, France
[b] THEMA, Université de Cergy-Pontoise, 33, Boulevard du Port, 95000 Cergy-Pontoise, France
[c] OECD, Money and Finance Division, 2, rue André Pascal, 75775 Paris Cedex 16, France

Summary. We present a new methodology for handling hidden Markov models. It consists of providing a Bayesian joint estimation of the parameters and the number of distinct regimes that have appeared in the sample. We adapt this approach to a switching regression model, and consider its application to the information content of the yield curve regarding future inflation in G7 countries. In order to compute the corresponding estimates, we implement a particle filter algorithm.

Key-words: Hidden Markov models; Information content of the yield curve; Particle filter; State number determination; Switching regression models.

Résumé. Nous présentons une nouvelle méthodologie pour le traitement des modèles à chaîne de Markov cachée. Elle consiste à mener une estimation bayésienne jointe des para– mètres et du nombre de différents régimes apparus dans l'échantillon. Nous adaptons cette approche à un modèle de régression à bascule, et considérons son application au contenu informatif de la courbe des taux sur l'inflation future, dans les pays du G7. Afin de calculer les estimateurs correspondants, nous avons recours à un algorithme de filtre particulaire.

Mots-clés: Modèle à chaîne de Markov cachée; Contenu informatif de la courbe des taux; Filtre particulaire; Détermination du nombre d'état; Modèle de régression à bascule.

†Corresponding author. Tel: +331-4117-6526; Fax: +331-4117-3852; Email: chopin@ensae.fr

1. Introduction

Many macroeconomic variables or structural relationships undergo episodes in which their behavior seems to be characterized by instability or important changes. In this respect, one may define instability as a switch from one period to another. The idea was first introduced by Quandt (1972), in the case of independent switches in a regression model. Goldfeld and Quandt (1973) have extended the analysis to Markov-chain regime-dependent-switching probabilities. Since the seminal contribution of Hamilton (1989, 1990), economists'attention has been drawn to Markov-switching modeling of endogenous structural changes. Dynamic models with Markov switching have offered new perspectives in many economic areas such as business fluctuations and long-run trend in GNP (Hamilton, 1989; Lam, 1990), the behavior of foreign exchanges rates and real interest rates (Engel and Hamilton, 1990; Garcia and Perron, 1996), the evolution of stock returns (Kim, Nelson and Startz, 1998), etc. Kim and Nelson (1999) consider the advantages of a Bayesian approach when dealing with such models, and address the practical implementation through Gibbs sampling techniques (originally introduced by Geman and Geman, 1984). However, most of these papers invoke inference procedures which are valid for a *given* number of distinct regimes. In contrast, Chopin (2001) proposes a new general approach of discrete state-space models, which allows for a Bayesian joint estimation of the parameters *and* the number of distinct regimes featured by the studied data. We show in this paper how to adapt this approach to a switching regression model, and describe the corresponding implementation strategy, which relies on a particle filter algorithm.

More precisely, to model abrupt changes in a given structural relationship, we introduce an unobserved discrete process, which gives the state of the system (regime) at date t. The hidden process may reflect changes in monetary policies, exchange rate regimes or any change in the economic environment. We focus here on the case where the hidden process is a Markov chain (hidden Markov models), and the structural relationship is linear (switching regression models). The divergence between our results and those found in the literature reflects in part the use of the Bayesian framework. The latter is motivated by three shortcomings of the classical approach.

First, classical inference procedures often rely on asymptotic justifications, and therefore may show some fragility when applied to short series, notably in a regime-switching context, since it may be that only a few points of the studied sequence originate from a given regime. Second, in the classical approach, estimation of the state variables is conditional on the maximum likelihood estimates of the parameters. Here, the state variables, parameters and regimes are jointly distributed random variables, i.e. estimates of each appropriately reflect the uncertainty of the others.

Third, determining the number of regimes is quite an involved problem within a classical approach (see Hansen, 1992; Hamilton, 1996). This is because classical inference procedures are only valid for a given number of regimes, while in practice if this number is mispecified - for instance if it is greater than necessary - extra regimes with no clear interpretation are often artificially created. These extra regimes may be discarded afterwards through some testing procedures, which unfortunately are often difficult to implement and may make the whole estimation process quite time-consuming. In contrast, our Bayesian method provides in the same run (jointly with the other parameters) an estimate of the *actual* number of regimes, that is the number of distinct regimes that indeed have appeared in the studied sequence, provided it is supplied with a correct upper bound K of this number of regimes.

As an illustration, we apply the methodology to the information content of the yield curve about future inflation in G7 countries. This is of particular interest for at least two reasons. Firstly, yield curves or term structure of interest rates are often regarded as one of the financial variables to help predict information or to extract information on future interest rates and inflation development. Thus policy makers or central banks may be interested not only in the ability of the slope of the term structure to predict changes in inflation rates but also in the structural stability as well as the number of regimes embedded in the data. Secondly, the stability of the information content of the yield curve about future inflation are often reviewed in the literature. One approach is to estimate the structural relationship and to get at the issue of sub-sample stability proceeding with single unknown breakpoint tests: the supremum LM test proposed by Andrews (1993) and the supremum predictive test proposed by Ghysels, Guay and Hall (1997). Even if the date at which a structural change occurs is endogenous to the model, the interpretation of the aforementioned statistics is not clear. They may provide some insights about the stability of the model but they do not identify the source of instability. Therefore, the instability of the parameters could reflect either a true instability of the parameters or a change in other aspects of the model. In addition, due to many changes in the economic environment, there is no reason to believe that there is only one break in the data. Thus, our approach may deliver some more appealing results.

In the following section, we discuss the Bayesian approach used in this paper. Thus we explain the state number determination of hidden Markov regression models, the prior modeling as well as the computational implementation of the particle filter. Section 3 presents a simple hidden Markov regression model for the information content of the yield curve regarding future inflation, describes the data and discusses the results. Section 4 briefly summarizes the main findings.

2. A Bayesian approach

2.1. State number determination for Hidden Markov models

A hidden Markov model features an unobserved K-state Markov chain (s_t) (the *regimes*), with transitions probabilities $(p_{kl})_{1\leq k,l\leq K}$

$$P(s_{t+1}=l|s_t=k)=p_{kl}, \tag{1}$$

and an observed parametric process (y_t), whose behaviour at time t is indexed on the current state s_t, in the sense that

$$y_t|\{s_t=k, y_1,...,y_{t-1}\}\sim f_{\xi_k}(y_t|y_{1:t-1}), \tag{2}$$

where $\{f_\xi(.|.), \xi\in\Xi\}$ is some given parametric family of conditional densities, and $y_{1:t-1}$ stands for the sequence $y_1,...,y_{t-1}$ (similarly, a state sequence $s_1,...,s_t$ will be denoted $s_{1:t}$ throughout the paper). Equations (1) and (2) are commonly denoted, respectively, the *system equation*, and the *observation equation.* Apart from the regimes, the unknown quantities form a fixed vector parameter θ, which comprises the p_{kl}'s and the regime parameters $\xi_1,...,\xi_K$.

In this paper, we will mainly focus on a switching-regression model, with observation equation, given $s_t=k$,

$$y_t=\alpha_k+\beta_k x_t+\sigma_k w_t, \tag{3}$$

where (w_t) is a standard gaussian white noise. In this case, the observation density reduces to $f_{\xi_k}(y_t)=\phi(y_t;\alpha_k+\beta_k x_t,\sigma_k^2)$, with $\xi_k=(\alpha_k,\beta_k,\sigma_k^2)$, and $\phi(.;\mu,\sigma^2)$ is the gaussian density with mean μ and variance σ^2. Notational dependencies in (x_t) (which is assumed to be an exogeneous process) are omitted for convenience. Note however the methodology presented below would apply more generally to any hidden Markov model, as defined with (1) and (2).

The current parameterization does not show a full identifiability of the parameters, since it is invariant by permutation of the labels of the regimes. To overcome this, an ordering constraint (such as for instance $\alpha_1 < ... < \alpha_K$ in our application) is usually introduced. However, within such a formulation, a given sequence of observations $y_1, ..., y_t$ may not visit all K regimes. This is especially true if some of the components of the transition matrix are allowed to be null, i.e. for a given l, $p_{kl} = 0$ for any $k = 1, ..., K$. Therefore, rather than estimating all the K components, whereas the data may bring no information on some of them, it seems more sensible to jointly estimate the number m of regimes that *actually* have appeared in the studied sequence, along with these m components. To do so, and following the lines of the state number determination procedure proposed in Chopin (2001), we introduce a second discrete process (m_t), which indicates at time t the number of regimes that have appeared *for the time being*, and we label the regimes *by order of appearance*, that is regime 1 is the first regime to appear, and so on. We reformulate the system equation as

$$\begin{aligned} s_1 &= 1, \\ P(s_{t+1} = l | s_t = k, m_t = m) &= \begin{cases} p_{kl} & \text{if } k, l \leq m \leq K, \\ \sum_{l'=m+1}^{M} p_{kl'} & \text{if } l = m+1 \leq K, \\ 0 & \text{otherwise,} \end{cases} \\ m_1 &= 1, \\ m_{t+1} &= \max(m_t, s_{t+1}). \end{aligned}$$

The new formulation is equivalent to the first system equation (1): when at time t, with $z_t = k$ and $m_t = m$ $(k \leq m)$, the next regime can be either an already visited regime l $(l \leq m)$ with probability p_{kl}, or a new regime, which will be labeled $m + 1$. Since the remaining regimes are not distinguishable at time t, the probability of appearance is indeed $\sum_{l'=m+1}^{M} p_{kl'}$. If a new regime appears, we have $m_{t+1} = z_{t+1} = m + 1$, if not, $m_{t+1} = m$, hence in general $m_{t+1} = \max(m_t, z_{t+1})$.

Within a Bayesian framework, we have to consider a posterior distribution of the form

$$\pi(\theta, m_T, s_{1:T} | y_{1:T}), \tag{4}$$

in order to draw a joint inference on the states, the parameters and m_t, the *actual* number of components, for the considered sequence $y_1, ..., y_T$. For convenience, we may decompose this posterior in

$$\pi(m_T = m | y_{1:T}), \text{ and } \pi(\xi_1, ..., \xi_m, p_{1\dagger}, ..., p_{m\dagger}, s_{1:T} | m_T = m, y_{1:T}), \tag{5}$$

for $m = 1, ...K$, where $p_{k\dagger}$ stands for the kth-line $(p_{k1}, ..., p_{kK})$ of the transition matrix. Conditionally on $m_t = m$, the data brings no information upon $\xi_{m+1}, ...\xi_K, p_{(m+1)\dagger}, ..., p_{K\dagger}$, hence there is no point in incorporating these parameters into the conditional posterior distribution above.

Note finally that the vector $\tilde{s}_t = (s_t, m_t)$ is obviously discrete and Markov (with $K(K+1)/2$ states), i.e. the new formulation still defines a hidden Markov model, with observed process (y_t) and hidden Markov chain $(\tilde{s}_t)$. Therefore, any known method for estimating (the states and parameters of) a hidden Markov model can be made to perform a state number determination procedure. See §2.3 for implementation details.

2.2. Prior modeling

For convenience, assume at first the components $\xi_1, ..., \xi_K$ and the lines $p_{k\dagger}$ of the transition matrix to be pairwise prior independent, so that the global prior on the vector θ decomposes in

$$\pi(\theta) = \prod_{k=1}^{K} \pi(p_{k\dagger}) \prod_{k=1}^{K} \pi(\xi_k). \tag{6}$$

The mixture structure of hidden Markov models usually prevents a fully non-informative prior modeling, since an improper prior (for ξ) usually yields to an improper posterior distribution. Various solutions have been proposed to overcome this problem‡, but for the sake of simplicity, we only consider proper priors in the following. In particular, for the switching regression model presented above, a most reasonable prior distribution $\pi(\xi)$ for $\xi = (\alpha, \beta, \sigma^2)$ is the natural conjugate prior for regression models

$$(\alpha, \beta|\sigma^2) \sim \mathcal{N}_2(M_0, \sigma^2 S_0), \qquad 1/\sigma^2 \sim \Gamma(a_0, b_0), \tag{7}$$

where the hyper-parameters M_0, S_0, a_0, and b_0 are to be set by the decision-maker.

The state number determination procedure defined in the previous section, which aims at providing an estimate of the *actual* number of regimes m_t, is in fact strongly affected by the choice of prior for the transition probabilities. If we assign for instance a (natural conjugate) Dirichlet prior to the lines of the transition matrix

$$(p_{k1}, ..., p_{kK}) \sim \mathcal{D}(\alpha_{k1}, ..., \alpha_{kK}), \tag{8}$$

we implicitly affect a null prior probability to the event $p_{1l} = ... = p_{Kl} = 0$, for a given l, and therefore assume the number of distinct regimes that would finally appear (as $t \to +\infty$) is exactly K. This is somehow presumptuous, and in practice such a prior may strongly bias the estimation of m_t. A more flexible approach is to consider K as a mere upper bound of the (actual or final) state number, and assume with some (non-null) prior probability that the data may be produced by a hidden Markov model with κ states ($\kappa < K$). In this spirit, an appealing prior is

$$\kappa \sim \mathcal{U}[1, K], \text{ and, given } \kappa \tag{9}$$

$$(p_{k1}, ..., p_{k\kappa}) \sim \mathcal{D}(a^{\kappa}_{k1}, ..., a^{\kappa}_{k\kappa}),\, p_{k(\kappa+1)} = ... = p_{kK} = 0 \tag{10}$$

where $\mathcal{U}[1, K]$ stands for the uniform distribution on $[1, K]$, and a^{κ}_{kl}'s are hyper-parameters. While it is commonly suggested to set $a^{\kappa}_{kl} = 1$ for any k, l, κ, we rather advice to set for instance $a^{\kappa}_{kl} = 1$, for $k \neq l$, $a^{\kappa}_{kk} = 3(\kappa - 1)$, in order to incorporate the prior information that the probabilities p_{kk} cannot be too close to 0, since this would imply unrealistically high frequencies of switching, and hinder any interpretation of the regimes.

As it stands, our state number determination procedure is still ill-defined, since it may fall in creating redundant regimes, in order to artificially improve the data fit. Note that this "over-fit" artifact more generally affects most inference procedures in a mixture context, either in a classical or a Bayesian approach, and must be countered with some parsimony mechanism. However, within a Bayesian framework, the decision-maker has the ability

‡for independent mixture models, see Diebolt and Robert (1994), and Wasserman (2000), for dynamic mixture models, see Chopin (2001).

to *prior specify* the desired degree of parsimony, by introducing for instance in the prior distribution a *discriminating factor*, that is, the prior in (6) is replaced by

$$\pi(\theta) \propto \prod_{k=1}^{K} \pi(p_{k\dagger}) \prod_{k=1}^{K} \pi(\xi_k) d(\xi_1, ..., \xi_K), \tag{11}$$

where $d(\xi_1, ..., \xi_K)$ is a discrimating factor of the form, for example,

$$d(\xi_1, ..., \xi_K) = \prod_{1 \leq i < j \leq K} \left\{ 1 - e^{-[\Delta(\xi_i, \xi_j)/\delta]^\nu} \right\}, \tag{12}$$

with $\Delta(\xi_i, \xi_j)$ a given distance on Ξ, and δ and ν tuning hyper-parameters. Ideally, $\Delta(\xi_i, \xi_j)$ should provide an identifiability measure between ξ_i and ξ_j, or in other words indicate to which extent the two considered parameters imply a distinct behaviour for the observation process. For instance, in our switching-regression model, the (expected) Kullback-Leibler divergence§ between observations densities $f_{\xi_i}(y_t)$ and $f_{\xi_j}(y_t)$, with $\xi_i = (\alpha_i, \beta_i, \sigma_i)$ and $\xi_j = (\alpha_j, \beta_j, \sigma_j)$, seems a reliable choice, and is easily derived as

$$\begin{aligned} \Delta(\xi_i, \xi_j) = \frac{\sigma_i^2}{\sigma_j^2} + \frac{\sigma_j^2}{\sigma_i^2} - 2 + (1 + \frac{1}{\sigma_j^2} + \frac{1}{\sigma_i^2})[(\alpha_i - \alpha_j)^2 + \\ E(x_t)(\alpha_i - \alpha_j)(\beta_i - \beta_j) + E(x_t^2)(\beta_i - \beta_j)^2]. \end{aligned} \tag{13}$$

For convenience, we assume (x_t) to be a stationary process, so that $E(x_t^2)$ and $E(x_t)$ are constant in time. In practice, these two expectations can be replaced by the corresponding empirical averages over the studied sequence.

2.3. Computational implementation

Since its introduction in the early 1990's, MCMC (Monte Carlo Markov Chains) methods (see Robert and Casella, 1999, for a thorough presentation) have gained an increasing popularity in the Bayesian literature, and are now considered the most convenient numerical tools for managing a Bayesian analysis. In particular, most recent papers dealing with hidden Markov models (Kim and Nelson, 1999) prescribe Gibbs sampler techniques for estimating the states and parameters.

However, the implementation of such algorithms can be quite delicate in complex settings. For instance, the prior distribution we proposed in the previous section yields to rather an intricate posterior distribution (partly because of the discriminating factor, and the mixture prior of the transition probabilities), and therefore severely complicate the derivation of a proper Gibbs sampler. Moreover, a drawback of MCMC methodology is the difficulty to assess in practice if the algorithm has converged, and produces reliable estimates. More specifically when dealing with mixture models (whose hidden Markov models are a particular case), it happens that the Markov chain produced by the MCMC algorithm seems to have reached stationarity whereas it is in fact confined in a "trap-state" (a local mode of the posterior distribution). This difficulty is often referred as the "switching-label" phenomenon, since it is considered that the hidden states simulated within Gibbs iterations have mismatched labels (Celeux et al., 2000).

§the Kullback-Leibler divergence between two densities f and g is defined as $E_f[\log(f)/\log(g)] + E_g[\log(g)/\log(f)]$.

Rather, we use the Monte Carlo HMM filter (MCHMM filter) proposed by Chopin (2001), which is a particle filter algorithm (see Doucet et al, 2001, for a general overview of this class of algorithms) devoted to the estimation of hidden Markov models. A first advantage of this algorithm is its great flexibility, in that its code is mostly model-independent (up to the computation of the prior and likelihood densities, and the simulation of the initial particles), hence the adaptation to another model (obtained by changing the observation equation for instance) or to another prior requires little effort. A second advantage of the MCHMM filter is that it is hardly affected by label-switching (at least in our experiments), since instead of producing a single Markov chain which explores *locally* the distribution of interest, and may be trapped by a local mode, it follows the evolution of numerous, well-spread "particles" (Monte Carlo realizations) which are to accurately map the whole surface of the target distribution at the final stage of the algorithm.

We now give a short description of the MCHMM filter. For more details (notably on the implementation of a resample-move strategy for improving the algorithm, in the spirit of Gilks and Berzuini, 2001), we refer to Chopin (2001).

Following the lines of the approach developed above, the distribution of interest, for a given observed sequence $y_1, ..., y_T$, is the posterior distribution $\pi(\theta, s_{1:T}, m_T|y_{1:T})$. For convenience, we consider more generally the problem of evaluating $\pi(\theta, \widetilde{s}_{1:T}|y_{1:T})$, with $\widetilde{s}_t = (s_t, m_t)$. As we already stated, the process $(\widetilde{s}_t)$ is a discrete Markov chain with $K' = K(K+1)/2$ states, which we re-label for convenience in $[1, K']$.

An important methodological remark is that conditional distributions of the form $\pi(\widetilde{s}_t|\theta, y_1, ..., y_T)$ can be derived *exactly*. More precisely, let $S_{t'}^{t}(\theta)$ the vector of probabilities $P(\widetilde{s}_t = k|\theta, y_{1:t'})$, $k = 1, ..., K'$. These probabilities are usually referred as, respectively, forecasting, filtering, or smoothing probabilities, whether $t > t'$, $t = t'$ or $t < t'$. We have (Hamilton, 1989)

$$S_t^{t+1}(\theta) = P' S_t^t(\theta), \qquad S_{t+1}^{t+1}(\theta) \propto O_{t+1}(\theta) \otimes S_t^{t+1}(\theta), \tag{14}$$

where P is the transition matrix defined by the corresponding components of θ, $O_{t+1}(\theta)$ is the vector of observation densities $f_{\xi_k}(y_{t+1}|y_{1:t}, \theta)$, for $k = 1, ..., K$, and $\otimes$ denotes the element-by-element product of two vectors. In the latter equation, S_{t+1}^{t+1} is defined up to a multiplicative constant ($\propto$ means "proportional to") which is retrieved by normalization. Moreover, we have, for $k \geq 0$ (Kitagawa, 1987)

$$S_{t+1}^{t-k}(\theta) \propto S_{t-k}^{t-k}(\theta) \otimes \left\{ P \left[S_{t+1}^{t-k+1}(\theta) \oslash S_{t-k}^{t-k+1}(\theta) \right] \right\}, \tag{15}$$

where $S_{t+1}^{t-k+1}(\theta) \oslash S_{t-k}^{t-k+1}(\theta)$ denotes the element-by element division of $S_{t+1}^{t-k+1}(\theta)$ by $S_{t-k}^{t-k+1}(\theta)$. Formulae (14) and (15) are usually referred as the forward/backward formulae, or the HMM filter. Their matrix formulation is adapted from Ryden (2000).

The MCHMM filter must be seen as a Monte Carlo generalization of the HMM filter: it consists in running a number H of forward recursions (14), for various values θ_j ($j = 1, ..., H$), of the parameter. These θ_j are weighted "particles" (θ_j particle with weight w_j), which provide a Monte Carlo approximation of the marginal posterior distribution $\pi(\theta|y_{1:t})$ at the iteration t of the algorithm ($t = 1, ..., T$), in the sense that

$$\lim_{H \to +\infty} \frac{\sum_{j=1}^{H} w_j h(\theta^{(j)})}{\sum_{j=1}^{H} w_j} = E_{\pi(\theta|y_{1:t})}[h(\theta)] \text{ almost surely}, \tag{16}$$

holds at iteration t for any function h such that the limit term is defined.

More precisely, assume the particle system (θ_j, w_j) fulfills (16) at iteration t, and the corresponding filtering probabilities $S_t^t(\theta_j)$ are known. Apply (14) for each particle, in order to get the $S_{t+1}^{t+1}(\theta_j)$'s, and define $u_{t+1}(\theta_j) = P(y_{t+1}|\theta_j)$. As it is easily shown, $u_{t+1}(\theta_j)$ is at the same time the normalization constant related to the proportionality relation of (14) (and therefore is computed when deriving $S_{t+1}^{t+1}(\theta_j)$) and the ratio (up to a multiplicative constant):

$$u_{t+1}(\theta_j) \propto \frac{\pi(\theta|y_{1:t+1})}{\pi(\theta|y_{1:t})}. \tag{17}$$

Then, by updating the weights through

$$w_j \to w_j \times u_{t+1}(\theta_j) \tag{18}$$

we operate a sequential importance sampling step from distribution $\pi(\theta|y_{1:t})$ to distribution $\pi(\theta|y_{1:t+1})$, and (16) is now fulfilled at iteration (t+1) (Liu and Chen, 1998).

In this manner, the algorithm provides at its final stage a Monte Carlo estimate of any expectation of the form $E_{\pi(\theta|y_{1:T})}[h(\theta)]$, and therefore allows for a marginal inference on θ. We then iterate the backward formula (15) for each particle θ_j (note the $S_t^t(\theta_j)$'s were already computed in the previous iterations), in order to derive the probabilities $P(\widetilde{s}_t = k|\theta_j, y_{1:T})$, $t = 0, ..., T$, $k = 1, ..., K'$. Since

$$\pi(\widetilde{s}_t = k|y_{1:T}) = \int P(\widetilde{s}_t = k|\theta, y_{1:T})\pi(\theta|y_{1:t})\, d\theta, \tag{19}$$

we apply (16) with $h(\theta) = P(\widetilde{s}_t = k|\theta_j, y_{1:T})$ and get that

$$\frac{\sum_{j=1}^{H} w_j P(\widetilde{s}_t = k|\theta_j, y_{1:T})}{\sum_{j=1}^{H} w_j} \tag{20}$$

is a consistent estimate of the probability $\pi(\widetilde{s}_t = k|y_{1:T})$. In this way, we also get a marginal inference of the states $\widetilde{s}_1, ...\widetilde{s}_T$ (and therefore of the s_t's and the m_t's).

3. An application: The information content of the yield curve regarding future inflation

3.1. *A simple model*

We argue that the information content with respect to future inflation varies over time and depends on regimes that occur at different points of time. In effect, the variation and instability in the relationship between yield spreads and future inflation may be explained by changes in the structure of the economy (Kozicki, 1997, Schich, 1999), i.e. changes in monetary policy, exchanges rate regime, financial market regulation or the institutional structure of bond markets. Accordingly, the first m_1 observations may come from regime 1, the next m_2 from regime 2 and the next m_3 from regime 1 again, etc. The first regime may be characterized with a strong information content about future inflation and the second may only reflect changes in the real interest rate.

The theoretical basis for the information content regarding future inflation can be illustrated by a simple model with K regimes. The K regimes are not however necessarily fulfilled.

Assume that the inflation rate evolves over time according to the following

$$\pi_t = \sum_{k=1}^{K} \mathbf{1}_{\{s_t=k\}} \pi_{k,t}, \tag{21}$$

where $\mathbf{1}_{\{s_t=k\}}$ is the indicator function of the event $s_t = k$ (i.e. $\mathbf{1}_{\{s_t=k\}} = 1, 0$ whether $s_t = k$ or not).

Thus, the difference between the m-period (π_t^m) and the n-period (π_t^n) ahead inflation rate is given by

$$\pi_t^m - \pi_t^n = \sum_{k=1}^{K} \mathbf{1}_{\{s_t=k\}} (\pi_{k,t}^m - \pi_{k,t}^n). \tag{22}$$

Using the methodology of Tzavalis and Wickens (1996), the Fisher decomposition of the m-year nominal interest rate $i_{k,t}^m$ yields

$$i_{k,t}^m = r_{k,t}^m + E_t[\pi_{k,t}^m] + \phi_{k,\pi}^m, \tag{23}$$

where $r_{k,t}^m$ stands for the (ex-ante) m-year real interest rate, $\phi_{k,\pi}^m$ is the forward inflation risk premium, and $E_t[.]$ is the expectation operator conditional on information available at time t.

Assuming rational expectations the realized inflation rate $\pi_{k,t}^m$ can be written as the sum of the expected m-period inflation rate and a serial uncorrelated, zero-mean error term:

$$\pi_{k,t}^m = E_t[\pi_{k,t}^m] + \varepsilon_{k,t}^m. \tag{24}$$

Thus, changes in the future m-period inflation rate from the n-period inflation rate ($m > n$) is expressed as

$$\pi_{k,t}^m - \pi_{k,t}^n = E_t[\pi_{k,t}^m - \pi_{k,t}^n] + \varepsilon_{k,t}^m - \varepsilon_{k,t}^n. \tag{25}$$

Combining equations (23) and (25), we obtain:

$$\pi_{k,t}^m - \pi_{k,t}^n = (i_{k,t}^m - i_{k,t}^n) - (r_{k,t}^m - r_{k,t}^n) - (\phi_{k,\pi}^m - \phi_{k,\pi}^n) + \varepsilon_{k,t}^m - \varepsilon_{k,t}^n. \tag{26}$$

Considering that the difference between the real interest rates for maturities m, n equals some constant plus a zero-mean random variable $u_{k,t}^m$, and that the difference between the forward inflation risk premium equals a constant $\phi_k^{m,n}$, the difference between the m-period and n-period inflation rate (25) can be written in state $s_t = k$:

$$\pi_{k,t}^m - \pi_{k,t}^n = (\phi_k^{m,n} + \gamma_k^{m,n}) + \beta_k^{m,n}(i_{k,t}^m - i_{k,t}^n) + \varepsilon_{k,t}^m - \varepsilon_{k,t}^n + u_{k,t}^{m,n}. \tag{27}$$

Note that an identification problem exists, in that only the sum $\alpha_k^{m,n} = (\phi_k^{m,n} + \gamma_k^{m,n})$ is estimable. Thus, a switch in the constant term may reflect either a change in real interest rates or a change in risk premium, or even both.

Finally, combining equations (27) and (22), the non-centered Markov switching model is given by:

$$\begin{aligned}\pi_{k,t}^m - \pi_{k,t}^n = \sum_{k=1}^{K} \mathbf{1}_{\{s_t=k\}} [\alpha_k^{m,n} + \beta_k^{m,n}(i_{k,t}^m - i_{k,t}^n) \\ + \varepsilon_{k,t}^m - \varepsilon_{k,t}^n + u_{k,t}^{m,n}].\end{aligned} \tag{28}$$

The model can be equivalently written in terms of equation (3) and then analyzed through the methodology developed in §2.

3.2. Empirical evidence

We estimate the information content of the term structure regarding future inflation in G7 countries. The interest rate data is end of month observations from one to five years for Canada (starting January 72 and ending May 95), France (January 1980, October 1998), Germany (January 67, January 1999), Italy (October 1985, January 1998), Japan (March 1980, May 1998), the United Kingdom (January 1970, February 1999) and the United States (June 1961, February 1998). We construct all the term spread where m (respectively n) varies from one to five (respectively from one to four). The data is obtained from national central banks and the beginning dates are dictated by data availability. The inflation rate is based on the Consumption Price Index (CPI) from the monthly OECD Analytical Database (ADBM). The actual forward inflation rate over the next m-period is computed as $\log(P_{t+12m}/P_t)/m$ with P_t denoting the CPI in month t. The relationship between yield curves and future inflation changes is estimated on the different maturities combinations.

The hyper-parameters of the prior distribution (see §2.2) were set to, respectively, $M_0 = (0,0)'$, $S_0 = \text{diag}(2.5, 5)$ (that is the diagonal matrix with diagonal terms 2.5, 5) , $a_0 = 1/2$, $b_0 = 1/8$, $\delta = 12.5$, $\nu = 4$, $a^{\kappa}_{kl} = 1$ for $k \neq l$, and $a^{\kappa}_{kk} = 3(\kappa - 1)$. The upper bound K of the number of components is set to four.

Table 1 reports the posterior expectations of the parameters in the case of the United States. First, results show that the four components are always fulfilled (with a probability equals to one). The large number of components indicates a great variability in the structural relationship between the yield curves and the future inflation across time. However the number of regimes is always four whatever the maturities. These results differ from the standard literature in the sense that breakpoint tests (Schich, 1999) only identify a single structural change in the United States. Figure 1 represents the evolution of the filtered expectations $E(z_t|y_{1:t})$ and $E(m_t|y_{1:t})$, which respectively estimate the current state s_t and the current number of components m_t (at time t). In addition, the expected duration in each regime defined as $1/(1-p_{kk})$ suggests asymmetric dynamics: expected durations are (on average) 14, 9, 20, 6 months in regimes 1 to 4 respectively. Second, according to the positive sign of the slope parameter, the information content of the yield curves regarding future inflation varies with the studied segment of the yield curve and the relevant regime. To test for the information content in regime k, we derive the Bayes factor for $\beta_k > 0$ (see Table 1).

This confirms previous empirical studies. Among others, Mishkin (1990a, 1990b), Jorion and Mishkin (1991) and Schich (1999) have shown that the information of the yield curve is subject to change over time using sub-sample OLS estimates. Third, as shown by figure 2, the source of instability is due to switches of the slope parameter as well as the constant term. In the latter, this may reflect important changes in the evolution of the real interest rates and/or the inflation risk premium. Notably, the switch from regime 2 to 3 features a rather higher variation of the constant term than the spread parameter. Fourth, the regression standard deviations in each regime are remarkably smaller than the standard deviation of the OLS estimate computed on the whole sample, hence within a given regime the structural relationship is stronger than previously thought.

Table 2 reports selected results for other G7 countries. First, the upper bound of the number of components K is not reached in every case. We obtain two components for France, Italy and Japan, three components for Germany, four components for Canada, and the United Kingdom. This reflects that our procedure provides the actual number of regimes, i.e. the number of distinct regimes that have truly appeared in the studied sequence. Figure 3 gives the evolution in time of the filtered expectations of the current state s_t on the current number of components m_t in Germany. Second, the information content of the yield curve regarding future inflation varies across countries. In agreement with earlier studies, the term structure in Canada, Germany appears more informative than in France, Italy, Japan and the United Kingdom about future inflation. In the case of Canada, Day and Lange (1997) show that the yield curve has some information content about future inflation for different maturities combinations. Gerlach (1997), Schich (1999) find evidence for a significant information content in the German term structure. However, contrary to the aforementioned papers, our results suggest that the information content depends on each regime and thus the predictive content may appear unstable. The negative sign of the constant term and its variations are in favor of variations of the real interest rates and/or the risk premium. Figure 3 shows that a shift from regime 1 to 2 is due to both a change in the constant term and the slope parameter whereas a shift from regime 2 to 3 reflects a lower information content of the yield curve. In the case of France, Jondeau and Ricart (1997) fail to find any significant information in the term structure but obtain some empirical evidence using sub-samples. Our results confirms such a predictive content in regime 1. In Japan, Koedijk and Kool (1995) failed to find any empirical support for the information content of the yield curves. Figure 5 shows that there are two apparent regimes. The negative sign of the slope parameter also indicates however a lack of the information content and thereby the predominance of the variations of the real interest rate and/or the forward inflation risk premium. Finally, while Jorion and Mishkin (1991) find some empirical support in the English term structure, we fail to obtain such results in the four identified regimes. Breedon and Chadha (1996) obtain similar results.

In summary, the information content of the term structure regarding future inflation varies over time, across countries, and the different combinations of studied maturities. This confirms earlier empirical results. In addition, we get that the predict content of the yield curve varies among regimes. Finally, the number of regimes (which is determined endogeneously in our method) is remarkably stable for different maturities combinations in a given country.

4. Conclusion

This paper gives new insights into the Bayesian analysis of hidden Markov models, with a particular focus on switching regression models. Mainly, the determination of the number of regimes that actually appear in a given sequence is more easily identified in our approach. This allows for a more precise characterization of the instability of a given structural relationship, as illustrated by our yield curve against future inflation data. Our method may show handful in a variety of other economic applications.

Acknowledgments

We are grateful to Frédérique Bec, Alain De Serres, Mike Kennedy, Christian P. Robert and Sebastian Schich for their helpful comments, and Philippe Donnay for his computing assistance. We also wish to thank Sebastian Schich for having made available to us the data used in Schich (1999).

Table 1: Inflation switching regressions in the United States

Spread (m-n)	Probabilities				Constant				Slope parameter				Standard deviation			
	p_{11}	p_{22}	p_{33}	p_{44}	α_1	α_2	α_3	α_4	β_1	β_2	β_3	β_4	σ_1	σ_2	σ_3	σ_4
2-1	0.903	0.873	0.928	0.896	-0.03	0.60	-1.10	1.02	0.07^{+}	0.35^{++}	0.92^{+++}	2.55^{+++}	0.21	0.26	0.51	0.44
3-1	0.910	0.882	0.941	0.898	0.16	0.94	-1.45	1.98	-0.11^{-}	0.57^{+++}	1.20^{+++}	1.35^{+++}	0.32	0.35	0.90	0.81
4-1	0.929	0.885	0.951	0.836	1.40	0.46	-1.44	4.68	-1.28^{---}	2.53^{+++}	1.68^{+++}	-1.03^{---}	0.36	0.51	1.21	0.47
5-1	0.884	0.941	0.883	0.937	0.99	1.44	-0.34	-2.13	-0.62^{---}	1.32^{+++}	-0.31^{--}	1.66^{+++}	0.24	0.58	0.33	1.43
4-2	0.935	0.931	0.920	0.923	0.73	-0.55	1.60	-2.05	-0.13^{-}	0.22^{++}	1.45^{+++}	0.81^{+++}	0.29	0.37	0.59	0.62
5-2	0.924	0.886	0.951	0.921	1.03	0.12	-1.54	2.30	-0.01^{-}	-0.50^{---}	1.30^{+++}	0.45^{++}	0.29	0.37	1.15	0.48
4-3	0.941	0.942	0.911	0.887	0.38	-0.37	1.24	-1.22	-0.33^{--}	-0.04^{-}	-0.06^{-}	1.68^{+++}	0.18	0.19	0.36	0.41
5-3	0.937	0.952	0.913	0.896	0.60	-0.61	1.91	-2.35	-0.02^{-}	0.09^{+}	-0.30^{-}	0.20^{++}	0.26	0.32	0.37	0.44
5-4	0.941	0.949	0.903	0.915	0.31	-0.31	1.01	-0.95	-0.23^{-}	0.09^{++}	0.23^{++}	-0.70^{--}	0.17	0.14	0.23	0.41

Note:

+, ++, +++, -, - -, - - - denote the slope parameters whose Bayes factor (for testing $\beta > 0$) is greater than 1, 10 and 100, smaller than 1, 0.1, 0.01 respectively.

Source: Authors' calculations

Table 2: Selected results in other G7 countries

Spread[1] (m-n)	Probabilities				Constant				Slope parameter				Standard deviation			
	p_{11}	p_{22}	p_{33}	p_{44}	α_1	α_2	α_3	α_4	β_1	β_2	β_3	β_4	σ_1	σ_2	σ_3	σ_4
Canada	0.875	0.858	0.908	0.892	-0.56	0.16	0.97	-1.11	-0.05^{-}	-0.19^{-}	0.03^{+}	0.72^{++}	0.16	0.21	0.42	0.23
France	0.953	0.991	..	..	-1.88	-0.16	..	..	0.17^{+}	-0.39^{-}	..	..	0.24	0.25	..	..
Germany	0.958	0.920	0.961	..	1.10	-0.72	-1.52	..	0.28^{+++}	0.88^{+++}	0.42^{+++}	..	0.37	0.32	0.29	..
Italy	0.952	0.953	..	..	0.25	-0.94	..	..	0.01^{+}	0.34^{++}	..	..	0.32	0.25	..	..
Japan	0.963	0.945	..	..	-0.38	0.42	..	..	-0.26^{-}	-1.18^{-}	..	..	0.29	0.28	..	..
United Kingdom	0.883	0.888	0.940	0.943	1.37	3.96	-0.17	-2.13	-0.13^{-}	-0.82^{---}	-0.07^{-}	-0.23^{-}	0.49	0.58	0.44	0.76

Notes:

1. The spreads are 4-3 for Canada, 5-3 for France, 5-2 for Germany, 4-2 for Italy, 3-2 for Japan and 5-3 for the United Kingdom, respectively.
2. +, ++, +++, -, - -, - - - denote the slope parameters whose Bayes factor (for testing $\beta > 0$) is greater than 1, 10 and 100, smaller than 1, 0.1, 0.01 respectively.

Source: Authors' calculations

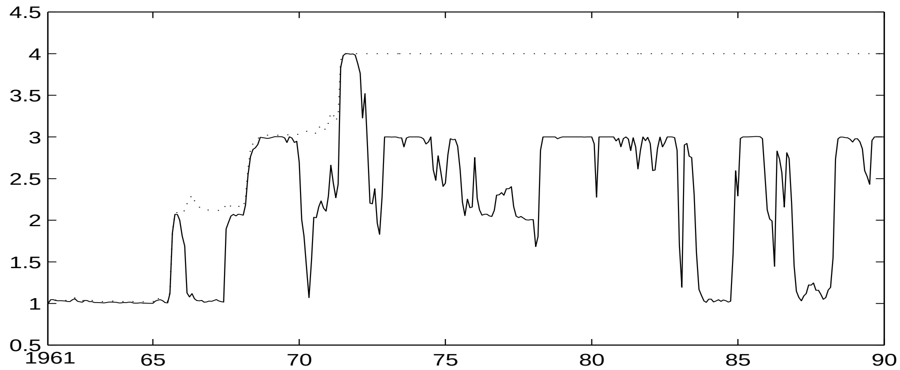

Fig. 1. USA 4-1, evolution in time of the filtered expectations of the current state $E(s_t|y_{1:t})$ (solid line) and the current number of regimes $E(m_t|y_{1:t})$ (dotted line).

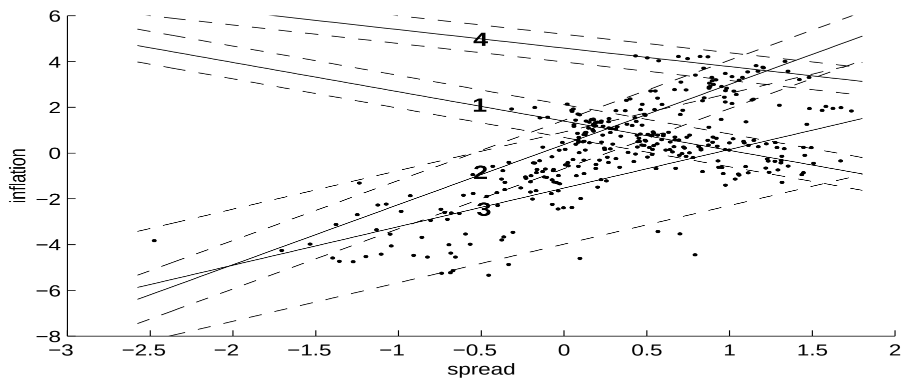

Fig. 2. USA 4-1, yield curve against future inflation between June 1961 and January 1990, regression lines for each regime (solid lines), regression lines translated by two times the standard deviation (dashed lines).

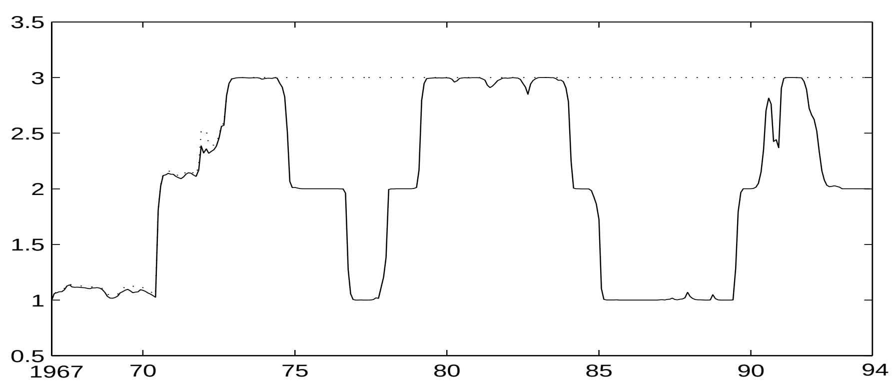

Fig. 3. Germany 5-2, evolution in time of the filtered expectations of the current state $E(s_t|y_{1:t})$ (solid line) and the current number of regimes $E(m_t|y_{1:t})$ (dotted line).

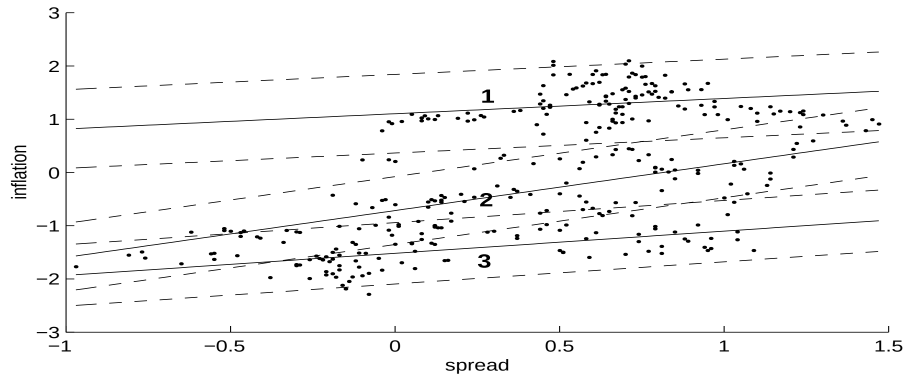

Fig. 4. Germany 5-2, yield curve against future inflation between January 1967 and January 1990 and the corresponding regression lines for each regime, same conventions as figure 1.

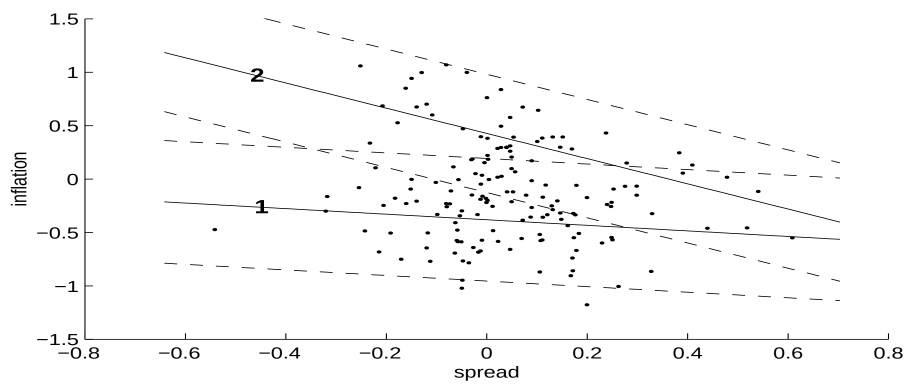

Fig. 5. Japan 3-2, yield curve against future inflation between March 1980 and March 1998 and the corresponding regression lines for each regime, same conventions as figure 1.

References

Andrews, D.W.K, 1993, Tests for parameter instability and structural change with unknown change point, Econometrica, 61, 821-856.

Breedon, F. and J. Chadha, 1996, Index-linked bonds are worth the trouble, Mimeo, Bank of England, April.

Celeux, G., M. Hurn and C. P. Robert, 2000, Computational and inferential difficulties with mixture posterior distributions, Journal of the American Statistical Association, 95, 957-970.

Chopin, N. , 2001, Sequential inference and state number determination for discrete state-space models through particle filtering, CREST Working Paper 2001-34, INSEE, Paris.

Day, J. and R. Lange, 1997, The structure of interest rates in Canada: information content about medium-term inflation, Bank of Canada, Working paper, 97-10.

Diebolt, J. and C. P. Robert, 1994, Estimation of finite mixture distributions through bayesian sampling, Journal of the Royal Statistical Society B, 56, 363-375.

Doucet, A., N. de Freitas and N.J. Gordon, 2001, Sequential Monte Carlo Methods in Practice , Springer-Verlag.

Engel, C., and J.D. Hamilton, 1990, Long swings in the dollar: Are they in the data and do the markets know it?, American Economic Review, 80, 689-713.

Garcia, R. and P. Perron, 1996, An analysis of the real interest rate under regimes shifts, Review of Economics and Statistics, 78, 11-125.

Geman, S. and D. Geman, 1984, Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741.

Gerlach, S., 1997, The information content of the term structure: evidence from Germany, Empirical Economics, 22, 161-179.

Ghysels, E., A. Guay and A. Hall, 1997, Predictive tests for structural change with unknown breakpoint, Journal of Econometrics, 92, 209-233.

Gilks, W.R. and C. Berzuini, 2001, Following a Moving Target-Monte Carlo Inference for Dynamic Bayesian Models, Journal of the Royal Statistical Society B, 63, 127-146.

Goldfeld, S. M. and R. E. Quandt, 1973, A Markov model for switching regression, Journal of Econometrics, 1, 3-16.

Hamilton, J. D. , 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica, 57(2), 357-384.

Hamilton, J. D., 1990, Analysis of times series subject to changes in regime, Journal of Econometrics, 45, 39-70.

Hamilton, J.D., 1996, Specification testing in Markov-switching time series models, Journal of Econometrics, 70, 127-57.

Hansen, B., 1992, The likelihood ratio test under non-standard conditions: testing the Markov switching model to GNP, Journal of Applied Econometrics, 7, 61-82.

Jondeau, E. and R. Ricart, 1997, Le contenu en information de la pente des taux: application au cas des titres publics français, Notes d'Etudes et de Recherche, 43, Banque de France, June.

Jorion, P. and F. Mishkin, 1991, A multicountry comparison of term-structure forecasts at long horizons, Journal of Financial Economics, 29, 59-80.

Kim, C. and C. Nelson, 1999, State-space models with regime switching : classical and Gibbs-sampling approaches with applications, MIT press, Cambridge.

Kim, C., C. R. Nelson and R. Startz, 1998, Testing for mean reversion in heteroskedastic data based on Gibbs-Sampling-Augmented Randomization , Journal of Empirical Finance, 5, 131-154.

Kitagawa, G., 1987, Non-Gaussian State-Space Modeling of Nonstationary Time Series, Journal of the American Statistical Association, 82, 1032–1063.

Koedijk, K.G. and C.J.M. Kool,1995, Future inflation and the information in international term structures, Empirical Economics, 20, 217-242.

Kozicki, S., 1997, Predicting real growth and inflation with the yield spread, Federal Reserve Bank of Kansas City, Economic Review, 82, 39-58.

Lam, P., 1990, The Hamilton model with a general autoregressive component: estimation and comparison with other models of economic times series, Journal of Monetary Economics, 26, 409-432.

Liu, J. and R. Chen, 1998, Sequential Monte Carlo methods for dynamic systems, Journal of the American Statistical Association, 93:1032–1044.

Mishkin, F. S., 1990a, What does the yield curve tell us about future inflation?, Journal of Monetary economics, 25, 77-95.

Mishkin, F. S., 1990b, The information in the longer maturity term structure about future inflation, The Quarterly Journal of Economics, August, 815-828.

Mishkin, F. S., 1991, A multi-country study of the information in the shorter maturity term structure about future inflation, Journal of International Money and Finance, 10, 2-22.

Ryden, T. , 2000, Statistics for hidden markov models, Lecture at the ENSAE, PARIS.

Robert, C. P. and G. Casella, 1999, Monte Carlo Statistical Methods, Springer-Verlag, New-York.

Quandt, R. E., 1972, A new approach to estimating switching regressions, Journal of the American Statistical Association, 67, 306-310.

Schich, S. T., 1999, What the yield curves say about inflation: does it changes over time?, OECD working papers nř227,OECD Paris.

Tzavalis, E. and M.R Wickens, 1996, Forecasting inflation from the term structure, Journal of Empirical Finance, 3, 103-122.

Wasserman, L. , 2000, Asymptotic inference for mixture models using data dependent priors, Journal of the Royal Statistical Society B, 62:159–180.