Bayesian hidden Markov models for financial data

Rosella Castellano and Luisa Scaccia
Dipartimento di Istituzioni Economiche e Finanziarie
Università di Macerata
Via Crescimbeni, 20, 62100 Macerata
castellano@unimc.it, scaccia@unimc.it

Abstract: Hidden Markov Models can be considered as an extension of mixture models, which allows for dependent observations and makes them suitable for financial applications. In a hierarchical Bayesian framework, we show how reversible jump Markov chain Monte Carlo techniques can be used to estimate the parameters of the model, as well as the number of regimes. An application to exchange rate dynamics modeling is presented.

Keywords: Bayes factor, Exchange rates, Markov chain Monte Carlo, Reversible jump

1 Introduction

A Hidden Markov Model (HMM) or Markov Switching Model is a mixture model whose mixing distribution is a finite state Markov chain. HMM have been successfully applied to financial data. Engel and Hamilton (1990) modeled segmented time-trends in the US dollar exchange rates via HMMs. HMMs reproduce most of the stylized facts about daily series of returns (Rydén *et al.*, 1998) and accurately estimate stochastic volatility (Rossi and Gallo, 2006).

The main problem associated with HMMs is to choose the number of states, i.e. the number of generating data processes, which differ one from another just for the value of the parameters. In a classical perspective, choosing the number of states would require hypothesis testing with nuisance parameters, identified only under the alternative. Thus, the regularity conditions necessary to apply asymptotic theory do not hold and the limiting distribution of the likelihood ratio test must be approximated by simulation, an approach demanding enormous computational efforts. Penalized likelihood methods such as the Akaike and Bayesian information criteria are less demanding, though, they produce no number quantifying the confidence in the results, such as a p-value.

In a Bayesian context there are different suggestions for choosing the number of states in a HMM. For example, Otranto and Gallo (2002) adopt a Bayesian nonparametric approach, based on Dirichlet processes. Following Robert *et al.* (2000) and Richardson and Green (1997), we use a fully Bayesian analysis, based on the Reversible Jump (RJ) algorithm, developed in Green (1995), which allows for the change of dimension of the parameter space, changing the number of states from one iteration to the other.

The paper is organized as follows: details of the prior modeling are given in Section 2; Section 3 describes Bayesian estimation of the parameters and model selection.

An application is in progress. We consider the well-known studies about exchange rate dynamics (Otranto and Gallo, 2002; Engel and Hamilton, 1990) in order to compare our results with those obtained in the past.

2 The model

Let $y = (y_t)_{t=1}^T$ be the observed changes in the exchange rate. We assume that the marginal distribution for an observation y_t is

$$y_t|\pi, \mu, \sigma \sim \sum_{i=1}^{k} \pi_i \phi(\cdot; \mu_i, \sigma_i^2) \qquad \text{for } t = 1, 2, \ldots, T, \tag{1}$$

conditional on weights $\pi = (\pi_i)_{i=1}^k$, means $\mu = (\mu_i)_{i=1}^k$ and standard deviations $\sigma = (\sigma_i)_{i=1}^k$, where $\phi(\cdot; \mu_i, \sigma_i^2)$ is the density of the $N(\mu_i, \sigma_i^2)$. We further postulate the existence of an unobserved variable, denoted $s = (s_t)_{t=1}^T$, that takes on values from 1 to k. This variable characterizes the "state" or "regime" of the process at any time t: if $s_t = i$, y_t is assumed to be drawn from a $N(\mu_i, \sigma_i^2)$ and the trend in the exchange rate is given by μ_i. We then assume a Markov chain for the evolution of the state variable and, thus, the process for s_t is presumed to depend on the past realizations of y and s only through s_{t-1}:

$$p(s_t = j|s_{t-1} = i) = \lambda_{ij} \tag{2}$$

The vector of weights π is simply the stationary vector of the transition matrix $\Lambda = (\lambda_{ij})$ and thus satisfies $\pi'\Lambda = \pi'$. Thus the model in (1) can be analogously expressed as

$$y_t|s, \mu, \sigma \sim \phi(\cdot; \mu_{s_t}, \sigma_{s_t}^2). \tag{3}$$

Integrating out s_t in (3), using its stationary distribution, leads back to (1). Finally, we assume that the number of components (regimes) k is unknown and subject to inference. In a Bayesian context, we also assume that:

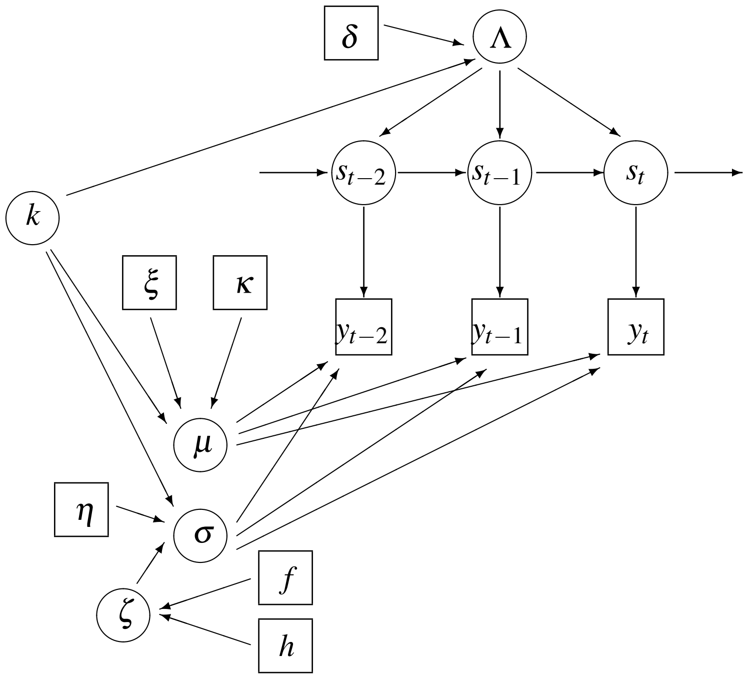
- the number of components k is *a priori* uniform on the values $\{1, 2, \ldots, K\}$;
- the rows of the transition matrix have a Dirichlet distribution: $\lambda_{ij} \sim D(\delta_j)$, for $i = 1, \ldots, k$ where $\delta_j = (\delta_{ij})_{i=1}^k$;
- μ_i and σ_i^{-2} are drawn independently, with priors $\mu_i \sim N(\xi, \kappa^{-1})$ and $\sigma_i^{-2} \sim \text{Ga}(\eta, \zeta)$, the latter parametrised so that the mean and the variance are η/ζ and η/ζ^2;
- ζ follows a Gamma distribution with parameters f and h.

These settings generate the hierarchical model presented in Figure 1.

3 Bayesian estimation and model selection

The complexity of the mixture model presented in this paper requires Markov chain Monte Carlo (MCMC) methods to approximate the posterior joint distribution of all the parameters. Details of these computational methods can be found in Tierney (1994).
To generate realizations from the posterior joint distribution, we alternate the following moves at each sweep of our MCMC algorithm: **a)** updating the transition matrix Λ, **b)** updating the state variables s, **c)** updating the means μ, **d)** updating the standard deviations σ, **e)** updating the hyperparameter ζ, **f)** updating the number of regimes k.
The first five moves are fairly standard and all performed through Gibbs sampling. In particular: (a) and (b) follow Robert *et al.* (1993). In (a), the i-th row of Λ is sampled from a Dirichlet $D(\delta_{i1} + n_{i1}, \ldots, \delta_{ik} + n_{ik})$, where $n_{ij} = \sum_{t=1}^{T-1} I\{s_t = i, s_{t+1} = j\}$ is the number of jumps from regime i to regime j and $I\{\cdot\}$ denotes the indicator function.

Figure 1: *Directed acyclic graph for the complete hierarchical model.*



In (b), $s_1, \ldots, s_T$ are sampled one at a time from $t = 1$ to $t = T$, with conditional probabilities $p(s_t = i|\cdots) \propto \lambda_{s_{t-1}i}\phi(y_t; \mu_i, \sigma_i^2)\lambda_{is_{t+1}}$ where '$\cdots$' denotes 'all the other variables'; for $t = 1$ the first factor is replaced by π_i and, for $t = T$, the last factor is replaced by 1.
Moves (c)-(e) follow Richardson and Green (1997) and, for identifiability purpose, we also adopt a unique labeling in which the μ_i are in increasing numerical order. As a consequence the joint prior distribution of μ_i is $k!$ times the product of the individual normal densities, restricted to the set $\mu_1 < \mu_2 < \ldots < \mu_k$. The μ_i can be drawn independently from $\mu_i|\cdots \sim N\left((\sigma_i^{-2}\sum_{t:s_t=i} y_t + \kappa\xi)/(\sigma_i^{-2}n_i + \kappa), (\sigma_i^{-2}n_i + \kappa)^{-1}\right)$, where $n_i = \#\{t : s_t = i\}$ is the number of observations currently allocated to the i-th regime. In order to preserve the constraints on the μ_i, the move is accepted provided that the order is unchanged.
In (d) we update each component of the vector σ^2 independently, drawing it from its full conditional $\sigma_i^{-2}|\cdots \sim \text{Ga}\left(\eta + n_i/2, \zeta + \sum_{t:s_t=i}(y_t - \mu_i)^2/2\right)$.
In (e) we sample ζ from its full conditional: $\zeta|\cdots \sim \text{Ga}\left(f + k\eta, h + \sum_{i=1}^{k}\sigma_i^{-2}\right)$.
Updating the value of k implies a change of dimensionality for the components μ and σ, the state variables s and the transition probability matrix Λ. We follow the approach used by Richardson and Green (1997) consisting in a random choice between splitting an existing state into two, and merging two existing states into one. The probabilities of these alternatives are b_k and $d_k = 1 - b_k$, respectively, when there are currently k states. Of course, $d_1 = 0$ and $b_K = 0$, and otherwise we choose $b_k = d_k = 0.5$, for $k = 2, 3, \ldots, K-1$.
For the *combine proposal* we randomly choose a pair of states (i_1, i_2) that are adjacent in terms of the current value of their means. These two states are merged into a new one, labeled $i^\star$, reducing k by 1. We then reallocate all those y_t with $s_t = i_1$ or $s_t = i_2$ to the new state $i^\star$ and create values for $\mu_{i^\star}$, $\sigma_{i^\star}^2$, $\pi_{i^\star}$ and for the transition probabilities from and to the states involved in the move in such a way that the new hidden Markov model and the old one both have the same first and second moments.
The *split proposal* starts by choosing a state $i^\star$ at random, which is then split into two new ones labeled i_1 and i_2, augmenting k by 1. Then we reallocate all those y_t with

$s_t = i^\star$ between the two new states, and create values for π_{i_1}, π_{i_2}, μ_{i_1}, μ_{i_2}, σ_{i_1}, σ_{i_2} and the transition probabilities from and to the states involved in the move. The aim is to split $i^\star$ in such a way that the dynamics of the hidden Markov chain are essentially preserved. We accomplish this generating appropriate vectors in the same manner proposed by Robert *et al.* (2000), with some straightforward adjustments (Castellano and Scaccia, 2007).
According to the RJ framework, the split and combine move are accepted with a probability computed to preserve the *reversibility* between the states of the MCMC algorithm.
After allowing for a burn-in period, the algorithm provides us with draws from the joint posterior distribution of $(\Lambda, \mu, \sigma^2, \zeta, k)$. From this, we can estimate the posterior probability of k as the ratio between the number of times the algorithm visited the model with k regimes and the total number of sweeps.
We can also verify the existence of long swings in the exchange rates, as hypothesized by Engel and Hamilton (1990), by testing $k = 2$ against $k = 1$ using the Bayes factor $B_{21} = \frac{p(k=2|y)}{p(k=1|y)} / \frac{p(k=2)}{p(k=1)}$ where $p(k = c)$ and $p(k = c|y)$ are respectively the prior and the posterior probabilities of the model with c regimes. The larger B_{21}, the greater the evidence provided by the data in favour of the model with 2 regimes.
Conditioning on a certain number of regimes, from the MCMC output we can easily make inferences on any other parameter of the model. However, if the purpose is to produce out-of-sample forecasts of the exchange rate, rather than conditioning on a particular model, we may prefer averaging over all the possible models, taking then into account our uncertainty about the true generating model. This is straightforward from the MCMC output; see Castellano and Scaccia (2007) for further details.

References

Castellano R. and Scaccia L. (2007) Bayesian hidden Markov models with an unknown number of regimes, Technical report, Università di Macerata.

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

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

Otranto E. and Gallo G. (2002) A non parametric bayesian approach to detect the number of regimes in Markov switching models, *Econometric Reviews*, 21, 477–496.

Richardson S. and Green P.J. (1997) On bayesian analysis of mixtures with an unknown number of components (with discussion), *Journal of the Royal Statistical Society, B*, 59, 731–792.

Robert C.P., Celeux G. and Diebolt J. (1993) Bayesian estimation of hidden Markov chains: a stochastic implementation, *Statistics and Probability letters*, 16, 77–83.

Robert C.P., Rydén T. and Titterington D.M. (2000) Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method, *Journal of the Royal Statistical Society, B*, 62, 57–75.

Rossi A. and Gallo G. (2006) Volatility estimation via hidden Markov models, *Journal of Empirical Finance*, 13, 203–230.

Rydén T., Teräsvirta T. and Åsbrink S. (1998) Stylized facts of daily return series and the hidden Markov model, *Journal of Applied Econometrics*, 13, 217–244.

Tierney L. (1994) Markov chains for exploring posterior distributions, *Annals of Statistics*, 22, 1701–1764.