Signal Processing 171 (2020) 107511

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# Semi-supervised optimal recursive filtering and smoothing in non-Gaussian Markov switching models

Fei Zheng [a], Stéphane Derrode [a], Wojciech Pieczynski [b,*]

[a] École Centrale de Lyon, Université de Lyon, LIRIS, CNRS UMR 5205, Écully, France
[b] Telecom SudParis, SAMOVAR, Institut Polytechnique de Paris, CNRS UMR 5157, Évry, France

ARTICLE INFO

*Article history:*
Received 27 April 2019
Revised 21 January 2020
Accepted 28 January 2020
Available online 29 January 2020

*Keywords:*
Markov switching model
Non-Gaussian non-linear system
Copulas
Model identification
CMSHLM
GICE-GLS
Semi-supervised filtering
Semi-supervised smoothing

ABSTRACT

Filtering and smoothing in switching state-space models are important in numerous applications. The classic family of conditionally Gaussian linear state space models (CGLSSMs) is a natural extension of the Gaussian linear system by introducing its dependence on switches. In spite of their simplicity, recursive filtering and smoothing are no longer feasible in CGLSSMs and approximate methods must be used. Conditionally Markov switching hidden linear models (CMSHLMs) are alternative models which allow recursive optimal exact filtering and smoothing. We introduce an original family of CMSHLMs defined with copulas and we address the problem of their identification. The proposed identification method chooses a model in a family of admissible parametric models and estimates the parameters. It is applied to a learning sample containing observations and states, while the switches are unknown. The interest of the proposed "semi-unsupervised" filtering and smoothing is validated via experiments on simulated data.



## 1. Introduction

We introduce a general switching model based on copulas and we propose an algorithm for its semi-supervised identification. The identification is performed from a learning sample set including states and observations, the switches being unknown. It consists of solving two problems:

(i) find the appropriate model in a set of possible parametric models;
(ii) estimate the parameters.

Then the recursive exact filtering and smoothing can work based on the identified models, and we show the interest of the whole procedure via simulation studies.

A switching model contains three random sequences: $\mathbf{X}_1^N = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$, $\mathbf{R}_1^N = (R_1, \ldots, R_N)$ and $\mathbf{Y}_1^N = (\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$. For $n = 1, \cdots, N$, $\mathbf{X}_n$ takes its values in $\mathbb{R}^s$, $R_n$ takes its value in $\mathbf{\Omega} = \{1, \ldots, K\}$, and $\mathbf{Y}_n$ takes its values in $\mathbb{R}^q$. For $n = 1, \cdots, N$, let $\mathbf{T}_n = (\mathbf{X}_n, R_n, \mathbf{Y}_n)$ and let us consider $\mathbf{T}_1^N = (\mathbf{T}_1, \ldots, \mathbf{T}_N)$. For some occasions, $\mathbf{T}_1^N$ will be also denoted as $\mathbf{T}_1^N = \left(\mathbf{X}_1^N, \mathbf{R}_1^N, \mathbf{Y}_1^N\right)$. The final restoration problem dealt with is to estimate both the hidden $\left(\mathbf{X}_1^N, \mathbf{R}_1^N\right) = \left(\mathbf{x}_1^N, \mathbf{r}_1^N\right)$ from observed $\mathbf{Y}_1^N = \mathbf{y}_1^N$.

To be concise, we will note different probability distributions with the same letter $p$. So the distribution of $\mathbf{X}_1^N$ will be denoted with $p\left(\mathbf{x}_1^N\right)$, the distribution of $R_n$ conditional on $\mathbf{Y}_n = \mathbf{y}_n$ will be denoted with $p(r_n|\mathbf{y}_n.)$ and so on. For discrete variables, like $R_1$, $p(r_1)$ is a probability, for continuous ones, like $\mathbf{Y}_n$, $p(\mathbf{y}_n)$ is a probability density function (pdf), and for mixed case, like $\mathbf{T}_1 = (\mathbf{X}_1, R_1, \mathbf{Y}_1)$, we have $p(\mathbf{x}_1, r_1, \mathbf{y}_1) = p(r_1)p(\mathbf{x}_1, \mathbf{y}_1|r_1)$, with $p(r_1)$ probability and $p(\mathbf{x}_1, \mathbf{y}_1|r_1.)$ pdf.

Let us consider "Conditionally Markov switching hidden linear model" (CMSHLM [1]) defined as:

$$\mathbf{T}_1^N = (\mathbf{T}_1, \cdots, \mathbf{T}_N) \text{ is Markov;} \tag{1}$$

$$p(r_{n+1}|\mathbf{x}_n, r_n, \mathbf{y}_n) = p(r_{n+1}|r_n); \tag{2}$$

$$p(r_{n+1}, \mathbf{y}_{n+1}|\mathbf{x}_n, r_n, \mathbf{y}_n) = p(r_{n+1}, \mathbf{y}_{n+1}|r_n, \mathbf{y}_n); \tag{3}$$

$$\begin{aligned}\mathbf{X}_{n+1} &= A_{n+1}\left(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1}\right)\mathbf{X}_n + B_{n+1}\left(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1}\right) \\ &\quad + C_{n+1}\left(\mathbf{R}_n^{n+1}, \mathbf{Y}_n^{n+1}\right)W_{n+1},\end{aligned} \tag{4}$$

where $A_{n+1}$, $B_{n+1}$, $C_{n+1}$ are some vector functions whose dimensions of the range is respectively $s \times s$, $s$ and $s$, and $W_2$, $\cdots$, $W_N$ is a sequence of centred variables with unit variance and such that $W_{n+1}$ is independent from $(\mathbf{T}_1, \cdots, \mathbf{T}_n)$ for each $n = 1, \cdots, N-1$.

* Corresponding author.
*E-mail address:* wojciech.pieczynski@telecom-sudparis.eu (W. Pieczynski).

https://doi.org/10.1016/j.sigpro.2020.107511
0165-1684/

Let us note that (2) implies the Markovianity of $\mathbf{R}_1^N$, while (3) implies the Markovianity of $\left(\mathbf{R}_1^N, \mathbf{Y}_1^N\right)$.

We propose the following contributions:

(i) CMSHLM in which $p\left(\mathbf{r}_1^N, \mathbf{y}_1^N\right)$ is copulas based one [2] is original;
(ii) $p\left(\mathbf{r}_1^N, \mathbf{y}_1^N\right)$ is identified from $\mathbf{Y}_1^N = \mathbf{y}_1^N$ through an original variant of the "generalized iterative conditional estimation" (GICE [3,4]);
(iii) identification and parameter estimation of $A_{n+1}$ and $B_{n+1}$ with a new "GICE with generalized least-squares" (GICE-GLS);
(iv) general copulas based CMSHLM identification provided with points (ii)-(iii) leads to semi-supervised (in learning sample $\left(\mathbf{X}_1^N, \mathbf{Y}_1^N\right) = \left(\mathbf{x}_1^N, \mathbf{y}_1^N\right)$ are known while $\mathbf{R}_1^N = \mathbf{r}_1^N$ are not) recursive exact filtering and smoothing.

Let us remark that CMSHLM with $\left(\mathbf{X}_1^N, \mathbf{Y}_1^N\right)$ Gaussian conditionally on $\mathbf{R}_1^N$ leads to "Conditionally Gaussian observed Markov switching models" (CGOMSMs [5–10]) which thus allow exact filtering and smoothing and can be seen as an alternative to the widely used "Conditionally Gaussian linear state space models" (CGLSSMs [11,12], among others).

More generally, filtering in non-Gaussian non-linear systems is widely applied in different problems and particle filters – which are asymptotically optimal – are very efficient when the number of particles is sufficient [11–17], among others. Approximating such stationary non-Gaussian non-linear systems (NGNLSs) with general CMSHLM proposed in the paper – as carried out using CGOMSMs in [9] – opens rich perspectives of dealing with stationary NGNLSs when particles based methods fail because of the excessively large number of particles needed.

Furthermore, smoothing in switching systems is a hard problem and using particles is often faced with the degeneracy problem. Researchers are very active in the field, [16,18–21] among others. Such problems do not occur in CMSHLMs and smoothing is even quite straightforward. Let us remark that although similar to smoothing methods in CGOMSMs described in [9], those presented in this paper are new.

The remaining of the paper is organized as follows. In section 2 we present the new copulas based CMSHLM (CB-CMSHLM), and specify filtering and smoothing. Section 3 is devoted to the proposed CB-CMSHLM identification method termed GICE-GLS. Some experiments are provided in Section 4 and the last Section 5 concludes the work and sets out the perspectives.

## 2. Filtering and smoothing in copulas based CMSHLMs

### 2.1. Copulas based CMSHLM

Let $(Y^1, \cdots, Y^d)$ be a random vector valued in $\mathbb{R}^d$, $F(y^1, \cdots, y^d) = P[Y^1 \le y^1, \cdots, Y^d \le y^d]$ its cumulative density function (CDF), and $F_1, \cdots, F_d$ CDFs of $Y^1, \cdots, Y^d$ respectively. Furthermore, a copula $C$ is a CDF defined on $[0, 1]^d$ such that marginal CDFs $C_1(y^1), \cdots, C_d(y^d)$ are identities on [0, 1]. According to Sklar's theorem [23], for given $F$ there exists a unique copula $C$ such that:

$$F\left(y^1, \cdots, y^d\right) = C\left(F_1\left(y^1\right), \cdots, F_d\left(y^d\right)\right). \tag{5}$$

Assuming differentiable $F$ and $C$, setting

$$c\left(y^1, \cdots, y^d\right) = \frac{\partial^d}{\partial y^1 \cdots \partial y^d} C\left(y^1, \cdots, y^d\right), \tag{6}$$

and taking derivative of (5), we obtain the probability density function (PDF) of $(Y^1, \cdots, Y^d)$:

$$f\left(y^1, \cdots, y^d\right) = c\left[F_1\left(y^1\right), \cdots, F_d\left(y^d\right)\right] \prod_{i=1}^{d} f_i\left(y^i\right), \tag{7}$$

with $f_i, \cdots, f_d$ PDFs of $Y^1, \cdots, Y^d$ respectively. Let us return to CMSHLM defined by (1)–(4). In addition, we will consider the following commonly used assumptions:

$$p\left(\mathbf{y}_{n+1} \middle| \mathbf{r}_n^{n+1}\right) = p(\mathbf{y}_{n+1} | r_{n+1}); \tag{8}$$

$$p\left(\mathbf{y}_n \middle| \mathbf{r}_n^{n+1}\right) = p(\mathbf{y}_n | r_n); \tag{9}$$

Applying (7) to $p\left(\mathbf{y}_n, \mathbf{y}_{n+1} \middle| \mathbf{r}_n^{n+1}\right)$ and using (8), (9), there exists a copula $C_{n+1}\left(\mathbf{r}_n^{n+1}\right)$ such that:

$$p\left(\mathbf{y}_n, \mathbf{y}_{n+1} \middle| \mathbf{r}_n^{n+1}\right) = p(\mathbf{y}_n | r_n) p(\mathbf{y}_{n+1} | r_{n+1}) \\ c_{n+1}\left(\mathbf{r}_n^{n+1}\right) (F_n(\mathbf{y}_n | r_n), F_{n+1}(\mathbf{y}_{n+1} | r_{n+1})) \tag{10}$$

and thus

$$p\left(\mathbf{y}_{n+1} \middle| \mathbf{r}_n^{n+1}, \mathbf{y}_n\right) = p(\mathbf{y}_{n+1} | r_{n+1}) c_{n+1}\left(\mathbf{r}_n^{n+1}\right) \\ (F_n(\mathbf{y}_n | r_n), F_{n+1}(\mathbf{y}_{n+1} | r_{n+1})) \tag{11}$$

Markovianity of $\mathbf{R}_1^N$ and $\left(\mathbf{R}_1^N, \mathbf{Y}_1^N\right)$ joined to (11) indicate that the distribution of $\left(\mathbf{R}_1^N, \mathbf{Y}_1^N\right)$ is given by Markov distribution of $\mathbf{R}_1^N$, margins $p(\mathbf{y}_1 | r_1.), \cdots, p(\mathbf{y}_N | r_N.)$, and copulas $c_2\left(\mathbf{r}_1^2\right), \cdots, c_N\left(\mathbf{r}_{N-1}^N\right)$. Let us notice that CB-CMSHLM so obtained is not necessarily stationary: margins and copulas can depend on $n$.

### 2.2. Filtering in copulas based CMSHLM

Recalling that, for each $n = 1, \ldots, N$, $\mathbf{X}_n$ takes its values in $\mathbb{R}^s$, and $\mathbf{Y}_n$ takes its values in $\mathbb{R}^q$, we have

$$\mathbf{X}_n = \begin{bmatrix} X_n^1 \\ \cdots \\ X_n^s \end{bmatrix}, \mathbf{Y}_n = \begin{bmatrix} Y_n^1 \\ \cdots \\ Y_n^q \end{bmatrix},$$

with $X_n^1, \ldots, X_n^s, Y_n^1, \ldots, Y_n^q$ real. We will classically note $\mathbf{X}_n^T = \left(X_n^1, \ldots, X_n^s\right)$, $\mathbf{Y}_n^T = \left(Y_n^1, \ldots, Y_n^q\right)$. Similarly to conditional distributions, conditional expectation of a random vector $A$ knowing a realization $B = b$ of a random vector $B$ will be denoted with $\mathbb{E}[A|b]$. The filtering problem consists of recursively computing $p\left(r_{n+1} \middle| \mathbf{y}_1^{n+1}\right)$, $\mathbb{E}\left[\mathbf{X}_{n+1} \middle| r_{n+1}, \mathbf{y}_1^{n+1}\right]$, and $\mathbb{E}\left[\mathbf{X}_{n+1}, \mathbf{X}_{n+1}^\intercal \middle| r_{n+1}, \mathbf{y}_1^{n+1}\right]$ from $p\left(r_n \middle| \mathbf{y}_1^n\right)$, $\mathbb{E}\left[\mathbf{X}_n \middle| r_n, \mathbf{y}_1^n\right]$, $\mathbb{E}\left[\mathbf{X}_n \mathbf{X}_n^\intercal \middle| r_n, \mathbf{y}_1^n\right]$, $p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n)$, and $\mathbf{y}_{n+1}$. We have

$$p\left(r_{n+1} \middle| \mathbf{y}_1^{n+1}\right) = \frac{\sum_{r_n} p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n) p\left(r_n \middle| \mathbf{y}_1^n\right)}{\sum_{r_{n+1}} \sum_{r_n} p(r_{n+1}, \mathbf{y}_{n+1} | r_n, \mathbf{y}_n) p\left(r_n \middle| \mathbf{y}_1^n\right)}; \tag{12}$$

$$\mathbb{E}\left[\mathbf{X}_{n+1} \middle| r_{n+1}, \mathbf{y}_1^{n+1}\right] = \sum_{r_n} \left[A_{n+1}\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) \mathbb{E}[\mathbf{X}_n | r_n, \mathbf{y}_1^n] \\ + B_{n+1}\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right)\right] p\left(r_n \middle| r_{n+1}, \mathbf{y}_1^{n+1}\right) \tag{13}$$

$$\mathbb{E}\left[\mathbf{X}_{n+1} \mathbf{X}_{n+1}^\intercal \middle| r_{n+1}, \mathbf{y}_1^{n+1}\right] = \sum_{r_n} F_{n+1}\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) p\left(r_n \middle| r_{n+1}, \mathbf{y}_1^{n+1}\right), \tag{14}$$

with

$$\begin{aligned} &F_{n+1}\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) \\ &= A_{n+1}\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) \mathbb{E}[\mathbf{X}_n \mathbf{X}_n^\intercal | r_n, \mathbf{y}_1^n] A_{n+1}^\intercal\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) \\ &+ B_{n+1}\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) B_{n+1}^\intercal\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) \\ &+ C_{n+1}\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) C_{n+1}^\intercal\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) \\ &+ A_{n+1}\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) \mathbb{E}[\mathbf{X}_n | r_n, \mathbf{y}_1^n] B_{n+1}^\intercal\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) \\ &+ B_{n+1}\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right) \mathbb{E}^\intercal[\mathbf{X}_n | r_n, \mathbf{y}_1^n] A_{n+1}^\intercal\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right), \end{aligned} \tag{15}$$

$$p\left(r_n\middle|r_{n+1},\mathbf{y}_1^{n+1}\right)=\frac{p(r_{n+1},\mathbf{y}_{n+1}|r_n,\mathbf{y}_n)p\left(r_n\middle|\mathbf{y}_1^n\right)}{\sum_{r_n}p(r_{n+1},\mathbf{y}_{n+1}|r_1,\mathbf{y}_n)p\left(r_n\middle|\mathbf{y}_1^n\right)}. \tag{16}$$

Let us briefly justify (12)–(16). (12) and (16) come from the Markovianity of $\left(\mathbf{R}_1^N,\mathbf{Y}_1^N\right)$, which implies

$$p\left(\mathbf{r}_n^{n+1},\mathbf{y}_1^{n+1}\right)=p(r_{n+1},\mathbf{y}_{n+1}|r_n,\mathbf{y}_n)p(r_n,\mathbf{y}_1^n) \tag{17}$$

To justify (13), we write $\mathbb{E}[\mathbf{X}_{n+1}|r_{n+1},\mathbf{y}_1^{n+1}]=\sum_{r_n}\mathbb{E}[\mathbf{X}_{n+1}|r_n,r_{n+1},\mathbf{y}_1^{n+1}]p(r_n|r_{n+1},\mathbf{y}_1^{n+1})=\sum_{r_n}[A_{n+1}(\mathbf{r}_n^{n+1},\mathbf{y}_1^{n+1})\mathbb{E}[\mathbf{X}_n|r_n,r_{n+1},\mathbf{y}_1^{n+1}]+B_{n+1}(\mathbf{r}_n^{n+1},\mathbf{y}_n^{n+1})]p(r_n|r_{n+1},\mathbf{y}_n^{n+1})$, and then apply $\mathbb{E}[\mathbf{X}_n|r_n,r_{n+1},\mathbf{y}_1^{n+1}]=\mathbb{E}[\mathbf{X}_n|r_n,\mathbf{y}_1^n.]$ – which comes from Eq. 3, according to which $\mathbf{X}_n$ and $(R_{n+1},\mathbf{Y}_{n+1})$ are independent conditionally on $(R_n,\ \mathbf{Y}_n)$. (14)–(15) are obtained in a similar way by replacing $\mathbf{X}_{n+1}$ by $\mathbf{X}_{n+1}\mathbf{X}_{n+1}^{\intercal}$.

**Remark 1.** As $p\left(r_{n+1}\middle|\mathbf{y}_1^{n+1}\right)$ and $\mathbb{E}\left[\mathbf{X}_{n+1}\middle|r_{n+1},\mathbf{y}_1^{n+1}\right]$ are computed from $p\left(r_n\middle|\mathbf{y}_1^n\right)$, $\mathbb{E}\left[\mathbf{X}_n\middle|r_n,\mathbf{y}_1^n\right]$, and $\mathbf{y}_{n+1}$ without using $C_{n+1}\left(\mathbf{R}_n^{n+1},\mathbf{Y}_n^{n+1}\right)$, (4) can be actually extended to

$$\mathbf{X}_{n+1}=A_{n+1}\left(\mathbf{R}_n^{n+1},\mathbf{Y}_n^{n+1}\right)\mathbf{X}_n+B_{n+1}\left(\mathbf{R}_n^{n+1},\mathbf{Y}_n^{n+1}\right)+W_{n+1}, \tag{18}$$

with the only hypotheses that $\mathbb{E}[W_{n+1}]=0$ and $W_{n+1}$ is independent from $(\mathbf{T}_1,\ \cdots,\mathbf{T}_n)$ for each $n=1,\cdots,N-1$. However, the variance of the filter depends on $\mathrm{Var}[W_{n+1}]$ would thus be unknown.

### 2.3. Smoothing in copulas based CMSHLM

Optimal smoothing consists of computation of $\mathbb{E}\left[\mathbf{X}_n\middle|r_n,\mathbf{y}_1^N\right]$ for each $n=1,\cdots,N$. Under CMSHLM they are not complicated to get from already calculated $\mathbb{E}\left[\mathbf{X}_n\middle|r_n,\mathbf{y}_1^n\right]$ in filtering given in the previous paragraph. We have:

$$\begin{aligned}\mathbb{E}\left[\mathbf{X}_n\middle|\mathbf{y}_1^N\right]&=\sum_{r_n}p\left(r_n\middle|\mathbf{y}_1^N\right)\mathbb{E}\left[\mathbf{X}_n\middle|r_n,\mathbf{y}_1^N\right]\\&=\sum_{r_n}p\left(r_n\middle|\mathbf{y}_1^N\right)\mathbb{E}[\mathbf{X}_n|r_n,\mathbf{y}_1^n],\end{aligned} \tag{19}$$

the second equality being due to the fact that $\mathbf{X}_n$ and $\mathbf{Y}_{n+1}^N$ are independent conditionally on $(R_n,\ \mathbf{Y}_n)$. $\left(\mathbf{R}_1^N,\mathbf{Y}_1^N\right)$ being Markov, $p\left(r_n\middle|\mathbf{y}_1^N\right)$ is classically obtained by recursive calculation of "forward" and "backward" probabilities $\alpha_n(r_n)=p\left(r_n,\mathbf{y}_1^n\right)$, $\beta_n(r_n)=p\left(\mathbf{y}_{n+1}^N|r_n\right)$ with:

$$\alpha_1(r_1)=p(r_1,\mathbf{y}_1);\quad \alpha_{n+1}(r_{n+1})=\sum_{r_n}p(r_{n+1},\mathbf{y}_{n+1}|r_n,\mathbf{y}_n)\alpha_n(r_n), \tag{20}$$

$$\beta_N(r_N)=1;\quad \beta_n(r_n)=\sum_{r_{n+1}}p(r_{n+1},\mathbf{y}_{n+1}|r_n,\mathbf{y}_n)\beta_{n+1}(r_{n+1}), \tag{21}$$

We have $p\left(r_n,\mathbf{y}_1^N\right)=\alpha_n(r_n)\beta_n(r_n)$, and thus

$$p\left(r_n\middle|\mathbf{y}_1^N\right)=\frac{\alpha_n(r_n)\beta_n(r_n)}{\sum_{r_n}\alpha_n(r_n)\beta_n(r_n)}. \tag{22}$$

$\mathbb{E}\left[\mathbf{X}_n\middle|\mathbf{y}_1^N\right]$ in smoothing does not require $C_{n+1}\left(\mathbf{R}_n^{n+1},\mathbf{Y}_n^{n+1}\right)$ as in filtering, while $\mathbb{E}\left[\mathbf{X}_n,\mathbf{X}_n^{\intercal}\middle|\mathbf{y}_1^N\right]$ can be calculated in a similar way and gives:

$$\mathbb{E}\left[\mathbf{X}_n,\mathbf{X}_n^{\intercal}\middle|\mathbf{y}_1^N\right]=\sum_{r_n}p\left(r_n\middle|\mathbf{y}_1^N\right)\mathbb{E}[\mathbf{X}_n,\mathbf{X}_n^{\intercal}|r_n,\mathbf{y}_1^n], \tag{23}$$

with $\mathbb{E}\left[\mathbf{X}_n,\mathbf{X}_n^{\intercal}\middle|r_n,\mathbf{y}_1^N\right]$ from (14), (15).

## 3. CB-CMSHLM identification

We tackle the identification problem of a CB-CMSHLM from a learning sample set of observations $\mathbf{Y}_1^N=\mathbf{y}_1^N$ data and space data $\mathbf{X}_1^N=\mathbf{x}_1^N$, while $\mathbf{R}_1^N=\mathbf{r}_1^N$ remains unknown. In CB-CMSHLM $\mathbf{T}_1^N=\left(\mathbf{X}_1^N,\mathbf{Y}_1^N,\mathbf{R}_1^N\right)$, the considered couple $\left(\mathbf{R}_1^N,\mathbf{Y}_1^N\right)$ is stationary, so that its distribution can be defined by $p(r_1,r_2,\mathbf{y}_1,\mathbf{y}_2)=p(r_1,r_2)p(\mathbf{y}_1,\mathbf{y}_2|r_1,r_2)$, equal to the distributions $p(r_n,r_{n+1},\mathbf{y}_n,\mathbf{y}_{n+1})$, $n=2,\cdots,N-1$. Furthermore, $A_{n+1}$ and $B_{n+1}$ in (4) are time independent from $n=2,\cdots,N-1$. To summarize, the model identification problem which we are facing is threefold:

(i) Estimate the distribution $p(r_1,\ r_2)$;
(ii) Find forms of copulas and margins, as well as related parameters, of the distributions $p(\mathbf{y}_1,\ \mathbf{y}_2|r_1,\ r_2.)$;
(iii) Find forms and parameters of $A\left(\mathbf{r}_1^2,\mathbf{y}_1^2\right)$, and $B\left(\mathbf{r}_1^2,\mathbf{y}_1^2\right)$ defining $p(\mathbf{x}_1,\ \mathbf{x}_2|r_1,\ r_2,\ \mathbf{y}_1,\ \mathbf{y}_2.)$.

For each $(r_1,\ r_2)\in\Omega^2$, forms of copulas $c(r_1,\ r_2)$, forms of margins $p(\mathbf{y}_1|r_1.)$, and forms of $A\left(\mathbf{r}_1^2,\mathbf{y}_1^2\right)$, $B\left(\mathbf{r}_1^2,\mathbf{y}_1^2\right)$, will be searched for in given sets of possible forms.

### 3.1. Generalized Iterative Conditional Estimation (GICE)

To solve (i) and (ii) we use an original variant of Generalized Iterative Conditional Estimation (GICE). GICE is a family of methods extending ICE to cases where the parameterized forms of distributions are unknown, but belong to a given family of candidate forms. Introduced in the frame of hidden discrete Markov (with correlated noise) models in [4] GICE can be applied here to identify the distribution of $\left(\mathbf{R}_1^N,\mathbf{Y}_1^N\right)$ from $\mathbf{Y}_1^N$ only: $\mathbf{X}_1^N$ can be temporarily set aside here. The new GICE variant we propose is as follows.

Let $\mathbf{Y}_1^N=\mathbf{y}_1^N$ be a sample, and for simplifying the notations, let us denote $f_{jk}(\mathbf{y}_1,\mathbf{y}_2)=p(\mathbf{y}_1,\mathbf{y}_2|r_1=j,r_2=k)$, $f_j(\mathbf{y}_1)=p(\mathbf{y}_1|r_1=j)$, $f_k(\mathbf{y}_2)=p(\mathbf{y}_2|r_1=k)$, and $c_{jk}\left(F_j(\mathbf{y}_1),F_k(\mathbf{y}_2)\right)=c(r_1=j,r_2=k)(F(\mathbf{y}_1|r_1=j),F(\mathbf{y}_2|r_2=k))$. So that:

$$f_{jk}(\mathbf{y}_1,\mathbf{y}_2)=f_j(\mathbf{y}_1)f_k(\mathbf{y}_2)c_{jk}\left(F_j(\mathbf{y}_1),F_k(\mathbf{y}_2)\right) \tag{24}$$

Furthermore, the switch probabilities are written as $p_{jk}=p(r_1=j,r_2=k)$.

For each $j\in\Omega$, the form $f_j$ is unknown, but we assume that it belongs to a known set of possible forms $\boldsymbol{H}=\{H_1,\cdots,H_L\}$. Each form $H_l$, $l=1,\cdots,L$, is a parametric set of probability distributions $H_l=\left\{f_{\theta(l)}\right\}_{\theta(l)\in\Theta(l)}$. Similarly, for each $j,\ k\in\Omega$, the form of $c_{jk}$ is unknown, but it is assumed to belong to a known set of possible forms $\boldsymbol{G}=\{G_1,\cdots,G_M\}$, each of which being a parametric set of copulas $G_m=\left\{c_{\alpha(m)}\right\}_{\alpha(m)\in\mathrm{A}(m)}$.

Thus to identify margins means to find (from $\mathbf{Y}_1^N=\mathbf{y}_1^N$) for each $j\in\Omega$, the right form $H_l^j$ in $\boldsymbol{H}$ and to estimate parameters $\theta_j(l)$. To identify copulas, the problem is to find, for each $j,\ k\in\Omega$, the right form $G_m^{jk}$ in $\boldsymbol{G}$, and to estimate parameters $\alpha_{jk}(m)$.

To achieve these goals by GICE, we further assume:

1. For each $j,\ k\in\Omega$, $l=1,\cdots,L$, and $m=1,\cdots,M$ there exist estimators $\hat{\theta}_j(l)$, $\hat{\alpha}_{jk}(m)$;
2. There is a rule $D^1$ which decides for each set of distributions $f_{\theta(l)}\in H_1,\ \cdots,f_{\theta(L)}\in H_L$ the best one which fits the given sample $\mathbf{y}^1=\left(y_1^1,\cdots,y_{Q_1}^1\right)$, with $Q_1$ denoting the sample size;
3. There exists a rule $D^2$ which decides for each set of copulas $c_{\alpha(1)}\in G_1,\ \cdots,\ c_{\alpha(M)}\in G_M$ the best one which fits the given samples $\mathbf{y}^2=\left(y_1^2,\cdots,y_{Q_2}^2\right)$, with $Q_2$ denoting the sample size.

Then, GICE iteratively runs the following steps to figure out forms of margins, forms of copulas, and related parameters (with superscript $i$ denoting the iteration number).

1. Initialize GICE with $\left(p^0_{jk}, f^0_j, c^0_{jk}\right)$, for $j,\ k \in \Omega$;
2. Find $\left(p^{i+1}_{jk}, f^{i+1}_j, c^{i+1}_{jk}\right)$ from $\left(p^{i}_{jk}, f^{i}_j, c^{i}_{jk}\right)$ and $\mathbf{y}_1^N$ by the sub-steps below:
   (a) for $n = 1, \cdots, N-1$, compute $p^i\left(r_n = j, r_{n+1} = k|\mathbf{y}_1^N\right)$ from $\left(p^{i}_{jk}, f^{i}_j, c^{i}_{jk}\right)$ and $\mathbf{y}_1^N$ with
   $$p^i\left(r_n = j, r_{n+1} = k|\mathbf{y}_1^N\right) = \frac{\alpha_n(j)p(r_{n+1}=k, \mathbf{y}_{n+1}|r_n=j, \mathbf{y}_n)\beta_{n+1}(k)}{\sum_{j^*,k^*\in\Omega}\alpha_n(j)p(r_{n+1}=k^*, \mathbf{y}_{n+1}|r_n=j^*, \mathbf{y}_n)\beta_{n+1}(k)},$$
   where $\alpha_n(j)$ and $\beta_{n+1}(k)$ are obtained by applying (20), (21). Then update $p_{jk}$ with $p^{i+1}_{jk} = \frac{1}{N-1}\sum_{n=1}^{N-1} p^i\left(r_n = j, r_{n+1} = k|\mathbf{y}_1^N\right)$;
   (b) sample $\left(\mathbf{r}_1^N\right)^{i+1} = \left(r_1^{i+1}, \cdots, r_N^{i+1}\right)$ according to $p\left(\mathbf{r}_1^N\middle|\mathbf{y}_1^N\right)$ based on current parameters $\left(p^{i}_{jk}, f^{i}_j, c^{i}_{jk}\right)$ (recall that $p\left(\mathbf{r}_1^N\middle|\mathbf{y}_1^N\right)$ is Markov with $p\left(r_1 = j\middle|\mathbf{y}_1^N\right) = \frac{\alpha_1(j)\beta_1(j)}{\sum_{k\in\Omega}\alpha_1(k)\beta_1(k)}$, and $p\left(r_{n+1} = j\middle|r_n = k, \mathbf{y}_1^N\right) = \frac{p(r_{n+1}=j, \mathbf{y}_{n+1}|r_n=k, \mathbf{y}_n)\beta_{n+1}(j)}{\beta_n(k)}$ for $n = 1, \cdots, N-1$);
   (c) for each $j,\ k \in \Omega$ consider $\left(\mathbf{y}_1^N\right)^{i+1}_j$ the sub-sequence of $\mathbf{y}_1^N$ formed with $\mathbf{y}_n$ such that $r_n^{i+1} = j$, and $\left(\mathbf{y}_1^N\right)^{i+1}_{jk}$ the sub-sequence of couples $(\mathbf{y}_n, \mathbf{y}_{n+1})$ in $\mathbf{y}_1^N$ such that $r_n^{i+1} = j$ and $r_{n+1}^{i+1} = k$. For $l = 1, \cdots, L$ and $m = 1, \cdots, M$, calculate $\theta^{i+1}_j(l) = \hat{\theta}_j(l)\left[\left(\mathbf{y}_1^N\right)^{i+1}_j\right]$ and $\alpha^{i+1}_{jk}(m) = \hat{\alpha}_{jk}(m)\left[\left(\mathbf{y}_1^N\right)^{i+1}_{jk}\right]$;
   (d) for each $j \in \Omega$, choose from $\left\{f_{\theta^{i+1}_j(1)}, \cdots, f_{\theta^{i+1}_j(L)}\right\}$ an element $f^{i+1}_j$ by applying tule $D^1$ to the sample $\mathbf{y}^1 = \left(\mathbf{y}_1^N\right)^{i+1}_j$. Similarly, for $j,\ k \in \Omega$ chose from $\left\{c_{\alpha^{i+1}_{jk}(1)}, \cdots, c_{\alpha^{i+1}_{jk}(M)}\right\}$ an element $c^{i+1}_{jk}$ by applying rule $D^2$ to the sample $\mathbf{y}^2 = \left(\mathbf{y}_1^N\right)^{i+1}_{jk}$;
3. Stop according to some criterion.

For the initialization in step 1, K-means is applied to group $\mathbf{y}_1^N$ and find the initial guess of switches $\left(\mathbf{r}_1^N\right)^0$, then $\left(p^0_{jk}, f^0_j, c^0_{jk}\right)$ can be initialized from $\left(\mathbf{r}_1^N, \mathbf{y}_1^N\right)^0$ so as for the sub-steps (c) and (d) (replacing the iteration index "$i+1$" with "0"). As GICE is a general estimation frame, different parameter estimators and decision rules for assumptions 1-3 can be applied. In this work, Maximum Likelihood (ML) estimators are chosen for both $\hat{\theta}_j(l)$ and $\hat{\alpha}_{jk}(m)$ while in [4] $\hat{\alpha}_{jk}(m)$ were obtained by mean of the empirical estimation of Kendall's tau. Besides, we adopt the minimization of Kolmogorov distance as decision rule $D^1$, while GICE in [4] is based on Pearson's system of distributions. Let us note that $D^1$ considered here is valid for every set of distributions while the Pearson system used in [4,24] is limited to a set containing fixed eight possible forms.

The algorithm will stop when it is considered to be converged according to some criterion. For example, it stops when no change of form is observed for the estimation of both margins and copulas, and the difference of the estimated parameters from $D^1$ and $D^2$ between 2 iterations is within some predefined threshold.

### *3.2. Least-square estimation for non-linear switching model*

The last problem (iii) left is to find forms and parameters of $A\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right)$, $B\left(\mathbf{y}_n^{n+1}, \mathbf{y}_n^{n+1}\right)$, and $C\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right)$ defining $\mathbf{X}_{n+1}$ from $\mathbf{X}_n$ with (4), and being independent from $n = 1, \cdots, N$. We have seen that $C\left(\mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right)$ intervenes neither in filtering nor in smoothing, thus we concentrate on dealing with $A\left(\mathbf{r}_1^2, \mathbf{y}_1^2\right)$ and $B\left(\mathbf{r}_1^2, \mathbf{y}_1^2\right)$. Let us temporarily assume that their forms are given and for each $\mathbf{r}_1^2 = (j, k)$, they depend on parameters $a_{jk}$ and $b_{jk}$ respectively: $A\left(r_1 = j, r_2 = k, \mathbf{y}_1^2\right) = A_{a_{jk}}\left(\mathbf{y}_1^2\right)$, $B\left(r_1 = j, r_2 = k, \mathbf{y}_1^2\right) = B_{b_{jk}}\left(\mathbf{y}_1^2\right)$. When $p\left(\mathbf{r}_1^N\middle|\mathbf{y}_1^N\right)$ is given, the parameter estimation of the Gaussian $p\left(\mathbf{x}_{n+1}\middle|\mathbf{x}_n, \mathbf{y}_n^{n+1}, \mathbf{r}_n^{n+1}\right)$ can be considered as the estimation of a multi-regimes switching regression, and the Least-Square (LS) is an efficient method to deal with this. Extending the Ordinary Least-Square (OLS) to the non-Gaussian case that we deal with, estimates $\hat{a} = \left(\hat{a}_{jk}\right)_{j,k\in\Omega}$ and $\hat{b} = \left(\hat{b}_{jk}\right)_{j,k\in\Omega}$ are obtained by minimizing with respect to $(a_{jk})_{j,k\,\in\,\Omega}$, $(b_{jk})_{j,k\,\in\,\Omega}$ the quantity

$$e^2 = \frac{1}{N-1}\sum_{n=1}^{N-1}\left\{\mathbf{x}_{n+1} - \sum_{(j,k)} p\left(\mathbf{r}_n^{n+1} = (j,k)\middle|\mathbf{y}_1^N\right)\left[A_{a_{jk}}\left(\mathbf{y}_n^{n+1}\right)\mathbf{x}_n + B_{b_{jk}}\left(\mathbf{y}_n^{n+1}\right)\right]\right\}^2, \tag{25}$$

As previously done for copulas and margins, let us assume that the form of $A\left(\mathbf{r}_1^2, \mathbf{y}_1^2\right)$ is not known but belongs to a given set of forms $\{K_1, \cdots, K_Q\}$, with each form $K_q$ being parameterized by $a^q = \left(a^q_{jk}\right)_{j,k\in\Omega}$. Similarly, the form of $B\left(\mathbf{r}_1^2, \mathbf{y}_1^2\right)$ is not known but belongs to a given set of forms $\{L_1, \cdots, L_S\}$, with each form $L_s$ being parameterized by $b^s = \left(b^s_{jk}\right)_{j,k\in\Omega}$. Then, minimization of (25) is applied to each couple of forms $(K_q,\ L_s)$, giving estimated $\hat{a}^q = \left(\hat{a}^q_{jk}\right)_{j,k\in\Omega}$ and $\hat{b}^s = \left(\hat{b}^s_{jk}\right)_{j,k\in\Omega}$. Then the couple of forms finally kept is the couple $\left(\hat{K}_q, \hat{L}_s\right)$ for which the related $\left(\hat{a}^q, \hat{b}^s\right)$ obtains the minimum of (25) (comparing to other $\left(\hat{a}^{q^*}, \hat{b}^{s^*}\right)$) related to other couples $\left(K_{q^*}, L_{s^*}\right)$.

**Example 1.** Let us consider the linear case $A_{jk}\left(\mathbf{y}_n^{n+1}\right) = a_{jk}g_1\left(\mathbf{y}_n^{n+1}\right)$, $B_{jk}\left(\mathbf{y}_n^{n+1}\right) = b_{jk}g_2\left(\mathbf{y}_n^{n+1}\right)$, with $g_1$, $g_2$ given functions. The explicit solution (the vector stacking all $a_{jk}$ and $b_{jk}$) of the minimization of (25) is:

$$\widehat{\boldsymbol{\beta}}(\mathbf{x}) = (\mathrm{L}^\intercal(\mathbf{x})\mathrm{L}(\mathbf{x}))^{-1}\mathrm{L}^\intercal(\mathbf{x})\mathbf{x}, \tag{26}$$

with $\mathbf{x} = [\mathbf{x}_2\ \cdots\ \mathbf{x}_N]^\intercal$, and matrix given with

$$\mathrm{L} = \begin{bmatrix} p^1_{1,1}g^1_{1,2} & \cdots & p^1_{1,K}g^1_{1,2} & \cdots & g^1_{K,K}g^1_{1,2} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p^{N-1}_{1,1}g^{N-1}_{1,2} & \cdots & p^{N-1}_{1,K}g^{N-1}_{1,2} & \cdots & g^{N-1}_{K,K}g^{N-1}_{1,2} \end{bmatrix}, \tag{27}$$

where $p^n_{jk} = p\left(\mathbf{r}_n^{n+1} = (j,k)\middle|\mathbf{y}_1^N\right)$, and $g^n_{1,2} = \left[\mathbf{x}_n^\intercal g_1^\intercal\left(\mathbf{y}_n^{n+1}\right)\ g_2^\intercal\left(\mathbf{y}_n^{n+1}\right)\right]$.

For the general case we can turn to various numerical algorithms to minimize the error. A potential solution can be the Gauss-Newton method with linear approximation of the functions, the Powell's Dog Leg method with a control of trust region, or some other hybrid methods introduced in [25–27] respectively. In experiments of the next section we adopt the Levenberg–Marquardt (LM) algorithm, which is a damped Gauss–Newton method as proposed in [28] and completed in [29–31].

Combining the two identification steps above, the entire Schema of GICE-GLS for CB-CMSHLM identification is given in Fig. 1.

Concerning computation complexity of GICE-LS, it is proportional to the number of possible margins forms $L$, and the possible copulas forms $M$. In the general GICE-GLS it is also proportional

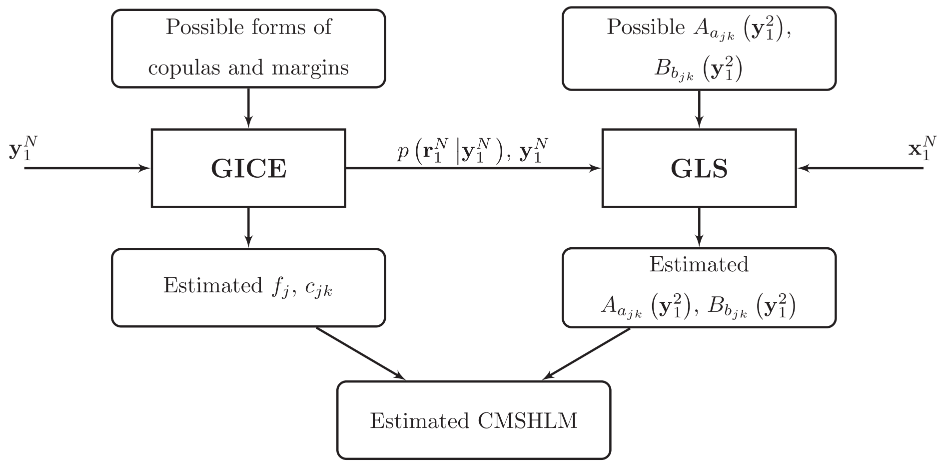



**Fig. 1.** Schema of CMSHLM estimation from learning sample $\left(\mathbf{x}_1^N, \mathbf{y}_1^N\right)$ through GICE-GLS.

to the number of possible forms of $A_a\left(\mathbf{y}_1^2\right)$ and $B_b\left(\mathbf{y}_1^2\right)$. Finally, the algorithm is proportional to the number of iterations. However, it is worthwhile to highlight that similar to the classic ICE in simple hidden Gaussian Markov chains, complexity time of ICE-GLS is linear in sample size $N$.

Let us remark that theoretical study of the GICE-GLS sequence is not tackled here and would probably be somewhat hard. However, we can note that it is a Markov chain from its very construction so that, when it converges in distribution, the limit does not depend on initial values, under some conditions to be verified. In different experiments we performed we noticed that indeed, results are very little sensitive to initializations. Next, we provide a short study of four different initializations (see experiments for Series 1).

## 4. Experiments

We present two series of experiments on simulated data and test a simplified version of GICE-GLS, called GICE-LS, in which the parameterized forms $A_{a_{jk}}\left(\mathbf{y}_1^2\right)$ and $B_{b_{jk}}\left(\mathbf{y}_1^2\right)$ are known, and the problem lies only in estimation of their parameters.

In the first series, the learning sample and data to be restored are simulated according to a CB-CMSHLM. After having identified the CB-CMSHLM through GICE-LS from the learning sample, filtering and smoothing obtained results are compared to the results got from the other two methods. The first one uses parameters estimated by ICE-LS and data are restored by exact restoration considering Gaussian margins and copulas. In the second one, identification and restoration are performed through CGOMSM-ABF proposed in [8,9]. The aim is to show that when data are not Gaussian considering them as Gaussian can significantly degrade the filtering and smoothing results.

In the second series, data are sampled with respect to a CGOMSM. The aim is to verify that when data follows the simpler Gaussian CGOMSM, which is a particular case of CB-CMSHLM, GICE-LS based filtering and smoothing provides a result comparable to those obtained with ICE-LS and CGOMSM- ABF.

The considered CB-CMSHLM is defined as follows.

- Both hidden states and observations are scalar;
- The Markov chain $\mathbf{R}_1^N$ is stationary and has $K = 2$ jumps;
- The margins are of six possible forms (see Appendix for details):

$$H = \{H_1, \cdots, H_6\} = \{\text{Gamma, Fisk, Gaussian, Laplace, Beta, Beta prime}\}, \tag{28}$$

- The copulas are of seven possible forms (all of them – except Product – belong to one-parameter copula families (see Appendix for details):

$$G = \{G_1, \cdots, G_7\} = \{\text{Gumble, Gaussian, Clayton, FGM, Arch12, Arch14, Product}\}, \tag{29}$$

- All estimators $\hat{\theta}^j(l)$ are the Maximum Likelihood ones;
- Rule $D^1$ consists of minimizing the Kolmogorov distance between empirical distribution $\hat{F}$ and candidates $F_1 \in H_1, \cdots, F_L \in H_L$. The Kolmogorov distance between two CDFs $F$, $F'$ is defined as

$$d\left(F, F'\right) = \sup_{\mathbf{y}\in\mathbb{R}^d} \left|F(\mathbf{y}) - F'(\mathbf{y})\right|, \tag{30}$$

where the notation |.| denotes the absolute value, and thus for a sample $\boldsymbol{\mu}_1^Q = \left(\mu_1, \cdots, \mu_Q\right)$ the chosen CDF $D^1\left(\mu_1^Q\right)$ among candidates $F_1 \in H_1, \cdots, F_L \in H_L$, is defined with:

$$D^1\left(\boldsymbol{\mu}_1^Q\right) = \arg\inf_{l\in\{1,\cdots,L\}} \left[d\left(F_l, \hat{F}\right)\right], \tag{31}$$

where empirical CDF $\hat{F}$ is given by:

$$\hat{F}(\mu) = \frac{1}{Q}\sum_{n=1}^{Q} \mathbb{1}_{[\mu_n \le \mu]}, \tag{32}$$

- Estimators $\hat{a}_{jk}$ are obtained with the method presented in [32]. For a sample $\mu_1^{2Q} = \left((\mu_1, \mu_2), \cdots, \left(\mu_{2Q-1}, \mu_{2Q}\right)\right)$, we have:

$$\hat{a}\left(\mu_1^{2Q}\right) = \arg\max_a \left[\sum_{n=1}^{N-1} \log\left(c_a\left(\hat{F}(\mu_n), \hat{F}(\mu_{n+1})\right)\right)\right], \tag{33}$$

where, $\hat{F}(\mu_n)$, $\hat{F}(\mu_{n+1})$ are empirical CDFs calculated from $\left(\mu_1, \cdots, \mu_{2Q-1}\right)$ and $(\mu_2, \cdots, \mu_{2Q})$ respectively. Let us remark that other copula estimation methods [33–35] could replace the applied ones.

- Finally, the rule $D^2$ is the maximum of pseudo-likelihood: for a sample $\mu_1^{2Q} = \left((\mu_1, \mu_2), \cdots, \left(\mu_{2Q-1}, \mu_{2Q}\right)\right)$, copula $\hat{c}$ related to each distribution $p(\mu_{2n-1}, \mu_{2n})$ is chosen among candidates $c_1 \in G_1, \cdots, c_M \in G_M$ with:

$$D^2\left(\mu_1^{2Q}\right) = \arg\sup_{m\in\{1,\cdots,M\}} \prod_{n=1}^{N-1} c_m\left(\left[\hat{F}(\mu_n), \hat{F}(\mu_{n+1})\right]\right), \tag{34}$$

with $\hat{F}(\mu_n)$, $\hat{F}(\mu_{n+1})$ being empirical CDFs as above.

**Table 1**
True margins, copulas, and their estimates from GICE-LS (extracted from cases in which true copulas and margins are perfectly found).

| Margins and parameters | $f_1(\theta_1)$ (Gamma) | $f_2(\theta_2)$ (Fisk) | Copulas and parameters | $c_{11}(\alpha_{11})$ (Gumbel) | $c_{22}(\alpha_{22})$ (Clayton) | $c_{12}(\alpha_{12}) = c_{21}(\alpha_{21})$ (Gaussian) |
|---|---|---|---|---|---|---|
| True $\theta_i$ | 16.00 | 4.00 | True $\alpha_{jk}$ | 1.10 | 4.67 | 0.45 |
| Estimated $\theta_i$ | 13,72 | 3.93 | Estimated $\alpha_{jk}$ | 1.15 | 4.46 | 0.46 |

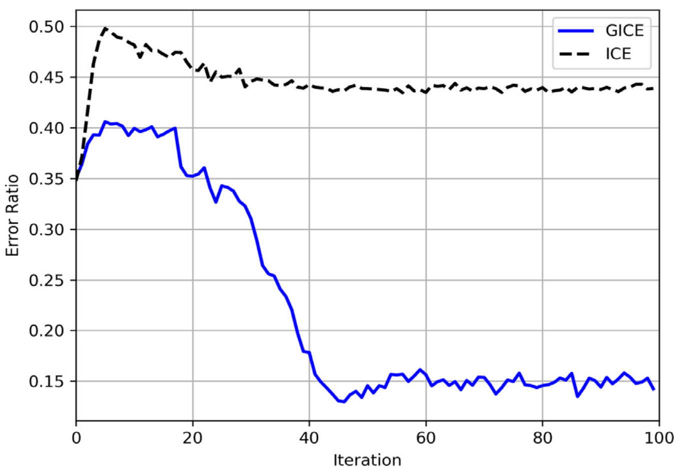



**Fig. 2.** Error ratio tendency of estimated $\mathbf{R}_1^N$ according to GICE and ICE iterations in Series 1.

**Table 2**
Error ratios and MSEs of optimal (based on true parameters) restorations, as well as the GICE-LS, ICE-LS and CGOMSM-ABF based ones (average of 100 independent experiments).

| | | Optimal | GICE-LS | ICE-LS | CGOMSM-ABF |
|---|---|---|---|---|---|
| Filtering | Error | 0.139 | 0.156 | 0.404 | 0.462 |
| | MSE | 2.380 | 2.771 | 5.762 | 9.353 |
| Smoothing | Error | 0.084 | 0.103 | 0.378 | 0.456 |
| | MSE | 2.290 | 2.631 | 5.750 | 9.273 |

**Table 3**
True $a_{jk}$, $b_{jk}$, $d_{jk}$ and their estimates (extracted from cases in which true copulas and margins are found).

| | True | | | | Estimates | | | |
|---|---|---|---|---|---|---|---|---|
| (j, k) | (1,1) | (1,2) | (2,1) | (2,2) | (1,1) | (1,2) | (2,1) | (2,2) |
| $a_{j,k}$ | 0.20 | 0.40 | 0.60 | 0.80 | 0.27 | 0.41 | 0.69 | 0.81 |
| $b_{j,k}$ | 0.70 | 0.50 | 0.60 | 0.90 | 0.69 | 0.56 | 0.63 | 0.90 |
| $d_{j,k}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 | -0.13 | -0.01 |

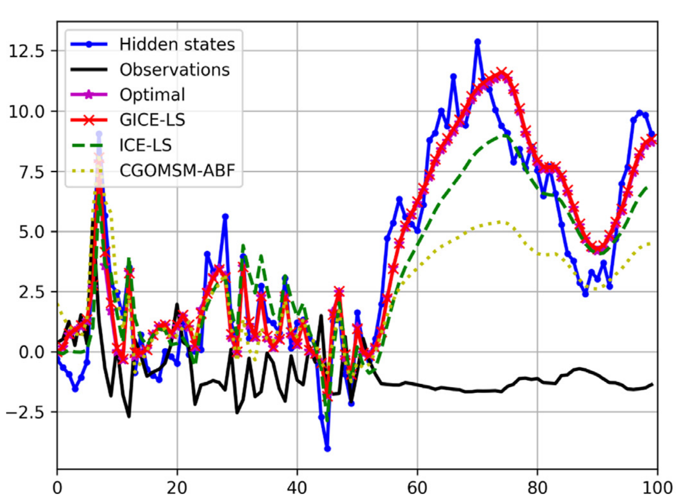
**Table 4**
Margins identification error ratio. $f_1$ is Gamma, and $f_2$ is Fisk - see Table 1.

| | Gamma | Fisk | Gaussian | Laplace | Beta | Beta prime |
|---|---|---|---|---|---|---|
| Identified $f_1$ | 87% | 12% | - | 1% | - | - |
| Identified $f_2$ | 1% | 99% | - | - | - | - |

***Series 1.***

In both series the probabilities $p_{jk} = p(r_1 = j, r_2 = k)$ defining the distribution of stationary $\mathbf{R}_1^N$ are $p_{11} = p_{22} = 0.45$, $p_{12} = p_{21} = 0.05$. A set of $N = 5000$ simulated $(\mathbf{x}_n,\ \mathbf{y}_n)$ is taken as a learning sample used for the model identification, and another set of $N = 1000$ simulated data is taken for testing form identification, parameter estimation, and related filtering and smoothing based on real and estimated models. The margins and copulas in $p\left(\mathbf{r}_1^N, \mathbf{y}_1^N\right)$ are set in Table 1. $p\left(\mathbf{x}_{n+1} \middle| \mathbf{x}_n, r_n = j, r_{n+1} = k, \mathbf{y}_n^{n+1}\right)$ are Gaussian with means $a_{jk}\mathbf{x}_n + B_{jk}\left(\mathbf{y}_n^{n+1}\right)$ – where $B_{jk}\left(\mathbf{y}_n^{n+1}\right) = b_{jk}\mathbf{y}_n\mathbf{y}_{n+1} + d_{jk}$ are non-linear in $\mathbf{y}_n$, $\mathbf{y}_{n+1}$, and the variances $\sigma_{jk}^2 = \left[C_{n+1}\left(r_n = j, r_{n+1} = k, \mathbf{y}_n^{n+1}\right)\right]^2$, which are thus independent from $n$ and $\mathbf{y}_n^{n+1}$. Let us recall that variances $\sigma_{jk}^2$ are only used to sample data and neither interfere in filtering nor smoothing. They are taken as $\sigma_{11}^2 = \sigma_{22}^2 = 1.0$, and $\sigma_{12}^2 = \sigma_{21}^2 = 0.8$. Restoration results of all four methods are indicated in Table 2. From the results and those of other similar experiments performed, we can advance the following conclusions:

(1) GICE-LS based filtering and smoothing are quite efficient for the data which follows CB-CMSHLM, with MSE close to the optimal one;
(2) ICE-LS provides better results than CGOMSM-ABF. Both of them wrongly assume that $p\left(\mathbf{y}_1^N \middle| \mathbf{r}_1^N\right)$ is Gaussian; the difference lies in the fact that CGOMSM-ABF also assumes $p\left(\mathbf{x}_n^{n+1} \middle| \mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right)$ Gaussian, while ICE-LS limits the Gaussian assumption to



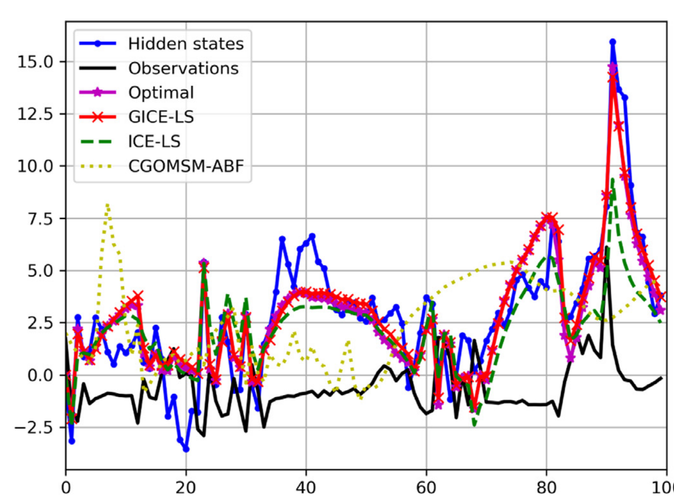




**Fig. 3.** Two examples of trajectories from Series 1 experiment (100 samples, smoothing).

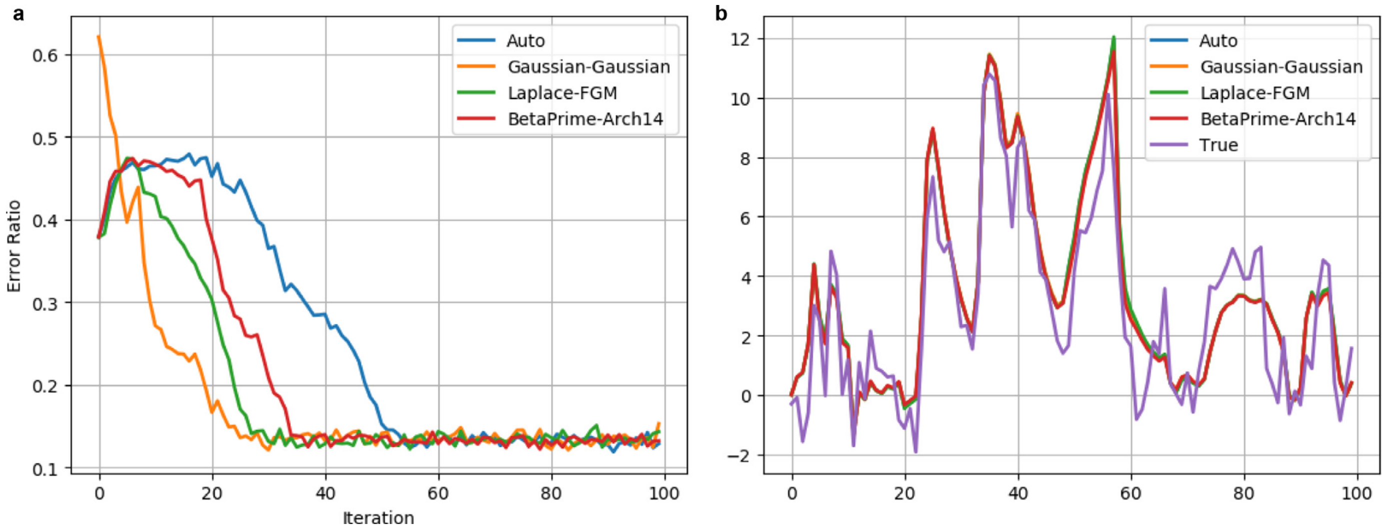



**Fig. 4.** Error ratio evolution while initializing with different margins and copulas (a). Trajectories evolution while initializing with different copulas (b). Auto corresponds to the GICE-LS initialization, Gaussian-Gaussian means all Gaussian margins and all Gaussian copulas, Laplace-FGM means all Laplace margins and all FGM copulas, and BetaPrime-Arch14 refers to all BetaPrime margins and all Arch14 copulas.

**Table 5**
Copulas identification error ratio. $c_{11}$ is Gumbel, $c_{12} = c_{21}$ are Gaussians, and $c_{22}$ is Clayton - see Table 1.

| | Gumbel | Gaussian | Clayton | FGM | Arch12 | Arch14 | Product |
|---|---|---|---|---|---|---|---|
| Identified $c_{11}$ | 96% | 2% | 1% | - | 1% | - | - |
| Identified $c_{12} = c_{21}$ | 34% | 58% | 4% | 8% | - | - | - |
| Identified $c_{22}$ | 2% | - | 96% | 1% | 1% | - | - |

**Table 6**
Error ratio and smoothing MSE while initializing with different margins and copulas. Auto corresponds to the GICE-LS initialization, Gaussian-Gaussian means all Gaussian margins and all Gaussian copulas, Laplace-FGM means all Laplace margins and all FGM copulas, and BetaPrime-Arch14 refers to all BetaPrime margins and all Arch14 copulas.

| Initialization | Auto | Gaussian-Gaussian | Laplace-FGM | BetaPrime-Arch14 |
|---|---|---|---|---|
| Error ratio | 0.062 | 0.060 | 0.063 | 0.055 |
| MSE | 3.013 | 3.142 | 2.930 | 3.439 |

**Table 7**
Error ratios and Mean Square Errors (MSEs) of optimal (based on true parameters) filtering and smoothing, and GICE-LS, ICE-LS, and CGOMSM-ABF based ones. Data sampled with CGOMSM with parameters given in Tables 8 and 9.

| | | Optimal | GICE-LS | ICE-LS | CGOMSM-ABF |
|---|---|---|---|---|---|
| Filtering | Error | 0.245 | 0.289 | 0.249 | 0.247 |
| | MSE | 1.037 | 1.047 | 1.044 | 1.044 |
| Smoothing | Error | 0.211 | 0.261 | 0.215 | 0.213 |
| | MSE | 1.032 | 1.044 | 1.039 | 1.040 |

$p\left(\mathbf{x}_{n+1} \middle| \mathbf{x}_n, \mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right)$. Thus, the non-linearity of $B_{jk}(\mathbf{y}_{n+1}) = b_{jk}\mathbf{y}_n\mathbf{y}_{n+1} + d_{jk}$ is better taken into account by ICE-LS;

(3) GICE can select false margins and copula, especially for $c_{12} = c_{21}$ (see Tables 4 and 5). However, this does not significantly degrade the optimal filtering and smoothing results;

(4) The estimates of $p_{jk}$ from GICE-LS are quite close to the true ones: $\hat{p}_{11} = 0.474$, $\hat{p}_{22} = 0.445$, $\hat{p}_{12} = \hat{p}_{21} = 0.040$;

(5) According to Fig. 2, where the error ratio of unsupervised switches estimation is concerned, GICE is much more effective than ICE;

(6) The two trajectory examples displayed in Fig. 3 clearly illustrate the superiority of GICE-LS over the other methods on the restoration of general CB-CMSHLM data considered;

(7) According to Table 3 estimates of $a_{jk}$, $b_{jk}$, and $d_{jk}$ are quite correct.

As stated at the end of the previous section, the sequence of margins, copulas, and parameters produced with GICE-GLS is a Markov chain and thus, under some ergodic theorem hypotheses, their evolution "forgets" initialization. Let us present a short study of four different initializations:

- Auto refers to the GICE-LS initialization described above;
- Gaussian-Gaussian means an initialization with all Gaussian margins and all Gaussian copulas;
- Laplace-FGM refers to an initialization with all Laplace margins and all FGM copulas;
- BetaPrime-Arch14 means an initialization with all BetaPrime margins and all Arch14 copulas.

Table 6 reports the error ratio and smoothing MSE while initializing with different margins and copulas, whereas Fig. 4(a) gives the error ratio evolution of unsupervised MPM search of $\mathbf{R}_1^N = \mathbf{r}_1^N$ from $\mathbf{Y}_1^N = \mathbf{y}_1^N$ as a function of GICE iterations. It appears that after 50 iterations, the initialization has little influence. Similarly, semi-supervised GICE-LS based smoothing results presented in Fig. 4(b) show that Auto, Gaussian-Gaussian, and Laplace-FGM initializations give the same plot (hidden behind the green one), the green curve is almost identical to the red one which corresponds to a BetaPrime-Arch14 initialization.

**Table 8**
True margins, copulas (Gaussian) and their estimates (extracted from cases in which true copulas and margins are found).

| | Margins | $f_1(\theta_1)$ | $f_2(\theta_2)$ | Copulas | $c_{11}(\alpha_{11})$ | $c_{22}(\alpha_{22})$ | $c_{12}(\alpha_{12}) = c_{21}(\alpha_{21})$ |
|---|---|---|---|---|---|---|---|
| | True $\theta_i$ | 0.00 | 1.00 | True $\alpha_{jk}$ | 0.80 | 0.20 | 0.45 |
| ICE | Estimated $\theta_i$ | 0.01 | 1.00 | Estimated $\alpha_{jk}$ | 0.79 | 0.20 | 0.42 |
| GICE | Estimated $\theta_i$ | -0.04 | 0.99 | Estimated $\alpha_{jk}$ | 0.78 | 0.23 | 0.49 |

**Table 9**
True $a_{jk}$, $b_{jk}$, $e_{jk}$, $d_{jk}$ and their estimates (extracted from cases in which true copulas and margins are found).

| | True | | | | ICE-LS estimates | | | | GICE-LS estimates | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (*j*, *k*) | (1,1) | (1,2) | (2,1) | (2,2) | (1,1) | (1,2) | (2,1) | (2,2) | (1,1) | (1,2) | (2,1) | (2,2) |
| $a_{jk}$ | 0.30 | 0.50 | 0.50 | 0.70 | 0.30 | 0.52 | 0.48 | 0.69 | 0.34 | 0.56 | 0.47 | 0.67 |
| $b_{jk}$ | 0.61 | 0.05 | 0.25 | -0.19 | 0.60 | 0.03 | 0.25 | -0.16 | 0.50 | 0.05 | 0.20 | -0.11 |
| $e_{jk}$ | 0.30 | 0.70 | 0.30 | 0.70 | 0.31 | 0.71 | 0.31 | 0.71 | 0.39 | 0.78 | 0.27 | 0.64 |
| $d_{jk}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.04 | 0.08 | -0.01 | 0.01 | 0.07 | 0.01 | 0.02 |

**Table 10**
Margins identification error ratio. $f_1$ and $f_2$ are Gaussians.

| | Gamma | Fisk | Gaussian | Laplace | Beta | Beta prime |
|---|---|---|---|---|---|---|
| Identified $f_1$ | 2% | 1% | 86% | 11% | - | - |
| Identified $f_2$ | 5% | 3% | 54% | 1% | - | 37% |

**Table 11**
Copulas identification error ratio. $c_{11}$ is Gumbel, $c_{12} = c_{21}$ are Gaussians, and $c_{22}$ is Clayton.

| | Gumbel | Gaussian | Clayton | FGM | Arch12 | Arch14 | Product |
|---|---|---|---|---|---|---|---|
| Identified $c_{11}$ | 1% | 43% | 2% | - | 3% | 51% | - |
| Identified $c_{12} = c_{21}$ | 32% | 52% | 10% | 4% | - | 2% | - |
| Identified $c_{22}$ | 14% | 60% | 4% | 19% | - | 3% | - |

Performing on a 3.6 GHz CPU, GICE-LS in this series takes about 180 s (for 100 iterations), while filtering and smoothing both takes around 0.03 s

***Series 2.***

In this second series, data is sampled with respect to a CGOMSM. The aim is to verify whether more complex GICE-LS, which considers six possible margins and seven possible copulas, is competing compared to ICE-LS, which uses just the right Gaussian margins and copulas. Thus in this series, both $p\left(\mathbf{y}_n^{n+1}\middle|\mathbf{r}_n^{n+1}\right)$ and $p\left(\mathbf{x}_{n+1}\middle|\mathbf{x}_n, \mathbf{r}_n^{n+1}, \mathbf{y}_n^{n+1}\right)$ are set to be Gaussian with $A_{jk}\left(\mathbf{y}_n^{n+1}\right) = a_{jk}$ and $B_{jk}\left(\mathbf{y}_n^{n+1}\right) = b_{jk}\mathbf{y}_n + e_{jk}\mathbf{y}_{n+1} + d_{jk}$, with $a_{jk}$, $b_{jk}$, $e_{jk}$ and their estimates specified in Table 9. Estimated switching joint probabilities from GICE are $p_{11} = 0.485$, $p_{22} = 0.421$, $p_{12} = p_{21} = 0.047$; while from ICE, they are $p_{11} = 0.489$, $p_{22} = 0.419$, $p_{12} = p_{21} = 0.046$.

According to Table 7 GICE-LS based filtering and smoothing results are comparable to ICE-LS and CGOMSM-ABF based ones, all of them being close to the optimal results. As in the previous series, GICE cannot always find Gaussian margins and Gaussian copulas – see Tables 10 and 11. However, this does not affect the restoration seriously since the found distributions are close to Gaussian ones, at least where filtering and smoothing are concerned.

## 5. Conclusion

We introduce an identifiable general switching CMSHLM model with copulas, called copulas based CMSHLM (CB-CMSHLM), and propose a family of methods called "generalized Iterative conditional estimation with generalized least squares" (GICE-GLS) for its identification from a set of admissible family of models. Recursive exact filtering and smoothing are then possible using CB-CMSHLM in a semi-unsupervised way. The high adaptable identification ability of GICE-LS, which is a particular simplified GICE-GLS, has been verified by experiments on both Gaussian linear and non-Gaussian non-linear data.

There are many perspectives for further work:

(1) Include the estimation of $C_{n+1}$ in (4), when dealing with the parameter estimations of $A_{n+1}$ and $B_{n+1}$; possibly by weighted least-square;
(2) Other alternative parameter estimation methods under the GICE frame are worth trying to improve the performance in specific situations. For example, the moments method could replace ML as the estimator for margins, while for copulas, a popular way is to estimate their Kendall's tau. Moreover, instead of the semi-parametric estimation applied in our work, parametric or non-parametric methods [33,35] are probably also worth a test;
(3) The model and methods proposed are easy to extend to higher dimensional state-spaces, at least when parameters are known. Their interest with respect to Markov chain Monte Carlo (MCMC) based methods is expected to increase when the state-space dimension grows, since under high dimension circumstance, a large amount of particles will be required by MCMC methods, therefore it loads us with the burden of calculation;
(4) The proposed GICE-GLS identification for CB-CMSHLMs is semi-supervised, for which a sample containing observations $\mathbf{Y}_1^N$ and states $\mathbf{X}_1^N$ is required, while switches $\mathbf{R}_1^N$ are unknown. Extending the method to a fully unsupervised one, which would work from the $\mathbf{Y}_1^N$ only, is an important perspective for applications. One possible idea to explore solutions could be inspired by the "double EM" algorithm proposed in the Gaussian case in [36];
(5) There are many possible variations over the several known copulas, margins, and functions $A_{n+1}$, $B_{n+1}$ and $C_{n+1}$ in (4). Choosing the best model for a given concrete problem opens a huge field of perspectives. In particular, stochastic volatility is an important item in finance [22,37–39] among others. Some

first approximation studies with CGOMSMs turned out to work well [8,40], and thus applying more complex CMSHLMs is a possible perspective for further works;

(6) Markov chains dealt in this paper are the simplest Markov graphical models and extensions of proposed CB-CMSHLMs to other Markov graphical models, for example those studied in [41], is another perspective to view. Some rare applications of hidden particular Markov graphical models with copulas to image processing have been proposed in hidden Markov trees [42], or hidden Markov fields [43,44]; however, copulas are still rarely used in hidden Markov models because the observations are, in general, assumed to be independent conditionally on hidden states;

(7) Classic switches considered in this paper could possibly be extended to "fuzzy" switches, as recently proposed in [45,46], which results in as many possibilities of extensions of the proposed models.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix.

Six standard forms of margin distributions and related parameters used in experiments are:

- Gamma: Setting $\Gamma(\theta)=\int_0^{+\infty} t^{\theta-1}\exp(-t)dt$ and $\gamma(\theta,y)=\int_0^y t^{\theta-1}\exp(-t)dt$, CDF $F$ and PDF $f$ are $F(y)=\frac{\gamma(\theta,y)}{\Gamma(\theta)}$, $f(y)=\frac{y^{\theta-1}\exp(-y)}{\Gamma(\theta)}$ (for $\theta>0$);
- Fisk (also known as log-logistic distribution): $F(y)=\frac{1}{1+y^{-\theta}}$, $f(y)=\frac{\theta y^{\theta-1}}{(1+y^{\theta})^2}$ (for $\theta>0$);
- Gaussian: Setting $erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x\exp(-t^2)dt$, $F(y)=\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{y}{\sqrt{2}}\right)\right)$ and $f(y)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)$;
- Laplace: $F(y)=\begin{cases}1-\frac{1}{2}\exp(-y) & \text{if } y\geq 0\\ \frac{1}{2}\exp(y) & \text{if } y<0\end{cases}$, $f(y)=\frac{1}{2}\exp(-|y|)$;
- Beta: setting $B(\theta_1,\theta_2)=\int_0^1 t^{\theta_1-1}(1-t)^{\theta_2-1}dt$, $I(x,\theta_1,\theta_2)=\int_0^x t^{\theta_1-1}(1-t)^{\theta_2-1}dt$, $F(y)=\frac{I(y,\theta_1,\theta_2)}{B(\theta_1,\theta_2)}$, and $f(y)=\frac{\Gamma(\theta_1+\theta_2)y^{\theta_1-1}(1-y)^{\theta_2-1}}{\Gamma(\theta_1)\Gamma(\theta_2)}$ (for $\theta_1>0$, $\theta_2>0$);
- Beta prime (also called beta distribution of the second kind or inverted beta distribution): $F(y)=I\left(\frac{y}{1+y},\theta_1,\theta_2\right)$, and $f(y)=\frac{y^{\theta_1-1}(1+y)^{-\theta_1-\theta_2}}{B(\theta_1,\theta_2)}$ (for $\theta_1>0$, for $\theta_2>0$).

Seven forms of copulas and related parameters used in experiments are:

- Gumbel copula: Setting $U_1=(-\ln(\mu_1))^{\alpha}$, $U_2=(-\ln(\mu_2))^{\alpha}$, CDF $C$ and PDF $c$ are (for $\alpha\in[1,+\infty[$) $C(\mu_1,\mu_2)=\exp\left(-(U_1+U_2)^{1/\alpha}\right)$, $c(\mu_1,\mu_2)=\frac{U_1}{\mu_1\ln(\mu_1)}\frac{U_2}{\mu_2\ln(\mu_2)}(\alpha-1+U_1+U_2)^{1/\alpha}(U_1+U_2)^{1/\alpha-2}\exp\left[-(U_1+U_2)^{1/\alpha}\right]$;
- Gaussian copula: Setting $\phi$ standard Gaussian PDF (mean 0 and variance 1), $\xi=\begin{bmatrix}\phi^{-1}(\mu_1)\\ \phi^{-2}(\mu_2)\end{bmatrix}$, $I=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$, $\rho=\begin{bmatrix}1 & \alpha\\ \alpha & 1\end{bmatrix}$ (for $\alpha\in]-1,1[$), $C(\mu_1,\mu_2)=\int_0^{\mu_1}\phi\left(\frac{\phi^{-1}(\mu_2)-\alpha\phi^{-1}(\mu_1)}{\sqrt{1-\alpha^2}}\right)d\mu$, $c(\mu_1,\mu_2)=\frac{1}{1-\alpha^2}\exp\left(-\frac{1}{2}\xi^{\intercal}(\rho-I)\xi\right)$;
- Clayton copula: $C(\mu_1,\mu_2)=\left(\mu_1^{-\alpha}+\mu_2^{-\alpha}-1\right)^{1/\alpha}$, $c(\mu_1,\mu_2)=(1+\alpha)\mu_1^{-1-\alpha}\mu_2^{-1-\alpha}\left(\mu_1^{-\alpha}+\mu_2^{-\alpha}-1\right)^{-(1/\alpha)-2}$ (for $\alpha\in[0,+\infty[$);
- FGM (Farlie–Gumbel–Morgenstern) copula: $C(\mu_1,\mu_2)=\mu_1\mu_2(1+\alpha(1-\mu_1)(1-\mu_2))$, $c(\mu_1,\mu_2)=1+\alpha\ (1-2\mu_1)(1-2\mu_2)$;
- Arch 12 (Archimedean of order 12) copula: Setting $U_1=\left(\frac{1}{u_1}-1\right)^{\alpha}$, $U_2=\left(\frac{1}{u_2}-1\right)^{\alpha}$ (for $\alpha\in[1,+\infty[$), $C(\mu_1,\mu_2)=\left(1+(U_1+U_2)^{1/\alpha}\right)^{-1}$, $c(\mu_1,\mu_2)=\frac{U_1}{u_1(1-u_1)}\frac{U_2}{u_2(1-u_2)}\frac{\left[\alpha-1+(1+\alpha)(U_1+U_2)^{1/\alpha}\right](U_1+U_2)^{(1/\alpha)-2}}{\left[1+(U_1+U_2)^{1/\alpha}\right]^3}$;
- Arch 14 (Archimedean of order 14) copula: Setting $U_1=\left(u_1^{-1/\alpha}-1\right)^{\alpha}$, $U_2=\left(u_2^{-1/\alpha}-1\right)^{\alpha}$ (for $\alpha\in[1,+\infty[$), $C(\mu_1,\mu_2)=\left(1+(U_1+U_2)^{1/\alpha}\right)^{-\alpha}$, $c(\mu_1,\mu_2)=U_1U_2(U_1+U_2)^{(1/\alpha)-2}\left[1+(U_1+U_2)^{1/\alpha}\right]^{-2-\alpha}\frac{\left[\alpha-1+2\alpha(U_1+U_2)^{1/\alpha}\right]}{\alpha u_1u_2\left(u_1^{1/\alpha}-1\right)\left(u_2^{1/\alpha}-1\right)}$;
- Product copula: $C(\mu_1,\mu_2)=\mu_1\mu_2$, $c(\mu_1,\mu_2)=1$.

## CRediT authorship contribution statement

**Fei Zheng:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **Stéphane Derrode:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing - review & editing. **Wojciech Pieczynski:** Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing - review & editing.

## References


[1] W. Pieczynski, Exact filtering in conditionally Markov switching hidden linear models, Comptes Rendus Mathématique 349 (9-10) (2011) 587–590.
[2] N. Brunel, W. Pieczynski, Unsupervised signal restoration using hidden Markov chains with copulas, Signal Process. 85 (12) (2005) 2304–2315.
[3] J.P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, IEEE Trans. Signal Process. 45 (10) (1997) 2613–2615.
[4] S. Derrode, W. Pieczynski, Unsupervised classification using hidden Markov chain with unknown noise copulas and margins, Signal Process. 128 (2016) 8–17.
[5] N. Abbassi, D. Benboudjema, W. Pieczynski, Kalman filtering approximations in triplet Markov Gaussian switching models, in: Statistical Signal Processing Workshop (SSP), IEEE, 2011, pp. 77–80.
[6] N. Abbassi, D. Benboudjema, S. Derrode, W. Pieczynski, Optimal filter approximations in conditionally Gaussian pairwise Markov switching models, IEEE Trans. Autom. Control 60 (4) (2015) 1104–1109.
[7] S. Derrode, W. Pieczynski, Exact fast computation of optimal filter in Gaussian switching linear systems, IEEE Signal Process. Lett. 20 (7) (2013) 701–704.
[8] I. Gorynin, S. Derrode, E. Monfrini, W. Pieczynski, Fast filtering in switching approximations of nonlinear Markov systems with applications to stochastic volatility, IEEE Trans. Autom. Control 62 (2) (2017) 853–862.
[9] I. Gorynin, S. Derrode, E. Monfrini, W. Pieczynski, Fast smoothing in switching approximations of non-linear and non-Gaussian models, Comput. Stat. Data Anal. 114 (2017) 38–46.
[10] Y. Petetin, F. Desbouvries, A class of fast exact Bayesian filters in dynamical models with jumps, IEEE Trans. Signal Process. 62 (14) (2014) 3643–3653.
[11] A. Doucet, A.M. Johansen, A tutorial on particle filtering and smoothing: fifteen years later, Handbook of Nonlinear Filtering 12 (656-704) (2009) 3.
[12] O. Cappé, E. Moulines, T. Ryden, Inference in Hidden Markov Models (Springer Series in Statistics), Springer-Verlag, Berlin, Heidelberg, 2005.
[13] A. Doucet, N.J. Gordon, V. Krishnamurthy, Particle filters for state estimation of jump Markov linear systems, IEEE Trans. Signal Process. 49 (3) (2001) 613–624.
[14] A. Doucet, C. Andrieu, Iterative algorithms for state estimation of jump Markov linear systems, IEEE Trans. Signal Process. 49 (6) (2001) 1216–1227.
[15] P. Fearnhead, P. Clifford, On-line inference for hidden Markov models via particle filters, J. R. Stat. Soc. B 65 (4) (2003) 887–899.
[16] W. Li, Y. Jia, Rao–Blackwellised particle filtering and smoothing for jump Markov non-linear systems with mode observation, IET Signal Process. 7 (4) (2013) 327–336.
[17] F. Lindsten, P. Bunch, S. Särkkä, T.B. Schön, S.J. Godsill, Rao–Blackwellized particle smoothers for conditionally linear Gaussian models, IEEE J. Sel. Top. Signal Process. 10 (2) (2015) 353–365.
[18] T.N.M. Nguyen, S. Le Corff, E. Moulines, On the two-filter approximations of marginal smoothing distributions in general state-space models, Adv. Appl. Probab. 50 (1) (2017) 154–177.

[19] N.M. Nguyen, S. Le Corff, E. Moulines, Particle rejuvenation of Rao-Blackwellized sequential Monte-Carlo smoothers for conditionally linear and Gaussian models, EURASIP J. Adv. Signal Process. 2017 (1) (2017) 54.
[20] J. Olsson, J. Westerborn, et al., Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm, Bernoulli 23 (3) (2017) 1951–1996.
[21] S. Särkkä, P. Bunch, S.J. Godsill, A backward-simulation based Rao-Blackwellized particle smoother for conditionally linear Gaussian models, IFAC Proc. Vol. 45 (16) (2012) 506–511.
[22] V. Genon-Catalot, T. Jeantheau, C. Larédo, et al., Stochastic volatility models as hidden Markov models and statistical applications, Bernoulli 6 (6) (2000) 1051–1079.
[23] A. Sklar, Fonctions de Répartition à N Dimensions Et Leurs Marges, Université Paris 8, 1959. https://books.google.fr/books?id=nreSmAEACAAJ
[24] S. Derrode, W. Pieczynski, Unsupervised data classification using pairwise Markov chains with automatic copulas selection, Comput. Stat. Data Anal. 63 (2013) 81–98.
[25] A. Björck, Numerical methods for least squares problems, SIAM, 1996.
[26] K.M. Brown, J.E. Dennis, A new algorithm for nonlinear least-squares curve fitting, Math. Softw. (2014) 391.
[27] K. Madsen, H.B. Nielsen, O. Tingleff, Methods for Non-Linear Least Squares Problems (2nd ed.), Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2004.
[28] K. Levenberg, A method for the solution of certain non-linear problems in least squares, Q. Appl. Math. 2 (2) (1944) 164–168.
[29] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, J. Soc. Ind. Appl.Math. 11 (2) (1963) 431–441.
[30] C. Kanzow, N. Yamashita, M. Fukushima, Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints, J. Comput. Appl. Math. 172 (2) (2004) 375–397.
[31] M. Osborne, Nonlinear least squares - the Levenberg algorithm revisited, ANZIAM J. 19 (3) (1976) 343–357.
[32] G. Kim, M.J. Silvapulle, P. Silvapulle, Comparison of semiparametric and parametric methods for estimating copulas, Comput. Stat. Data Anal. 51 (6) (2007) 2836–2850.
[33] A.K. Atiampo, G.L. Loum, Unsupervised image segmentation with pairwise Markov chains based on nonparametric estimation of copula using orthogonal polynomials, Int. J. Image Graphics 16 (4) (2016) 1–15, doi:10.1142/S0219467816500200.
[34] D. Huard, G. Évin, A.-C. Favre, Bayesian copula selection, Comput. Stat. Data Anal. 51 (2) (2006) 809–822.
[35] G. Kauermann, C. Schellhase, D. Ruppert, Flexible copula density estimation with penalized hierarchical B-splines, Scand. J. Stat. 40 (4) (2013) 685–705.
[36] F. Zheng, S. Derrode, W. Pieczynski, Parameter estimation in switching Markov systems and unsupervised smoothing, IEEE Trans. Autom. Control 64 (4) (2018) 1761–1767.
[37] A.C. Harvey, N. Shephard, Estimation of an asymmetric stochastic volatility model for asset returns, J. Bus. Econ. Stat. 14 (4) (1996) 429–434.
[38] F. Li, Identifying asymmetric comovements of international stock market returns, J. Financ. Econom. 12 (3) (2014) 507–543.
[39] A.J. Patton, Modelling asymmetric exchange rate dependence, Int. Econ. Rev. 47 (2) (2006) 527–556.
[40] G.L.C. Yap, Optimal filter approximations for latent long memory stochastic volatility, Comput. Econ. (2019) 1–22.
[41] J.J. Dabrowski, J.-P. de Villier, C. Beyers, Naive Bayes switching linear dynamical system: amodel for dynamic system modelling, Inf. Fusion 42 (2018) 75–100.
[42] F. Flitti, C. Collet, E. Slezak, Image fusion based on pyramidal multiband multiresolution Markovian analysis, Signal Image Video Process. 3 (3) (2009) 275–289.
[43] V.A. Krylov, G. Moser, S.B. Serpico, J. Zerubia, Supervised high-resolution dual-polarization SAR image classification by finite mixtures and copulas, IEEE J. Sel. Top. Signal Process. 5 (3) (2011) 554–566.
[44] A. Voisin, V.A. Krylov, G. Moser, S.B. Serpico, J. Zerubia, Supervised classification of multisensor and multiresolution remote sensing images with a hierarchical copula-based approach, IEEE Trans. Geosci. Remote Sens. 52 (6) (2013) 3346–3358.
[45] Z. Bouyahia, S. Derrode, W. Pieczynski, Filtering in Gaussian linear systems with fuzzy switches, IEEE Transactions on Fuzzy Systems, 2019 (2019), doi:10.1109/TFUZZ.2019.2921944. In press
[46] M. Zhang, C. Shen, Z.-G. Wu, D. Zhang, Dissipative filtering for switched fuzzy systems with missing measurements, IEEE Transactions on Cybernetics, 2019 (2019), doi:10.1109/TCYB.2019.2908430. In press