ORIGINAL RESEARCH PAPER

An MCMC Computational Approach for A Continuous Time State-dependent Regime Switching Diffusion Process

El Houcine Hibbah [a], Hamid El Maroufy[a], Christiane Fuchs[bcd] and Taib Ziad[e]

[a]Department of Applied Mathematics, Faculty of Sciences and Technics, Sultan Mouly Slimane University, Morocco;
[b] Bielefeld University, Faculty of Business Administration and Economics, Bielefield, Germany;
[c] Helmholtz Zentrum Munchen, German Research Center for Environmental Health GmbH, Institute of Computational Biology, Neuherberg, Germany;
[d] Chair of Mathematical Modeling of Biology Systems, Technisché Universität München, Garching, Germany;
[e] Faculty of Mathematics, Chalmers University of Technology, Göteborg, Sweden.

ARTICLE HISTORY
Compiled December 25, 2019

ABSTRACT
State-dependent regime switching diffusion processes or hybrid switching diffusion processes (HSD) are hard to simulate with classical methods which leads us to adopt an MCMC Bayesian approach very convenient to estimate complicated models such as the HSD one. In the HSD, the diffusion component is dependent on the switching discrete hidden regimes and the transition rates of the regime switching are dependent on the diffusion observations. Since in reality phenomena are only observed in discrete times, data imputation is called for to create more observations so as to have good approximations for the density of the diffusion process. Three categories of entities will be computed in a Bayesian context: The latent imputed observations, the regime switching states, and the parameters of the models. The latent imputed data is updated at random time intervals in block using a Metropolis Hastings algorithm. The switching states are computed by an adaptation of a forward filtering backward smoothing algorithm to a the HSD model. The parameters are estimated after prior specifications and conditional posterior densities formulation using Gibbs sampler or Metropolis Hastings algorithm.

KEYWORDS
Data imputation; Hidden states computation; Hybrid switching diffusion model; Latent state estimation; Metropolis-Hastings algorithm; Random time imputation; State dependent transition rate matrix.

1. Introduction

Understanding phenomena dynamic such as disease progression is very challenging because it incorporates behavioural switching in continuous time; which requires us to apply a methodology that is suitable when data is observed at irregular times. Indeed, one should appeal to a rich class of continuous time models. Those models are usually taken to be diffusion processes such as the Ornstein-Uhlenbeck models [13] . Even

CONTACT El Houcine Hibbah. Email: maroufy.hibbah@gmail.com

though the usefulness of diffusion processes and the many advantages they can offer, they can't handle all the dynamics caused by some complex phenomena. In many cases, phenomena change behaviour or evolution depending on intermediate indicators or variables. For example in economics the states can be the economic situations (growth, stability, crisis)and the developed variable can be options pricing or any economic factors that are changing attitude due to different economics situations. Also, in ecology [4] , animal movement has been shown to change behaviour depending of the state of whether the animal is resting at a nest, foraging, or heading for a nest. Our modeling can be well present in disease progression such as HIV progression [23] , where an observed bio-marker called CD4 counts decline is varying from one HIV disease stage to another. Thus, we recourse for diffusion processes but with the extension of regime switching. By adding the regime switching characteristics, the regime switching model (RSM) will be combining two components: a continuous diffusion observation component and an unobserved component most of the time discrete and supposed to be Markovian. Consequently, RSM are usually supposed to be Markovian and hence called Markov regime switching models (MRS). Their underlying idea is the switching mechanism who is supposed to be latent (unobserved or unknown). The simplest model supposes a two-dimensional Markov processes, where the first component is continuous and real valued and depending on the second, and the second has discrete values and acts monotonously from the first process. Such models have attracted a lot of attention in diverse fields or sciences such as in population dynamics [28] , in pattern recognition [17] , or in ecology [5,6] . However in our case, we will go a little further and direct our attention toward a more advanced switching diffusion process that supposes the dependence of the rate of the switching or of the transition on the observations or covariates. A fundamental theoretical description and practical significance related to this modelling is provided in [8]. A recent example with the Bayesian estimation methodology is presented in [4], where animal movement is an adaptive movement since it depends and follows the habitat (resting or foraging). Those kinds of models are called HSD models and highlight the coexistence of continuous dynamics and discrete events: one component describes the continuous dynamics, whereas the other is a switching process representing discrete events with the switching part depending on the continuous dynamics. The exception to RSM is that the rates of transition for the discrete states in HSD process depend on the continuous dynamics [21,27] .

In this paper, we will provide a Markov chain Monte Carlo (MCMC) method to estimate the parameters of a HSD model. While we could find classical methods to compute the parameters of a RSM as in [22,24] where a maximum likelihood and expectation-maximization procedures are adopted, likelihood inference is very challenging for HSD. In fact, the diffusion component is non linear most of the time and most of the diffusion phenomena of practical interest are nonlinear in their nature [32]. Furthermore, the transition rates of the regime switching are dependent on the diffusion observations [4,8]. Moreover, the observations are available only at scattered discrete times or are subject to measurement errors [20] . This leads us to impute data between successive observations to find an efficient approximation to the unavailable transition density of the HSD using an Euler approximation in the case of HSD as in [33] . The method of imputation aims at augmenting the likelihood with the imputed data. In the Bayesian context, the procedure alternates between updating the imputed data and updating the parameters. Unfortunately the update of the imputed data can suffer from poor mixing such as the single site update of [16], or has a long mixing times of MCMC algorithm such as the block update of [14] if the number of imputation

data is so large. The challenge of our modelling is that we have to augment the likelihood with the hidden states beside the imputed data and the parameters. Also, unlike regular state independent Markov switching diffusion process, our model supposes the dependence of the transition rate matrix on the diffusion process. Furthermore, instead of fixed times imputation, we opt for the algorithm of random time points update as described in [35] and applied to HSD in a Bayesian context in [4] . These will avoid any discretization error, where the times and the imputed data proposed are accepted or rejected using a Metropolis Hastings algorithm (MHA). Thus, we will appeal to the Bayesian inference through MCMC methods to sample the posterior distribution for the parameters, the latent imputed data and the hidden states for the uni-dimensional regime switching process since the discrete states of the switches are supposed unknown in this model. We will extend the Bayesian inference and data imputation of [19, Ch. 7] for non-linear diffusion to uni-dimensional HSD using the random time intervals for the imputation as in [4].
This paper is organized as follows: In section 2, we describe the model considered. Section 3 provides the Bayesian inference for the latent imputed data, the hidden states, and the parameters including the parameters of transition rate matrix. In section 4, we see more details on how priors are chosen and how posteriors densities are computed with an illustration by an important non linear HSD process, and an application to disease progression. Finally we give a discussion and a conclusion.

2. Model description and notations

We consider a continuous time HSP of the following form

$$\begin{aligned} &dX(t) = \mu(X(t), S(t))dt + \sigma(X(t), S(t))dW_t. \\ &X(t) = X_0, \text{ and } S(t) = S_0, \end{aligned} \tag{1}$$

with $P(S(t+\delta t) = j | S(t) = i, X(s), S(s), s \leq t) = q_{ij}(X(t))\delta t + o(\delta t)$, $i \neq j$, $\mu()$ and $\sigma(.)$ are appropriate real valued functions satisfying certain regularity conditions . W is a uni-dimensional standard Brownian motion. $S(.)$ is a switching process taking discrete values in $\{1, .., a\}$ with a is an integer and with the dynamics of $S(.)$ depending on $X(t)$. Specifically, the spontaneous transition rate matrix $Q(x,t) = (q_{kl}(x,t))_{1 \leq k,l \leq a}$ for the switching process has the following properties:

(i) $q_{kl}(x,t) \geq 0$, for $k \neq l$, for any x and time t.
(ii) $\sum_l q_{kl}(x,t) = 0$, for any x and time t.

[45] has summarized the difference between continuous state dependent switching diffusion process and Markovian switching models in term of properties of solutions of the process and numerical procedures.
We suppose we have N individuals with n_i the number of observation times for each individual i such that the trajectory for each individual i is given by $x_i = (x_{i1}, ..., x_{in_i})$. We consider t_{ij} the times of observations for individual i, and the latent switching states will be denoted s_{ij} for individual i at time t_{ij}. Time points are usually different between subjects, also n_i may differ between individuals. Θ is the set of the parameters in the model, while θ is the set of all parameters except the parameters related to the transition rate matrix. Hence, by applying the Markov property and the conditional

probability and Bayes rules, the likelihood could be written as:

$$L(X,S;\Theta) \propto \prod_{i=1}^{N}(P(x_{i1},s_{i1},\Theta)\prod_{j=2}^{n_i}P(x_{ij},s_{ij}|x_{ij-1},s_{ij-1},\Theta)$$

$$\propto \prod_{i=1}^{N}(P(x_{i1},s_{i1},\Theta)\prod_{j=2}^{n_i}P(x_{ij}|x_{ij-1},s_{ij},s_{ij-1},\Theta)P(s_{ij}|x_{ij-1},s_{ij-1},\Theta)$$

$$\propto \prod_{i=1}^{N}(P(x_{i1},s_{i1},\Theta)\prod_{j=2}^{n_i}P(x_{ij}|x_{ij-1},s_{ij-1},\Theta)P(s_{ij}|x_{ij-1},s_{ij-1},\Theta)$$

Applying exponential matrix to the switching process gives:

$$L(X,S;\Theta) \propto \prod_{i=1}^{N}(P(x_{i1},s_{i1},\Theta)\prod_{j=2}^{n_i}P(x_{ij}|x_{ij-1},s_{ij-1},\Theta)\times$$

$$[\exp(Q(x_{ij-1})\Delta t_{ij})]_{s_{ij-1},s_{ij}}$$

$P(s_{ij}|x_{ij-1},s_{ij-1},\Theta)$ is the element (s_{ij-1},s_{ij}) of the transition probability matrix evaluated at the diffusion location x_{ij-1}. Since the transition matrix is obtained from the exponential of the generator Q evaluated at x_{ij-1}, and we obtain the elements $[\exp(Q(x_{ij-1})\Delta t_{ij})]_{s_{ij-1},s_{ij}}$, with $\Delta t_{ij} = t_{ij} - t_{ij-1}$. This exponential approximation supposes constant x_t for a short period of time Δt_{ij}.

3. Bayesian inference

Direct estimation of HSD processes is very difficult because the transition density of the diffusion component is most of the time unavailable in a closed form and we need to approximate it. Thus, to overcome the issue of low frequency data and consequently the density approximation can be used, we need to introduce intermediate data between successive observations. Thus, we appeal to Bayesian data imputation [15,16,34] . Hence an $m_{ij}-1$ missing observations are imputed between successive observations x_{ij} and x_{ij-1} to obtain for each individual i a vector : $X_i = (x_{i1}, X_{i1}^2, ..., X_{i1}^{m_{i1}}, x_{i2}, X_{i2}^2 ....., x_{in_{i-1}}, X_{in_{i-1}}^2, ..., X_{in_{i-1}}^{m_{in_i}}, x_{in_i})$, with $X_{ij}^1 = x_{ij}$ and X_{ij}^k will have corresponding S_{ij}^k at the time point t_{ij}^k (the k^{th} imputation time after observation x_{ij}) . Since we have three categories of variables to estimate: The parameters Θ, the latent switching state S and the imputed (latent) data X, our imputed joint posterior density is formulated as follows:

$$P(\Theta,X,S|x) \propto P(\Theta)\prod_{i=1}^{N}(P(X_{i1}^1,S_{i1}^1|\Theta)\times$$

$$\prod_{j=1}^{n_i-1}P(X_{ij+1}^1|X_{ij}^{m_{ij}},S_{ij}^{m_{ij}},\Theta)\left[\exp(Q(X_{ij}^{m_{ij}})\Delta t_{ij+1}^1)\right]_{S_{ij}^{m_{ij}},S_{ij+1}^1}\times$$

$$\prod_{j=1}^{n_i-1}\prod_{k=1}^{m_{ij-1}}P(X_{ij}^{k+1}|X_{ij}^k,S_{ij}^k,\Theta)\left[\exp(Q(X_{ij}^k)\Delta t_{ij}^{k+1})\right]_{S_{ij}^k,S_{ij}^{k+1}},$$

To evaluate the conditional posterior, we use a numerical Euler approximation [25] for the small time t_{ij}^k (the k^{th} imputation time between observation x_{ij} and x_{ij+1}, with the exception that $\Delta t_{ij+1}^1 = t_{ij+1}^1 - t_{ij}^{m_{ij}}$ and t_{ij}^1 is the time at the observation x_{ij}) , and we have: $\pi^{Euler}(X_{ij}^{k+1}|X_{ij}^k, S, \Theta) \approx \phi(X_{ij}^{k+1}|X_{ij}^k + \mu(X_{ij}^k, S_{ij}^k, \Theta)\Delta t_{ij}^{k+1}, \sigma(X_{ij}^k, S_{ij}^k, \Theta)\Delta t_{ij}^{k+1})$, with S_{ij}^k the regime switching state value for individual i at time imputation t_{ij}^k, with $\Delta t_{ij}^{k+1} = t_{ij}^{k+1} - t_{ij}^k$ and $\phi(z|\nu, \Lambda)$ is a normal distribution with mean ν and variance Λ. $P(\Theta)$ is the prior distribution for the set of parameters Θ. The MCMC algorithm will generate a Markov chain targeting the augmented posterior $P(\Theta, X, S|x)$ under some mild regularity conditions. In fact, the data augmentation procedure of [40] inference may proceed by alternating between the simulation of the parameters conditional on the augmented data, and the simulation of the augmented data given the observed data and the current state of the model parameters. A discussion of convergence issues related to non linear diffusion processes with latent data is provided in [20]. Such issues include the update choice for the imputed data and that the m_{ij} shouldn't be so large in order to avoid the slow mixing of the targeted chain in the MCMC algorithm. Hence, we perform our simulation by alternating between drawing from the following conditional posteriors given all other quantities (denoted by .):

(i) Draw the latent(imputed) observations from $P(X|.)$
(ii) Draw the latent switching states from $P(S|.)$
(iii) Draw the parameters from $P(\Theta|.)$

3.1. *Sampling the latent data*

While it is possible to update each latent(imputed) data separately, we will adopt a block update for the whole sequence of imputed data for each individual observation X_i. Since it is difficult to come with an analytical form for the posterior distribution of the imputed data, a MHA [19, Ch. 7] is utilized to draw the new imputed data X_i^{new} from an old X_i^{old}, for $i = 1, ..., n$. Though for the imputation of auxiliary data, we opt for the algorithm of random time points update as described in [35] and applied to HSD in a bayesian context in [4]. These will avoid any discretization error. Suppose that the transition rates are all bounded above and let the times will be generated from a Poisson process with parameter $\kappa > \max_{i,x,t}\{-Q_{ii}(x,t)\}$, and the algorithm proceeds for each individual $i = 1, ..., N$ as follow:

Algorithm. 1 :

(i) For $j = 1, ..., n_{i-1}$
(ii) For each interval $[t_{ij}, t_{ij+1}]$, the times are generated from a Poisson process with parameter κ, and we obtain the $t_{ij}^{new,k}$ time points for $k = 1, ..., m_{ij}$.
(iii) for $k = 1$: $X_{ij}^{new,1} = X_{ij}^{old,1} = x_{ij}$ and $S_{ij}^{new,1} = S_{ij}^{old,1}$.
(iv) for each $k = 2, ..., m_{ij}$,
  (a) we determine whether there is a switch in time with probability $\frac{-Q_{S_{ij}^{new,k-1}, S_{ij}^{new,k-1}}(X_{ij}^{new,k-1})}{\kappa}$.
  (b) If there is a switch, we sample the new state $S_{ij}^{new,k} = C$ with probability

$$\frac{Q_{S_{ij}^{new,k-1},C}(X_{ij}^{new,k-1})}{-Q_{S_{ij}^{new,k-1},S_{ij}^{new,k-1}}(X_{ij}^{new,k-1})}$$

(c) We propose the new imputed $X_{ij}^{new,k}$ following the last update of time $t_{ij}^{new,k}$ and $S_{ij}^{new,k-1}$.

(v) We ran a MHA to decide on the acceptance of the new time proposal $T_i^{new} = \{t_{ij}^{new,k},\, k = 1, ..., m_{ij},\, j = 1, ..., n_{i-1}\}$ as well as the proposed imputed data X_i^{new} .

So by introducing $m_{ij} - 1$ new observations between two successive observations x_{ij-1} and x_{ij}, the MHA acceptance ratio will be:

$$\zeta(X_i^{new}, T_i^{new}; X_i^{old}, T_i^{old}) = 1 \wedge \frac{P(X_i^{new}|D, S, \Theta)\psi(X_i^{old}|X_i^{new}, D, S, \Theta)}{P(X_i^{old}|D, S, \Theta)\psi(X_i^{new}|X_i^{old}, D, S, \Theta)}.$$

Where P denote the posterior density, ψ the proposal density, and D the observed, not the imputed data. Due to the Markov property in HSD, we have:

$$\begin{aligned}
&\frac{P(X_i^{new}|D, S, \Theta)}{P(X_i^{old}|D, S, \Theta)} \\
&= \prod_{j=1}^{n_i-1} \frac{P(X_{ij+1}^1|X_{ij}^{new,m_{ij}}, S, \Theta)}{P(X_{ij+1}^1|X_{ij}^{old,m_{ij}}, S, \Theta)} \prod_{j=1}^{n_i-1} \prod_{k=1}^{m_{ij-1}} \frac{P(X_{ij}^{new,k+1}|X_{ij}^{new,k}, S, \Theta)}{P(X_{ij}^{old,k+1}|X_{ij}^{old,k}, S, \Theta)} \\
&= \prod_{j=1}^{n_i-1} \frac{P(\Delta t_{ij+1}^{new,1}, X_{ij}^{new,m_{ij}}, X_{ij+1}^1)}{P(\Delta t_{ij+1}^{old,1}, X_{ij}^{old,m_{ij}}, X_{ij+1}^1)} \prod_{j=1}^{n_i-1} \prod_{k=1}^{m_{ij-1}} \frac{P(\Delta t_{ij}^{new,k+1}, X_{ij}^{new,k}, X_{ij}^{new,k+1})}{P(\Delta t_{ij}^{old,k+1}, X_{ij}^{old,k}, X_{ij}^{old,k+1})}
\end{aligned}$$

$\Delta t_{ij}^{k+1} = t_{ij}^{k+1} - t_{ij}^k$ is very small to permit an Euler approximation for P by $P^{Euler}(X_{ij}^{k+1}|X_{ij}^k, S, \Theta) \approx \phi(X_{ij}^{k+1}|X_{ij}^k + \mu(X_{ij}^k, S_{ij}^k, \Theta)\Delta t_{ij}^{k+1}, \sigma(X_{ij}^k, S_{ij}^k, \Theta)\Delta t_{ij}^{k+1})$, with S_{ij}^k the regime switching state value for individual i at time imputation t_{ij}^k , and $\phi(z|\nu, \Lambda)$ is normal distribution with mean ν and variance Λ.
While to the choice of the proposal density ψ, one could choose an Euler proposal or a double-sided Euler proposal, but due to the dependency between X^{new} and X^{old}, we will adopt the modified bridge proposal ψ_{MB} of [14] ; supposed to overcome the issue of dependence between successive draws; as adopted in the Bayesian context by [10]. Note that $X_{ij}^{new,1} = X_{ij}^{old,1} = x_{ij}$ and $S_{ij}^{new,1} = S_{ij}^{old,1}$, and the proposed path will be accepted with probability:

$$\begin{aligned}
\zeta(X_i^{new}, T_i^{new}; X_i^{old}, T_i^{old}) = 1 \wedge &\left(\prod_{j=1}^{n_i-1} \frac{P(X_{ij+1}^1|X_{ij}^{new,m_{ij}}, S, \Theta)}{P(X_{ij+1}^1|X_{ij}^{old,m_{ij}}, S, \Theta)} \right) \times \\
&\left(\prod_{j=1}^{n_i-1} \prod_{k=1}^{m_{ij-1}} \frac{P(X_{ij}^{new,k+1}|X_{ij}^{new,k}, S, \Theta)}{P(X_{ij}^{old,k+1}|X_{ij}^{old,k}, S, \Theta)} \right) \times \frac{\psi_{MB}(X_i^{old}|X_{i1}, X_{in_i}, S, \Theta)}{\psi_{MB}(X_i^{new}|X_{i1}, X_{in_i}, S, \Theta)}, \quad (2)
\end{aligned}$$

where $\frac{\psi_{MB}(X_i^{old}|X_{i1}, X_{in_i}, S, \Theta)}{\psi_{MB}(X_i^{new}|X_{i1}, X_{in_i}, S, \Theta)} = \prod_{j=1}^{n_i-1} \prod_{k=1}^{m_{ij-1}} \frac{\psi_{MB}(X_{ij}^{old,k+1}|X_{ij}^1, X_{ij+1}^1, S, \Theta)}{\psi_{MB}(X_{ij}^{new,k+1}|X_{ij}^1, X_{ij+1}^1, S, \Theta)}$

with $\psi_{MB}(X_{ij}^{new,k+1}|X_{ij}^1, X_{ij+1}^1, S, \Theta) =$

$\phi(X_{ij}^{new,k+1}|X_{ij}^{new,k} + \frac{X_{ij}^{m_{ij}} - X_{ij}^{new,k}}{t_{ij+1}^{new,1} - t_{ij}^{new,k}} \Delta t_{ij}^{new,k+1}, ...$

$\frac{t_{ij+1}^{new,1} - t_{ij}^{new,k+1}}{t_{ij+1}^{new,1} - t_{ij}^{new,k}} \sigma(X_{ij}^{new,k}, S_{ij}^{new,k}, \Theta) \Delta t_{ij}^{new,k+1})$.

3.2. *Sampling the switching hidden states*

Instead of simulating each hidden state separately, Chib [9] developed a method called block update of the hidden states in which we simulate the full latent data for each individual i, $i = 1, 2, ..., N$. Its algorithm was adapted to a multivariate autoregressive hidden Markov model in [30]. We will adapt this algorithm to the HSD process, with the modification that our model suppose observations with different lengths and non equidistant intervals. For clarity of representation and for good understanding of this so called forward filtering backward smoothing (FFBS) algorithm, we stack the imputed observations with the observations in one vector of size N_i (The staked data contains both the observation data x_i " and the imputed terms X_{ij}^k to give the new stacked vector $\tilde{X}_i$) $\tilde{X}_{i.} = (\tilde{X}_{i1}, ..., \tilde{X}_{iN_i})$, with the corresponding switching hidden vector $\tilde{S}_{i.} = (\tilde{S}_{i1}, ..., \tilde{S}_{iN_i})$, and the times $\tilde{t}_{i.} = (\tilde{t}_{i1}, ..., \tilde{t}_{iN_i})$. The times contains both the original as well as the new times staked in a new vector and we use the forward filtering backward smoothing to sample the whole sequence of the hidden states. A similar procedure is described in [35]. Let denote $\tilde{X}_i^{-j} = (\tilde{X}_{i1}, ..., \tilde{X}_{ij})$, $\tilde{X}_i^j = (\tilde{X}_{ij}, ..., \tilde{X}_{iN_i})$, $\tilde{S}_i^{-j} = (\tilde{S}_{i1}, ..., \tilde{S}_{ij})$, and $\tilde{S}_i^j = (\tilde{S}_{ij}, ..., \tilde{S}_{iN_i})$. Now, we write the joint conditional distribution for the hidden states as:

$$P(\tilde{S}_{i.}|\tilde{X}_{i.}, \Theta) = \prod_{j=1}^{N_i} P(\tilde{S}_{ij}|\tilde{X}_{i.}, \tilde{S}_i^{j+1}).$$

Hence the states computation is based on the term $P(\tilde{S}_{ij}|\tilde{X}_{i.}, \tilde{S}_i^{j+1})$ which will be evaluated in the backward pass after running the forward filtering that proceeds as :

Algorithm. 2:

(i) Initialize for the time $j = 1$.

(ii) For $j = 2, ..., N_i$, and $k = 1, ..., a$, compute and alternate between :

(a) $P(\tilde{S}_{ij} = k|\tilde{X}_i^{-(j-1)}, \Theta) \propto \sum_{l=1}^{a} P(\tilde{S}_{ij} = k|\tilde{S}_{ij-1} = l, \tilde{X}_{ij-1}) \times$

$P(\tilde{S}_{ij-1}|\tilde{X}_i^{-(j-1)}, \Theta)$

(b) $P(\tilde{S}_{ij} = k|\tilde{X}_i^{-j}, \Theta) \propto P(\tilde{S}_{ij} = k|\tilde{X}_i^{-(j-1)}, \Theta) * f(\tilde{X}_{ij}|\tilde{X}_{ij-1}), \tilde{S}_{ij-1}, \Theta)$.

Later on, the backward smoothing proceeds by:

(i) Initialize for the time $j = N_i$ from the last forward quantity $P(\tilde{S}_{iN_i} = k|\tilde{X}_{i.}, \Theta)$.
(ii) For $j = N_i - 1, ..., 2, 1$, and $k = 1, ..., a$, compute and alternate between :
  (a) $P(\tilde{S}_{ij+1}|\tilde{S}_{ij}, \tilde{X}_{ij})$ is the element $(\tilde{S}_{ij}, \tilde{S}_{ij+1})$ of the transition probability matrix evaluated at the diffusion location $\tilde{X}_{ij}$. Since the transition matrix is obtained from the exponential of the generator Q evaluated at $\tilde{X}_{ij}$, and we obtain the element $\left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]_{\tilde{S}_{ij},\tilde{S}_{ij+1}}$ for small time $\Delta\tilde{t}_{ij+1}$, assuming constant $\tilde{X}_{ij}$ for a short period of time $\Delta\tilde{t}_{ij+1}$.
  (b) $P(\tilde{S}_{ij}|\tilde{X}_{i.}, \tilde{S}_i^{j+1}) \propto P(\tilde{S}_{ij} = k|\tilde{X}_i^{-j}, \Theta) * P(\tilde{S}_{ij+1}|\tilde{S}_{ij}, \tilde{X}_{ij})$.
  (c) Use those last probabilities to draw the hidden states.

Again, let's remind that in the HSD model, the transition rates are related to the observed diffusion component.

3.3. *Sampling from* $P(\theta|\tilde{X}, \tilde{S})$

Given the estimation of $\tilde{X}$ and $\tilde{S}$ in the previous subsection, the conditional posterior of θ is proportional to the prior $P(\Theta)$ multiplied by the likelihood, which gives:

$$P(\theta|\tilde{X}, \tilde{S}) \propto P(\theta)P(X_{i1}^1, S_{i1}^1|\Theta) \times \prod_{i=1}^{N}\prod_{j=1}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta)\left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]_{\tilde{S}_{ij},\tilde{S}_{ij+1}}$$

Usually, we don't come with a standard distribution to sample from by using Gibbs sampler, thus we will appeal to a MHA. After using proportionality and dropping the terms of the transition rate matrix since they are independent of θ. Hence, we come with a MHA [19, Ch. 7] with an acceptance probability ratio for every new proposal θ^* as:

$$\zeta(\theta, \theta^*) = 1 \wedge \frac{P(\theta^*)}{P(\theta)} \times \left(\frac{\prod_{i=1}^{N}\prod_{j=1}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta^*)}{\prod_{i=1}^{N}\prod_{j=1}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta)}\right) \times \frac{\psi(\theta|\theta^*, \tilde{X}, \tilde{S})}{\psi(\theta^*|\theta, \tilde{X}, \tilde{S})}$$

As before, using Euler approximation we have: $P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta) = P^{Euler}(\tilde{X}_{ij+1}|\tilde{X}_{ij}, S, \theta) \approx \phi(\tilde{X}_{ij+1}|\tilde{X}_{ij+1} + \mu(\tilde{X}_{ij+1}, \tilde{S}_{ij}, \theta)\Delta\tilde{t}_{ij}, \sigma(\tilde{X}_{ij}, \tilde{S}_{ij}, \theta)\Delta\tilde{t}_{ij+1})$, ψ is a proposal density to draw a new Θ^*, it depends on the form of the likelihood and the hypothesis of the HSD. Some times, we could use simply a random walk proposal which is independent of $\tilde{X}$ and $\tilde{S}$ and it is only related to the old draw $\Theta^{(old)}$, and we simply propose from a gaussian distribution for some $\theta_j^{new} \sim \mathcal{N}(\theta_j^{old}, \epsilon)$. ϵ is the random walk step that can be adjusted to improve convergence. For positive values,

we could propose from the log-normal distribution: $log\,\theta_j^{new} \sim \mathcal{LN}(log\,\theta_j^{old}, \epsilon)$.
Finally, and since we have a HSD, the parameters will be dependent on the switching process $\tilde{S}$, so for every given parameter θ_j, we will have to estimate $(\theta_j^1, ..., \theta_j^k, ..., \theta_j^a)$, and the posterior for θ_j^k is

$$P(\theta_j^k|\tilde{X}, \tilde{S}) \quad \propto \quad P(\theta) \times \prod_{i=1}^{N} \prod_{j=1, \tilde{S}_{ij}=k}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \theta). \tag{3}$$

4. Numerical example with application to disease progression

We will consider here an example of a non linear HSD process and give more details on how we can estimate each parameter of the model. In fact, a non linear drift and a non linear volatility would allow the HSD process to dispose of elements to represent clearly any complex problem and to be more flexible. Considering the non-linear process of [1] with an extra addition of the regime switching and we get in our case $d\tilde{X}_{ij} = \left[\alpha_{0,k} + \alpha_{1,k}\tilde{X}_{ij} + \alpha_{2,k}\tilde{X}_{ij}^2 + \frac{\alpha_{3,k}}{\tilde{X}_{ij}}\right] d\tilde{t}_{ij+1} + \sigma\tilde{X}_{ij}^{\eta_k} dW$, for the hidden switching state $k \in \{1, ..., a\}$. This general form with the regime switching gives many explanations for majority of phenomena, where a negative $\alpha_{2,k}$ guarantees ergodicity and second order stationarity for a volatility function $\sigma_k(\tilde{X}_{ij}) = \sigma\tilde{X}_{ij}^{\eta_k}$. Than the first order Euler approximation of this model for a state $\tilde{S}_{ij} = k$:

$$\tilde{X}_{ij+1} - \tilde{X}_{ij} = \left[\alpha_{0,k} + \alpha_{1,k}\tilde{X}_{ij} + \alpha_{2,k}\tilde{X}_{ij}^2 + \frac{\alpha_{3,k}}{\tilde{X}_{ij}}\right] \Delta\tilde{t}_{ij+1} + \sigma\tilde{X}_{ij}^{\eta_k}\sqrt{\Delta\tilde{t}_{ij+1}}\epsilon,\ \epsilon \sim \mathcal{N}(0,1), \tag{4}$$

4.1. *Computation of α_k and σ^2:*

To simulate our parameters α_k and σ^2 using the MCMC algorithm, first let mention that when the posterior density is not known we have to use a MHA as described in parameters update 3, otherwise if we come up with a known posterior density to draw from it directly using Gibbs sampler as it is the case here. For this reason, let's pose $Y_{ij} = \frac{\tilde{X}_{ij+1} - \tilde{X}_{ij}}{\tilde{X}_{ij}^{\eta_k}}$, $\beta_k = (\alpha_{0,k},\, \alpha_{1,k},\, \alpha_{2,k},\, \alpha_{0,k})$ and $y_{ij} = \left(\frac{\sqrt{\Delta\tilde{t}_{ij+1}}}{\tilde{X}_{ij}^{\eta_k}}, \frac{\tilde{X}_{ij}\sqrt{\Delta\tilde{t}_{ij+1}}}{\tilde{X}_{ij}^{\eta_k}}, \frac{\tilde{X}_{ij}^2\sqrt{\Delta\tilde{t}_{ij+1}}}{\tilde{X}_{ij}^{\eta_k}}, \frac{(1|\tilde{X}_{ij})\sqrt{\Delta\tilde{t}_{ij+1}}}{\tilde{X}_{ij}^{\eta_k}}\right)$. In matrix form, the Euler discretization may be represented as in [38]: $Y = y\beta_k + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma^2)$, which is the formulation for a regression model. Consequently, the parameters of the HSD can be easily computed using the Bayesian approach for regime switching regression model. Hence for each $\beta_k, k = 1, ..., a$, the posterior is proportional to the prior multiplied by the likelihood

$$P(\beta_k|\tilde{X}, \tilde{S}) \quad \propto \quad P(\beta) \times \prod_{i=1}^{N} \prod_{j=1, \tilde{S}_{ij}=k}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \Theta).$$

As in ([18, p:251]); by supposing a normal and an inverse gamma prior respectively for β_k and σ^2: Consequently with a conjugate normal prior for $\beta_k \sim \mathcal{N}_4(b_{0k}, \sigma^2 B_{0K})$,

we have $\beta_k|. \sim \mathcal{N}_4(b_k, B_K)$, where
$B_k = (B_{0K}^{-1} + y'y)^{-1}$ and $b_K = B_k(B_{0K}^{-1}b_{0k} + y_k^{'}Y_k)$; Y_k and y_k corresponds only to the observations where $\tilde{S}_{ij} = k$.
Similarly, under an inverted conjugate prior for $\sigma^2 \sim \mathcal{IG}(c_0, C_0)$, the posterior density for σ^2 given the observations, the imputed data, the other parameters and the switching states is $\sigma^2|. \sim \mathcal{IG}(c, C)$, where
$c = (c_0 + \frac{M}{2})$, and $C = (C_0 + \frac{1}{2}\sum_{k=1}^{a} b_{0k}^{'}B_{0K}^{-1}b_{0k} + \frac{1}{2}\sum_{k=1}^{a}(y_k^{'}y_k - b_k^{'}B_K^{-1}b_k))$, M is the number for all individuals observations.

Ridge regression:

One of the problem that can be faced here is the effects of multi-collinearity on the Bayesian regression estimation. Consequently, we recourse to the Bayesian Ridge Regression approach as in [43] who supposes the same priors as earlier for β_k and σ^2 and we get agian similar posteriors as before, the only difference is that in the ridge regression we have $B_{0K} = Diagonal(R_1, R_2, R_3, R_4)$. Moreover, to overcome the issue of choosing fixed value for R_j; we suppose the prior $P(R_1, R_2, R_3, R_4) \sim \prod_{j=1}^{4} R_j^{\frac{A}{2}-1}\exp(-\frac{b}{2}R_j)$. Hence, we alternate between updating the following posteriors, for each $k = 1, ..., a$:

$$
\begin{aligned}
R_{jj} &\sim Gamma(\frac{A + M_k}{2}, \frac{b + \beta_j^2}{2}),\ M_k: \text{the number of observations where } \tilde{S}_{ij} = k. \\
B_{0K} &= Diag(R_1, R_2, R_3, R_4). \\
\beta_k &\sim \mathcal{N}_4(b_{0k}, \sigma^2 B_{0K}). \qquad (5) \\
\sigma^2 &\sim \mathcal{IG}(c, C). \qquad (6)
\end{aligned}
$$

4.2. *Sampling the posterior dependent parameters on the transition rate matrix Q:*

In this example, we will suppose that the intensities of the transition rate matrix are dependent on the observed diffusion process through the relation of Gompertz model as in [44]. The consideration of the Gompertz model comes from the fact that this function has been used longer in insurance, in biology such as tumor evolution or bacteria growth and in many other fields [41]. Also the Gompertz model has interpretable parameters, and we have: $q_{kl}(\tilde{X}_{ij}) = \lambda_{kl}\exp(-\gamma_{kl}\tilde{X}_{ij})$, $\lambda_{kl}, \gamma_{kl} > 0$, for $k \neq l \in \{1, ..., a\}$, $i = 1, ..., N$; $and\, j = 1, ..., N_i$. We will suppose prior independence between the parameters of Θ. While the $\lambda's$ and the $\gamma's$ can be updated in block as in [29], here we adopt an approach similar to [37], where each element is updated conditional on the other since they are correlated. Hence by supposing a Gamma prior $\mathcal{G}(0.01, 0.01)$ for the λ_{kl}, the posterior of λ_{kl} will be:

$$
P(\lambda_{kl}|\tilde{X}, \tilde{S}) \propto P(\lambda_{kl})\prod_{i=1}^{N}\prod_{j=1}^{N_i-1}\left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right)\right]_{\tilde{S}_{ij},\tilde{S}_{ij+1}}.
$$

As we can see, all the other parameters especially θ are omitted due to the Bayes rule. Consequently, we obtain a non standard posterior and we have to adopt a random walk MHA to draw a new λ_{kl}^{new} (the new MCMC iteration) from an old value λ_{kl}^{old} (the

previous MCMC iteration) with an acceptance probability :

$$\zeta(\lambda_{kl}^{new}, \lambda_{kl}^{old}) = \\ 1 \wedge \left(\frac{P(\lambda_{kl}^{new}) \prod_{i=1}^{N} \prod_{j=1}^{N_i-1} \left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right) \right]_{\tilde{S}_{ij},\tilde{S}_{ij+1}}^{new}}{P(\lambda_{kl}^{old}) \prod_{i=1}^{N} \prod_{j=1}^{N_i-1} \left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right) \right]_{\tilde{S}_{ij},\tilde{S}_{ij+1}}^{old}} \right) \times \frac{\lambda_{kl}^{new}}{\lambda_{kl}^{old}}, \quad (7)$$

Where we propose the λ_{kl}^{new} from a log-normal distribution $\mathcal{LN}(\lambda_{kl}^{old}, \epsilon_\lambda)$ so as to keep operating on real positive values with ϵ_λ the random walk step. Similarly, a no close form is obtained for the posterior of γ_{kl}. So, we consider again a random walk MHA. With a Gamma prior $\mathcal{G}(0.01, 0.01)$ on γ_{kl} and a random walk proposal $\mathcal{LN}(\gamma_{kl}^{old}, \epsilon_\gamma)$, our MHA acceptance probability is:

$$\zeta(\gamma_{kl}^{new}, \gamma_{kl}^{old}) = 1 \wedge \left(\frac{P(\gamma_{kl}^{new}) \prod_{i=1}^{N} \prod_{j=1}^{N_i-1} \left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right) \right]_{\tilde{S}_{ij},\tilde{S}_{ij+1}}^{new}}{P(\gamma_{kl}^{old}) \prod_{i=1}^{N} \prod_{j=1}^{N_i-1} \left[\exp\left(Q(\tilde{X}_{ij})\Delta\tilde{t}_{ij+1}\right) \right]_{\tilde{S}_{ij},\tilde{S}_{ij+1}}^{old}} \right) \times \frac{\gamma_{kl}^{new}}{\gamma_{kl}^{old}}, \quad (8)$$

4.3. *Computation of η_k :*

To compute η_k, the posterior is proportional to the prior multiplied by the likelihood. By supposing a Gamma prior $\mathcal{G}(0.01, 0.01)$, we come with a non standard posterior, and we call for random walk MHA with the acceptance probability :

$$\zeta(\eta_k^{new}, \eta_k^{old}) = 1 \wedge \left(\frac{P(\eta_k^{new}) \prod_{i=1}^{N} \prod_{j=1,\tilde{S}_{ij}=k}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \eta_k^{new}, \Theta_{-\eta_k})}{P(\eta_k^{old}) \prod_{i=1}^{N} \prod_{j=1,\tilde{S}_{ij}=k}^{N_i-1} P(\tilde{X}_{ij+1}|\tilde{X}_{ij}, \tilde{S}_{ij}, \eta_k^{new}, \Theta_{-\eta_k})} \right) \times \frac{\eta_k^{new}}{\eta_k^{old}}. \quad (9)$$

$\Theta_{-\eta_k}$ represent all the parameters except η_k. Each η_k must be positive. As before we propose from a log-normal distribution $\mathcal{LN}(\eta_k^{old}, \epsilon_\eta)$ with step ϵ_η chosen concisely.

4.4. *Numerical implementation and simulation:*

To assess the accuracy of our finding, we will simulate observations for $N = 50$ individuals with follow up size between 16 and 20 for every individual. We suppose we have $a = 3$ regime switching states. the parameters of the non linear HSD are : $\alpha_0 = (0.3, 0.6, 0.9)$, $\alpha_1 = (0.02, 0.04, 0.06)$, $\alpha_2 = (-0.08, -0.06, -0.04)$, $\alpha_3 = (0.01, 0.05, 0.09)$, $\eta = (0.4, 0.5, 0.6)$, $\sigma^2 = 0.04$, and

$$\lambda = \begin{pmatrix} 0 & 0.2 & 0.4 \\ 2 & 0 & 3 \\ 0.3 & 0.9 & 0 \end{pmatrix}, \gamma = \begin{pmatrix} 0 & 0.2 & 0.04 \\ 0.01 & 0 & 4 \\ 1 & 0.8 & 0 \end{pmatrix}$$

The simulation algorithm work with the help of the Euler approximation for every individual i for $i = 1, ..., 50$, and we have:

Algorithm. 3:

(i) Choose n_i uniformly in $[16, 20]$ for $j = 1$
  (a) Choose the regime switching state $S_{i1} = k$ uniformly for k in $\{1, 2, 3\}$.
  (b) Initialize $X_{i,1} \sim \mathcal{N}(\alpha_{0,k}, \sigma^2)$.
  (c) Compute Q using formulation $Q_{kl}(X_{i1}) = \lambda_{kl} \exp(\gamma_{kl} X_{i1})$, $\lambda_{kl} > 0$, for $k \neq l \in \{1, 2, 3\}$, and $Q_{kk}(X_{i1}) = -\sum_{l \neq k} Q_{kl}(X_{i1})$.
  (d) Compute t_{i2} from $\exp(-Q_{kk}(X_{i1}))$

  For $j = 2, ..., n_i$:

(ii) Compute S_{ij} using the line S_{ij-1} of $Q(X_{ij-1})$

(iii) Compute X_{ij} for S_{ij-1} using Euler approximation for the model 4.

(iv) Compute Q using formulation $Q_{kl}(X_{ij}) = \lambda_{kl}\, exp(\gamma_{kl} X_{ij})$, $\lambda_{kl} > 0$, for $k \neq l \in \{1, 2, 3\}$, and $Q_{kk}(X_{ij}) = -\sum_{l \neq k} Q_{kl}(X_{ij})$.

(v) Compute t_{ij+1} from $\exp(-Q_{kk}(X_{ij}))$, if $j < n_i$.

To assess the efficiency of our methods, we will see how our MCMC algorithm can estimate the true values (values used to generate the simulated data). Before providing the MCMC algorithm, we should point out that we don't opt for the usual regularly spaced points imputation procedure that can drive the Bayesian estimation to break down if the amount of imputation is large [20]. In fact, it has been shown dependence between the unknown parameters in the diffusion and the missing data while adopting this imputation. This can result in slow rates of convergence of naive sampling or could conduct to identifiability problem as in the single update of [16] or the block update of [14]. Thus we call for random time imputation that allow exact estimation as in [4]. After the generation of the simulated data and to check the accuracy of the estimation, we ran the MCMC algorithm for a large number of iterations. During this ran there is a burnings period (where the algorithm hasn't converged yet). The burnings period varies depending on the volatility complexity, the number of parameters in the model, or the number of imputation as well as the latent data. After the burnings period, the MCMC converges and the inference is based on the last iterations of the MCMC algorithm. Our algorithm proceeds for 8000 iterations (number of iterations found in our case to be convenient for the MCMC to converge so as we can draw inference from the MCMC output) as follow:

Algorithm. 4:

(i) Initialize Θ

(ii) for $m = 2, ..., 8000$:
  (a) Propose the new times T_i^{new} using a Poisson process with parameter κ.
  (b) Propose the new imputed data X_i^{new} Using Euler approximation and Modified brownian proposal.
  (c) Accept the new proposal of the times as well as the imputed data using MHA (2).

(d) Stack the new data and the new times in new vector of data and times: $\tilde{X}$ and $\tilde{T}$.
(e) Simulate the regime switching states $P(\tilde{S}|.)$ using the FFBS algorithm .
(f) Simulate the parameters of the transition rate matrix Q: λ and γ using the random walk MHA (7) and (8) respectively .
(g) Compute $\beta_k|. \sim \mathcal{N}_4(b_k, B_K)$ from (5) for $k = 1, 2, 3$.
(h) Compute $\sigma^2|. \sim \mathcal{IG}(c, C)$ from (6).
(i) Simulate η_k using MHA (9), for $k = 1, 2, 3$.

Table 1 gives the posterior statistics of our algorithm such as the posterior mean, the standard deviation and the 95% credible intervals, where estimations are of good approximation to the true values even though that this complicated process has a large number of parameters. Consequently, parameter estimations have been rendered very simple. Moreover the convergence issues have been checked for every parameter of the model by the graphical inspection of the trace-plot, the kernel density, and the autocorrelation function plot. It is revealed that we get a good mixing of the MCMC chain (figure 1), a perfect density shape (figure 2) and autocorrelations that decay immediately after a few lag (figure 3) . Hence convergence for this MCMC method is achieved. In fact, the MCMC methods shows that they are appropriate especially that we have opted for random time imputation. Our algorithm works well if we choose different values for this switching one factor model, or ran a bootstrap simulation. Hence, the algorithm was efficient in approximating the true values.

4.5. *Application to disease progression:*

Many models have been proposed to model disease progression through markers observations, among them [46] proposed a deterministic differential equation model (DEM) to model markers in disease such as Alzheimers disease, Huntingtons disease, or Parkinsons disease. Taking the stochastic version of this DEM by adding the regime switching, we get a process similar to the previous regime switching one factor model with the following expression:
$d\tilde{X}_{ij} = \left[\alpha_{0,k} + \alpha_{1,k}\tilde{X}_{ij} + \alpha_{2,k}\tilde{X}_{ij}^2\right] d\tilde{t}_{ij+1} + \sigma dW$, for $k = 1, 2, 3$.
We applied this process to model marker observations from a slow developed disease: COPD (Chronic Obstructive Pulmonary Disease). Doctors use Gold stages to address the stage of the COPD (the severity of the disease). Knowing how severe the COPD is in a patient helps to choose the best treatment. The idea is to unravel the stages of the disease using one marker or a combination of many markers. In fact, there are many clinical markers for COPD disease among them: Chest hyperinflation, low body mass index (BMI), the use of accessory muscles of respiration, and prolonged expiration. Since, we are interested in one dimensional diffusion process, we take into consideration as a marker the FEV1 (how much air one can exhales from his lungs in one second, measured in Liter). We will see how the FEV1 observations allow to estimate the parameters of the HSD process. We have supposed that our model has three hidden states; $a = 3$ (Mild, Moderate, and severe) [39]. We extracted our FEV1 observations using data from the Danish Lung Cancer Screening Trial (DLCST) [36] where 2052 current or ex-smokers aged $50 - 70$ years having FEV1 measured annually for 5 years (2005-2009). We have for each subject, five FEV1 measurements with the

date of each measurement. While having more than 10 marker observations by patient would give more parameters precision and identification, disposing here of only 5 marker observations by patient is found to be sufficient for giving accurate results. From the database, it can be seen that he values of the FEV1 marker decreases with the severity of the COPD disease (from values that are greater than $3\,L$ in mild stages to approximately less than $1.5\,L$ for severe stages).
We ran our MCMC algorithm to fit the HSD model and make the inference after the burnings period (inference is based on the last 3000 samples after the burnings). In fact, starting from good initial values for the parameters would help in reducing the time of the burnings period and hence accelerate the MCMC convergence. Such initial values can be obtained from other estimation methods such as the maximum likelihood computation.
Also, while the $\alpha' s$ and σ are computed here using Gibbs sampler via a ridge regression approach, in the general case when the posterior is not a known one we call for the MHA such as for the parameter of the transition rate matrix: the $\lambda' s$ and $\gamma' s$ that have been computed via MHA.
Furthermore, this MCMC uses the random time imputation mechanism. Hence, we have seen that between successive observations, it could happen no imputation or 1 imputation, or 5, or more than 10 time points.
Finally, table 2 gives the posterior computations for each one of the parameters of the model. It shows how the marker can lead to the estimation of the parameters depending on the hidden states. We could see that most of parameters (the linear effect parameters and the quadratic effect parameters) have negative values, which goes with the attitudes of the FEV1 marker to decrease with the time. Moreover, and while here we fitted a model with 3 hidden states one can consider the cases where $a = 4\, or\, 5$, and uses the Bayesian information criterion (BIC) to choose the best model. Also, we could take other forms of stochastic differential equations and choose which equation fits well the data using the BIC.

5. Discussion

Bayesian approach is very efficient in simulating complicated models such as the non-linear diffusion processes. In fact, one can incorporate any prior information or knowledge in the likelihood through the prior specification and this is possible because the posterior of the parameters is proportional to prior multiplied by the likelihood. Also, and while MCMC algorithm can converge even when starting in dispersed initial values for the parameters, we can take use of classical method inference on data to get a good starting values such as maximum likelihood or expectation-maximization algorithms.
Another issue that should be pointed here too is that this model uses observation intervals that are non equi-distant, though we get accurate estimate using The Euler discretization; and why not should we try to improve this accuracy in the future by using the Milstein discretization [42]. Moreover, the number of observations imputed was chosen using the random time imputation which gives exact simulation. With this way, for small intervals we don't impute any data while for large intervals we could impute data.
Other ideas that can attract attention, is the use of the random walk MHA; that could have many problems such as the moving step. Indeed, a bad choice for the moving step can lead to bad mixing or create high correlated draws. Thus, we should some times avoid the random walk MHA and find good proposal density for every new

Table 1. MCMC estimation for the parameters of the model(8000 iterations)

Posterior computations				
Parameters	True value	Mean	Standard deviation	Credible interval (95%)
$\alpha_{0,1}$	0.3	0.3026	0.0011	(0.3004,0.3048)
$\alpha_{0,2}$	0.6	0.6028	0.0201	(0.5636,0.6426)
$\alpha_{0,3}$	0.9	0.9048	0.0047	(0.8956,0.9140)
$\alpha_{1,1}$	0.02	0.0211	$3.5015.10^{-04}$	(0.0204,0.0218)
$\alpha_{1,2}$	0.04	0.0401	0.0036	0.0331,0.0472)
$\alpha_{1,3}$	0.06	0.0602	$6.0245.10^{-04}$	(0.0590,0.0614)
$\alpha_{2,1}$	-0.08	-0.0722	$6.2750.10^{-04}$	(-0.0734,-0.0710)
$\alpha_{2,2}$	-0.06	-0.0593	0.0038	(-0.0671,-0.0522)
$\alpha_{2,3}$	-0.04	-0.0395	$5.575210.^{-04}$	(-0.0406,-0.0384)
$\alpha_{3,1}$	0.01	0.0101	$3.4369.10^{-04}$	(0.0094,0.0107)
$\alpha_{3,2}$	0.05	0.0503	0.0037	(0.0431,0.0577)
$\alpha_{3,3}$	0.09	0.0901	$7.1083.10^{-04}$	(0.0887,0.0915)
λ_{12}	0.2	0.1839	0.0040	(0.1911,0.2070)
λ_{13}	1	0.9108	0.0195	(0.9413,1.0181)
λ_{21}	0.3	0.2790	0.0059	(0.2895,0.3126)
λ_{23}	2	1.8189	0.0392	(1.8738,2.0267)
λ_{31}	0.4	0.3708	0.0080	(0.3846,0.4161)
λ_{32}	7	6.5094	0.1432	(6.7576,7.3138)
γ_{12}	0.2	0.1827	0.0040	(0.1902,0.2056)
γ_{13}	0.04	0.03681	$7.7337.10^{-04}$	(0.0381,0.0411)
γ_{21}	0.3	0.2899	0.0063	(0.3021,0.3268)
γ_{23}	0.06	0.0538	0.0012	(0.0558,0.0604)
γ_{31}	0.4	0.3841	0.0082	(0.3948,0.4264)
γ_{32}	0.08	0.0742	0.0016	(0.0778,0.0842)
η_1	0.3	0.2909	0.0201	(0.2510,0.3298)
η_2	0.4	0.3788	0.0430	(0.3098,0.4694)
η_3	0.5	0.5324	0.0199	(0.4933,0.5716)
σ^2	0.04	0.0377	0.0023	(0.0333,0.0425)

draw for each parameter or adopting more efficient algorithm such as the accept-reject MHA [11]. Hopefully here the ridge regression has allowed to sample many parameters through Gibbs sampler.
Finally, even that the HSD model here adopts an homoscedastic σ^2, it could be easily extended to σ_k^2 depending on the hidden states $k = 1, .., a$ or we can take the stochastic variance as in the case of ARCH and GARCH model [2,7,12,31].

6. Conclusion

This work provides a Bayesian approach for the simulation of a state-dependent switching diffusion process; one of the process usually hard to handle in a classical framework. We used Euler discretization to overcome the issue of dispersed observations as it is the case for most diffusion models. We have adapted the random time data imputation to the HSD model. We have let the transition rate matrix to depend on the diffusion observations. Hence, we have lead to the estimation of three categories of variables:

Table 2. MCMC output for fitting a HSD Process via FEV1 marker in COPD disease progression

Posterior computations			
Parameters	Mean	Standard deviation	Credible interval (95%)
$\alpha_{0,1}$	0.0030	0.0084	(-0.0128,0.0172)
$\alpha_{0,2}$	0.0169	0.0231	(-0.0300,0.0618)
$\alpha_{0,3}$	0.0164	0.0127	(-0.0113,0.0402)
$\alpha_{1,1}$	-0.0539	0.0209	(-0.0952,-0.0136)
$\alpha_{1,2}$	-0.0226	0.0656	(-0.1545,0.1105)
$\alpha_{1,3}$	-0.0240	0.0395	(-0.1037,0.0497)
$\alpha_{2,1}$	-0.0264	0.0242	(-0.0730, 0.0202)
$\alpha_{2,2}$	-0.0094	0.0744	(-0.1557,0.1452)
$\alpha_{2,3}$	-0.0120	0.0450	(-0.0991,0.0761)
λ_{12}	0.2657	0.0058	(0.2538,0.2782)
λ_{13}	0.7661	0.0369	(0.7224,0.8706)
λ_{21}	0.0458	7.55×10^{-4}	(0.0435,0.0470)
λ_{23}	0.6294	0.0114	(0.6152,0.6584)
λ_{31}	0.0708	0.0011	(0.0688,0.0733)
λ_{32}	0.2223	0.0034	(0.2144,0.2263)
γ_{12}	0.0122	4.79×10^{-4}	(0.0113,0.0131)
γ_{13}	0.312	0.012	(0.2889,0.3369)
γ_{21}	2.0406	0.081	(1.8855,2.2007)
γ_{23}	0.0087	3.59×10^{-4}	(0.0081, 0.0095)
γ_{31}	4.592	0.187	(4.2384,4.9777)
γ_{32}	3.122	0.126	(2.8812,3.3773)
σ^2	0.0086	7.78×10^{-4}	(0.0073,0.0103)

The imputed data, the hidden switching states, and the parameters of the diffusion process; the sampling of the hidden state has been realized by FFBS algorithm adapted to the HSD. Overall, even though the complexity of the switching one factor model, the MCMC algorithm has shown its efficiency to estimate the parameters accurately.

References

[1] Ait-Sahalia, Y. (1996b). Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies.* **9(2)** 385426.

[2] Adesi, G.B., Rasmussen, H., and Ravanelli, C. (2005), An option pricing formula for the GARCH diffusion model, *Computational Statistics & Data Analysis.* **49(2)** 287-310.

[3] Anga, A. , and Bekaertb, G. (2002), Short rate nonlinearities and regime switches, *Journal of Economic Dynamics & Control.* **26** 1243-1274.

[4] Blackwell, P. G., Niu, M., Lambert, M. S. and LaPoint, S. D. (2016). Exact Bayesian inference for animal movement in continuous time. *Methods Ecol Evol.* **7** 184195. doi:10.1111/2041-210X.12460.

[5] Blackwell, P. G. (1997). Random diffusion models for animal movement. *Ecological Modelling.* **100** 87-102.

[6] Blackwell, P. G. (2003). Bayesian Inference for Markov Processes with Diffusion and Discrete Components. *Biometrika.* **90(3)** 613-27.

[7] Bollerslev, T., Chou,R.Y. , and Kroner, K.F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* **52**(1-2) 559.

[8] Berman, S. M. (1994). A bivariate markov process with diffusion and discrete components. *Communications in Statistics. Stochastic Models* **10**(2) 271-308. DOI:10.1080/15326349408807297.

[9] Chib, S. (1996). Calculating posterior distributions and model estimates in Markov Mixtures models. *Journal of Econometrics* **75** 79–98.

[10] Chib S., and Shephard, N. (2002) . [Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes]: comment. *J Bus Econom Stat* **20(3)** 32527.

[11] Chib, S., and Jeliazkov, I. (2005). Accept-reject Metropolis-Hastings sampling and marginal likelihood estimation. *Statistica Neerlandica.* **59**3044.

[12] Chou, R.Y. (1988). Volatility persistence and stock valuations: Some empirical evidence using GARCH. *Journal of Applied Econometrics.* **3(4)** 279294.

[13] Dunn, J.E. , and Gipson, P.S. (1977). Analysis of Radio Telemetry Data in Studies of Home Range. *Biometrics.* **33(1)** 85-101.

[14] Durham, G.B., Gallant, A.R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econ. Stat.* **20** 297338.

[15] Elerian, O., S. Chib, and Shephard, N. (2001). Likelihood inference for discretely observed non-linear diffusion. *Econometrica* **69(4)** 959-993.

[16] Eraker, B. (2001). MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business & Economic Statistics.* **19(2)**. 177-191.

[17] Fink, G.A. (2008) *Markov Models for Pattern Recognition.* Springer-Verlag, Berlin.

[18] Fruhwirth-Schnatter, S. (2006) . *Finite Mixture and Markov Switching Models.* Springer Series in Statistics. New York:NY.

[19] Fuchs, C. (2013). *Inference for Diffusion Processes: With Applications in Life Sciences.* Springer-Verlag, Berlin Heidelberg.

[20] Golightly, A., and Wilkinson, D. (2008). Bayesian Inference for Nonlinear Multivariate Diffusion Models Observed with Error. *Computational Statistics and Data Analysis* **52** 1674-1693.

[21] Ghosh, M.K. ,Arapostathis, A. , and Marcus, S.I. (1993). Optimal control of switching diffusions with application to flexible manufacturing systems, *SIAM J. Control Optim..* **31** 11831204.

[22] Goutte, S., Ismail, A. and Pham, H. (2017). Regime-switching stochastic volatility model: estimation and calibration to VIX options. *Applied Mathematical Finance.* **24(1)** 38-75.

[23] Guihenneuc-Jouyaux, C. , Richardson, S. , Longini, IM. Jr. (2000). Modeling markers of disease progression by a hidden Markov process: application to characterizing

CD4 cell decline. *Biometrics*. **56(3)** 733-41.

[24] Janczura, J. and Weron, R. (2011).Efficient estimation of Markov regime-switching models: An application to electricity wholesale market prices. *AStA Adv. Stat. Anal*. **96** 385-407.

[25] Kloeden, P.E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*. Applications of Mathematics (New York) 23, Springer-Verlag, Berlin, doi:10.1007/978-3-662-12616-5.

[26] Korbitz, P. (2013): Aspects of estimation uncertainty in risk management. Open Access Repositorium der Universitt Ulm. Dissertation. http://dx.doi.org/10.18725/OPARU-2556.

[27] Krystul, J. and Blom, H.A.P. (2004). *Monte Carlo simulation of rare events in hybrid systems.*Technical Report WP8, Deliverable D8.3. HYBRIDGE

[28] Li, X. , Jiang, D. , and Mao, X. (2009). Population dynamical behavior of Lotka-Volterra system under regime switching. *J. Comput. Appl. Math.*. **232** 427448.

[29] Ma, J. (2013). *A Bayesian approach to longitudinal categorical data in a continuous time Markov chain model. Ph.D. Thesis.* The University of Texas School of Public Health, Houston, TX.

[30] Maroufy, E. H., Hibbah, E. H., Zyad, A., and Ziad, T. (2017). Bayesian Estimation of Multivariate Autoregressive Hidden Markov Model with Application to Breast Cancer Biomarker Modeling. In Prieto, J. T., editor, *Bayesian Inference, Chapter 7.* **http://dx.doi.org/10.5772/intechopen.70053.**

[31] Nelson D.B. (1990), ARCH Models as Diffusion Approximations, *Journal of Econometrics*. **45** 738.

[32] OGAWA, S. (1992), Monte Carlo Simulation of Non-linear Diffusion Processes, *Japan J. Indust. Appl. Math.* **9** 25-33.

[33] Parton, A. , and Blackwell, P. G. (2016), Bayesian Inference for Multistate Step and Turn Animal Movement in Continuous Time, *Journal of Agricultural, Biological, and Environmental Statistics*. **22(3)** 373-392.

[34] Roberts, G.O. and Stramer, O.. (2001) . On Inference for Partially Observed Nonlinear Diffusion Models Using the Metropolis-Hastings Algorithm. *Biometrika*. **88(3)** 603-621.

[35] Rao, V. and Teh, Y. W. (2013) . Fast MCMC Sampling for Markov Jump Processes and Extensions. *Journal of Machine Learning Research*. **14** 3295-3320

[36] Shaker, S.B. , Dirksen, A. , Lo, P. , Skovgaard, L.T. , de Bruijne, M. , and Pedersen J.H. (2012) . Factors influencing decline in lung density in a Danish lung cancer screening cohort. *Eur Respir J.*. **40:5** 1142-8. doi: 10.1183/09031936.00207911.

[37] Siekmann, I., Wagner, L.E.II, Yule, D., Fox, C., Bryant, D., Crampin, E.J. and Sneyd, J. (2011) . Fast MCMC Sampling for Markov Jump Processes and Extensions. *Biophys. J.* **100** 1919-1929.

[38] Sorwar, G. (2011) . Estimating single factor jump diffusion interest rate models. *Applied Financial Economics*. **21:22** 1679-1689, DOI: 10.1080/09603107.2011.591729.

[39] Stockley, R. A. (2014). Biomarkers in chronic obstructive pulmonary disease: confusing or useful?. *International Journal of COPD*. **4:9** 163-177

[40] Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82** 528-550.

[41] Tjørve,K. M. C. and Even Tjørve, T. (2017). The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PLOS ONE*. **12(6)**: e0178691. https://doi.org/10.1371/journal. pone.017869

[42] TSE, Y.K. , YU, J. , and CHANG, X.B. (2002). Estimation of Hyperbolic Diffusion Using Mcmc Method. *Research Collection.* School of Economics. Available at: http://ink.library.smu.edu.sg/soe_research_all/5.

[43] Tsionas, E., Tassiopoulos, A. (2016). Bayesian implications of ridge regression and Zellners g-prior (Tech. Rep.). Athens: Athens University of Economics and Business - Department of Economics. SSRN: http://ssrn.com/abstract=238758.

[44] van den Hout, A., Fox, J.P., and Klein Entink, R.H. (2015). Bayesian inference for an illness-death model for stroke with cognition as a latent time-dependent risk factor. *Statistical Methods in Medical Research* , **24(6)**, 769-87. DOI: 10.1177/0962280211426359.

[45] Yin, G. and Zhu, C. (2010). Properties of solutions of stochastic differential equations with continuous state-dependent switching. *J. Differential Eqs.* **249** 24092439.

[46] Young, A.L. , Oxtoby, N.P. , Ourselin, S. , Schott J.M. , Alexander, D.C. (2015). A simulation system for biomarker evolution in neurodegenerative disease. *Med Image Anal.* **26(1)** 47-56.doi: 10.1016/j.media.2015.07.004.

.

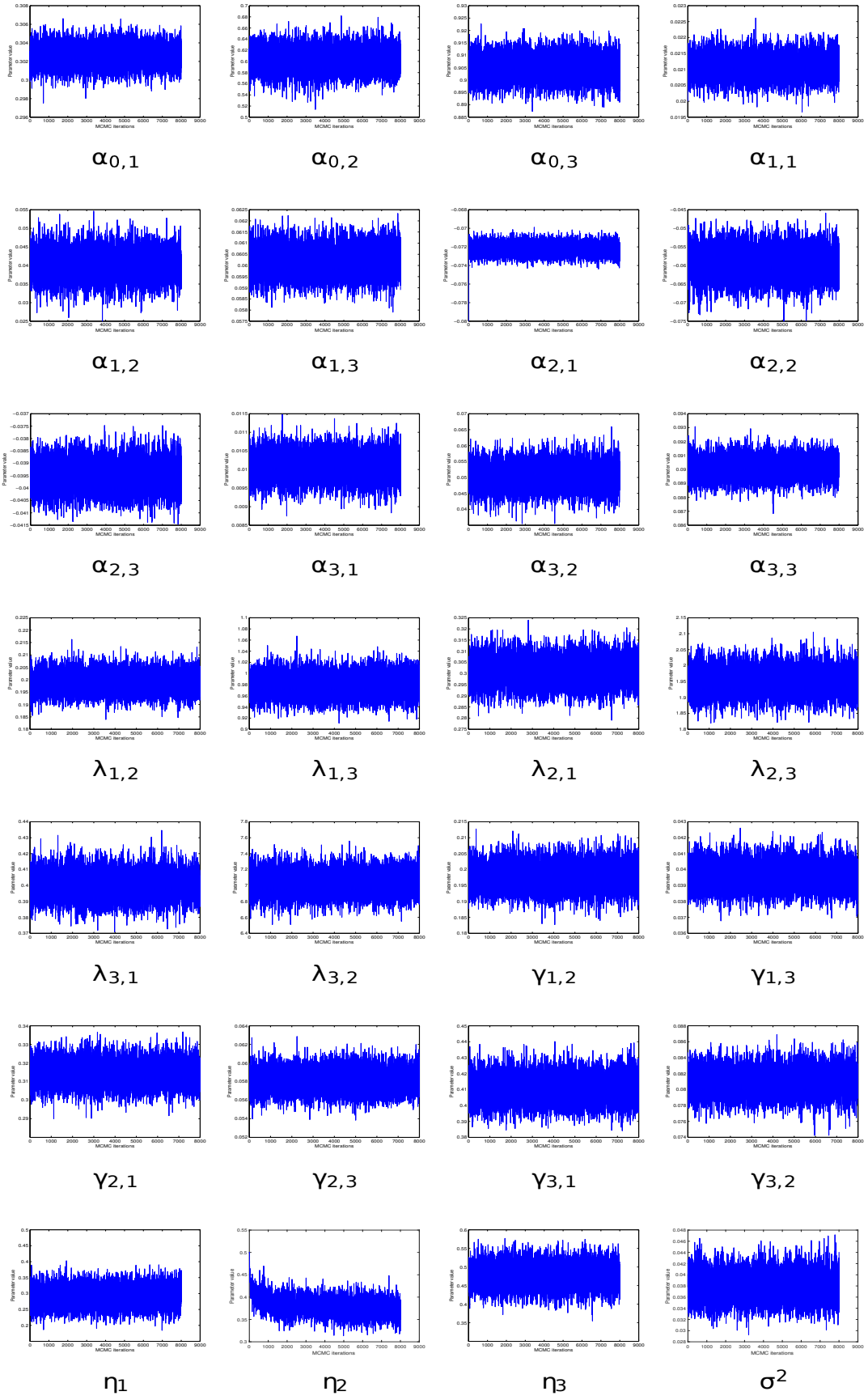

Figure 1. 8000 MCMC iteration plots for the parameters of the model

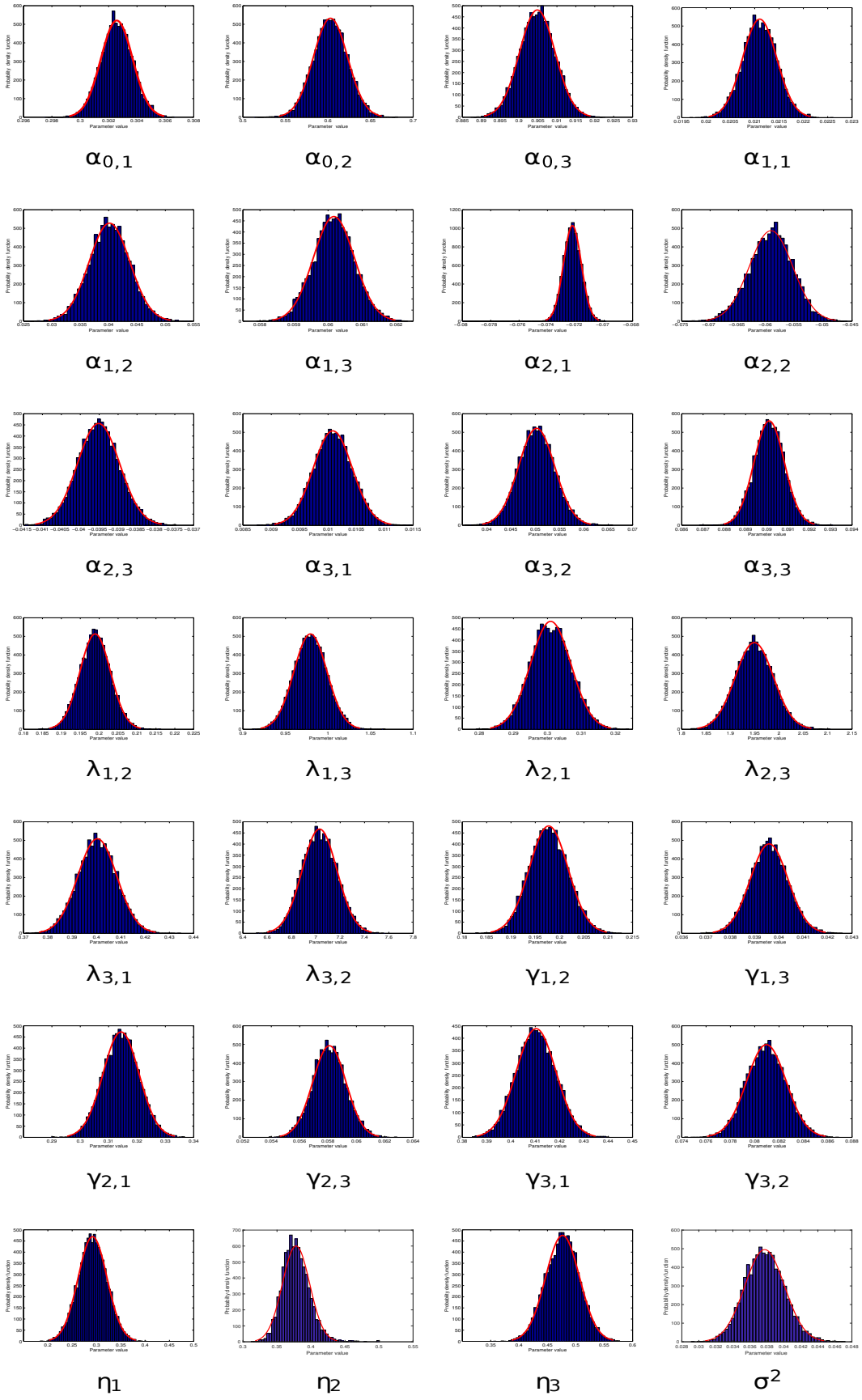

Figure 2. Posterior density plots for the parameters of model

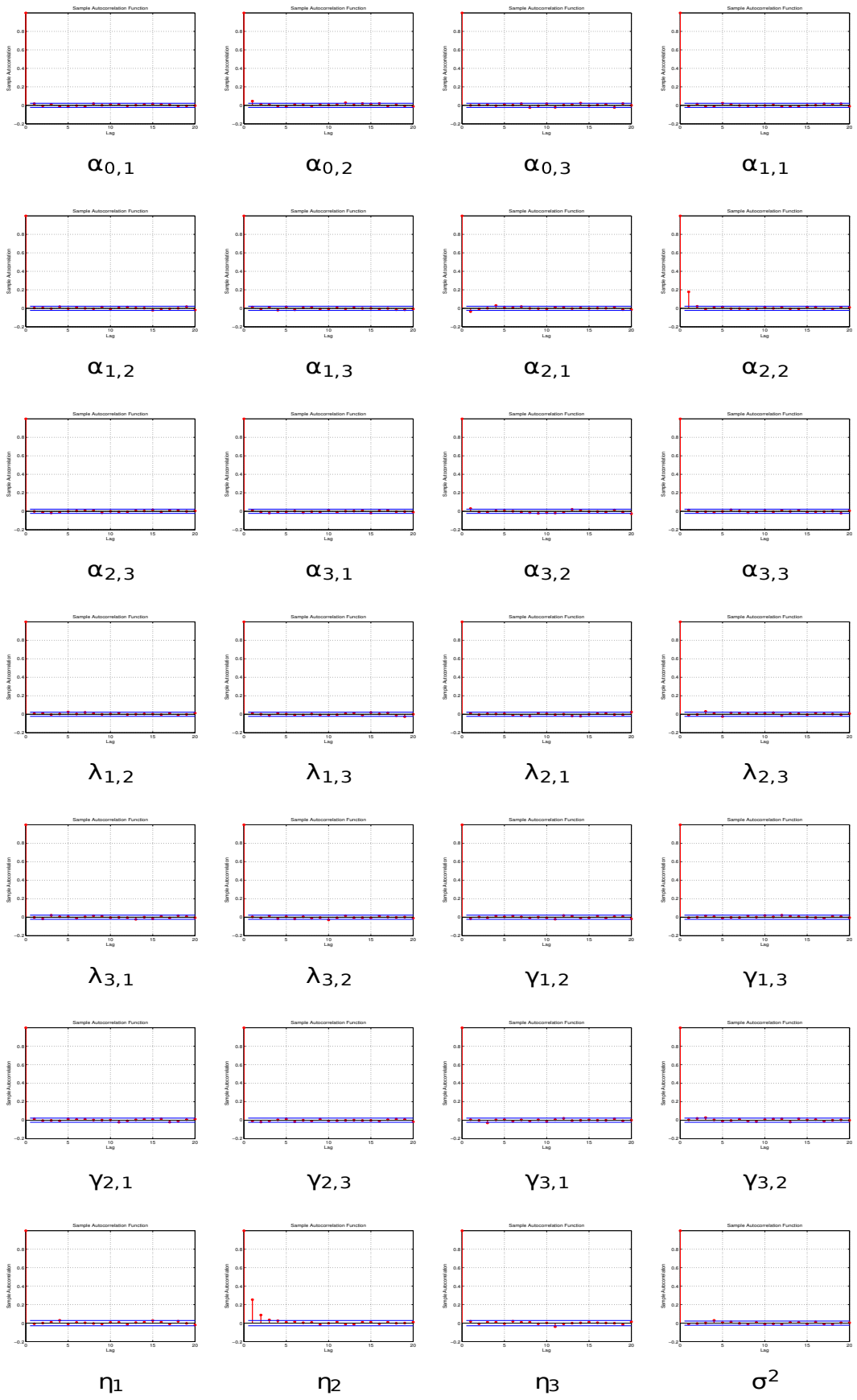

Figure 3. Autocorrelation sample plots for the parameters of model