

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Bayesian Estimation of Multivariate Autoregressive Hidden Markov Model with Application to Breast Cancer Biomarker Modeling

Hamid El Maroufy, El Houcine Hibbah,
Abdelmajid Zyad and Taib Ziad

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.70053>

Abstract

In this work, a first-order autoregressive hidden Markov model (AR(1)HMM) is proposed. It is one of the suitable models to characterize a marker of breast cancer disease progression essentially the progression that follows from a reaction to a treatment or caused by natural developments. The model supposes we have observations that increase or decrease with relation to a hidden phenomenon. We would like to discover if the information about those observations can let us learn about the progression of the phenomenon and permit us to evaluate the transition between its states (supposed discrete here). The hidden states governed by the Markovian process would be the disease stages, and the marker observations would be the depending observations. The parameters of the autoregressive model are selected at the first level according to a Markov process, and at the second level, the next observation is generated from a standard autoregressive model of first order (unlike other models considering the successive observations are independent). A Markov Chain Monte Carlo (MCMC) method is used for the parameter estimation, where we develop the posterior density for each parameter and we use a joint estimation of the hidden states or block update of the states.

Keywords: autoregressive hidden Markov model, breast cancer progression marker, Gibbs sampler, hidden states joint estimation, Markov Chain Monte Carlo

1. Introduction

The main motivation behind this work is to characterize progression in breast cancer. In fact, disease progression cannot be assessed correctly without the use of biomarkers, which would

effectively monitor the evolution of the patient health state, and this is the case for breast cancer. The major challenge in this matter for researchers and clinicians is to unravel the stage of the disease, so as to tailor the treatment for each patient and to monitor the response of a patient to a treatment.

Currently, studies have shown that there is a correlation between the levels of certain markers such as cancer antigen CA15-3, carcinoembryonic antigen (CEA), and serum HER2 Neu with the stage of the disease [1]. This gives an opportunity of using a hidden Markov model (HMM) to predict the stage of the disease based on biomarker data and to address the effectiveness of the treatments in their influence on the transition of the cancer from one state to another. In HMM, we have two constituents: the Markovian hidden process suitable to represent the breast cancer stage and the observation process given by the biomarker data. By the way, we can learn about the disease transition rates and how it progresses from primary breast cancer to advanced cancer stage, for example.

Indeed, HMM is a useful tool for tackling numerous concrete problems in many fields but some possible applications of HMM are in speech processing [2], in biology [3], in disease progression [4], in economics [5, 6], and in gene expression [7]. For a complete review of HMM, the reader is referred to Zucchini and MacDonald [8], in which properties and definitions of HMM are presented in a plausible way with both classical estimation by maximum likelihood method and expectation maximization (EM) algorithm and the new Bayesian inference is addressed.

The model we consider here is a variation of the regular hidden Markov model, since we will use extensions to incorporate dependence among successive observations, suggesting autoregressive dependence among continuous observations. Consequently, we have relaxed the conditional independence assumption from a standard HMM, because we would like to add some dynamics to the patient disease progression and because in reality the current patient biomarker observation is dependent on the past one. In fact, the autoregressive assumption in HMM has shown its advantage over regular HMM that cannot catch the strong dependence between successive observations (e.g., Ref. [9]). A similar model to ours can be found in Ref. [10]. This kind of models, which were first proposed in Ref. [11] to describe econometrics time series, is generalization of both HMM and autoregressive models, will be effective in representing multiple heterogeneous dynamics such as the disease progression dynamics, and can be even generalized to a regime-switching ARMA models such as in Ref. [12].

Moreover, Our model can also be viewed as an extension of the multivariate double-chain Markov model (DCMM) developed by Ref. [13], where there are two discrete Markov chains of first order: the first Markov chain is observed and the second one is hidden. In contrast to this DCMM, our multivariate first-order autoregressive hidden Markov model (MAR(1)HMM) will lead to continuous observations, where each observation conditional on the hidden process will depend on the previous observation according to an autoregressive process of first order. This dynamic is promising for continuous observed disease biomarkers.

Parameter estimation is very challenging for HMM family models since the likelihood is not available in a closed form most of the time. Thus, we call for a Markov Chain Monte Carlo (MCMC) procedure instead of a maximum likelihood-based approach. This choice rises from the fact that the Bayesian analysis uses prior knowledge about the process being measured,

and it allows direct probability statements and an approximation of posterior distributions for the parameters. Instead in the maximum likelihood approach, we cannot have declared prior or have exact distribution for the parameters when the likelihood is untractable or when we have missing data (e.g., Refs. [14–16]).

Since the realization of HMM includes two separate entities: the parameters and the hidden states, the Bayesian computation is carried out after augmenting the likelihood by the missing hidden states [17]. The hidden states are sampled using a Gibbs sampler adopting a joint estimation of the hidden states or block update of the states (instead of a single update of each state separately) by means of a forward filtering/backward smoothing algorithm. Given the hidden states, we can compute the autoregressive parameters and the transition probabilities of the Markov chain by Gibbs sampler from their posterior densities after specifying conjugate priors for the parameters. Hence, the MCMC algorithm will alternate between simulating the hidden states and the parameters. Finally, we can obtain posteriors statistics such as the means, standard deviations and confidence intervals after assessing the convergence of the MCMC algorithm.

This chapter is organized as follows: after a preliminary on HMM, a description of the model is given in Section 3. In Section 4, we will give the Bayesian estimation of the parameters and the hidden states and provide the details of the MCMC algorithm, before presenting the results of a simulation studies in Section 5 and we will finish by a conclusion.

2. Preliminary

Since the model suggested is of the HMM type, we will describe HMM in more detail: an HMM is a stochastic process $\{X_t, Y_t\}_{t=0}^T$, where $\{X_t\}_{t=0}^T$ is a hidden Markov chain (unobservable) and $\{Y_t\}_{t=0}^T$ is a sequence of observable independent random variables such that Y_t depends only on X_t for the time $t = 0, 1, \dots, T$. Here the process $\{X_t\}_{t=0}^T$ evolves independently of $\{Y_t\}_{t=0}^T$ and is supposed to be a homogeneous finite Markov chain with probability transition matrix Π of dimension $a \times a$, where a indicates the number of the hidden states and $\Pi_0 = (\Pi_{01}, \dots, \Pi_{0a})$ is the initial state distribution.

We denote the probability density function of $Y_t = y_t$ given $X_t = k$ for $k \in \{1, \dots, a\}$ with $P_{x_t}(y_t, \theta_k)$, where θ_k refers to the parameters of P when $X_t = k$. We suppose further that the processes $Y_t | X_t$ and $Y_{t'} | X_{t'}$ are independent for $t \neq t'$. Let $\Theta = (\theta_1, \dots, \theta_a)$ and $\theta = (\Pi_0, \Pi, \Theta)$, and then, the HMM can be described as follows: First, the likelihood of the observations and the hidden states can be decomposed to $P(y_0, \dots, y_T, x_0, \dots, x_T, \theta) = P(y_0, \dots, y_T | x_0, \dots, x_T, \theta)P(x_0, \dots, x_T, \theta)$. Since $\{X_t\}_{t=0}^T$

is a Markov chain, $P(x_0, \dots, x_T, \theta) = \Pi_0(x_0) \prod_{t=1}^T \Pi(x_t | x_{t-1})$. Under the conditional independence of

the observations given the hidden states, $P(y_0, \dots, y_T | x_0, \dots, x_T, \theta) = P_{x_0}(y_0 | \theta_{x_0}) \prod_{t=1}^T P_{x_t}(y_t | \theta_{x_t})$.

Consequently, the likelihood function for the hidden states and the observations is given by

$$P(y_0, y_1, \dots, y_T, x_0, x_1, \dots, x_T, \theta) = \Pi(x_0) P_{x_0}(y_0 | \theta_{x_0}) \prod_{t=0}^T \Pi(x_t | x_{t-1}) P_{x_t}(y_t | \theta_{x_t}).$$

3. Model description and specification

The MAR(1)HMM model we consider in this work is a hidden Markov model, where conditionally on the latent states, the observations are not independent like it is the case for a regular hidden Markov model. Instead, the current observation is allowed to depend on the previous observation according to an autoregressive model of first order. As in an HMM model, the latent states evolve according to a discrete first-order time homogeneous Markov model. We consider data of n continuous random variables observed over time, each of potentially different lengths, i.e., for each individual $i = 1, 2, \dots, n$, we observe a vector $y_{i,\cdot} = (y_{i,u_i}, \dots, y_{i,m_i})^T$, with $u_i < m_i$.

Define $u_0 = \min_{1 \leq i \leq n} \{u_i\}$ and $M = \max_{1 \leq i \leq n} \{m_i\}$ and note that the times u_i and m_i may vary over the entire observation period from u_0 to M with the restriction that $u_i - m_i \geq 1$, for $i = 1, 2, \dots, n$.

We assume, for $i = 1, 2, \dots, n$ for integer time $t = u_i, \dots, m_i$, that the random variable $Y_{i,t}$ taking nonnegative values depends only on the states X_t and the previous observation $Y_{i,t-1}$, and based on the model developed by Farcomeni and Arima [10], we get the following model:

$$Y_{i,t}|X_t=x_t = \beta^{(x_t)} Y_{i,t-1} + \mu^{(x_t)} + \varepsilon_{i,t}. \quad (1)$$

The choice of the autoregressive part of the model is motivated by the fact that successive biomarker observations are most of the time correlated from many diseases unlike the hypothesis of independence between observations in HMMs.

We interpret x as the vector of the hidden health states of the patients; in the case of breast cancer, those states would be localized or advanced metastatic breast cancer for example, while y is the vector of the biomarkers observed and measured for the patients. The $\varepsilon_{i,t}$ are normal variables with mean 0 and variance σ^2 such that $\varepsilon_{i,t}$ and $\varepsilon_{i',t'}$ are uncorrelated, $(i, t) \neq (i', t')$.

The parameters $\beta^{(x_t)}$ and $\mu^{(x_t)}$ are parameters taking values in \mathbb{R} for each hidden state and $\sigma^2 \in \mathbb{R}^+$.

Similar to Ref. [13], the transition matrix of the Markov chain Π is time homogeneous with dimension $a \times a$ where a is the number of hidden states, and $\Pi = (\Pi_{gh}, g = 1, \dots, a; h = 1, \dots, a)$ where $\Pi_{gh} = P(X_t = h | X_{t-1} = g)$, for $g = 1, 2, \dots, a; h = 1, 2, \dots, a$; and $t = u_0 + 1, \dots, M$. We let the first state X_{u_0} to be selected from a discrete distribution with vector of probabilities $r = (r_1, \dots, r_a)$. Also we consider the time of initial observation u_i , the initial observed state y_{i,u_i} , and the number of consecutive time points that were observed $m_i - u_i + 1$. Let $\mu = (\mu^{(1)}, \dots, \mu^{(a)})$, $\beta = (\beta^{(1)}, \dots, \beta^{(a)})$, and $\theta = (\mu, \beta, \sigma^2, r, \Pi)$ be the set of all parameters in the model. We suppose that the individuals, i.e., $Y_{i,t}$, behave independently conditionally on X . Therefore, for $i = 1, \dots, n$, $P(y_{i,\cdot} | y_{i,u_i}, x, \theta) =$

$$\prod_{t=u_i+1}^{m_i} P(y_{i,t} | y_{i,t-1}, x_t, \theta) \text{ and } P(x | \theta) = P(x_{u_0}) \prod_{t=u_0+1}^M P(x_t | x_{t-1}, \Pi), \text{ where } P(x_t | x_{t-1}, \Pi) = P(X_t =$$

$x_t | X_{t-1} = x_{t-1}, \Pi) = \Pi_{x_{t-1}, x_t}$. Then, the likelihood density for the observations of all individuals $y = (y_1, \dots, y_n)$ given first time vector of observations $y_0 = (y_{1,u_1}, \dots, y_{n,u_n})$, x , and θ is

$$P(y | y_0, x, \theta) = \prod_{i=1}^n P(y_{i,\cdot} | y_{i,u_i}, x, \theta),$$

This is due to the conditional independence of the $y_{i,t}$ given x and θ . The joint mass of each $y_{i,t}$ and x given y_{i,u_i} and θ can be written as follows: $P(y_{i,t}, x | y_{i,u_i}, \theta) = P(y_{i,t} | y_{i,u_i}, x, \theta) \times P(x | y_{i,u_i}, \theta)$. Using the Markov property of the hidden process, we have after simplification $P(x | y_{i,u_i}, \theta) \propto P(y_{i,u_i} | x, \theta) P(x | \theta) = P(y_{i,u_i} | x_{u_i}, \theta) r_{x_{u_0}} \prod_{x_{u_0}, x_{u_0+1}} \times \dots \times \prod_{x_{M-1}, x_M}$. In addition, $P(y_{i,t} | y_{i,u_i}, x, \theta) = \prod_{t=u_i+1}^{m_i} P(y_{i,t} | y_{i,t-1}, x, \theta)$, and consequently,

$P(y_{i,t}, x | y_{i,u_i}, \theta) \propto r_{x_{u_0}} P(y_{i,u_i} | x_{u_i}, \theta) \prod_{t=u_0+1}^M \prod_{x_{t-1}, x_t} \prod_{t=u_i+1}^{m_i} P(y_{i,t} | y_{i,t-1}, x, \theta)$. Finally, under the hypothesis of normal error distribution for the autoregressive parameters of the model (Eq. (1)) and the Chapman-Kolmogorov property, the joint distribution of $y_{i,t}$ and x given y_{i,u_i} and θ can be simplified to:

$$P(y_{i,t}, x | y_{i,u_i}, \theta) \propto P(y_{i,u_i} | x_{u_i}, \theta) \prod_{h=1}^a r_h^{\chi_{\{x_{u_0}\}}(h)} \prod_{t=u_0+1}^M \prod_{g=1}^a \prod_{h=1}^a \Pi_{g,h}^{\chi_{\{x_t, x_{t-1}\}}(g,h)} \times \prod_{t=u_i+1}^{m_i} \prod_{h=1}^a \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,t} - \mu^{(h)} - \beta^{(h)} y_{i,t-1}}{\sigma} \right) \right]^{\chi_{\{x_t\}}(h)},$$

where ϕ denotes the density of a standard normal distribution $\mathcal{N}(0, 1)$ and $\chi_{\{A\}}(x)$ is the usual indicator function of a set A . Finally, the joint distribution of y and x has the following form:

$$P(y, x | y_0, \theta) \propto \prod_{h=1}^a r_h^{\chi_{\{x_{u_0}\}}(h)} \prod_{t=u_0+1}^M \prod_{g=1}^a \prod_{h=1}^a \Pi_{g,h}^{\chi_{\{x_t, x_{t-1}\}}(g,h)} \times \prod_{i=1}^n \prod_{l=1}^a \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,u_i} - \mu^{(l)}}{\sigma} \right) \right]^{\chi_{\{x_t\}}(l)} \times \prod_{i=1}^n \prod_{t=u_i+1}^{m_i} \prod_{h=1}^a \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,t} - \mu^{(h)} - \beta^{(h)} y_{i,t-1}}{\sigma} \right) \right]^{\chi_{\{x_t\}}(h)}. \tag{2}$$

4. Bayesian estimation of the model parameters

We will use a Bayesian approach to estimate the model parameters. Inference in the Bayesian framework is obtained through the posterior density, which is proportional to the prior multiplied by the likelihood. The posterior distribution for our model, as in most cases, cannot be derived analytically, and we will approximate it through MCMC methods specifically designed for working with the augmented likelihood with the hidden states. In fact, MCMC methods start by specifying the prior density $\Pi(\theta)$ for the parameters. Since the data Y are available, the general sampling methods work recursively by alternating between simulating the full conditional distributions X given y and θ given x and y .

4.1. Prior distributions

Under the assumption of independence between the parameters $\theta = (\mu, \beta, \sigma^2, r, \Pi)$, the prior density could be written as $P(\theta) = P(r)P(\Pi)P(\mu)P(\beta)P(\sigma^2)$. r is the parameters of a multinomial distribution; hence, the natural choice for the prior would be a Dirichlet distribution

$r \sim \mathbb{D}(\alpha_{01}, \dots, \alpha_{0a})$. Later on, $\sum_{j=1}^a \Pi_{ij} = 1$, and we assume that $\Pi_i \sim \mathbb{D}(\delta_{i1}, \dots, \delta_{ia})$ for each row i

of the transition matrix. This choice of the Dirichlet prior can be even the default $\mathbb{D}(1, \dots, 1)$ as recently discussed in Ref. [18]. In fact, a Dirichlet prior is justified because the posterior density of each row of the transition matrix is proportional to the density of a Dirichlet distribution, and hence, choosing a Dirichlet prior would give a posterior Dirichlet. This can be justified as follows for a given set of parameters $\lambda = (\lambda_1, \dots, \lambda_a)$ from a discrete or from a multinomial density:

$$\pi(x_1, \dots, x_a, \lambda_1, \dots, \lambda_a) = \frac{n!}{x_1! \dots x_a!} \lambda_1^{x_1} \dots \lambda_a^{x_a} \text{ for the nonnegative integers } x_1, \dots, x_a, \text{ with } \sum_{i=1}^a x_i = n.$$

This probability mass function can be expressed, using the gamma function Γ , as

$$\pi(x_1, \dots, x_a, \lambda_1, \dots, \lambda_a) = \frac{\Gamma\left(\sum_{i=1}^a x_i + 1\right)}{\prod_{i=1}^a \Gamma(x_i + 1)} \prod_{i=1}^a \lambda_i^{x_i}. \text{ This form shows its resemblance to the Dirichlet}$$

distribution, and by starting from supposing the prior $\lambda \propto \mathbb{D}(\alpha_0, \dots, \alpha_a)$, the posterior is $P(\lambda|x) \propto P(\lambda)P(x|\lambda) \propto \prod_i \lambda_i^{x_i} \prod_i \lambda_i^{\alpha_i - 1} \propto \prod_i \lambda_i^{x_i + \alpha_i - 1} \propto \mathbb{D}(x_1 + \alpha_1, \dots, x_a + \alpha_a)$.

Furthermore, concerning the priors for parameters of the autoregressive model, we suppose for $h = 1, \dots, a$: $\mu^{(h)} \sim \mathcal{N}(\alpha_h, \tau_h)$, $\beta^{(h)} \sim \mathcal{N}(b_h, c_h)$, and inverse gamma (\mathbb{IG}) prior for $\sigma^2 \sim \mathbb{IG}(\epsilon, \zeta)$. $\alpha_h, \tau_h, b_h, c_h, \epsilon, \zeta$ are hyperparameters to be specified. For more details on Bayesian inference and prior selection in HMM, the reader is referred to Ref. [19]. In our case, prior distributions for the autoregressive parameters were proposed by Ref. [20] for a mixture autoregressive model, who points out that they are conventional prior choices for mixture models.

4.2. Sampling the posterior distribution for the hidden states

Chib [21] developed a method for the simulation of the hidden states from the full joint distribution for the univariate hidden Markov model case. We will describe his full Bayesian algorithm for the univariate hidden Markov model before a generalization to our MAR(1) HMM.

4.2.1. Chib's algorithm for the univariate hidden Markov model for estimation of the states

Suppose we have an observed process $Y_n = (y_1, \dots, y_n)$ and the hidden states $X_n = (x_1, \dots, x_n)$, θ are the parameters of the model. We adopt for simplicity $X_t = (x_1, \dots, x_t)$ the history of the states up to time t and $X^{t+1} = (x_{t+1}, \dots, x_n)$ the future from $t + 1$ to n . We use the same notation for Y_t and Y^{t+1} .

For each state $x_t \in \{1, 2, \dots, a\}$ for $t = 1, 2, \dots, n$, the hidden model can be described by a conditional density given the hidden states $\pi(y_t | Y_{t-1}, x_t = k) = \pi(y_t | Y_{t-1}, \theta_k)$, $k = 1, \dots, a$, with x_t

depending only on x_{t-1} and having transition matrix Π and initial distribution Π_0 , and the parameters for $\pi(\cdot)$ are $\theta = (\theta_1, \dots, \theta_a)$.

Chib [21] shows that it is preferable to simulate the full latent data $X_n = (x_1, \dots, x_n)$ from the joint distribution of $x_1, \dots, x_n | Y_n, \theta$, in order to improve the convergence property of the MCMC algorithm because instead of n additional blocks if each state is simulated separately, only one additional block is required. First, we write the joint conditional density as

$$P(X_n | Y_n, \theta, \Pi) = P(x_n | Y_n, \theta) P(x_{n-1} | Y_n, x_n, \theta, \Pi) \times \dots \times P(x_1 | Y_n, X^2, \theta, \Pi).$$

For sampling, it is sufficient to consider the sampling of x_t from $P(x_t | Y_n, X^{t+1}, \theta, \Pi)$. Moreover, $P(x_t | Y_n, X^{t+1}, \theta, \Pi) \propto P(x_t | Y_t, \theta, \Pi) P(x_{t+1} | x_t, \Pi)$. This expression has two ingredients: the first is $P(x_{t+1} | x_t, \Pi)$, which is the transition matrix from the Markov chain. The second is $P(x_t | Y_t, \theta, \Pi)$ that would be obtained by recursively starting at $t = 1$.

The mass function $P(x_{t-1} | Y_{t-1}, \theta, \Pi)$ is transformed into $P(x_t | Y_t, \theta, \Pi)$, which is in turn transformed into $P(x_{t+1} | Y_{t+1}, \theta, \Pi)$ and so on. The update is as follows: for $k = 1, \dots, a$, we could write

$$P(x_t = k | Y_t, \theta, \Pi) = \frac{P(x_t = k | Y_{t-1}, \theta, \Pi) \pi(y_t | y_{t-1}, \theta_k)}{\sum_{l=1}^a P(x_t = l | Y_{t-1}, \theta, \Pi) \pi(y_t | y_{t-1}, \theta_l)}.$$

These calculations are initialized at $t = 0$, by setting $P(x_1 | Y_0, \theta)$ to be the stationary distribution of the Markov chain. Precisely, the simulation proceeds for $k = 1, \dots, a$, recursively by first simulating $P(x_1 = k | Y_0, \theta)$, from the initial distribution $\Pi_0(k)$ and $P(x_1 = k | Y_1, \theta, \Pi) \propto P(x_1 = k | Y_0, \theta, \Pi) \pi(y_1 | Y_0, \theta_k)$. Then, we get by forward calculation $P(x_t = k | Y_{t-1}, \theta) = \sum_{l=1}^a \Pi_{lk} P(x_{t-1} = l | Y_{t-1}, \theta)$, for each $t = 2, \dots, n$, where Π_{lk} is the transition probability and $P(x_t = k | Y_t, \theta) \propto P(x_t = k | Y_{t-1}, \theta, P) \Pi(y_t | Y_{t-1}, \theta_k)$. The last term in the forward computation $P(x_n = k | Y_n, \theta)$ would serve as a start for the backward pass, and we get recursively for each $t = n - 1, \dots, 1$; $P(x_t = k | Y_n, X^{t+1}, \theta) \propto P(x_t | Y_{t-1}, \theta, \Pi) P(x_{t+1} | x_t = k, \Pi)$, which permits the obtention of $X_n = (x_1, \dots, x_n)$.

4.2.2. Simulating the hidden states for the MAR(1)HMM

Returning to our model, and adopting notations and algorithm developed by Fitzpatrick and Marchev, f will denote the observation density for the MAR(1)HMM, and for $u_0 < t < M$;

$$x_{-t} = (x_{u_0}, \dots, x_t), \quad x^t = (x_t, \dots, x_M), \quad y(t) = (y_{i,t}, i = 1, 2, \dots, n), \quad y_{,t} = \bigcup_{i:u_i < t} \{y_{i,u_i}, \dots, y_{i,\min\{t, m_i\}}\},$$

and $y^t = \bigcup_{i:t < m_i} \{y_{i,\max\{t+1, u_i\}}, \dots, y_{i, m_i}\}$. The posterior distribution of the hidden state could be

written as: $P(x_{-M} | y_{,M}, \theta) = P(x_M | y_{,M}, \theta) \times \dots \times P(x_{u_0} | y_{,M}, x^{u_0+1}, \theta)$. So we could sample the whole sequence of states by sampling from $P(x_t | y_{,M}, x^{t+1}, \theta)$. Hence, the estimation of the hidden states is performed recursively by first initializing

$$P(x_{u_0}|y_{,u_0}, \theta) \propto P(y_{,u_0}|x_{u_0})P(x_{u_0}|r); y_{,u_0} = \{y_{i,u_i}, u_i = u_0, i = 1, \dots, n\}.$$

$$P(x_{u_0+1} = k|y_{,u_0}, \theta) = \sum_{l=1}^a \Pi_{lk} P(x_{u_0} = l|Y_{,u_0}); k = 1, \dots, a.$$

$$P(x_{u_0+1} = k|y_{,u_0+1}, \theta) \propto P(x_{u_0+1} = k|y_{,u_0}, \theta) f(y_{(u_0)}|y_{,u_0}, \theta_k).$$

We perform a similar calculation for every state at time t , and we conclude by calculating $P(x_M = k|y_{,M-1}, \theta) = \sum_{l=1}^a \Pi_{lk} P(x_{M-1} = l|Y_{,M-1}, \theta)$, and $P(x_M = k|y_{,M}, \theta) \propto P(x_M = k|y_{,M-1}, \theta) f(y_{(M)}|y_{,M-1}, \theta_k)$. Later on, we get $P(x_M = k|y_{,M}, \theta)$, which permits the simulation of $P(x_M|y_{,M}, \theta)$. Finally, by backward calculation, we simulate from the probabilities $P(x_t|y_{,M}, x^{t+1}, \theta) \propto P(x_{t+1}|x_t, \Pi)P(x_t|y_{,t}, \theta)$ for each time $t = M - 1, \dots, u_0$. Those backward probabilities would permit the simulation of the latent states.

4.3. Sampling from $P(\theta|x, y)$

4.3.1. Sampling Π

Under the prior assumption of Dirichlet prior for each row of the transition matrix $P(\Pi_i) \propto \mathbb{D}(\delta_{i1}, \dots, \delta_{ia})$, and the independence assumption between those rows, the posterior distribution for Π_i can be developed using Eq. (2) as follows: Let n_{ij} denote the number of single transitions from state i to state j , so

$$\begin{aligned} P(\Pi_i|y, x) &\propto P(\Pi_i) \prod_{t=u_0+1}^M \prod_{j=1}^a \Pi^{X_{\{x_{t-1}, x_t\}}(i,j)} \propto P(\Pi_i) \prod_{j=1}^a \Pi_{ij}^{n_{ij}} \propto \prod_{j=1}^a \Pi_{ij}^{\delta_{ij} + n_{ij} - 1} \\ &\propto \mathbb{D}(\delta_{i1} + n_{i1}, \dots, \delta_{ia} + n_{ia}). \end{aligned}$$

4.3.2. Sampling posterior distribution for initial distribution

Let $n_{0l} = \chi_{x_{u_0}}(l)$, for $l = 1, \dots, a$. Using (2), under Dirichlet prior $\mathbb{D}(\delta_{01}, \dots, \delta_{0a})$ for the parameter r , we obtain $P(r|x, y) \propto P(r) \prod_{l=1}^a r_l^{X_{\{x_{u_0}\}}(l)} \propto \prod_{l=1}^a r_l^{\delta_{0l} + n_{0l} - 1} \propto \mathbb{D}(\delta_{01} + n_{01}, \dots, \delta_{0a} + n_{0a})$.

4.3.3. Sampling posterior distribution for the autoregressive parameters μ, β, σ^2

When a complete conditional distribution is known such as the normal distribution or beta distribution, we use the Gibbs sampler to draw the random variable. This is the case for our model. Let us define $n_{u_i}(l) = \sum_{i=1}^n \chi_{\{x_{u_i}=l\}}$, $n_l = \sum_{i=1}^n \sum_{t=u_i+1}^{m_i} \chi_{\{x_t=l\}}$, $N = \sum_{l=1}^a n_l$, $n_{0l} = \chi_{\{x_{u_0}\}}(l)$. So for $l = 1, 2, \dots, a$; by supposing $\mathcal{N}(\alpha_l, \tau_l)$ as prior distribution and using Eq. (2), the conditional posterior distribution of $\mu^{(l)}$ is:

$$P(\mu^{(l)}|y, x) \propto P(\mu^{(l)}) \prod_{i=1}^n \left[\frac{1}{\sigma} \phi \left(\frac{y_{i, u_i} - \mu^{(l)}}{\sigma} \right) \right]^{x_{\{x_{u_i}\}}^{(l)}} \times \prod_{i=1}^n \prod_{t=u_i+1}^{m_i} \left(\frac{1}{\sigma} \phi \left(\frac{y_{i,t} - \mu^{(l)} - \beta^{(l)} y_{i,t-1}}{\sigma} \right) \right)^{x_{\{x_t\}}^{(l)}}.$$

$$\propto \exp \frac{-1}{2} \left\{ \frac{(\mu^{(l)} - \alpha_l)^2}{\tau_l} + \sum_{i=1, x_{u_i}=l}^n \left(\frac{y_{i, u_i} - \mu^{(l)}}{\sigma} \right)^2 + \sum_{i=1}^n \sum_{t=u_i+1, x_t=l}^{m_i} \left(\frac{y_{i,t} - \mu^{(l)} - \beta^{(l)} y_{i,t-1}}{\sigma} \right)^2 \right\}.$$

then $\mu^{(l)}/y, x \sim \mathcal{N}(\tilde{\tau}_l, \tilde{\alpha}_l)$ with inverse mean $\tilde{\tau}_l^{-1} = \frac{n_{u_i}^{(l)} + n_l}{\sigma^2} + \frac{1}{\tau_l}$ and variance

$$\tilde{\alpha}_l = \tilde{\tau}_l \left(\frac{\sum_{i=1, x_{u_i}=l}^n y_{i, u_i} + \sum_{i=1}^n \sum_{t=u_i+1, x_t=l}^{m_i} (y_{i,t} - \beta^{(l)} y_{i,t-1})}{\sigma^2} + \frac{\alpha_l}{\tau_l} \right).$$

For $\beta^{(l)}, l = 1, \dots, a$, and similar to $\mu^{(l)}$, $\mathcal{N}(b_l, c_l)$ was proposed as prior choice to obtain:

$$P(\beta^{(l)}|y, x) \propto P(\beta^{(l)}) \prod_{i=1}^n \prod_{t=u_i+1}^{m_i} \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,t} - \mu^{(l)} - \beta^{(l)} y_{i,t-1}}{\sigma} \right) \right]^{x_{\{x_t\}}^{(l)}},$$

and therefore, $\beta^{(l)}/y, x \sim \mathcal{N}(\tilde{c}_l, \tilde{b}_l)$ with inverse mean $\tilde{c}_l^{-1} = \frac{1}{c_l} + \frac{\sum_{i=1}^n \sum_{t=u_i+1, x_t=l}^{m_i} y_{i,t-1}^2}{\sigma^2}$ and variance

$$\tilde{b}_l = \tilde{c}_l \left(\frac{b_l}{c_l} + \frac{\sum_{i=1}^n \sum_{t=u_i+1, x_t=l}^{m_i} (y_{i,t} - \mu^{(l)}) y_{i,t-1}}{\sigma^2} \right).$$

For the posterior distribution of σ^2 , by supposing $\mathbb{IG}(\varepsilon, \zeta)$ as prior, we deduce from Eq. (2)

$$P(\sigma^2|y, x) \propto (\sigma^2)^{-(\varepsilon+1)} \exp \left(-\frac{\zeta}{\sigma^2} \right) \prod_{i=1}^n \left[\frac{1}{\sigma} \phi \left(\frac{y_{i, u_i} - \mu^{(x_{u_i})}}{\sigma} \right) \right]^{x_{\{x_{u_i}\}}} \\ \times \prod_{i=1}^n \prod_{t=u_i+1}^{m_i} \left[\frac{1}{\sigma} \phi \left(\frac{y_{i,t} - \mu^{(x_t)} - \beta^{(x_t)} y_{i,t-1}}{\sigma} \right) \right]^{x_{\{x_t\}}},$$

consequently $\sigma^2/y, x \sim \mathbb{IG}(\tilde{\varepsilon}, \tilde{\zeta})$ with parameters $\tilde{\varepsilon} = \frac{n_{u_i} + N}{2} + \varepsilon$ and

$$\tilde{\zeta} = \frac{\sum_{i=1}^n (y_{i, u_i} - \mu^{(x_{u_i})})^2 + \sum_{i=1}^n \sum_{t=u_i+1}^{m_i} (y_{i,t} - \mu^{(x_t)} - \beta^{(x_t)} y_{i,t-1})^2}{2} + \zeta.$$

Finally, the algorithm is ran for $d = 1, \dots, D$ iterations by alternating between the following steps, where in each step we compute a conditional posterior for the given parameter:

The MCMC algorithm:

1. For $h = 1, 2, \dots, a$, give reference values for the hyperparameters α_{hr} , τ_{hr} , a_{hr} , b_{hr} , δ_{0hr} and δ_{ih} for $i = 1, 2, \dots, a$.
2. Initialization (Step $d = 1$ of the MCMC iterations): Initialize $\Pi^{(1)}$, $r^{(1)}$, $\mu^{(1)}$, $\beta^{(1)}$, and $\sigma^{2(1)}$.
3. Simulation of the hidden states:
 - a. Initialization of forward simulation: $P(x_{u_0}^{(d)} | y_{u_0}, \theta) \propto P(y_{u_0} | x_{u_0}^{(d)})P(x_{u_0}^{(d)} | r^{(d)})$, with $y_{u_0} = \{y_{i, u_i}, u_i = u_0, i = 1, \dots, n\}$.
 - b. Forward simulation: For $k = 1, \dots, a$ and $t = u_0 + 1, \dots, M$:

$$P(x_t^{(d)} = k | y_{t-1}, \theta) = \sum_{l=1}^a \Pi_{lk}^{(d)} P(x_{t-1}^{(d)} = l | Y_{t-1}, \theta) \text{ and}$$

$$P(x_t^{(d)} = k | y_{t'} \theta) = \frac{P(x_t^{(d)} = k | y_{t-1}, \theta_k) f(y(t) | y_{t-1}, \theta_k)}{\sum_{l=1}^a P(x_t^{(d)} = l | y_{t-1}, \theta) f(y(t) | y_{t-1}, \theta_l)}$$
 - c. Initialization of backward simulation: For $k = 1, \dots, a$, given $P(x_M^{(d)} = k | y_{M'}, \theta)$ from forward simulation, we get $P(x_M^{(d)} | y_{M'}, \theta)$.
 - d. Backward simulation: For $k = 1, \dots, a$ and $t = M - 1, \dots, u_0$:

$$P(x_t^{(d)} | y_{M'}, x^{t+1(d)}, \theta) \propto P(x_{t+1}^{(d)} | x_t^{(d)}, \pi) P(x_t^{(d)} | y_{t'}, \theta).$$
4. Estimation of the initial distribution and the transition distribution
 - a. for $l = 1, \dots, a$, $k = 1, \dots, a$. Calculate $n_{0l} = \chi_{\{x_{u_0}^{(d)}\}}(l)$ and $n_{kl} = \sum_{t=u_0+1}^M \chi_{\{x_{t-1}^{(d)}, x_t^{(d)}\}}(k, l)$.
 - b. Sample $(r_1^{(d+1)}, \dots, r_a^{(d+1)}) \propto \mathbb{D}(\delta_{01} + n_{01}, \dots, \delta_{0a} + n_{0a})$.
 - c. For $i = 1, \dots, a$; sample $(\Pi_{i1}^{(d+1)}, \dots, \Pi_{ia}^{(d+1)}) \propto \mathbb{D}(\delta_{i1} + n_{i1}, \dots, \delta_{ia} + n_{ia})$.
5. Simulation of μ : For $l = 1, \dots, a$,
 - a. $\tilde{\tau}_l^{-1} = \frac{n_{u_i(l)} + n_l}{\sigma_{(d)}^2} + \frac{1}{\tau_l}$.
 - b.
$$\tilde{\alpha}_l = \tilde{\tau}_l \left(\frac{\sum_{i=1, x_{u_i}=l}^n y_{i, u_i} + \sum_{i=1}^n \sum_{t=u_i+1, x_t=l}^{m_i} (y_{i, t} - \beta_{(d)}^{(l)} y_{i, t-1})}{\sigma_{(d)}^2} + \frac{\alpha_l}{\tau_l} \right).$$
 - c. Simulate $\mu_{(d+1)}^{(l)} / y, x \sim \mathcal{N}(\tilde{\alpha}_l, \tilde{\tau}_l)$.

6. Simulation of β : For $l = 1, \dots, a$,

a.
$$\tilde{c}_l^{-1} = \frac{1}{c_l} + \frac{\sum_{i=1}^n \sum_{t=u_i+1, x_t=l}^{m_i} y_{i,t-1}^2}{\sigma_{(d)}^2}.$$

b.
$$\tilde{b}_l = \tilde{c}_l \left(\frac{b_l}{c_l} + \frac{\sum_{i=1}^n \sum_{t=u_i+1, x_t=l}^{m_i} (y_{i,t} - \mu_{(d+1)}^{(l)}) y_{i,t-1}}{\sigma_{(d)}^2} \right).$$

c. Simulate $\beta_{(d+1)}^{(l)} / y, x \sim \mathcal{N}(\tilde{b}_l, \tilde{c}_l)$.

7. - Simulation of σ^2 :

a.
$$\tilde{\epsilon} = \frac{n_{u_i} + N}{2} + \epsilon.$$

b.
$$\tilde{\zeta} = \frac{\sum_{i=1}^n (y_{i,u_i} - \mu_{(d+1)}^{(x_{u_i})})^2 + \sum_{i=1}^n \sum_{t=u_i+1}^{m_i} (y_{i,t} - \mu_{(d+1)}^{(x_t)} - \beta_{(d+1)}^{(x_t)} y_{i,t-1})^2}{2} + \zeta.$$

c. Simulate $\sigma_{(d+1)}^2 / y, x \sim \mathbb{IG}(\epsilon, \tilde{\zeta})$.

5. Simulation study

In this section, we apply our results to the breast cancer model discussed earlier. The main reason behind our work is that the progression of breast cancer cannot be seen directly unless we use observations related to the disease that could characterize its progression; those observations here are quantities which could be measured; they are called biomarkers, where the word biomarker is used to designate any objective indication of a biological process or disease condition including during treatment and should be measurable. Furthermore, biomarkers are increasingly used in the management of breast cancer patients. One example is reported in Ref. [22], stating that there is correlation between elevation of CEA and/or CA15-3 and disease progression, in breast cancer patients. Also we use the autoregressive dependence among the observations to add more dynamics to the model unlike conventional HMMs where the successive observations given the Markov process are independent. We used the classification of breast cancer in three states: local where the disease is confined within the breast, the regional phase when the lymph nodes are involved, and the distant stage where the cancer is found in other parts of the body. We restrict ourselves to these three stages unlike other stage classifications that divide the progression in more than three stages such as the TNM (tumor, node, and metastasis) system. By lack of finding data about breast cancer biomarkers, we will confine ourselves to simulate an MAR(1)HMM model for observation time $M = 24$, and a number of individuals $n = 210$, $a = 3$ for Markov states number, with the length observation

time for each individual selected uniformly between 2 and M . The simulation process supposes we have for the autoregressive means $\mu = (\mu^{(1)}, \mu^{(2)}, \mu^{(3)}) = (12, 24, 36)$, since markers such as CA15–3 increase as the disease advances toward metastatic breast cancer. In addition, CA15–3 increase rapidly between successive observations, and thus, we take in the simulation the parameters $\beta = (\beta^{(1)}, \beta^{(2)}, \beta^{(3)}) = (0.2, 0.4, 0.8)$.

The algorithm of simulation works as follows:

1. For each individual $i = 1, \dots, n$, choose m_i the length of observation for that individual i .
2. Generate each discrete disease state x_t using transition matrix $\Pi = (0.7, 0.2, 0.1; 0.1, 0.6, 0.3; 0.2, 0.3, 0.5)$ for $t = u_{0+1}, \dots, M$.
3. Generate the observations $y_{i,t}$ for all individuals using our model 1.

We choose a prior $\sigma^2 \sim \mathbb{IG}(0.001, 0.001)$, a $\mathbb{D}(1, \dots, 1)$ prior for each row of Π , and Gaussian noninformative priors for the μ s and the β s. Having the hidden states and the observations, we ran our algorithm for 8000 MCMC iterations. MCMC algorithm convergence was assessed by analyzing MCMC iterations mixing plots that are shown in **Figure 1**, autocorrelation sample graphs checking as illustrated in **Figure 2**, and inspecting histograms of posterior densities for the parameters of the models in **Figure 3**. All parameters show good mixing of chains, autocorrelations that decay immediately after a few lags, and perfect posterior densities fitting. Also the Gelman [23] potential scale reduction factor (PSRF) was plot. The PSRF is measured for more than two MCMC chains (three chains in this works are considered), and it is measured for each parameter of the model; it should show how the chains have forgotten their initial values and that the output from all chains is indistinguishable. It is based on a comparison of within-chain and between-chains variances and is similar to a classical analysis of variance; when the PSRF is high (perhaps greater than 1.1 or 1.2), then we should run our chains out longer to improve convergence to the stationary distribution. Each PSRF declines to 1 as the number of iterations approaches infinity to confirm convergence. All the parameters have shown a PSRF less than 1.1 as the number of iteration increases and by the way a good sign of convergence (**Figure 4**). Moreover, we should point out that in the family of Markov switching model there is the so-called label switching problem (e.g., Ref. [24]) which arises identifiability problem, and hence, we would not estimate perfectly the parameters. In addition, the posterior densities could show evidence of multimodality. some authors postprocess the output of the MCMC to deal with the issue (e.g., [25]), while other uses a random permutation of the parameters in each iteration of the MCMC algorithm (e.g., [26]) or one can call for an invariant loss function method (e.g., [27]). In our case, no identifiability issue is noticed since we used well-separated prior hyperparameters. Even when we start from different initial values for the parameters, our algorithm converges immediately after a few iterations.

Finally and before giving our results, we should report that the simulation of the Dirichlet posterior was carried out following ([28, p. 22], [29, p. 155]) who reported that the posterior Dirichlet parameters should be simulated using the beta distribution approach. **Table 1** shows how the posterior values estimated from algorithm are very close to the true ones.

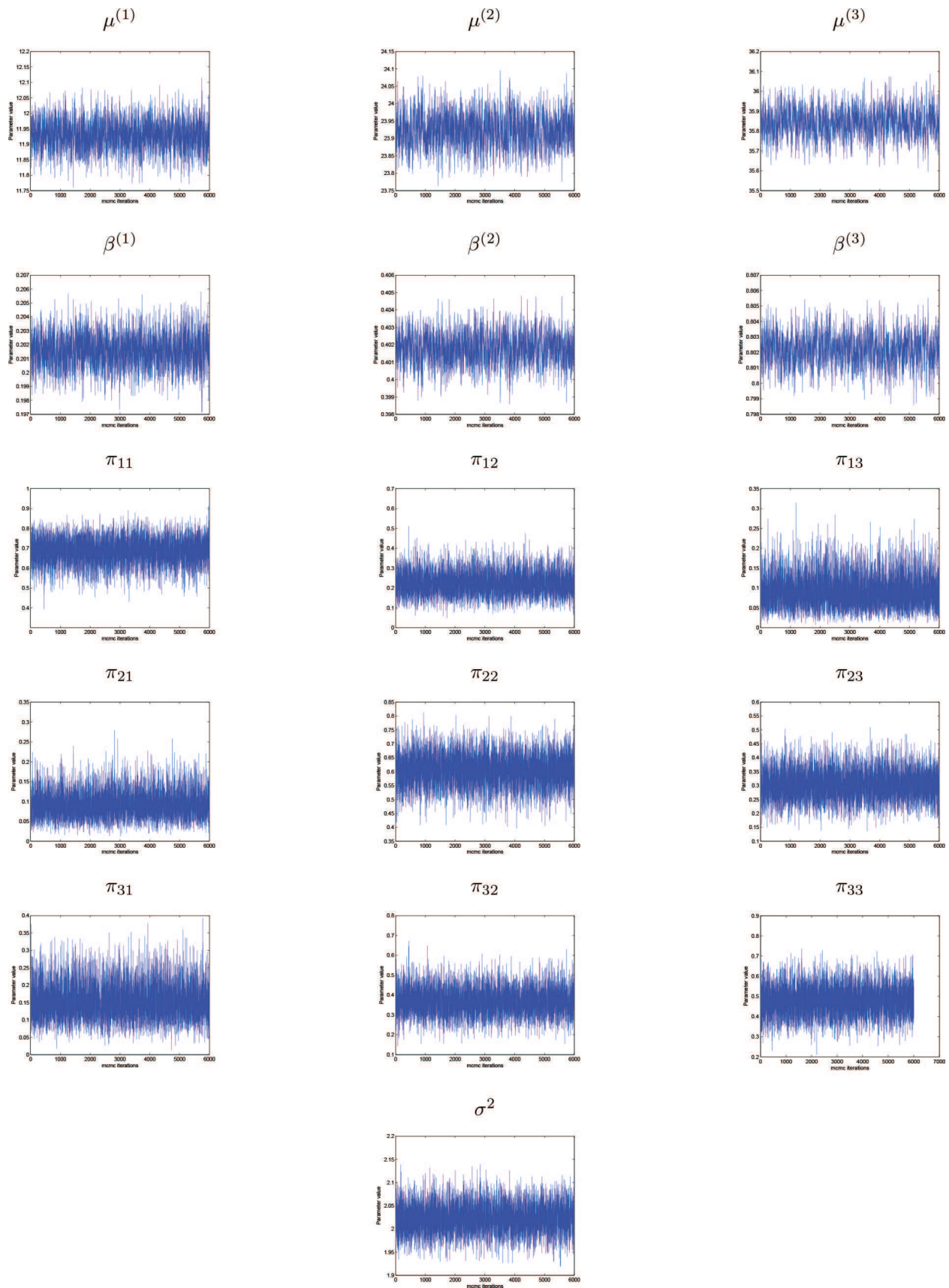


Figure 1. Markov chain mixing for each parameter through MCMC algorithm simulation.

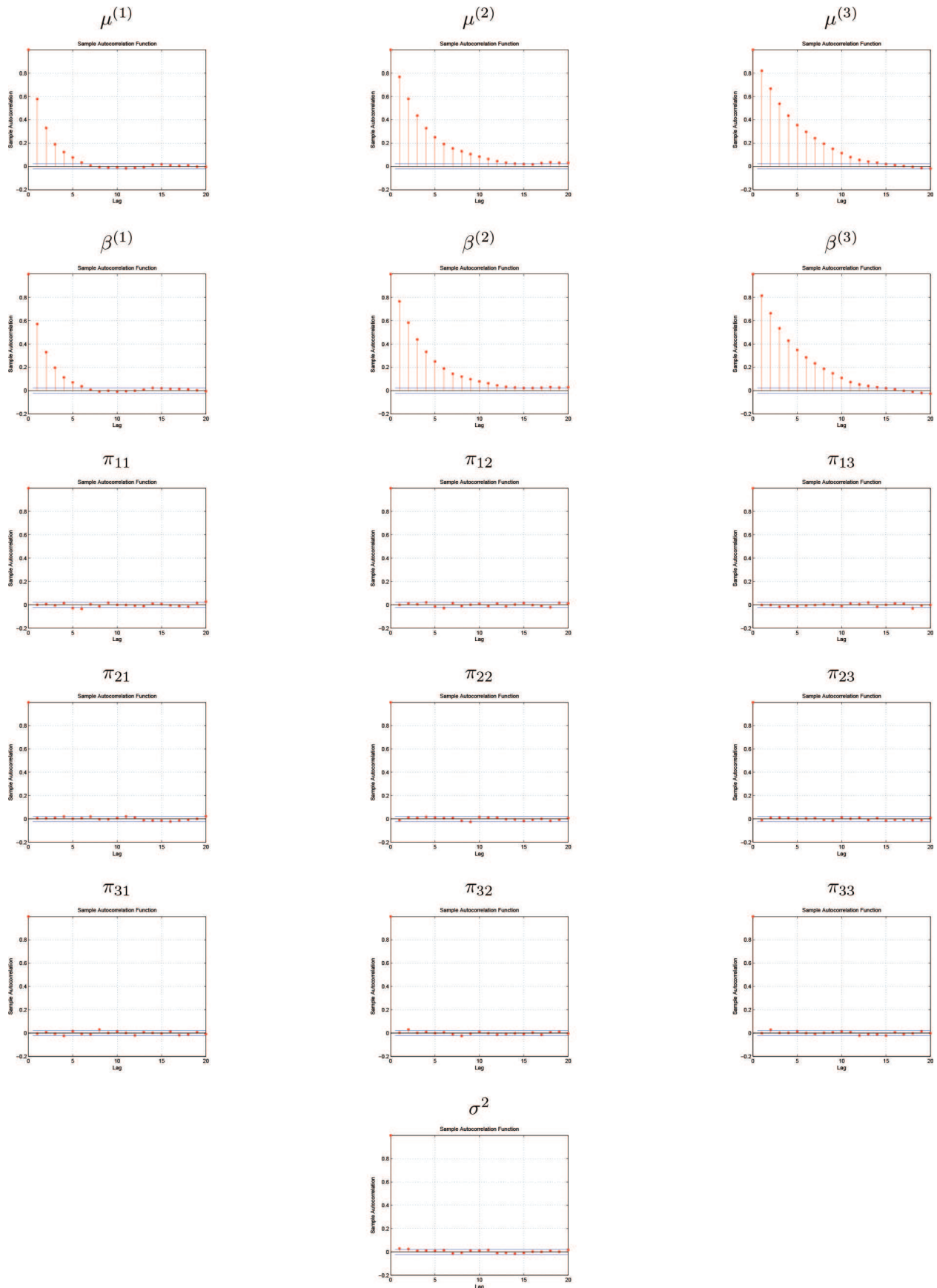


Figure 2. Autocorrelation sample plots for parameters of the model.

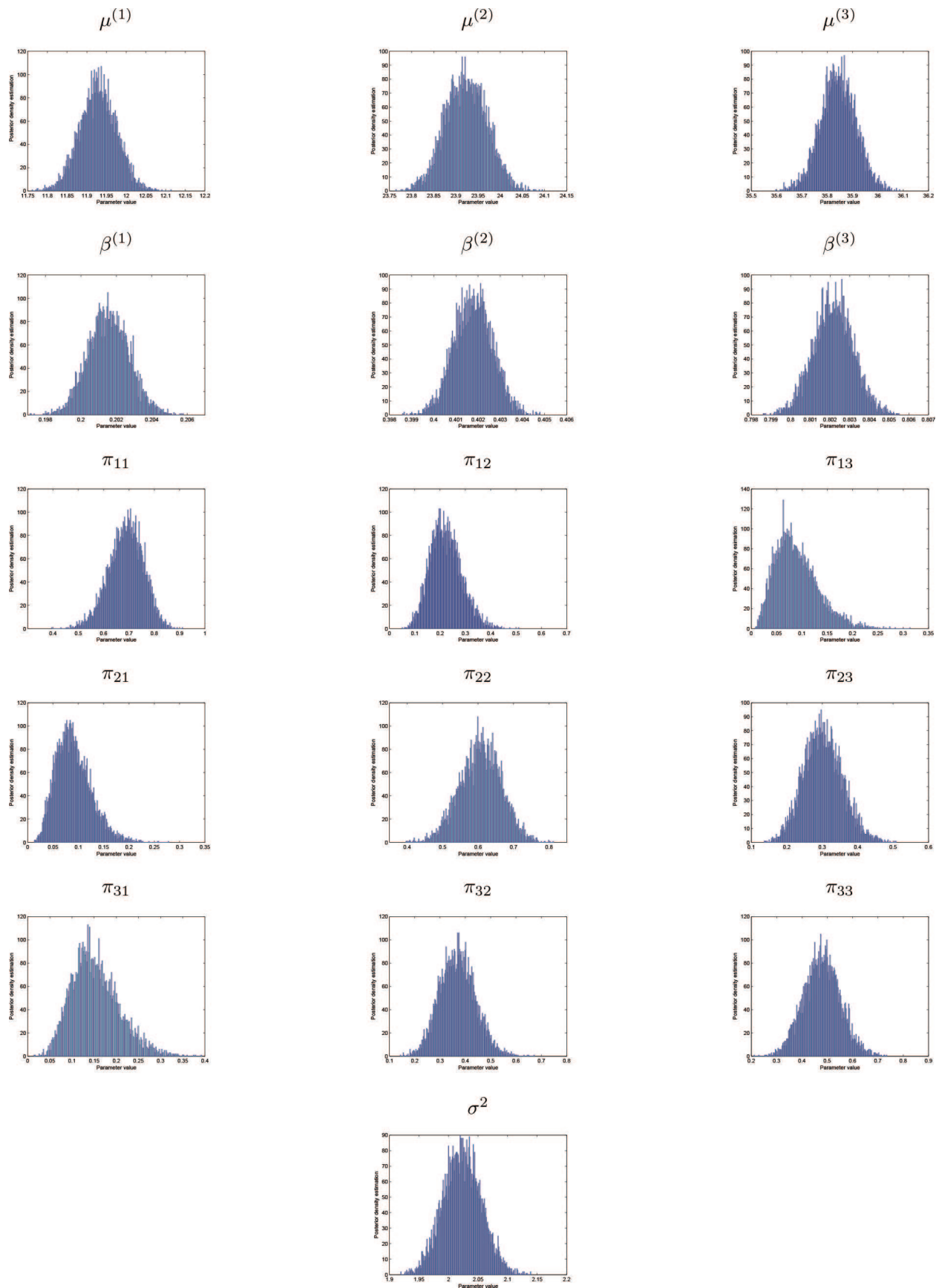


Figure 3. Posterior densities for the parameters of the model (after 8000 iterations).

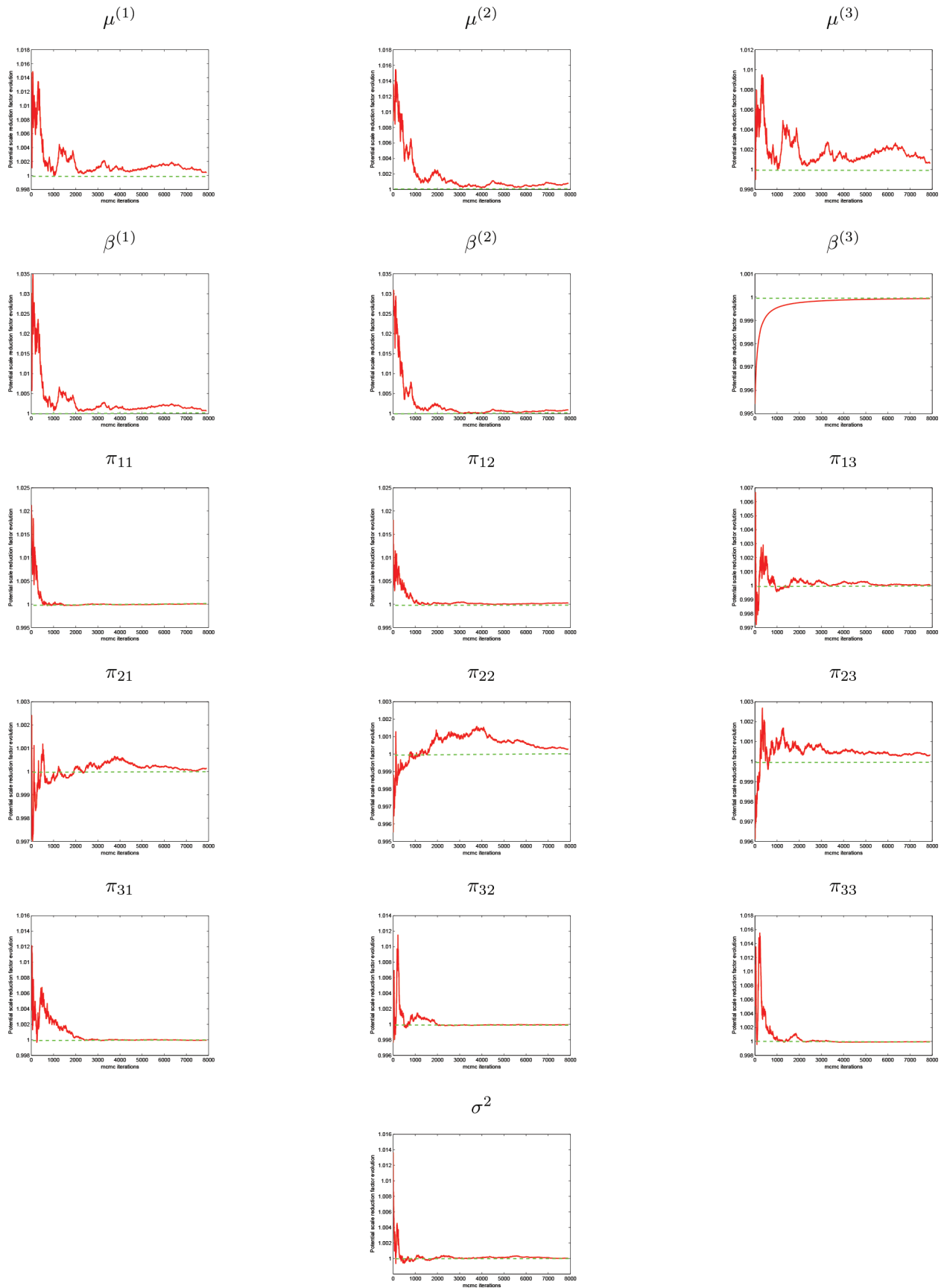


Figure 4. Potential scale reduction factor convergence to less than 1.02 with more iterations.

Parameter	True value	Posterior statistics		
		Mean	Standard deviation	Confidence interval (5%)
μ^1	12	11.929	0.047	(11.851–12.005)
μ^2	24	23.923	0.047	(23.847–24.000)
μ^3	36	35.843	0.070	(35.729–35.959)
β^1	0.2	0.2016	0.0012	(0.1997–0.2035)
β^2	0.4	0.4018	0.0009	(0.4004–0.4032)
β^3	0.8	0.8022	0.0010	(0.8005–0.8038)
π_{11}	0.7	0.688	0.068	(0.5715–0.797)
π_{12}	0.2	0.223	0.062	(0.129–0.332)
π_{13}	0.1	0.090	0.042	(0.032–0.17)
π_{21}	0.1	0.091	0.035	(0.041–0.154)
π_{22}	0.6	0.607	0.059	(0.507–0.701)
π_{23}	0.3	0.302	0.055	(0.214–0.397)
π_{31}	0.2	0.153	0.053	(0.075–0.250)
π_{32}	0.3	0.368	0.071	(0.257–0.488)
π_{33}	0.5	0.479	0.073	(0.358–0.599)
σ^2	2	2.023	0.032	(1.970–2.077)

Table 1. Posterior inference for the parameters of the MAR(1)HMM model.

6. Conclusion

We have extended the method of Chib [21] for block update estimation of the states to a MAR (1)HMM model. Furthermore, we would like to point out that our model can easily be extended to include missing observations, as we should only add an extra step in each MCMC iteration to estimate the missing observations. Also, we can estimate the autoregressive model for different values of the autoregressive order, $p \geq 1$, by evaluating the Bayesian information criterion to select the best order that fits the observations of the model. Our model would capture the complexity and the dynamics of the evolution of breast cancer by introducing the latent states; the probabilities of transition between the latent states allow to compare among the effects of treatments on slowing or accelerating the transition of the disease from one health stage to another the autoregressive parameter mean values corresponding to different stages of the disease would guide medical doctors and scientists to monitor patients in different phases of the disease. The model incorporates individual observations with different lengths.

Last but not least, we like to mention the utilities of switching diffusion processes in addressing and analyzing many complicated applications such as in finance and risk management. Our future work would be to apply these processes to explore disease progression, because they are characterized by the coexistence of continuous dynamics and discrete events as well as their interactions.

Acknowledgements

We would like to thank the editorial staff for the comments that helped in improving this work. Also, we would like to thank the supporters of this work: The Lalla Salma Foundation Prevention and Treatment of Cancer, Rabat, Morocco; and the Germano-Moroccan Program for Scientific Research PMARS 2015-060.

Author details

Hamid El Maroufy^{1*†}, El Houcine Hibbah^{1†}, Abdelmajid Zyad^{2†} and Taib Ziad³

*Address all correspondence to: h_elmaroufy@hotmail.com

1 Department of Mathematics, Faculty of Sciences and Technics, Sultan Moulay Slimane University, Béni Mellal, Morocco

2 Biological Engineering Laboratory, Team of Natural Substances, Cell and Molecular Immuno-Pharmacology, Sultan Moulay Slimane University, Morocco

3 Early Clinical Development, Astra Zeneca RD, Gothenburg, Mölndal, Sweden

[†]The three first authors acknowledge the financial support of the Lalla Salma Foundation of Cancer: Prevention and Treatment, Project 09/2013.

References

- [1] Samy N, Ragab HM, El Maksoud NA, Shaalan M. Prognostic significance of serum Her2/neu, BCL2, CA15-3 and CEA in breast cancer patients: A short follow up. *Cancer Biomarkers*. 2009;**6**:63-72
- [2] Benmiloud B, Piczunski W. Estimation des parametres dans les chaines de markov cachees et segmentation. *Traitement du Signal*. 1995;**12**:433-454
- [3] Boys R, Handerson D. A Bayesian approach to DNA sequence segmentation (with discussion). *Biometrics*. 2004;**60**:573-588
- [4] Guihenneuc-Jouyaux C, Richardson S, Longini IM Jr. Modeling disease progression by a hidden Markov process: Application to characterizing CD4 cell decline. *Biometrics*. 2000;**56**:733-741
- [5] Albert J, Chib S. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics*. 1993;**11**:1-15
- [6] Korolkiewickz M, Elliot J. A hidden Markov model of credit quality. *Journal of Economic Dynamics and Control*. 2008;**32**:3807-3819

- [7] Zeng Y, Frias J. A novel HMM-based clustering algorithm for the analysis of gene expression time-course data. *Computational Statistics and Data Analysis*. 2006;**50**:2472–2494
- [8] Zucchini W, MacDonald I. *Hidden Markov Models for Time Series: An Introduction Using R*. New York: Springer; 2009
- [9] Ailliot P, Monbet V. Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*. 2012;**30**:92–101
- [10] Farcomeni A, Arima S. A Bayesian autoregressive three state hidden Markov model for identifying switching monotonic regimes in microarray time course data. *Statistical Applications in Genetics and Molecular Biology*. 2013;**23**:467–480
- [11] Hamilton J. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*. 1989;**57**:357–384
- [12] Kim C, Kim J. Bayesian inference in regime-switching ARMA models with absorbing states: The dynamics of the ex-ante real interest rate under regime shifts. *Journal of Business and Economic Statistics*. 2015;**33**:566–578
- [13] Fitzpatrick, M. and Marchev, D. Efficient bayesian estimation of the multivariate double chain markov model. *Statistics and Computing*, 2013;**23**(4):467-480
- [14] Lindley D. The philosophy of statistics. *The Statistician*. 2000;**49**:293–337
- [15] Bolstad, W. M. (2007). *Introduction to Bayesian Statistics*. John Wiley and Sons, Inc., Hoboken, New Jersey; 2007
- [16] Gelman A, Shalizi CR. Philosophy and the practice of Bayesian statistics in the social sciences. *British Journal of Mathematical and Statistical Psychology*. 2013;**66**:8–38
- [17] Hobert, J.P. The data augmentation algorithm: Theory and methodology. In: Brooks S, Gelman A, Jones GL, Meng X-L, editors. *The Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: Chapman and Hall/CRC; 2011. p. 253
- [18] Tuyl F, Gerlach R, Mengersen K. Posterior predictive arguments in favor of the Bayes-Laplace priors as the consensus prior for the binomial and multinomial parameters. *Bayesian Analysis*. 2013;**4**:151–158
- [19] Cappe O, Moulines E, Ryden T. *Inference in Hidden Markov Models*. New York: Springer-Verlag; 2005
- [20] Sampietro S. Mixture of autoregressive components for modeling financial market volatility. *LIUC Papers. Serie Metodi quantitativi* 16. 2005
- [21] Chib S. Calculating posterior distributions and model estimates in Markov mixtures models. *Journal of Econometrics*. 1996;**75**:79–98
- [22] Laessig D, Nagel D, Heinemann V, Untch M, Kahlert S, Bauerfeind I, Stieber P. Importance of CEA and CA15-3 during disease progression in metastatic breast cancer patients. *Anticancer Research*. 2007;**27**:1963–1968

- [23] Gelman A. Inference and monitoring convergence. In: Gilks W, Richardson S, Spiegelhalter D, editors. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall/CRC; 1995. p. 131
- [24] Fruhwirth-Schnatter S. *Finite Mixture and Markov Switching Models*. New York: Springer; 2006
- [25] Celoux G. Bayesian inference for mixture: The label switching problem. In: Payne R, Green P, editors. *Proceedings in Computational Statistics*. Heidelberg: Physica; 1998. pp. 227–232
- [26] Fruhwirth-Schnatter S. Markov Chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*. 2001; **96**:194–209
- [27] Hurn M, Justel A, Rober C. Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*. 2003;**79**:55–79
- [28] Kim C, Nelson C. *State-Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. Cambridge, MA: MIT Press; 1999
- [29] Krozlig H. *Markov-Switching Vector Autoregressions*. Berlin, Heidelberg: Springer-Verlag; 1997