ELSEVIER

Journal of Econometrics 93 (1999) 229–255

JOURNAL OF Econometrics

www.elsevier.nl/locate/econbase

Bayesian estimation of switching ARMA models

M. Billio[a,b], A. Monfort[a,*], C.P. Robert[a,c]

[a]*CREST, INSEE, Timbre J301, Bâtiment Malakoff 2, 15 Bd Gabriel Péri, 92245 Malakoff Cedex, Paris, France*

[b]*University Ca' Foscari of Venice, Venice, Italy*

[c]*UPRES-A CNRS 6085, Université de ROUEN, France*

Received 1 January 1997; accepted 1 October 1998

Abstract

Switching ARMA processes have recently appeared as an efficient modelling to nonlinear time-series models, because they can represent multiple or heterogeneous dynamics through simple components. The levels of dependence between the observations are double: at a first level, the parameters of the model are selected by a Markovian procedure. At a second level, the next observation is generated according to a standard time-series model. When the model involves a moving average structure, the complexity of the resulting likelihood function is such that simulation techniques, like those proposed by Shephard (1994, Biometrika 81, 115–131) and Billio and Monfort (1998, Journal of Statistical Planning and Inference 68, 65–103), are necessary to derive an inference on the parameters of the model. We propose in this paper a Bayesian approach with a non-informative prior distribution developed in Mengersen and Robert (1996, Bayesian Statistics 5. Oxford University Press, Oxford, pp. 255–276) and Robert and Titterington (1998, Statistics and Computing 8(2), 145–158) in the setup of mixtures of distributions and hidden Markov models, respectively. The computation of the Bayes estimates relies on MCMC techniques which iteratively simulate missing states, innovations and parameters until convergence. The performances of the method are illustrated on several simulated examples. This work also extends the papers by Chib and Greenberg (1994, Journal of Econometrics 64, 183–206) and Chib (1996, Journal of Econometrics 75(1),

* Corresponding author. Tel.: + 331-4117-6081; fax: + 331-4117-6029.
E-mail address: monfort@ensae.fr (A. Monfort)

0304-4076/99/$ - see front matter
PII: S0304-4076(99)00010-X

79–97) which deal with ARMA and hidden Markov models, respectively.

JEL classification: C11

Keywords: Markov model; Moving average model; Hidden Markov chain; MCMC algorithm; ARMA model; Convergence control; Prior feedback

1. Introduction

In a fruitful generalization of ARMA(p, q) models proposed by Hamilton (1988), the parameters of these models are specified as a stochastic process driven by a state (unobservable or *hidden*) Markov chain. The likelihood functions of these switching ARMA(p, q) models can be recursively computed (see Kitagawa, 1987 or Hamilton, 1989) when the joint process made of the variables of interest and of the state process is Markovian, i.e. if q is equal to 0. When a non-trivial MA structure is present in the model, recursive algorithms fail because of the combinatorial explosion resulting from the computation. For similar reasons, Bayesian inference is also delicate in such setups. Additional difficulties stem from the restrictions imposed by stationarity issues, although Jones (1987) and Barnett et al. (1996) have shown that a reparameterization could greatly simplify this issue. Since the hindrance is mainly due to the non-Markovian nature of the model, a natural idea is to recover a Markovian framework by adding additional states which lead to a switching linear state space representation. In this setting, it is possible to propose simulated EM algorithms (Shephard, 1994) or importance sampling algorithms based on the partial Kalman filter (Gouriéroux and Monfort, 1996; Billio and Monfort, 1998).

In this paper, we follow a Bayesian approach based on the data augmentation principle of Tanner and Wong (1987) (see also Albert and Chib, 1993; Chib and Greenberg, 1994 and Jaquier et al., 1994). However, in order to minimize the size of the state variable, we do not attempt in this paper to get back in the Markov framework and we work directly with the initial model. We use two versions of Monte Carlo Markov Chain (MCMC) algorithms, which thus involve mixtures of Gibbs and Metropolis–Hastings algorithms. In both versions, the data augmentation steps are identical but the simulations of some parameters are different: a first version proposes Metropolis–Hastings steps based on approximated innovations while the other uses explicit Gibbs steps inspired from Chib and Greenberg (1994) which are based on polynomial long divisions.

The prior modelling on the parameters of the different models is non-informative, following some developments of Mengersen and Robert (1996) and Robert and Titterington (1998) on mixtures of distributions and hidden Markov

chains. The structure of these switching state models is indeed quite similar to mixture setups, with the additional difficulty of probabilistic dependence at both the observed and latent stages. The prior we select is a direct extension of the non-informative prior of Mengersen and Robert (1996), which relates the different means of the model through normal distributions and allows for improper priors on the variances, thus bypassing the requirement for proper conjugate priors imposed by Chib and Greenberg (1994).

The paper is organized as follows: Section 2 presents the setup, i.e. the general switching ARMA model and the corresponding prior modelling. Section 3 describes the general algorithms and Section 4 presents applications of these algorithms to various examples based on simulated series.

2. The general framework

2.1. The switching ARMA model

We consider the following general switching ARMA model:

$$y_t = \gamma_{s_t} + \sum_{i=1}^{p} \varphi_{i,s_t}(y_{t-i} - \gamma_{s_{t-i}}) + \sigma_{s_t}\varepsilon_t - \sum_{j=1}^{q} \theta_{j,s_t}\sigma_{s_{t-j}}\varepsilon_{t-j}, \quad t \geqslant 1, \tag{2.1}$$

where the ε_t's are independent $\mathcal{N}(0,1)$ and the s_t's constitute a two-state Markov chain starting from s_{1-p}, generated from the invariant distribution. The s_t's are independent from the ε_t's; we also assume that $y_0, \ldots, y_{1-p}$ are fixed and $\varepsilon_0 = \varepsilon_{-1} = \cdots = \varepsilon_{1-q} = 0$. We use the notations $\pi_{ij} = P(s_t = i | s_t = j)$ $(i, j = 0, 1)$. With lag-polynomial notations, the model can be written

$$\Phi_{s_t}(L)(y_t - \gamma_{s_t}) = \Theta_{s_t}(L)(\sigma_{s_t}\varepsilon_t), \tag{2.2}$$

where L is the lag operator and

$$\Phi_k(L) = 1 - \varphi_{1,k}L - \cdots - \varphi_{p,k}L^p, \quad k = 0, 1,$$

$$\Theta_k(L) = 1 - \theta_{1,k}L - \cdots - \theta_{q,k}L^q, \quad k = 0, 1.$$

We may also want to impose *stationarity* constraints within each regime, i.e. that the roots of the polynomials $\Phi_k(L)$ $(k = 0, 1)$ are outside the unit circle and *invertibility* constraints, i.e. that the roots of the polynomials $\Theta_k(L)$ $(k = 0, 1)$ are outside the unit circle.

While this model is fairly general, we could further generalize it in at least two directions. First, we could assume that the Markov chain has more than two

states and, second, we could assume that exogenous variables appear in the r.h.s. of Eq. (2.1). These generalizations could easily be included in our framework, the only drawback being in the use of more complex notations and, therefore, we do not consider them explicitly in this paper.

2.2. Prior modelling

For identifiability reasons, we assume that $\sigma_1 < \sigma_0$. The prior distribution on the parameter $\xi_1 = (\gamma_0, \gamma_1, \sigma_0, \sigma_1, \pi_{00}, \pi_{11})$ is chosen to be proportional to

$$f_1(\xi_1) = \sigma_0^{-3} \exp\{-(\gamma_0 - \gamma_1)^2 \zeta/\sigma_0^2\} \mathbb{I}_{\sigma_0 > \sigma_1} \mathbb{I}_{[0,1]}(\pi_{00}) \mathbb{I}_{[0,1]}(\pi_{11}). \tag{2.3}$$

In other words, σ_0 has a diffuse prior proportional to $1/\sigma_0$; conditionally on σ_0, σ_1 is uniform on $[0, \sigma_0]$; conditionally on (σ_0, σ_1), γ_0 has a flat prior and, conditionally on $(\sigma_0, \sigma_1, \gamma_0)$, γ_1 is normal $\mathcal{N}(\gamma_0, \sigma_0^2/\zeta)$ and π_{00} and π_{11} are independent uniforms on $[0, 1]$. This representation marginally allows for a flat prior on the γ_i's, while linking both regimes of Eq. (2.1). It is not fully non-informative in the sense that the precision hyperparameter ζ needs to be specified but it appears from simulation experiments that the range of values of ζ leading to the same estimates is wide enough to insure stability of the method when $\zeta \in [0.01, 1]$. Note also that the prior is invariant under affine transformations. Robert and Titterington (1998) show that this prior leads to a well-defined posterior distribution in the setup of mixtures and hidden Markov models.

Extensions to larger numbers of latent classes is straightforward, in the sense that the successive differences $(\gamma_i - \gamma_{i-1})$ are modeled as $\mathcal{N}(0, \sigma_{i-1}^2/\zeta)$ variables. (The identifying constraint is then $\sigma_1 > \sigma_2 \ldots$, as in Mengersen and Robert, 1996.)

If we now consider the prior on the $(\varphi_{i,k}, \theta_{j,k})$'s $(i = 1, \ldots, p, j = 1, \ldots, q, k = 0, 1)$, the constraints imposed by both stationarity and invertibility conditions imply a convoluted domain for the $\varphi_{i,k}$'s $(i = 1, \ldots, p)$. However, following Barnett et al. (1996), there exists a simple re-parameterization of this domain as

$$\varphi_{i,k} = \varphi_i^*(r_{1k}, \ldots, r_{pk}), \quad i = 1, \ldots, p, \quad k = 0, 1,$$

where $(r_{1k}, \ldots, r_{pk})$ is variation free in $[-1; 1]^p$ – these parameters can be interpreted as partial autocorrelations in the autoregressive case – and where $\varphi_i^*(r_1, \ldots, r_p)$ is linear in every r_j, when the other r_ℓ's $(\ell \neq j)$ are fixed. Therefore, we simply put an uniform prior on the r_{ik}'s, $(i = 1, \ldots, p, k = 0,1)$. In a similar way, we can reparameterize the $\theta_{j,k}$'s, $(j = 1, \ldots, q, k = 0, 1)$, by using two vectors $(\rho_{1k}, \ldots, \rho_{qk})$, in $[-1, 1]^q$, and impose uniform priors on these new parameters.

If we denote by ξ_2 the set of the $\{r_{ik},\ \rho_{jk}\}$'s, and by $f_1(\xi_1)$ the prior on ξ_1, the overall prior on ξ is thus proportional to

$$f(\xi) = f_1(\xi_1)\prod_{k=0}^{1}\ \{\mathbb{I}_{[-1,1]^p}(r_{1k},\ldots,r_{pk})\mathbb{I}_{[-1,1]^q}(\rho_{1k},\ldots,\rho_{qk})\}. \tag{2.4}$$

The joint probability density function (p.d.f.) of $[(y_t)_{t=1,\ldots,T},\ (s_t)_{t=1-p,\ldots,T},\ \xi]$ is therefore proportional to

$$\prod_{t=1}^{T}\exp\Bigg\{-\frac{1}{2\sigma_{s_t}^2}[y_t-\gamma_{s_t}-\varphi_{1,s_t}(y_{t-1}-\gamma_{s_{t-1}})-\cdots-\varphi_{p,s_t}(y_{t-p}-\gamma_{s_{t-p}})$$

$$+\theta_{1,s_t}\eta_{t-1}+\cdots+\theta_{q,s_t}\eta_{t-q}]^2\Bigg\}\sigma_{s_t}^{-1}$$

$$\prod_{t=2-p}^{T}\pi_{s_{t-1},s_t}\times g(s_{1-p})f(\xi), \tag{2.5}$$

where $g(\cdot)$ is the invariant p.d.f. of the Markov chain and where the $\eta_t(=\sigma_{s_t}\varepsilon_t)$ are recursively defined by

$$\eta_t = y_t-\gamma_{s_t}-\sum_{i=1}^{p}\varphi_{i,s_t}(y_{t-i}-\gamma_{s_{t-i}})+\sum_{j=1}^{q}\theta_{j,s_t}\eta_{t-j},\quad t\geqslant 1, \tag{2.6}$$

and $\eta_0=\eta_{-1}=\cdots=\eta_{1-q}=0$. As shown by Eq. (2.6), η_t is a function of $(y_{1-p},\ldots,y_t)$, $(s_{1-p},\ldots,s_t)$, γ_0, γ_1, (r_{jk}) and (ρ_{jk}), but not of σ_0, σ_1, π_{00} and π_{11}.

3. MCMC implementation

3.1. Simulating from the posterior distribution

The posterior density $f(\xi|y_1,\ldots,y_T)$ and the posterior expectation of ξ cannot be computed in a closed form. This complex setting thus requires a simulation-based approach, like Gibbs sampling (Gelfand and Smith, 1990), where we simulate from the joint posterior distribution $f(\xi,\ s_{1-p},\ldots,\ s_T|y_1,\ldots,\ y_T)$, to derive the distribution of interest as the marginal distribution of ξ and to approximate the posterior expectation by a sample average.

More precisely, we simulate successively from $f(s_{1-p},\ldots,\ s_T|\xi,\ y_1,\ldots,y_T)$ – this is the *data augmentation* step – and from $f(\xi^{(i)}|\xi^{(-i)},\ y_1,\ldots,y_T)$, $i=1,\ldots,2p+2q+6$, where $\xi^{(i)}$ is a component of ξ and $\xi^{(-i)}$ the set of the other components – this is the parameter simulation step. When conditional distributions cannot be directly simulated, as in the dataGaugmentation step, the corresponding steps are replaced by (single) Metropolis–Hastings steps. Such algorithms

are validated in Tierney (1994) (see also Gilks et al., 1996; Robert, 1996). These references provide detailed information about Metropolis–Hastings algorithms, which allow for MCMC simulation from arbitrarily complex distributions.

Although Chib and Greenberg (1994) have already developed an MCMC algorithm for non-switching MA and ARMA models which can be extended to this setup (see Section 3.4), we first propose a new approach which, comparatively, involves more Metropolis–Hastings steps and less Gibbs steps. According to the current tenets on the relative virtues of Gibbs and Metropolis–Hastings methods (see, e.g., Gilks et al., 1996; Liu, 1995; Robert, 1996, or Tierney, 1994), this approach is more robust, in particular because it increases the range of the chain excursions on the posterior surface, as shown below on simulations. (Note also that Chib and Greenberg (1994), do not condition on the first p observations.)

Switching ARMA models are quite complex, in particular because of the missing data structure, namely the non-observability of the latent states. As in mixture estimation (see e.g. Mengersen and Robert, 1996), once these states are known, the inferential problem gets simpler. The additional complexity in this specific case is that the likelihood is still involved when the s_t's are known, because of the moving average structure. The main ideas of the algorithm are the following: In the data augmentation (or state simulation) steps, candidate Metropolis–Hastings distributions are obtained by ignoring the influence of some future states in the full conditional distribution of each state. In the simulation of the parameters, candidate distributions are obtained by replacing the true innovations with the so-called 'pseudo-innovations' computed from the results of previous steps. Besides being theoretically well-grounded, these devices greatly simplify the simulation and do not exhibit a negative influence on the convergence of the algorithm.

In the following algorithms, the initial values for the parameter ξ and the latent states (s_t) are chosen at random, the location and scale for mean and variance being determined from the data.

3.2. The data augmentation step

The joint conditional distribution of the unobservable states $s_{1-p}, \ldots, s_T$ is

$$f(s_{1-p}, \ldots, s_T | y_1, \ldots, y_T, \xi) \propto \prod_{t=2-p}^{T} \pi_{s_{t-1}, s_t} g(s_{1-p}) \times \prod_{t=1}^{T} \exp\left\{ -\frac{1}{2\sigma_{s_t}^2} \left[y_t - \gamma_{s_t} - \sum_{i=1}^{p} \varphi_{i,s_t} (y_{t-i} - \gamma_{s_{t-i}}) + \sum_{j=1}^{q} \theta_{j,s_t} \eta_{t-j} \right]^2 \right\} \sigma_{s_t}^{-1}, \tag{3.1}$$

where the η_t's are defined in Eq. (2.6).

A direct Gibbs approach requires simulations from the various conditional distributions $f(s_t|s_{-t}, y_1, \ldots, y_T, \xi)$. However, s_t implicitly occurs in each term of the second product in Eq. (3.1) with index larger than t, since the η_τ's are functions of $s_\tau, s_{\tau-1}, \ldots, s_{1-p}$. Note that this complication is due to the MA structure, since, in the pure AR case, s_t occurs in (at most) $p+1$ terms for any t. Note also that s_t occurs in two terms of the first product of Eq. (3.1), that is, those corresponding to the indices t and $t+1$ $(1-p < t < T)$. (This feature is clearly related with the Markov structure of (s_t).)

Although the computation of the normalizing constant remains feasible, since the s_t's only take two possible values, it is time-consuming and this difficulty calls for the use of a Metropolis–Hastings step and the candidate distribution we use is obtained by simply ignoring the terms in the second product of Eq. (3.1) with index larger than t; note that this method could be generalized to higher order dependencies, namely to cases where we ignore the η_τ's for $\tau > t+h$ and a given $h \geqslant 0$.

We also adopt a global Metropolis–Hastings approach, in which we accept or reject globally the whole sequence $(s_{1-p}, \ldots, s_T)$. While this grouped approach may seem inferior to a component-wise acceptance of the s_t's, there are theoretical arguments (Liu et al., 1995; Müller, 1992; Shephard and Pitt, 1997) in its favor, related to both wider moves within the support of the joint distribution and smaller variance. Besides, the acceptance rate remains quite high, while the computation of the Metropolis–Hastings acceptance probability is much more involved in the component-wise case. (Nonetheless, in cases where the acceptance rate gets too low, a possible solution is to run instead the component-wise version.) Thus, if $(s_t^{(m)})$ denotes the vector of the simulated states and $\xi^{(m)}$ the vector of the simulated parameters after m iterations of the algorithm, the simulation of $(s_t^{(m+1)})$ proceeds as follows:

1. Set $\eta_0^{(m+1)} = \cdots = \eta_{1-q}^{(m+1)} = 0$ and generate $s_{1-p}^{(m+1)}$ from the invariant distribution of the chain associated with $(\pi_{00}^{(m)}, \pi_{11}^{(m)})$.
2. For $t = 2-p, \ldots, 0$, generate $s_t^{(m+1)}$ from

$$p(s_t) \propto \pi^{(m)}_{s_{t-1}^{(m+1)}, s_t} \pi^{(m)}_{s_t, s_{t+1}^{(m)}}.$$

3. For $t = 1, \ldots, T-1$, generate $s_t^{(m+1)}$ from

$$p(s_t) \propto \exp\left\{ -\left[y_t - \gamma_{s_t}^{(m)} - \sum_{i=1}^{p} \varphi_{i,s_t}^{(m)} (y_{t-i} - \gamma_{s_{t-i}^{(m+1)}}^{(m)}) + \sum_{j=1}^{q} \theta_{j,s_t}^{(m)} \eta_{t-j}^{(m+1)} \right]^2 \Big/ 2(\sigma_{s_t}^{(m)})^2 \right\} \times (\sigma_{s_t}^{(m)})^{-1} \pi^{(m)}_{s_{t-1}^{(m+1)}, s_t} \pi^{(m)}_{s_t, s_{t+1}^{(m)}}$$

and compute

$$\eta_t^{(m+1)} = y_t - \gamma_{s_t}^{(m)(m+1)} - \sum_{i=1}^{p} \varphi_{i,s_t}^{(m)(m+1)}(y_{t-i} - \gamma_{s_{t-i}}^{(m)(m+1)}) + \sum_{j=1}^{q} \theta_{j,s_t}^{(m)(m+1)}\eta_{t-j}^{(m+1)}.$$

4. For $t = T$, generate $s_T^{(m+1)}$ from

$$p(s_T) \propto \exp\left\{ -\left[y_t - \gamma_{s_T}^{(m)} - \sum_{i=1}^{p} \varphi_{i,s_T}^{(m)}(y_{T-i} - \gamma_{s_{T-i}}^{(m)(m+1)}) + \sum_{j=1}^{q} \theta_{j,s_T}^{(m)}\eta_{T-j}^{(m+1)} \right]^2 \Big/ 2(\sigma_{s_T}^{(m)})^2 \right\} (\sigma_{s_T}^{(m)})^{-1}\pi_{s_{T-1}^{(m+1)},s_T}^{(m)}.$$

Note that the path $(s_t^{(m)})$ resulting from the previous iteration occurs in the candidate p.d.f. only through the transition probability from s_t to $s_{t+1}^{(m)}$. The proposed path $(s_t^{(m+1)}, t = 1-p, \ldots, T)$ is accepted as the new value of (s_t) with probability $\min(\omega_s, 1)$, where ω_s is given in the Appendix.

3.3. *Simulation of* $(\pi_{00}, \pi_{11}, \sigma_0, \sigma_1)$

As noted above, the parameters $\pi_{00}, \pi_{11}, \sigma_0, \sigma_1$ do not appear in η_t expressed as a function of $(y_{1-p}, \ldots, y_t)$, $(s_{1-p}, \ldots, s_t)$ and the parameter ξ. Therefore, the full conditional distribution of any of these parameters, i.e. its conditional distribution given the other parameters and (s_t), (y_t), is potentially easier to compute and simulate, and genuine Gibbs steps can be used in this setup, as shown below.

The full conditional distribution of π_{00} is proportional to the product of the terms in Eq. (2.5) involving π_{00}, i.e. it is proportional to

$$\mathbb{I}_{[0,1]}(\pi_{00})\pi_{00}^{n_{00}}(1-\pi_{00})^{n_{01}},$$

where $(i, j = 0, 1)$

$$n_{ij} = \sum_{t=2-p}^{T} \mathbb{I}_{\{s_t^{(m+1)}=j\}}\mathbb{I}_{\{s_{t-1}^{(m+1)}=i\}}.$$

To be completely exact, the conditional distribution of π_{00} also involves the term corresponding to s_{1-p} through the stationary distribution, that is

$$g(s_{1-p}) = \frac{\pi_{01}^{s_{1-p}}\pi_{10}^{1-s_{1-p}}}{\pi_{01}+\pi_{10}}.$$

A natural way to take this term into account is to evaluate its influence through an additional Metropolis–Hastings step. However, the corresponding acceptance probability is

$$\left(\frac{\pi'_{01}}{\pi_{01}^{(m)}}\right)^{s_{1-p}^{(m)}}\left(\frac{\pi'_{10}}{\pi_{10}^{(m)}}\right)^{s_{1-p}^{(m)}}\frac{\pi_{10}^{(m)}+\pi_{01}^{(m)}}{\pi'_{10}+\pi'_{01}},$$

which is close to 1 since the weights π_{ij} do not change on a large scale from one iteration to the other.

Therefore, the (almost) complete conditional distribution of π_{00} is the beta distribution $\mathscr{B}e(n_{00}+1, n_{01}+1)$. Similarly the complete conditional distribution of π_{11} is $\mathscr{B}e(n_{11}+1, n_{10}+1)$ (up to the term involving $s_{1-p}^{(m+1)}$).

For the simulation of the parameters (σ_0, σ_1), consider

$$\tilde{y}_t^{(m)} = y_t - \gamma_{s_t^{(m+1)}}^{(m)} - \sum_{i=1}^{p} \varphi_{i,s_t^{(m+1)}}^{(m)}(y_{t-1} - \gamma_{s_{t-1}^{(m+1)}}^{(m)}) + \sum_{j=1}^{q} \theta_{j,s_t^{(m+1)}}^{(m)}\eta_{t-j}^{(m)}, \quad t \geqslant 1,$$

where the $\eta_{t-j}^{(m)}$ are computed from (y_t), $(s_t^{(m+1)})$, $\gamma_0^{(m)}$, $\gamma_1^{(m)}$, $(\varphi_{i,k})$, and $(\theta_{j,k})$. From Eq. (2.5), we see that the complete conditional p.d.f. of (σ_0, σ_1) is proportional to

$$\mathbb{I}_{(\sigma_0>\sigma_1)}\sigma_0^{-n_0^{(m+1)}-3}\exp\left\{-\frac{1}{2\sigma_0^2}\left(\sum_{t=1}^{T}\mathbb{I}_{\{s_t^{(m+1)}=0\}}(\tilde{y}_t^{(m)})^2 + \zeta(\gamma_0^{(m)}-\gamma_1^{(m)})^2\right)\right\}$$

$$\times\,\sigma_1^{-n_1^{(m+1)}}\exp\left\{-\frac{1}{2\sigma_1^2}\sum_{t=1}^{T}\mathbb{I}_{\{s_t^{(m+1)}=1\}}(\tilde{y}_t^{(m)})^2\right\},$$

where

$$n_0^{(m+1)} = \sum_{t=1}^{T}\mathbb{I}_{\{s_t^{(m+1)}=0\}}, \quad n_1^{(m+1)} = \sum_{t=1}^{T}\mathbb{I}_{\{s_t^{(m+1)}=1\}}.$$

Therefore, the complete conditional distribution of σ_0^{-2} is the truncated gamma

$$\mathbb{I}_{\{\sigma_0>\sigma_1^{(m)}\}}\mathscr{G}a\left[\frac{n_0^{(m+1)}+1}{2}, \frac{1}{2}\left(\sum_{t=1}^{T}\mathbb{I}_{\{s_t^{(m+1)}=0\}}(\tilde{y}_t^{(m)})^2 + \zeta(\gamma_0^{(m)}-\gamma_1^{(m)})^2\right)\right].$$

Similarly, the complete conditional distribution of σ_1^{-2} is the truncated gamma

$$\mathbb{I}_{\{\sigma_0^{(m+1)} > \sigma_1\}} \mathcal{G}a\left[\frac{n_1^{(m+1)} - 1}{2}, \frac{1}{2}\sum_{t=1}^{T} \mathbb{I}_{\{s_t^{(m+1)}=1\}}(\tilde{y}_t^{(m)})^2\right],$$

which can be simulated with the optimal accept–reject algorithm of Philippe (1997).

3.4. Simulation of (γ_0, γ_1)

The complete conditional distribution of (γ_0, γ_1) is proportional to

$$\prod_{t=1}^{T} \exp\left\{-\frac{1}{2(\sigma_{s_t^{(m+1)}}^{(m+1)})^2}\left[y_t - \gamma_{s_t^{(m+1)}} - \sum_{i=1}^{p} \varphi_{i,s_t^{(m+1)}}^{(m)}(y_{t-i} - \gamma_{s_{t-i}^{(m+1)}}) + \sum_{j=1}^{q} \theta_{j,s_t^{(m+1)}}^{(m)} \tilde{\eta}_{t-j}^{(m)}\right]^2 - \frac{(\gamma_0 - \gamma_1)^2 \zeta}{2(\sigma_0^{(m+1)})^2}\right\}, \tag{3.2}$$

where the $\tilde{\eta}_{t-j}^{(m)}$ are functions of (y_t), $(s_t^{(m+1)})$, (γ_0, γ_1), $(\varphi_{i,k}^{(m)})$, and $(\theta_{j,k}^{(m)})$, as in Eq. (2.6).

This distribution is quite involved since (γ_0, γ_1) appears in the $\tilde{\eta}_{t-j}^{(m)}$'s. We thus consider two possibilities: the first one uses a Metropolis–Hastings step based on pseudo-innovations, the second one, which is similar to that used by Chib and Greenberg (1994), is based on the exact distribution obtained from a long division.

3.4.1. A first version based on the pseudo-innovations

Since the difficulty stems from the innovation terms $\tilde{\eta}_{t-j}^{(m)}$, a natural idea is to use a Metropolis–Hastings step where the candidate distribution is obtained from a modification of Eq. (3.2) where the $\tilde{\eta}_{t-j}^{(m)}(\gamma_0, \gamma_1)$ have been replaced by the pseudo-innovations $\tilde{\eta}_{t-j}^{(m)}(\gamma_0^{(m)}, \gamma_1^{(m)})$. This means that we use the recursive formulae in Eq. (2.6) with the previous value of (γ_0, γ_1) and the current value of the other parameters and of (s_t). Once this replacement is made, the remaining function of (γ_0, γ_1) is the exponential of a quadratic form. Therefore, the two full conditional candidate distributions are Gaussian and it is a straightforward exercise to compute the mean and the variance of these distributions.

3.4.2. A second version based on a long division

Eq. (2.6) can be written

$$\Theta_{s_t}(L)\eta_t = \Phi_{s_t}(L)(y_t - \gamma_{s_t}).$$

A long division of 1 by $\Theta_{s_t}(L)$ gives

$$R_t(L)\Theta_{s_t}(L) = 1 + L^t Q_t(L),$$

where $R_t(L)$ is a polynomial of degree $t-1$. Therefore, given the nullity of the initial values of η_t, we get

$$\eta_t = R_t(L)\Phi_{s_t}(L)(y_t - \gamma_{s_t}).$$

The complete conditional distribution of (γ_0, γ_1), given in Eq. (3.2), can then be rewritten as being proportional to

$$\prod_{t=1}^{T} \exp\left\{ -\frac{1}{2(\sigma_{s_t^{(m+1)}}^{(m+1)})^2}[R_t^{(m)}\Phi_{s_t^{(m+1)}}^{(m)}(L)(y_t - \gamma_{s_t^{(m+1)}})]^2 - \frac{(\gamma_0-\gamma_1)^2\zeta}{2(\sigma_0^{(m+1)})^2}\right\}, \tag{3.3}$$

which is clearly the exponential transform of a quadratic form in (γ_0, γ_1). Therefore, the two full conditional distributions of γ_0 and γ_1 are Gaussian and their simulation is straightforward, although the computation of the moments is rather tedious. This provides us with two exact Gibbs steps.

3.5. Simulation of $(\varphi_{i,k})$

The complete conditional distribution of $(\varphi_{i,k})$, or equivalently of $(r_{i,k})$ defined in Section 2.2, is proportional to

$$\prod_{t=1}^{T} \exp\left\{ -\frac{1}{2(\sigma_{s_t^{(m+1)}}^{(m+1)})^2}\left[y_t - \gamma_{s_t^{(m+1)}}^{(m+1)} - \sum_{i=1}^{p} \varphi_{i,s_t^{(m+1)}}(y_i - \gamma_{s_{t-i}^{(m+1)}}^{(m+1)}) \right.\right.$$

$$\left.\left. + \sum_{j=1}^{q} \theta_{j,s_t^{(m+1)}}^{(m)} \eta_{t-j}^{*(m)} \right]^2 \right\} \prod_{k=0}^{1} \mathbb{I}_{[-1,1]^p}(r_{1k}, \ldots, r_{pk}), \tag{3.4}$$

where the $\varphi_{i,k}$'s are expressed in terms of the $r_{i,k}$'s and where the $\eta_{t-j}^{*(m)}$'s are functions of (y_t), $(s_t^{(m+1)})$, $(\gamma_0^{(m+1)}, \gamma_1^{(m+1)})$, $(\varphi_{i,k})$, and $(\theta_{j,k}^{(m)})$. Since the η_{t-j}^{*}'s depend on the $\varphi_{i,k}$'s, we again consider two different procedures.

3.5.1. A first version based on the pseudo-innovations

If we replace the $\eta_{t-j}^{*(m)}(\varphi_{i,s_t^{(m+1)}})$'s, by the pseudo-innovations $\eta_{t-j}^{*(m)}(\varphi_{i,s_t^{(m+1)}}^{(m)})$, as in Section 3.4.1, and if we recall the multilinearity of the functions $\varphi_{i,k} = \varphi_i^*(r_{1k}, \ldots, r_{pk})$, it follows that the candidate full conditional distributions of each r_{ik} is Gaussian restricted to $[-1, 1]$.

3.5.2. A second version based on a long division

Using the same device as in Section 3.4.2, the joint complete conditional distribution of the r_{ik}'s can be written as being proportional to

$$\prod_{t=1}^{T} \exp\left\{ -\frac{1}{2(\sigma_{s_t^{(m+1)}}^{(m+1)})^2}[R_t^{(m)}(L)\Phi_{s_t^{(m+1)}}(L)(y_t - \gamma_{s_t^{(m+1)}}^{(m+1)})]^2 \right\}$$

$$\times \prod_{k=0}^{1} \mathbb{I}_{[-1,1]^p}(r_{1k}, \ldots, r_{pk}). \tag{3.5}$$

Since $\Phi_{s_t^{(m+1)}}(L)$ is linear in each r_{ik}, the univariate complete conditional distribution of each r_{ik} is Gaussian truncated to $[-1, 1]$. While the computation of the moments of the Gaussian distribution is more complicated than in Section 3.5.1, it is thus formally possible to run exact Gibbs steps.

3.6. Simulation of $(\theta_{j,k})$

The complete conditional distribution of the $\theta_{j,k}$'s, or equivalently of the ρ_{jk}'s defined in Section 2.2, is proportional to

$$\prod_{t=1}^{T} \exp\left\{ -\frac{1}{2(\sigma_{s_t^{(m+1)}}^{(m+1)})^2}\left[y_t - \gamma_{s_t^{(m+1)}}^{(m+1)} - \sum_{i=1}^{p} \varphi_{i,s_t^{(m+1)}}^{(m+1)}(y_i - \gamma_{s_{t-i}^{(m+1)}}^{(m+1)}) \right.\right.$$

$$\left.\left. + \sum_{j=1}^{q} \theta_{j,s_t^{(m+1)}} \bar{\eta}_{t-j}^{(m)} \right]^2 \right\} \prod_{k=0}^{1} \mathbb{I}_{[-1,1]^q}(\rho_{1k}, \ldots, \rho_{qk}), \tag{3.6}$$

where the $\theta_{i,k}$'s are expressed in terms of the ρ_{ik}'s and where the $\bar{\eta}_{t-j}^{(m)}$'s are functions of (y_t), $(s_t^{(m+1)})$, $(\gamma_0^{(m+1)}, \gamma_1^{(m+1)})$, $(\varphi_{i,k}^{(m+1)})$ and $(\theta_{j,k})$ as in Eq. (2.6).

The first version of the algorithm described in Section 3.5.1 can be immediately transposed to the $\theta_{j,k}$'s, using the pseudo-innovations $\bar{\eta}_{t-j}^{(m)}(\theta_{j,s_t^{(m+1)}}^{(m)})$. On the contrary, the second version no longer works since, while the joint complete conditional distribution of the ρ_{jk}'s can be written

$$\prod_{t=1}^{T} \exp\left\{ -\frac{1}{2(\sigma_{s_t^{(m+1)}}^{(m+1)})^2}[R_t(L)\Phi_{s_t^{(m+1)}}^{(m+1)}(L)(y_t - \gamma_{s_t^{(m+1)}}^{(m+1)})^2] \right\}$$

$$\times \prod_{k=0}^{1} \mathbb{I}_{[-1,1]}(\rho_{1k}, \ldots, \rho_{qk}),$$

each ρ_{jk} appears in $R_t(L)$ in a nonlinear way and, therefore, normality is lost.

3.7. Implementation details

The above algorithms are started with random assignments of the observations to the latent state, that is random generation of the s_t's in $\{0, 1\}$, and random generation of the parameters on a scale derived from the sample, that is with normal and uniform distributions based on the sample mean, sample variance and sample correlation. The data augmentation step is then run first to correct for the purely random assignments of the observations.

We also introduce variance factors at the different Metropolis–Hastings stages, namely for the simulation of (γ_0, γ_1), $(\varphi_{i,k})$ and $(\theta_{j,k})$, in order to keep reasonable acceptance rates, that is acceptance rates far enough from both 0 and 1. Apart from (s_t), where the acceptance rate is usually high, these acceptance rates are between 0.3 and 0.7, which is the range advocated by Gelman et al. (1996). We observed that the choice of these variance factors has strong influence, not only on the acceptance rates in the simulation of the above parameters but also in the simulation of the s_t's.

Although no attempt is made at assessing convergence through stopping rules, as in Cowles and Carlin (1996), we run a preliminary batch of 1000 iterations and start registering the values of the averages for each parameter till graphical stabilizing. We also included assignment maps, as in Robert (1997), described below, to assess the mixing rate on the s_t's.

4. Applications

4.1. Switching MA models

The simplest case where the Markovian property is lost and, therefore, Hamilton's (1989) algorithm fails is the MA(1) case. More precisely, we consider the following switching MA(1) model:

$$y_t = \gamma_{s_t} + \sigma_{s_t}\varepsilon_t - \theta\sigma_{s_{t-1}}\varepsilon_{t-1}, \quad t \geqslant 1, \tag{4.1}$$

where the ε_t's ($t \geqslant 1$) are iid $\mathcal{N}(0, 1)$, independent of the s_t's, $\varepsilon_0 = 0$ and s_1 is generated from the invariant distribution of the Markov chain.

The different parameters of these models are described in Table 1 (first line), along with the Bayes posterior expectations obtained by the first algorithm (second line). The second algorithm leads to the same numerical results, except in Example 3, where convergence is not achieved after 50,000 iterations. Since we are in a simulation setup, we can also compare the estimated versus the true assignments. In all cases, the correct assignment rate is around 75%, except for Example 3 where it gets down to 55%. The treatment of Example 3 is more delicate because of the proximity of both components.

Table 1
Parameters of the different benchmarks (first row of each pair) used to test the MCMC method and the corresponding Bayes estimates (second row) approximated with 10,000 iterations of the first algorithm for the MA(1) model

Example	γ_0	γ_1	σ_0^2	σ_1^2	θ	π_{00}	π_{11}	n
1	0.0	0.0	1.59	0.109	− 0.85	0.86	0.61	500
	− 0.03	− 0.005	1.50	0.112	− 0.85	0.90	0.56	
2	0.0	0.2	1.35	0.185	0.55	0.26	0.11	2500
	0.01	0.19	1.27	0.206	0.52	0.27	0.12	
3	0.0	0.2	1.35	0.53	− 0.65	0.76	0.81	1000
	− 0.04	0.27	1.46	0.71	− 0.65	0.74	0.88	

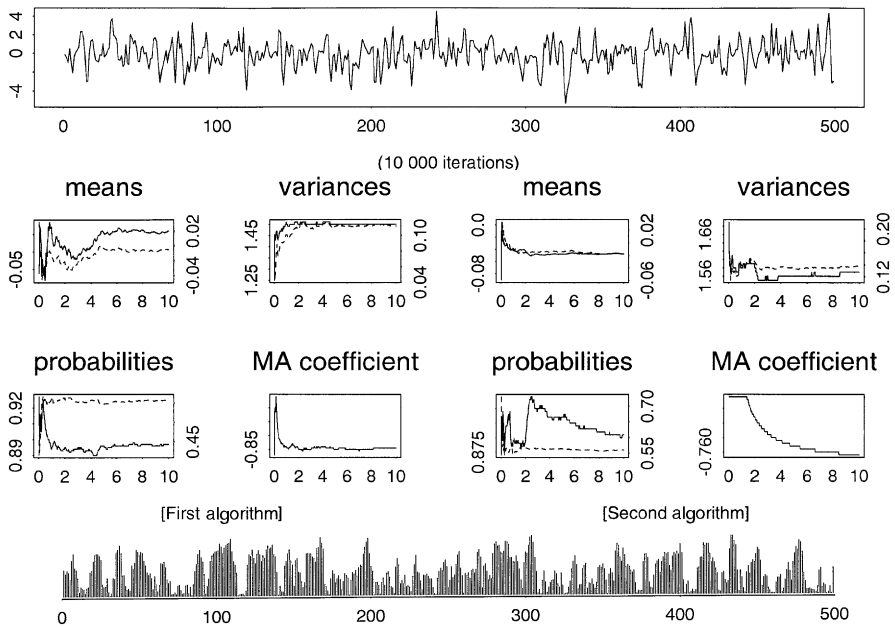

Fig. 1. Performances of both algorithms for Example 1 of Table 1. The full lines in the graphs of the cumulated averages of the different parameters correspond to the first component ($s_t = 0$) and the dotted lines to the second component ($s_t = 1$). The simulated dataset is provided at the top of the figure and the estimated probability of being in state 1 for each observation is given at the bottom.

As shown by the convergence graphs in Figs. 1, 3 and 4, the first algorithm provides a faster convergence rate than the second one since the different averages reach a stationary value more rapidly in the first case (note that the scales are smaller for most of the graphs on the left-hand side.) In these graphs,

the original sequence (y_t) is represented at the top, the convergence of each average being given for both algorithms. Note also, despite the different number of iterations, the lack of stability of the second algorithm for Example 3. Figs. 2 and 5 give another evaluation of the convergence properties of the first algorithm, by plotting the assignments of the different observations against the iterations. This device was introduced by Robert (1997) to evaluate the convergence of MCMC algorithms in mixture estimation. As can be seen from Fig. 5, the assignment of most observations remain quite stable along iterations.

We also considered an extension of the MA(1) model (4.1), namely the MA(2) model

$$y_t = \gamma_{s_t} + \sigma_{s_t}\varepsilon_t - \theta_1\sigma_{s_{t-1}}\varepsilon_{t-1} - \theta_2\sigma_{s_{t-2}}\varepsilon_{t-2}, \quad t \geqslant 1,$$

with the same assumptions on the ε_t's. The parameters chosen for the simulation are given in Table 2. Convergence occurs quite rapidly for the different parameters of the model, as shown by Fig. 6, while a stable assignment of the observations can be achieved by a proper scaling of the variance factors in the

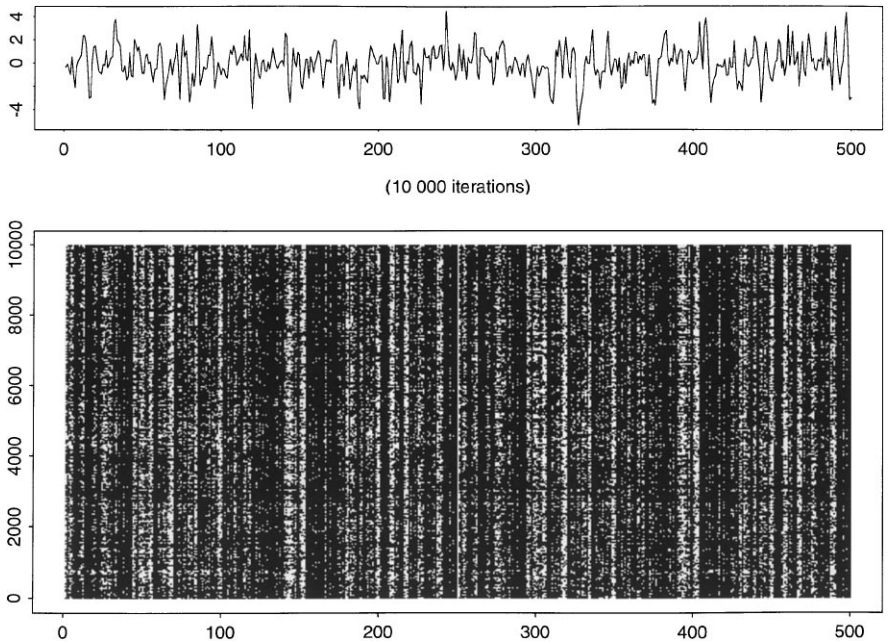

Fig. 2. Sample and assignment map for Example 1 of Table 1. The values of $s_t^{(m)}$ are represented for each t,m by a white dot if $s_t^{(m)} = 0$ and by a black dot otherwise. (The estimated probability at the bottom of Fig. 1 is the average per column of the assignment map.)

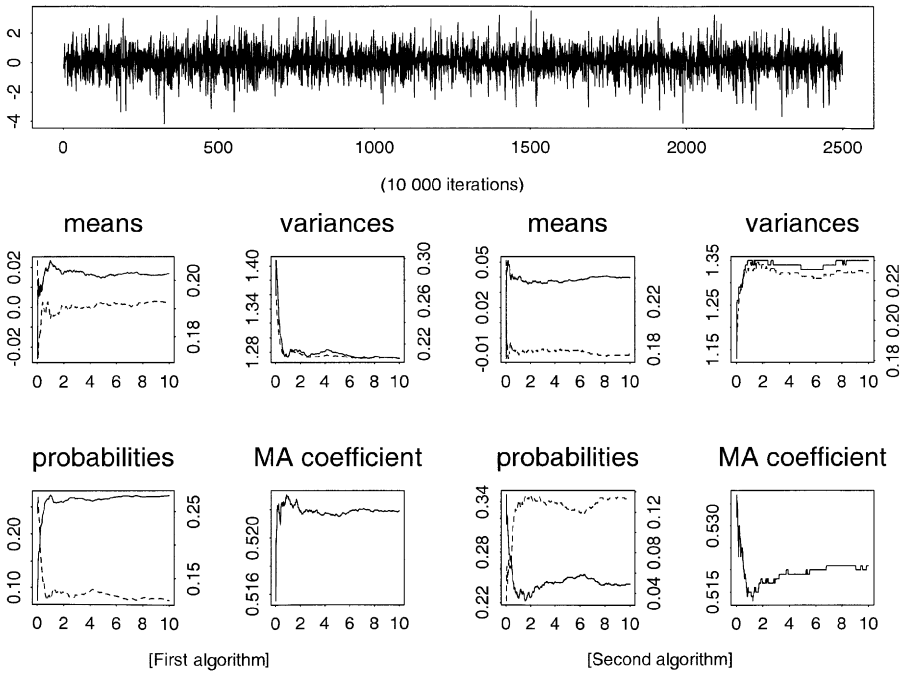

Fig. 3. Performances of both algorithms for Example 2 of Table 1.

Metropolis–Hastings stages. (See Fig. 7.) The correct assignment rates are 80% for Example 1 and 85% for Example 2.

4.2. *Switching ARMA (1, 1) models: A first type*

We first consider the following switching ARMA(1, 1) model:

$$y_t = \gamma_{s_t} + \varphi(y_{t-1} - \gamma_{s_{t-1}}) + \sigma_{s_t}\varepsilon_t - \theta\sigma_{s_{t-1}}\varepsilon_{t-1}, \quad t \geqslant 1, \tag{4.2}$$

where y_0 is fixed, $\varepsilon_0 = 0$ and s_0 is generated from the invariant distribution of the Markov chain.

As for the switching MA models, we selected various sets of parameters presented in Table 3, to study the behavior of the MCMC algorithm and of the Bayes estimates (posterior expectations). The two first examples show a quite satisfactory result since the averages stabilize reasonably quickly (see Figs. 8 and 9) and the corresponding Bayes estimates are close to the true parameters. The assignment maps also show a high degree of stability (see Fig. 10) and the correct assignment rate is, again, close to 75%.

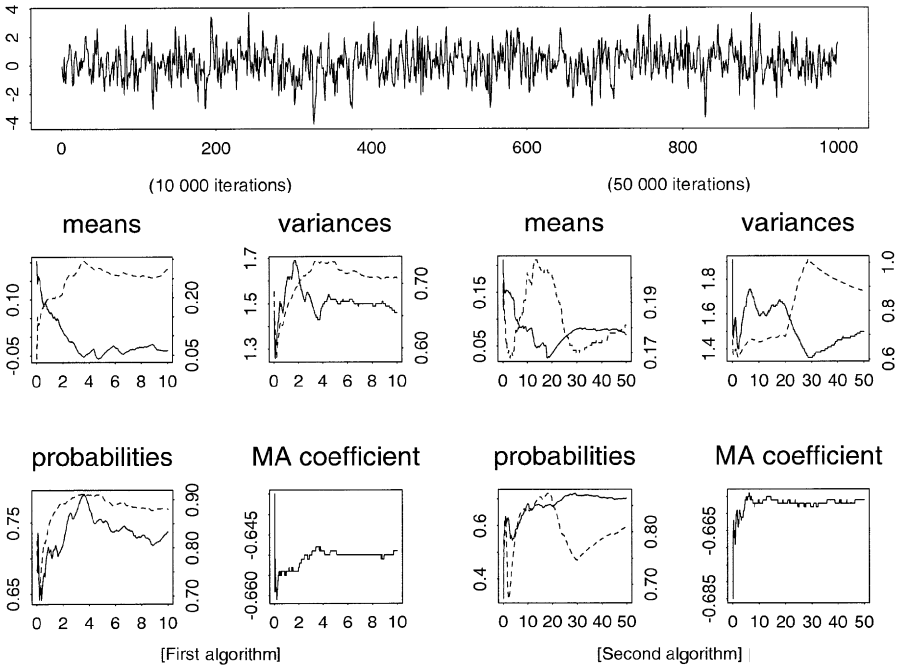

Fig. 4. Performances of both algorithms for Example 3.

The last example was chosen for its pathological features, since the equality $\varphi = \theta$ prevents identifiability for both parameters. As shown by Fig. 11, the result is equally satisfactory since the other parameters are correctly estimated, while the averages for φ and θ are equal for most of the simulation but take longer to stabilize. These parameters are not identifiable and we observe convergence to 0 for both θ and φ, which is due to the choice of symmetric priors on these parameters. (Note also that the other parameters are extremely well separated and equally well estimated.) A similar although more moderate phenomenon occurs in Figs. 8 and 9, where the graphs for the AR and MA coefficients have very close profiles (but correspond to different scales).

4.3. *Switching ARMA (1, 1) models: A second type*

We now consider the following switching ARMA(1, 1) model:

$$y_t = \mu + \varphi_{s_t}(y_{t-1} - \mu) + \sigma\varepsilon_t - \theta_{s_t}\sigma\varepsilon_{t-1}, \quad t \geqslant 1, \tag{4.4}$$

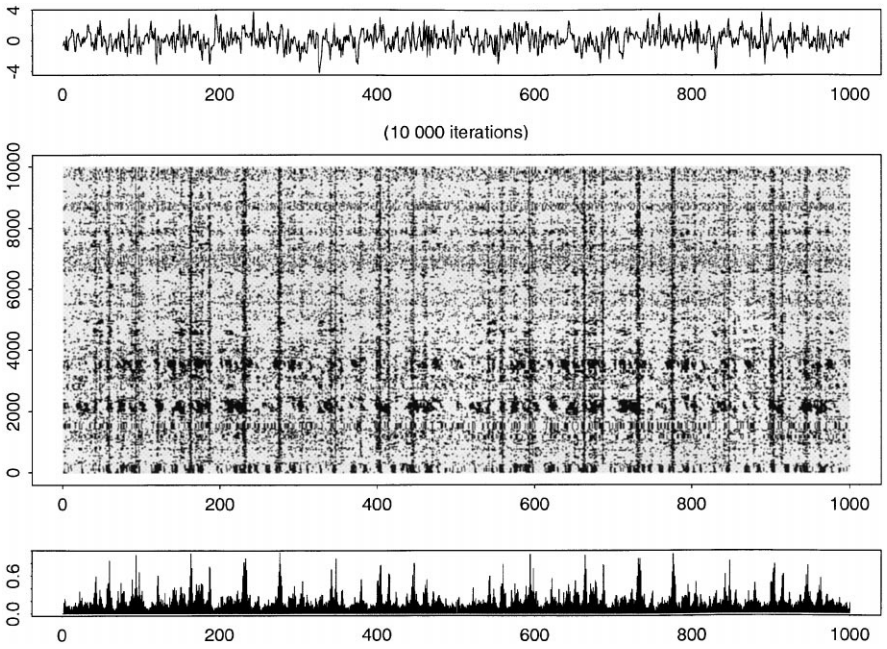

Fig. 5. Sample, assignment map and average assignment for Example 3 of Table 1.

Table 2
Parameters of the different benchmarks (first row of each pair) used to test the MCMC method and corresponding Bayes estimates (second row) for the MA(2) model

Example	γ_0	γ_1	σ_0^2	σ_1^2	θ_1	θ_2	π_{00}	π_{11}	n
1	1.4	1.3	1.69	0.16	0.3	−0.5	0.42	0.37	200
	1.33	1.27	1.56	0.227	0.259	−0.567	0.434	0.417	
2	1.8	1.3	1.44	0.64	−0.1	0.5	0.52	0.87	250
	1.57	1.23	1.53	0.487	−0.111	0.491	0.501	0.633	

As in Section 4.2, the performances of the algorithm are satisfactory, convergence being reached in a small number of iterations (from 5000 to 10,000 iterations). This second type of switching ARMA models being more involved, the estimation of the corresponding parameters is more delicate and requires larger sample sizes to ensure a close fit for a given sample. As can be seen in Table 4, which describes the results of three simulations, both location and scale parameters are well approximated in all cases, which

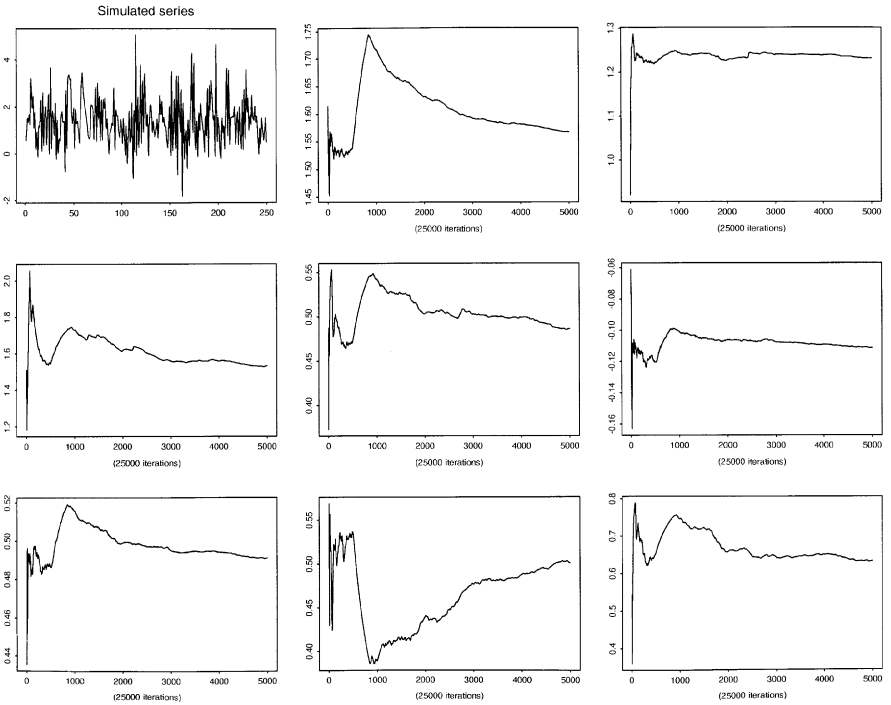

Fig. 6. Performances of the algorithm for Example 2 of Table 2. The true parameter values are given in Table 2. The order of the plotted parameters is γ_0, γ_1, σ_0^2, σ_1^2, θ_1, θ_0, π_{00}, π_{11}).

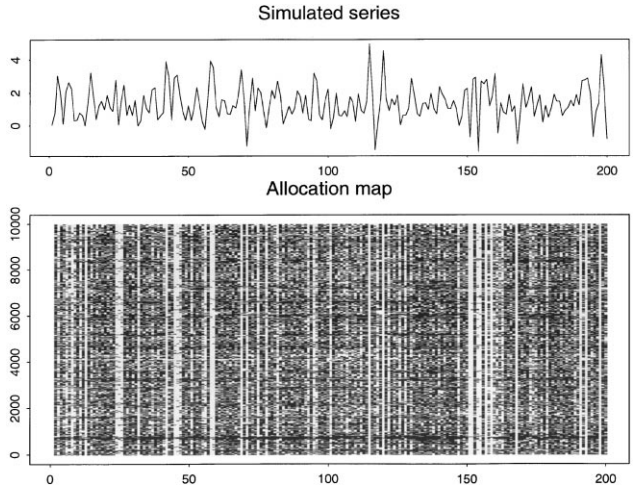

Fig. 7. Sample and assignment map for Example 2 of Table 2. The true parameter values are given in Table 2.

Table 3
Parameters of the different benchmarks (first row of each pair) used to test the MCMC method and corresponding Bayes estimates (second row)

Example	γ_0	γ_1	σ_0^2	σ_1^2	φ	θ	π_{00}	π_{11}	n
1	0.94	− 0.12	0.122	0.102	0.32	0.17	0.26	0.31	250
	0.88	− 0.11	0.147	0.132	0.22	0.06	0.28	0.31	
2	0.0	0.2	1.35	0.185	− 0.55	− 0.17	0.16	0.61	250
	0.07	0.16	1.86	0.165	− 0.52	− 0.15	0.15	0.62	
3	2.94	− 1.72	0.562	0.102	0.27	0.27	0.74	0.69	150
	2.97	− 1.75	0.454	0.112	− 0.01	− 0.02	0.73	0.71	

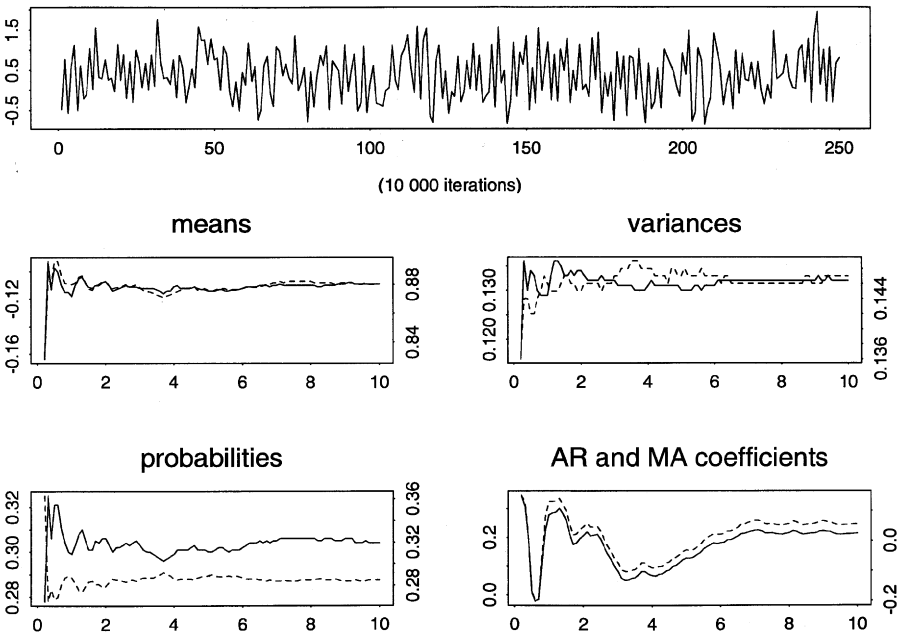

Fig. 8. Performances of the algorithm for Example 1 of Table 3. The true parameter values are given in Table 3. The full lines correspond to state 0 of the latent Markov process and the dotted lines to state 1, except for the last figure, where the full graph is for the AR coefficient and the dotted graph for the MA coefficient.

is not always the case for the AR and MA parameters or the transition probabilities (see Figs. 12–15). This can be attributed to the weak identifiability of these models where an interplay between the AR and MA coefficients does not deeply perturbate the predictive properties of the model. The

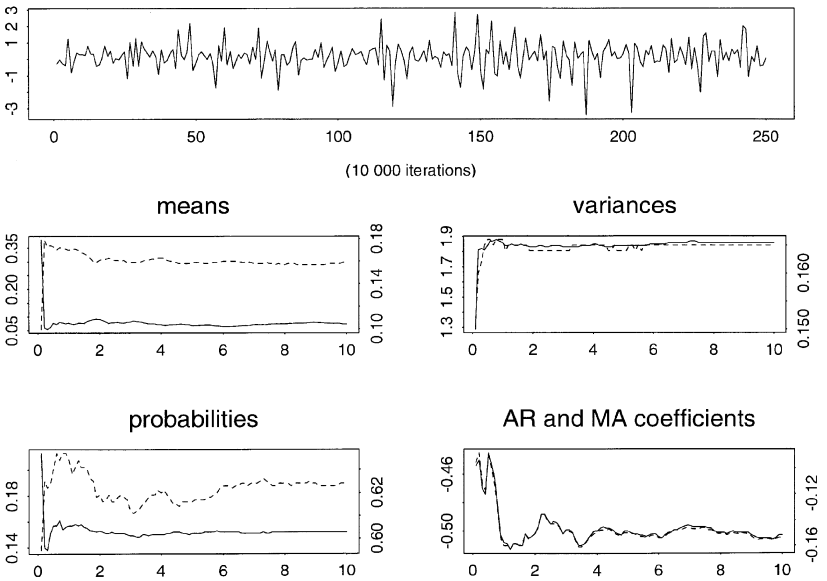

Fig. 9. Performances of the algorithm for Example 2 of Table 3.

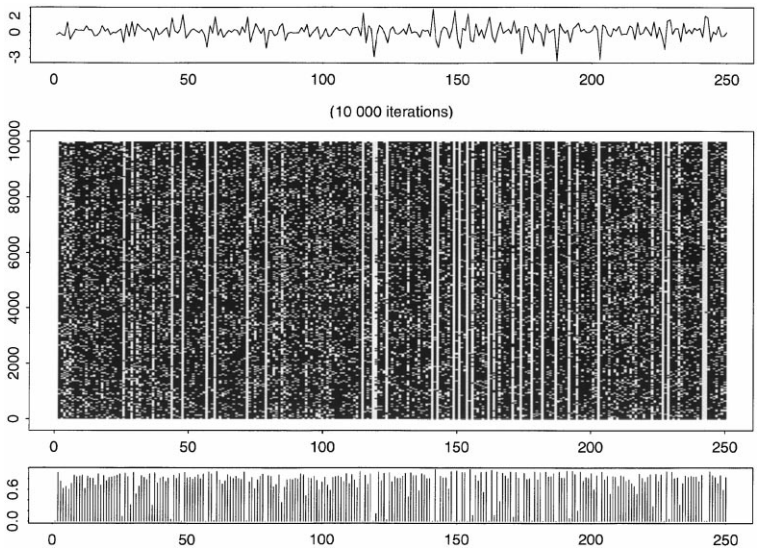

Fig. 10. Sample, assignment map and assignment average for Example 2 of Table 3. The values of $s_t^{(m)}$ are represented for each t, m by a white dot if $s_t^{(m)} = 0$ and by a black dot otherwise.

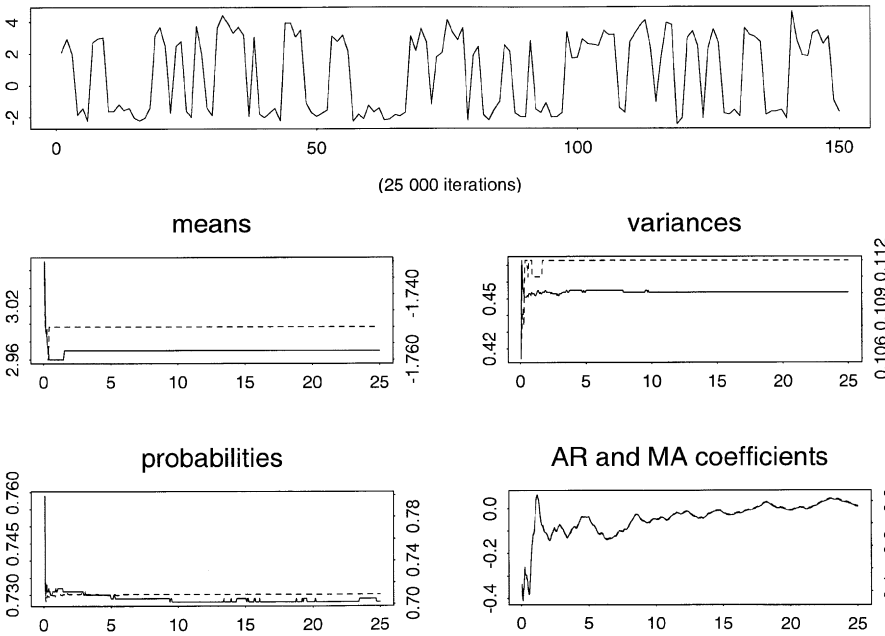

Fig. 11. Performances of the algorithm for Example 4 of Table 3.

Table 4
Parameters of the different benchmarks (first row of each pair) used to test the MCMC method and corresponding Bayes estimates (second row)

Example	μ	σ^2	φ_0	φ_1	θ_0	θ_1	π_{00}	π_{11}	n
1	0.00	1.00	0.02	0.95	− 0.77	− 0.07	0.79	0.94	500
	− 0.18	1.03	0.098	0.934	− 0.543	− 0.154	0.82	0.94	
2	0.00	1.00	0.32	− 0.57	0.02	0.25	0.51	0.26	5000
	0.01	1.02	0.30	− 0.66	− 0.02	0.08	0.49	0.25	
3	0.00	1.00	0.33	− 0.26	− 0.81	0.79	0.50	0.50	2500
	− 0.00	1.28	0.46	− 0.34	− 0.40	0.40	0.46	0.48	

correct assignment rate remains close to 75%, while the assignment map in Fig. 13 shows clearcut separations between different zones of the sample. Note in particular the occurrence of nearly constant assignment of observations to state $s_t = 0$ or $s_t = 1$ in the areas of important changes in the sample graph.

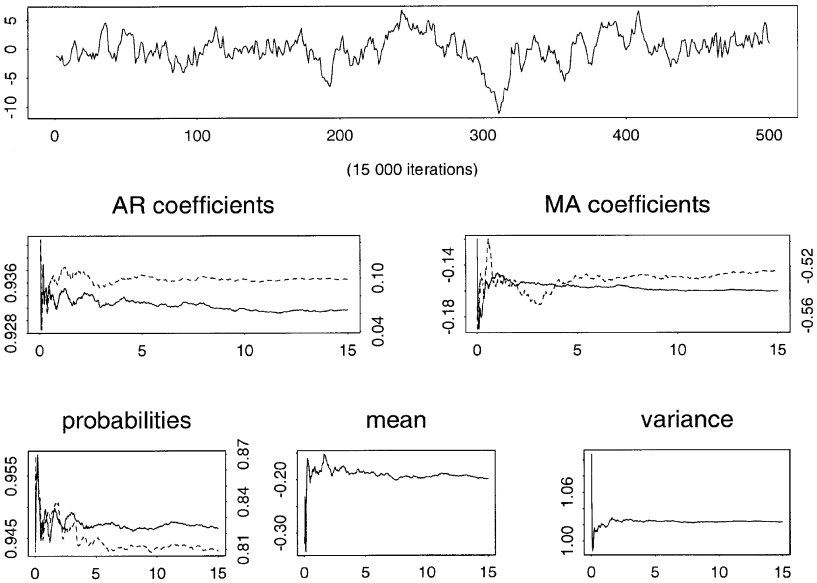

Fig. 12. Performances of the algorithm for Example 1 of Table 4. The true values are given in Table 4. Dotted graphs correspond to latent state 0.

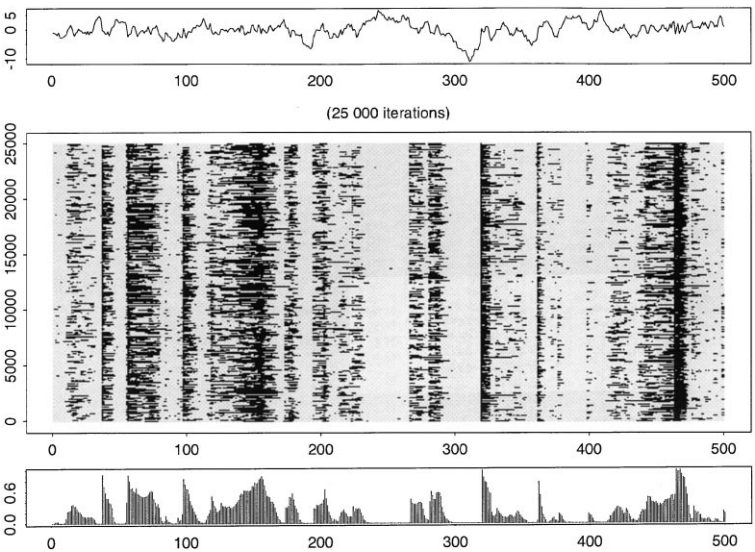

Fig. 13. Sample, assignment map and assignment average for Example 1 of Table 4. The values of $s_t^{(m)}$ are represented for each t, m by a white dot if $s_t^{(m)} = 0$ and by a black dot otherwise.

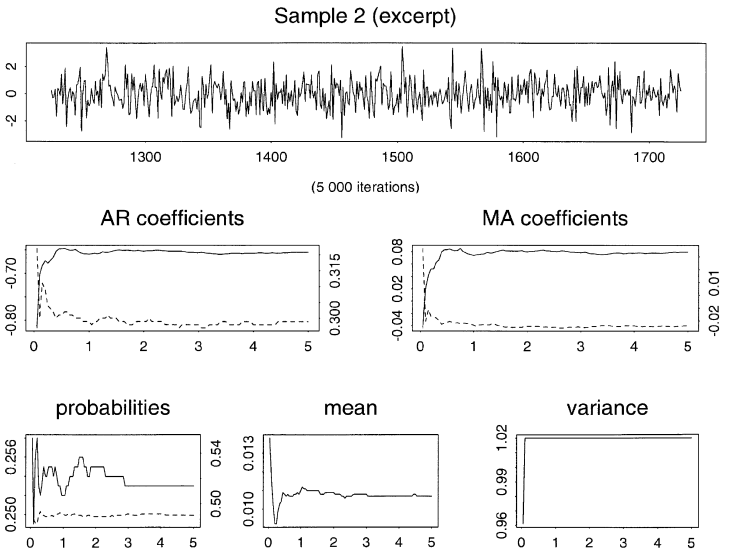

Fig. 14. Excerpt from the sample and performances of the algorithm for Example 2 of Table 4.

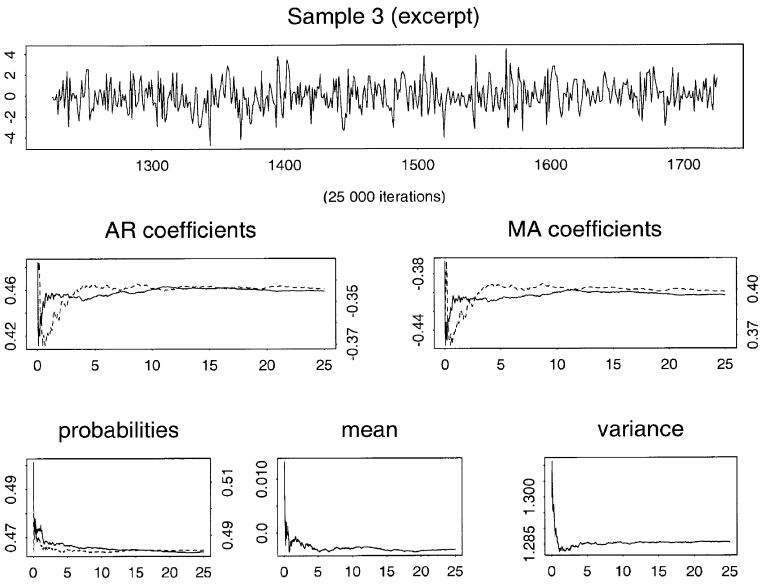

Fig. 15. Excerpt from the sample and performances of the algorithm for Example 3 of Table 4.

4.4. Conclusion

The above simulations have shown that Bayesian inference in different switching ARMA models is feasible and give satisfactory results, both in terms of estimation of the parameters, since the estimations obtained in Tables 1–4 are close to the true values of the parameters, and in terms of assignment of the observations. While the MCMC implementation of this inference is involved and is only of many possible alternatives, we observed, in Section 4.1, superior performances than those of an approach based on long divisions. Moreover, the fundamental strategy of using pseudo-innovations for the construction of proposal distributions can easily be transposed to other settings like ARCH models and stochastic volatility (see Billio et al., 1998). The assignment maps of Robert (1997) have also shown their relevance as convergence assessment tools, although more standard methods can also be used there (see Mengersen et al., 1998). As pointed out in Section 3.7, the choice of the scaling factors in the proposal distributions is quite influential in the performance of the resulting algorithm, in particular when the number of observations is large.

Acknowledgements

We are grateful to the associate editor and to two anonymous referees for their most valuable comments. The third author is thankful to Marilena Barbieri for an helpful discussion.

Appendix: The acceptance probability in the data augmentation step

We use the following notations: The true conditional p.d.f. of a path $(s_t)_{t=1-p,\dots,T}$ given $y=(y_t,\dots,y_T)$ and ξ is denoted by $f[(s_t)|y,\xi]$ and is proportional to the joint p.d.f. given in Eq. (2.5),

$$f[(s_t)|y,\xi] \propto g(s_{1-p}) \prod_{t=2-p}^{T} \pi_{s_{t-1},s_t} \prod_{t=1}^{T} \varphi\left[\frac{1}{\sigma_{s_t}}\left\{y_t-\gamma_{s_t}-\sum_{i=1}^{p}\varphi_{i,s_t}(y_{t-i}-\gamma_{s_{t-i}})\right.\right.$$

$$\left.\left.+\sum_{j=1}^{q}\theta_{j,s_t}\eta_{t-j}(\xi,(s_t))\right\}\right]\sigma_{s_t}^{-1},$$

where φ is the p.d.f. of $N(0,1)$ and $\eta_t(\xi,(s_t))$ is given by the recurrent Eq. (2.6). The candidate distribution at step $(m+1)$ is denoted by $p[(s_t)|(s_t^{(m)}),\xi^{(m)},y]$,

where $p[(s_t)|(s_t^{(m)}), \xi, y]$ is given by (denoting $\varphi_{0,k} = -1$).

$$g(s_{1-p}) \cdot \prod_{t=2-p}^{0} \frac{\pi_{s_{t-1},s_t}\pi_{s_t,s_{t+1}^{(m)}}}{\sum_{k=0}^{1}\pi_{s_{t-1},k}\pi_{k,s_{t+1}^{(m)}}}$$

$$\times \prod_{t=1}^{T-1} \frac{\varphi[\sigma_{s_t}^{-1}\{-\sum_{i=0}^{p}\varphi_{i,s_t}(y_{t-i}-\gamma_{s_{t-i}})+\sum_{j=1}^{q}\theta_{j,s_t}\eta_{t-j}(\xi,(s_t))\}](\pi_{s_{t-1},s_t}\pi_{s_t,s_{t+1}^{(m)}})/\sigma_{s_t}}{\sum_{k=0}^{1}\varphi[\sigma_k^{-1}\{-\sum_{i=0}^{p}\varphi_{i,k}(y_{t-i}-\gamma_{s_{t-i}})+\sum_{j=1}^{q}\theta_{j,k}\eta_{t-j}(\xi,(s_t))\}](\pi_{s_{t-1},k}\pi_{k,s_{t+1}^{(m)}})/\sigma_k}$$

$$\times \frac{\varphi[\sigma_{s_T}^{-1}\{-\sum_{i=0}^{p}\varphi_{i,s_T}(y_{T-i}-\gamma_{s_{T-i}})+\sum_{j=1}^{q}\theta_{j,s_T}\eta_{T-j}(\xi,(s_t))\}](\pi_{s_{T-1},s_T})/\sigma_{s_T}}{\sum_{k=0}^{1}\varphi[\sigma_k^{-1}\{-\sum_{i=1}^{p}\varphi_{i,k}(y_{T-i}-\gamma_{s_{T-i}})+\sum_{j=1}^{q}\theta_{j,k}\eta_{T-j}(\xi,(s_t))\}](\pi_{s_{T-1},k})/\sigma_k}.$$

The acceptance probability is then

$$\min(\omega_s, 1),$$

with

$$\omega_s = \frac{f[(s_t^{(m+1)})|y, \xi^{(m)}]p[(s_t^{(m)})|(s_t^{(m+1)}, \xi^{(m)}, y]}{f[(s_t^{(m)})|y, \xi^{(m)}]p[(s_t^{(m+1)})|(s_t^{(m)}, \xi^{(m)}, y]}.$$

Using the notation ξ instead of $\xi^{(m)}$, for simplicity's sake, ω_s is equal to

$$\omega_s = \prod_{t=2-p}^{0} \frac{\pi_{s_t^{(m)},s_{t+1}^{(m+1)}}}{\sum_{k=0}^{1}\pi_{s_{t-1,k}^{(m)}}\pi_{k,s_{t+1}^{(m+1)}}} \cdot \prod_{t=2-p}^{0} \frac{\sum_{k=0}^{1}\pi_{s_{t-1}^{(m+1)},k}\pi_{k,s_{t+1}^{(m)}}}{\pi_{s_t^{(m+1)},s_{t+1}^{(m)}}}$$

$$\times \prod_{t=1}^{T-1} \frac{\pi_{s_t^{(m)},s_{t+1}^{(m+1)}}}{\sum_{k=0}^{1}\varphi[\sigma_k^{-1}\{-\sum_{i=0}^{p}\varphi_{i,k}(y_{t-i}-\gamma_{s_{t-i}^{(m)}})+\sum_{j=1}^{q}\theta_{j,k}\eta_{t-j}(\xi,(s_t^{(m)}))\}](\pi_{s_{t-1}^{(m)},k}\pi_{k,s_{t+1}^{(m+1)}})/\sigma_k}$$

$$\times \prod_{t=1}^{T-1} \frac{\sum_{k=0}^{1}\varphi[\sigma_k^{-1}\{-\sum_{i=0}^{p}\varphi_{i,k}(y_{t-i}-\gamma_{s_{t-i}^{(m+1)}})+\sum_{j=1}^{q}\theta_{j,k}\eta_{t-j}(\xi,(s_t^{(m+1)}))\}](\pi_{s_{t-1}^{(m+1)},k}\pi_{k,s_{t+1}^{(m)}})/\sigma_k}{\pi_{s_t^{(m+1)},s_{t+1}^{(m)}}}$$

$$\times \frac{\sum_{k=0}^{1}\varphi[\sigma_k^{-1}\{-\sum_{i=0}^{p}\varphi_{i,k}(y_{T-i}-\gamma_{s_{T-i}^{(m+1)}})+\sum_{j=1}^{q}\theta_{j,k}\eta_{T-j}(\xi,(s_t^{(m+1)}))\}](\pi_{s_{T-1}^{(m+1)},k})/\sigma_k^{-1}}{\sum_{k=0}^{1}\varphi[\sigma_k^{-1}\{-\sum_{i=0}^{p}\varphi_{i,k}(y_{T-i}-\gamma_{s_{T-i}^{(m)}})+\sum_{j=1}^{q}\theta_{j,k}\eta_{T-j}(\xi,(s_t^{(m)}))\}](\pi_{s_{T-1}^{(m)},k})/\sigma_k}.$$

References

Albert, J., Chib, S., 1993. Bayesian inference for autoregressive time series with mean and variance subject to Markov jumps. Journal of Business Economic Statistics 11, 1–15.

Barnett, G., Kohn, R., Sheather, S., 1996. Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. Journal of Econometrics 74, 237–254.

Billio, M., Monfort, A., 1998. Switching state space models, likelihood function, filtering and smoothing. Journal of Statistical Planning and Inference 68, 65–103.

Billio, M., Monfort, A., Robert, C.P., 1998. The simulated likelihood ratio (SLR) method. Doc. travail CREST no. 9821, Insee, Paris.

Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. Journal of Econometrics 75 (1), 79–97.
Chib, S., Greenberg, E., 1994. Bayes inference for regression models with ARMA(p, q) errors. Journal of Econometrics 64, 183–206.
Cowles, M.K., Carlin, B.P., 1996. Markov Chain Monte-Carlo convergence diagnostics: A comparative study. Journal of the American Statistical Association 91, 883–904.
Gelfand, A.E., Smith, A.F.M., 1990. Sampling based approaches to calculating marginal densities. Journal of the American Statistical Association 85, 398–409.
Gelman, A., Gilks, W.R., Roberts, G.O., 1996. Efficient metropolis jumping rules. In: Berger, J.O., Bernardo, J.M., Dawid, A.P., Lindley, D.V., Smith, A.F.M. (Eds.), Bayesian Statistics 5. Oxford University Press, Oxford, pp. 599–608.
Gilks, W.R., Richardson, S., Spiegelhalter, D.I., 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall, London.
Gouriéroux, C., Monfort, A., 1996. Simulation Based Methods in Econometrics. Oxford University Press, Oxford.
Hamilton, J.D., 1988. Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. Journal of Economic Dynamics and Control 12, 385–423.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.
Jaquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models (with discussion). Journal of Business and Economic Statistics 12, 371–417.
Jones, M.C., 1987. Randomly choosing parameters from the stationarity and invertibility region of autoregressive-moving average models. Applied Statistics (Series C) 38, 134–138.
Kitagawa, G., 1987. Non-Gaussian state-space modelling of nonstationary time series. Journal of the American Statistical Association 82, 1032–1063.
Liu, J.S., 1995. Metropolized Gibbs Sampler: An improvement. Tech. Report, Dept. of Statistics, Stanford U., California.
Liu, J.S., Wong, W.H., Kong, A., 1995. Correlation structure and convergence rate of the Gibbs sampler with various scans. Journal of the Royal Statistical Society (Series B) 57, 157–169.
Mengersen, K.L., Robert, C.P., 1996. Testing for mixtures: a Bayesian entropic approach (with discussion). In: Berger, J.O., Bernardo, J.M., Dawid, A.P., Lindley, D.V., Smith, A.F.M. (Eds.), Bayesian Statistics 5. Oxford University Press, Oxford, pp. 255–276.
Mengersen, K.L., Robert, C.P., Guihenneuc-Jouyaux, C., 1998. MCMC convergence diagnostics: A 'review'. In: Berger, J.O., Bernardo, J.M., Dawid, A.P., Lindley, D.V., Smith, A.F.M. (Eds.), Bayesian Statistics 6. Oxford University Press, Oxford, 1999.
Müller, P., 1992. Alternatives to the Gibbs sampling scheme. Technical Report, Institute of Statistics and Decision Sciences, Duke Uni.
Philippe, A., 1997. Simulation of right- and left-truncated Gamma distributions by mixtures. Statistics and Computing 7, 173–182.
Robert, C.P., 1996. Méthodes de Monte Carlo par Chaînes de Markov. Economica, Paris.
Robert, C.P., 1997. Invited discussion of 'On Bayesian Analysis of Mixtures with an unknown Number of Components' by S. Richardson and P. Green. Journal of the Royal Statistical Society (Series B) 59, 758–764.
Robert, C.P., Titterington, M., 1998. Resampling schemes for hidden Markov models and their application for maximum likelihood estimation. Statistics and Computing 8 (2), 145–158.
Shephard, N., 1994. Partial non-Gaussian state space. Biometrika 81, 115–131.
Tanner, M., Wong, W., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–550.
Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Annals of Statistics 22, 1701–1786.