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

Fei Zheng(1), Stephane Derrode(1), and Wojciech Pieczynski(2)

(1)Ecole Centrale de Lyon, Université de Lyon, LIRIS, CNRS UMR 5205, Ecully, France.
(2)Telecom SudParis, SAMOVAR, Institut Polytechnique de Paris, CNRS UMR 5157, Evry, France.

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.

Key words: Markov switching models, Non-Gaussian non-linear system, Copulas, Model identification, CMSHLM, GICE-GLS, Semi-supervised filtering, Semi-supervised smoothing.

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: $X_1^N = (X_1, ..., X_N)$, $R_1^N = (R_1, ..., R_N)$, and $Y_1^N = (Y_1, ..., Y_N)$. For $n = 1, \ldots, N$, X_n takes its values in R^s, R_n takes its values in $\Omega = \{1, ..., K\}$, and Y_n takes its values in R^q. For $n = 1, \ldots, N$, let $T_n = (X_n, R_n, Y_n)$ and let us consider $T_1^N = (T_1, ..., T_N)$. For some occasions, T_1^N will be also denoted as $T_1^N = (X_1^N, R_1^N, Y_1^N)$. The final restoration problem dealt with is to estimate both the hidden $(X_1^N, R_1^N) = (x_1^N, r_1^N)$ from observed $Y_1^N = y_1^N$.
To be concise, we will note different probability distributions with the same letter p. So the distribution of X_1^N will be denoted with $p(x_1^N)$, the distribution of R_n conditional on $Y_n = y_n$ will be denoted with $p(r_n | y_n)$, and so on. For discrete variable, like R_1, $p(r_1)$ is a probability, for continuous one, like Y_n, $p(y_n)$ is a probability density function (pdf), and for mixed case, like $T_1 = (X_1, R_1, Y_1)$, we have $p(x_1, r_1, y_1) = p(r_1) p(x_1, y_1 | r_1)$, with $p(r_1)$ probability and $p(x_1, y_1 | r_1)$ pdf.

Let us consider "Conditionally Markov switching hidden linear model" (CMSHLM [41]) defined as:

$$T_1^N = (T_1, ..., T_N) \text{ is Markov;} \tag{1.1}$$

$$p(r_{n+1}|x_n, r_n, y_n) = p(r_{n+1}|r_n); \tag{1.2}$$

$$p(r_{n+1}, y_{n+1}|x_n, r_n, y_n) = p(r_{n+1}, y_{n+1}|r_n, y_n); \tag{1.3}$$

$$X_{n+1} = A_{n+1}(R_n^{n+1}, Y_n^{n+1})X_n + B_{n+1}(R_n^{n+1}, Y_n^{n+1}) + C_{n+1}(R_n^{n+1}, Y_n^{n+1})W_{n+1}, \tag{1.4}$$

where A_{n+1}, B_{n+1}, C_{n+1} are some functions and W_2, …, W_N is a sequence of centred variables with unity variance and such that W_{n+1} is independent from $(T_1, ..., T_n)$ for each $n = 1$, …, $N-1$. Let us note that (1.2) implies the Markovianity of R_1^N, while (1.3) implies the Markovianity of (R_1^N, Y_1^N).

We propose the following contributions:

(i) CMSHLM in which $p(r_1^N, y_1^N)$ is copulas based one [7] is original;

(ii) $p(r_1^N, y_1^N)$ is identified from $Y_1^N = y_1^N$ through an original variant of the "generalized iterative conditional estimation" (GICE [13]);

(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 $(X_1^N, Y_1^N) = (x_1^N, y_1^N)$ are known while $R_1^N = r_1^N$ are not) recursive exact filtering and smoothing.

Let us remark that CMSHLM with (X_1^N, Y_1^N) Gaussian conditionally on R_1^N leads to "Conditionally Gaussian observed Markov switching models" (CGOMSMs [1, 2, 11, 20, 21, 40]), 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 [7, 14, 28], 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 [8, 14-17, 29, 31], among others. Approximating such stationary non-Gaussian non-linear systems (NSNGSs) with general CMSHLM proposed in the paper – as carried out using CGOMSMs in [21] – opens rich perspectives of dealing with stationary NSNGSs 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, [29, 35, 36, 38, 42] 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 [19], those presented in this paper are new.

The rest 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.

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 named GICE-GLS. Some experiments are provided in Section 4 and the last Section 5 concludes the work and gives the perspectives.

2. Filtering and smoothing in copulas based CNSHLMs

2.1. Copulas based CMSHLM

Let $Y_1^d = (Y_1, ..., Y_d)$ be a random vector valued in R^d, $F(y_1, ..., y_d) = P[Y_1 \le y_1, ..., Y_d \le y_d]$ its cumulative density function (CDF), and F_1, …, F_d CDFs of Y_1, …, Y_d, respectively. Furthermore, a copula C is a CDF defined on $[0,1]^d$ such that marginal CDFs $C_1(y_1)$, …, $C_d(y_d)$ are identities on $[0,1]$. According to Sklar's theorem, for given F there exists a unique copula C such that:

$$F(y_1, ..., y_d) = C(F_1(y_1), ..., F_d(y_d)) . \tag{2.1}$$

Assuming differentiable F and C, setting

$$c(y_1, ..., y_d) = \frac{\partial^d}{\partial y_1 ... \partial y_d} C(y_1, ..., y_d) \tag{2.2}$$

and taking derivative of (2.1), we obtain the probability density function (PDF) of $Y_1^d = (Y_1, ..., Y_d)$:

$$f(y_1, ..., y_d) = \prod_{i=1}^{d} f_i(y_i) c[F_1(y_1), ..., F_d(y_d)] , \tag{2.3}$$

with f_1, …, f_d PDFs of Y_1, …, Y_d respectively.
Let us return to CMSHLM defined by (1.1)-(1.4). In addition, we will consider the following commonly used assumptions:

$$p(y_{n+1} | r_n^{n+1}) = p(y_{n+1} | r_{n+1}) ; \tag{2.4}$$

$$p(y_n | r_n^{n+1}) = p(y_n | r_n) ; \tag{2.5}$$

Applying (2.3) to $p(y_n, y_{n+1} | r_n^{n+1})$ and using (2.4), (2.5), there exists a copula $c_{n+1}(r_n^{n+1})$ such:

$$p(y_n, y_{n+1} | r_n^{n+1}) = p(y_n | r_n) \, p(y_{n+1} | r_{n+1}) c_{n+1}(r_n^{n+1})(F_n(y_n | r_n), (F_{n+1}(y_{n+1} | r_{n+1})) , \tag{2.6}$$

and thus

$$p(y_{n+1} \mid r_n^{n+1}, y_n) = p(y_{n+1} \mid r_{n+1}) c_{n+1}(r_n^{n+1})(F_n(y_n \mid r_n), F_{n+1}(y_{n+1} \mid r_{n+1})). \tag{2.7}$$

Markovianity of R_1^N and (R_1^N, Y_1^N) joined to (2.7) indicate that the distribution of (R_1^N, Y_1^N) is given by Markov distribution of R_1^N, margins $p(y_1 \mid r_1)$, …, $p(y_N \mid r_N)$, and copulas $c_2(r_1^2)$, …, $c_N(r_{N-1}^N)$. 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

The filtering problem consists of recursively computing $p(r_{n+1} \mid y_1^{n+1})$, $E[X_{n+1} \mid r_{n+1}, y_1^{n+1}]$, and $E[X_{n+1} X_{n+1}^T \mid r_{n+1}, y_1^{n+1}]$ from $p(r_n \mid y_1^n)$, $E[X_n \mid r_n, y_1^n]$, $E[X_n X_n^T \mid r_n, y_1^n]$, $p(r_{n+1}, y_{n+1} \mid r_n, y_n)$, and y_{n+1}. We have

$$p(r_{n+1} \mid y_1^{n+1}) = \frac{\sum_{r_n} p(r_{n+1}, y_{n+1} \mid r_n, y_n) p(r_n \mid y_1^n)}{\sum_{r_{n+1}} \sum_{r_n} p(r_{n+1}, y_{n+1} \mid r_n, y_n) p(r_n \mid y_1^n)}; \tag{2.8}$$

$$E[X_{n+1} \mid r_{n+1}, y_1^{n+1}] = \sum_{r_n} [A_{n+1}(r_n^{n+1}, y_n^{n+1}) E[X_n \mid r_n, y_1^n] + B_{n+1}(r_n^{n+1}, y_n^{n+1})] p(r_n \mid r_{n+1}, y_1^{n+1}) \tag{2.9}$$

$$E[X_{n+1} X_{n+1}^T \mid r_{n+1}, y_1^{n+1}] = \sum_{r_n} F_{n+1}(r_n^{n+1}, y_n^{n+1}) p(r_n \mid r_{n+1}, y_1^{n+1}), \tag{2.10}$$

with

$$\begin{aligned} F_{n+1}(r_n^{n+1}, y_n^{n+1}) &= A_{n+1}(r_n^{n+1}, y_n^{n+1}) E[X_n \mid r_n, y_1^n] E^T[X_n \mid r_n, y_1^n] A_{n+1}^T(r_n^{n+1}, y_n^{n+1}) + \\ & B_{n+1}(r_n^{n+1}, y_n^{n+1}) B_{n+1}^T(r_n^{n+1}, y_n^{n+1}) + C_{n+1}(r_n^{n+1}, y_n^{n+1}) C_{n+1}^T(r_n^{n+1}, y_n^{n+1}) + \\ & A_{n+1}(r_n^{n+1}, y_n^{n+1}) E[X_n \mid r_n, y_1^n] B_{n+1}^T(r_n^{n+1}, y_n^{n+1}) + B_{n+1}(r_n^{n+1}, y_n^{n+1}) E^T[X_n \mid r_n, y_1^n] A_{n+1}^T(r_n^{n+1}, y_n^{n+1}), \end{aligned} \tag{2.11}$$

$$p(r_n \mid r_{n+1}, y_1^{n+1}) = \frac{p(r_{n+1}, y_{n+1} \mid r_n, y_n) p(r_n \mid y_1^n)}{\sum_{r_n} p(r_{n+1}, y_{n+1} \mid r_n, y_n) p(r_n \mid y_1^n)}. \tag{2.12}$$

Let us briefly justify (2.8)-(2.12). (2.8) and (2.12) come from the Markovianity of (R_1^N, Y_1^N), which implies

$$p(r_n^{n+1}, y_1^{n+1}) = p(r_{n+1}, y_{n+1} \mid r_n, y_n) p(r_n, y_1^n) \tag{2.13}$$

To justify (2.9), we write $E[X_{n+1}|r_{n+1}, y_1^{n+1}] = \sum_{r_n} E[X_{n+1}|r_n, r_{n+1}, y_1^{n+1}] p(r_n|r_{n+1}, y_1^{n+1}) = \sum_{r_n} \{[A_{n+1}(r_n^{n+1}, y_n^{n+1}) E[X_n|r_n, r_{n+1}, y_1^{n+1}] + B_{n+1}(r_n^{n+1}, y_n^{n+1}] p(r_n|r_{n+1}, y_1^{n+1})\}$, and then we apply $E[X_n|r_n, r_{n+1}, y_1^{n+1}] = E[X_n|r_n, y_1^n]$ - which comes from (1.1), according to which X_n and (R_{n+1}, Y_{n+1}) are independent conditionally on (R_n, Y_n). (2.10)-(2.11) are obtained in similar way replacing X_{n+1} by $X_{n+1}X_{n+1}^T$.

Remark 2.1 As $p(r_{n+1}| y_1^{n+1})$ and $E[X_{n+1}|r_{n+1}, y_1^{n+1}]$ are computed from $p(r_n| y_1^n)$, $E[X_n|r_n, y_1^n]$, and y_{n+1} without using $C_{n+1}(R_n^{n+1}, Y_n^{n+1})$. So (1.4) can be actually extended to

$$X_{n+1} = A_{n+1}(R_n^{n+1}, Y_n^{n+1}) X_n + B_{n+1}(R_n^{n+1}, Y_n^{n+1}) + W_{n+1}, \tag{2.14}$$

with the only hypotheses that $E[W_{n+1}] = 0$ and W_{n+1} is independent from $(T_1, ..., T_n)$ for each $n = 1$, …, $N-1$. However, the filter's variance depends on $Var[W_{n+1}]$, and would thus be unknown.

2.4. Smoothing in copulas based CMSHLM

Optimal smoothing consists of computation of $E[X_n|r_n, y_1^N]$ for each $n = 1$, …, N. Under CMSHLM they are not complicated to get from already calculated $E[X_n|r_n, y_1^n]$ in filtering given in the previous paragraph. We have:

$$E[X_n| y_1^N] = \sum_{r_n} p(r_n|y_1^N) E[X_n|r_n, y_1^N] = \sum_{r_n} p(r_n|y_1^N) E[X_n|r_n, y_1^n], \tag{2.15}$$

the second equality being due to the fact that X_n and Y_{n+1}^N are independent conditionally on (R_n, Y_n). Both R_1^N and (R_1^N, Y_1^N) being Markov, $p(r_n|y_1^N)$ is classically obtained by recursive calculation of "forward" and "backward" probabilities $\alpha_n(r_n) = p(r_n, y_1^n)$, $\beta_n(r_n) = p(y_{n+1}^N|r_n)$ with:

$$\alpha_1(r_1) = p(r_1, y_1) \; ; \; \alpha_{n+1}(r_{n+1}) = \sum_{r_n} p(r_{n+1}, y_{n+1}|r_n, y_n) \alpha_n(r_n) \;, \tag{2.16}$$

$$\beta_N(r_N) = 1 \; ; \; \beta_n(r_n) = \sum_{r_{n+1}} p(r_{n+1}, y_{n+1}|r_n, y_n) \beta_{n+1}(r_{n+1}) \;. \tag{2.17}$$

We have $p(r_n, y_1^N) = \alpha_n(r_n)\beta_n(r_n)$, and thus:

$$p(r_n|y_1^N) = \frac{\alpha_n(r_n)\beta_n(r_n)}{\sum_{r_n} \alpha_n(r_n)\beta_n(r_n)} \;, \tag{2.18}$$

$E[X_n| y_1^N]$ in smoothing does not require $C_{n+1}(R_n^{n+1}, Y_n^{n+1})$ as in filtering, while $E[X_n X_n^T| y_1^N]$ can be calculated in a similar way and gives:

$$E[X_n X_n^T \mid y_1^N] = \sum_{r_n} p(r_n \mid y_1^N) E[X_n X_n^T \mid r_n, y_1^n], \tag{2.19}$$

with $E[X_n X_n^T \mid r_n, y_1^n]$ from (2.10)-(2.11).

3. CB-CMSHLM identification

We tackle the identification problem of a CB-CMSHLM from a learning sample set of observations $Y_1^N = y_1^N$ data and space data $X_1^N = x_1^N$, while $R_1^N = r_1^N$ remains unknown. In CB-CMSHLM $T_1^N = (X_1^N, R_1^N, Y_1^N)$, the considered couple (R_1^N, Y_1^N) is stationary, so that its distribution can be defined by $p(r_1, r_2, y_1, y_2) = p(r_1, r_2) p(y_1, y_2 \mid r_1, r_2)$, equal to the distributions $p(r_n, r_{n+1}, y_n, y_{n+1})$, $n = 2$, …, $N-1$. Furthermore, A_{n+1} and B_{n+1} in (1.4) are time independent from $n = 2$, …, $N-1$. 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 $p(y_1, y_2 \mid r_1, r_2)$;

(iii) Find forms and parameters of $A(r_1^2, y_1^2)$ and $B(r_1^2, y_1^2)$ defining $p(x_1, x_2 \mid r_1, r_2, y_1, y_2)$.

For each $(r_1, r_2) \in \Omega^2$, forms of copulas $c(r_1, r_2)$, forms of margins $p(y_1 \mid r_1)$, and forms of $A(r_1^2, y_1^2)$, $B(r_1^2, y_1^2)$, 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 [13], GICE can be applied here to identify the distribution of (R_1^N, Y_1^N) from Y_1^N only: X_1^N can be temporarily set aside here. The new GICE variant we propose is as follows.

Let $Y_1^N = y_1^N$ be a sample, and for simplifying the notations, let us denote $f_{jk}(y_1, y_2) = p(y_1, y_2 \mid r_1 = j, r_2 = k)$, $f_j(y_1) = p(y_1 \mid r_1 = j)$, $f_k(y_2) = p(y_2 \mid r_2 = k)$, and $c_{jk}(F_j(y_1), F_k(y_2)) = c(r_1 = j, r_2 = k)(F(y_1 \mid r_1 = j), F(y_2 \mid r_2 = k))$. So that:

$$f_{jk}(y_1, y_2) = f_j(y_1) f_k(y_2) c_{jk}(F_j(y_1), F_k(y_2)). \tag{3.1}$$

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

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

Thus to identify margins means to find (from $Y_1^N = y_1^N$) for each $j \in \Omega$, the right form H_l^j in 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 G, and to estimate parameters $\alpha^{jk}(m)$.

To achieve these goals by GICE, we further assume:
(1) For each $j,k \in \Omega$, $l = 1,...,L$, and $m = 1,...,M$ 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 $f_{\theta(1)} \in H_1$, …, $f_{\theta(L)} \in H_L$ the best one which fits the given sample $y^1 = (y_1^1,..., y_{Q_1}^1)$, with Q_1 denoting the sample size;
(3) There exists a rule D^2 which decides for each set of copulas $c_{\alpha(1)} \in G_1$, …, $c_{\alpha(M)} \in G_M$ the best one which fits the given sample $y^2 = (y_1^2,..., y_{Q_2}^2)$, 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 $(p_{jk}^0, f_j^0, c_{jk}^0)$, for $j,k \in \Omega$;
2. Find $(p_{jk}^{i+1}, f_{jk}^{i+1}, c_{jk}^{i+1})$ from $(p_{jk}^i, f_{jk}^i, c_{jk}^i)$ and y_1^N by the sub-steps below:

(a) for $n = 1$, …, $N-1$, compute $p^i(r_n = j, r_{n+1} = k \,|\, y_1^N)$ from $(p_{jk}^i, f_{jk}^i, c_{jk}^i)$ and y_1^N with $p^i(r_n = j, r_{n+1} = k \,|\, y_1^N) = \dfrac{\alpha_n(j) p(r_{n+1}, y_{n+1} | r_n, y_n) \beta_{n+1}(k)}{\sum_{j,k \in \Omega} \alpha_n(j) p(r_{n+1}, y_{n+1} | r_n, y_n) \beta_{n+1}(k)}$, where $\alpha_n(j)$ and $\beta_{n+1}(k)$ are obtained by applying (2.17), (2.18). Then update p_{jk} with $p_{jk}^{i+1} = \dfrac{1}{N-1} \sum_{n=1}^{N-1} p^i(r_n = j, r_{n+1} = k \,|\, y_1^N)$;

(b) sample $(r_1^N)^{i+1} = (r_1^{i+1}, ..., r_N^{i+1})$ according to $p(r_1^N | y_1^N)$ based on current parameters $(p_{jk}^i, f_{jk}^i, c_{jk}^i)$ (recall that $p(r_1^N | y_1^N)$ is Markov with $p(r_1 = j | y_1^N) = \dfrac{\alpha_1(j)\beta_1(j)}{\sum_{k \in \Omega} \alpha_1(k)\beta_1(k)}$, and $p(r_{n+1} = j | r_n = k, y_1^N) = \dfrac{p(r_{n+1} = j, y_{n+1} | r_n = k, y_n) \beta_{n+1}(j)}{\beta_n(k)}$) for $n = 1$, …, $N-1$);

(c) for each $j,k \in \Omega$ consider $(y_1^N)_j^{i+1}$ the subsequence of y_1^N formed with y_n such that $r_n^{i+1} = j$, and $(y_1^N)_{jk}^{i+1}$ the subsequence of couples (y_n, y_{n+1}) in y_1^N such that $r_n^{i+1} = j$ and $r_{n+1}^{i+1} = k$. For $l = 1$, …, L and $m = 1$, …, M, calculate $\theta_j^{i+1}(l) = \hat{\theta}_j(l)[(y_1^N)_j^{i+1}]$ and $\alpha_{jk}^{i+1}(m) = \hat{\alpha}_{jk}(m)[(y_1^N)_{jk}^{i+1}]$;

(d) for each $j \in \Omega$, choose from $\{f_{\theta_j^{i+1}(1)},..., f_{\theta_j^{i+1}(L)}\}$ an element f_j^{i+1} by applying rule D^1 to the sample $y^1 = (y_1^N)_j^{i+1}$. Similarly, for $j,k \in \Omega$ chose from $\{c_{\alpha_{jk}^{i+1}(1)},..., c_{\alpha_{jk}^{i+1}(M)}\}$ an element c_{jk}^{i+1} by applying rule D^2 to the sample $y^2 = (y_1^N)_{jk}^{i+1}$;

3. Stop according to some criterion.

For the initialization in step 1, K-means is applied to group y_1^N and find the initial guess of switches $(r_1^N)^0$, then $(p_{jk}^0, f_j^0, c_{jk}^0)$ can be initialized from $(r_1^N, y_1^N)^0$ so as for the sub-step (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 [13] $\hat{\alpha}_{jk}(m)$ 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 [13] 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 [13] 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, no change of form is observed for the estimation of both margins and copulas, and the difference of the objective function value of each decision (both 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 $A(r_n^{n+1}, y_n^{n+1})$, $B(r_n^{n+1}, y_n^{n+1})$, and $C(r_n^{n+1}, y_n^{n+1})$ defining X_{n+1} from X_n with (1.4), and being independent from $n = 1$, …, N. We have seen that $C(r_n^{n+1}, y_n^{n+1})$ intervenes neither in filtering nor in smoothing, thus we concentrate on dealing with $A(r_1^2, y_1^2)$ and $B(r_1^2, y_1^2)$. Let us temporarily assume that their forms are given and for each $r_1^2 = (j,k)$, and they depend on parameters a_{jk} and b_{jk} respectively: $A(r_1 = j, r_2 = k, y_1^2) = A_{a_{jk}}(y_1^2)$, $B(r_1 = j, r_2 = k, y_1^2) = B_{b_{jk}}(y_1^2)$. When $p(r_1^N | y_1^N)$ is given, the parameter estimation of the Gaussian $p(x_{n+1} | x_n, y_n^{n+1}, r_n^{n+1})$ 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 $\hat{a} = (\hat{a}_{jk})_{j,k\in\Omega}$ and $\hat{b} = (\hat{b}_{jk})_{j,k\in\Omega}$ are obtained by minimizing with respect to $(a_{jk})_{j,k\in\Omega}$, $(b_{jk})_{j,k\in\Omega}$ the quantity

$$e^2 = \frac{1}{N-1}\sum_{n=1}^{N-1}\{x_{n+1} - \sum_{(j,k)} p(r_n^{n+1} = (j,k) | y_1^N)[A_{a_{jk}}(y_n^{n+1})x_n + B_{b_{jk}}(y_n^{n+1})]\}^2 \text{ , (3.2)}$$

As previously done for copulas and margins, let us assume that the form of $A(r_1^2, y_1^2)$ is not known but belongs to a given set of forms $\{K_1, ..., K_Q\}$, with each form K_q being parameterized by $a^q = (a_{jk}^q)_{j,k\in\Omega}$. Similarly, the form of $B(r_1^2, y_1^2)$ is not known but belongs to a given set of forms $\{L_1, ..., L_S\}$, with each form L_s being parameterized by $b^s = (b_{jk}^s)_{j,k\in\Omega}$. Then, minimization of (3.2) is applied to each couple of forms (K_q, L_s), giving estimated $\hat{a}^q = (\hat{a}_{jk}^q)_{j,k\in\Omega}$ and $\hat{b}^s = (\hat{b}_{jk}^s)_{j,k\in\Omega}$. Then the couple of forms finally kept is the couple $(\hat{K}_q, \hat{L}_s)$ for which the related $(\hat{a}^q, \hat{b}^s)$ obtains the minimum (3.2) (comparing to other $(\hat{a}^{q^*}, \hat{b}^{s^*})$ related to other couples (K_{q^*}, L_{s^*})).

Example 3.1

Let us consider the linear case $A_{jk}(y_n^{n+1}) = a_{jk}g_1(y_n^{n+1})$, $B_{jk}(y_n^{n+1}) = b_{jk}g_2(y_n^{n+1})$, with g_1, g_2 given functions. The explicit solution (the vector stacking all a_{jk} and b_{jk}) of the minimization of (3.2) is:

$$\hat{\beta}(x) = (L^T L)^{-1} L^T x, \tag{3.3}$$

with $x = (x_2, ..., x_N)^T$, and L matrix given with

$$L = \begin{bmatrix} p_{1,1}^1 g_{1,2}^1 & \cdots & p_{1,K}^1 g_{1,2}^1 & \cdots & p_{1,K}^1 g_{1,2}^1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{1,1}^{N-1} g_{1,2}^1 & \cdots & p_{1,K}^{N-1} g_{1,2}^{N-1} & \cdots & p_{1,K}^{N-1} g_{1,2}^{N-1} \end{bmatrix}, \tag{3.4}$$

where $p_{j,k}^n = p(r_n^{n+1} = (j,k) | y_1^N)$, and $g_{1,2}^n = [g_1(y_n^{n+1}) x_n \quad g_2(y_n^{n+1})]$.

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 [4, 5, 31] 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 [24, 33, 37].

Combining the two identification steps above, the entire Schema of GICE-GLS for CB-CMSHLM identification is given in Figure 3.1.

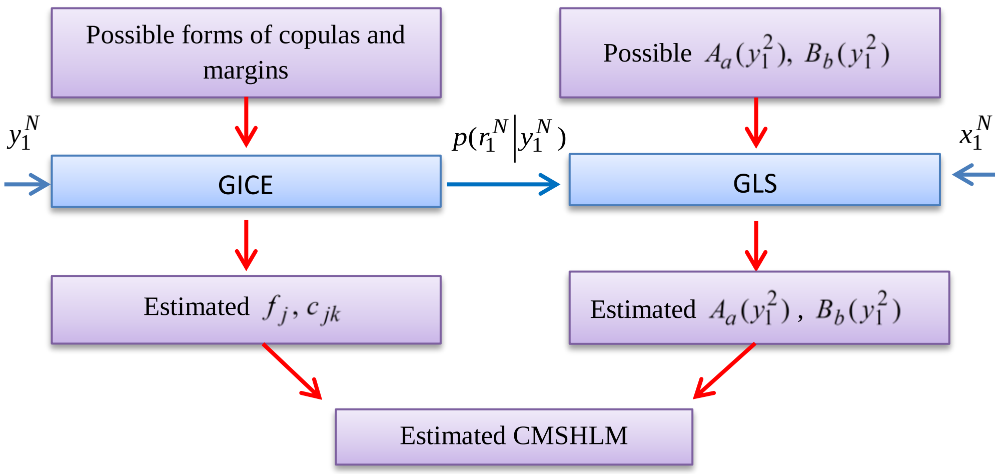

Figure 3.1. Schema of CMSHLM estimation from learning sample (x_1^N, y_1^N) Through GICE-GLS.

Concerning complexity time of ICE-GLS, we note that is linear in the sample size N. Besides, it is proportional to the number of possible margins forms L, and it is proportional to the number of possible copulas forms M. In the general GICE-GLS it would also be proportional to the number of possible forms of $A_a(y_1^2)$ and $B_b(y_1^2)$. Finally, it also is proportional to the number of iterations of GICE-GLS. However, important is that similarly to the classic ICE in simple hidden Gaussian Markov chains, complexity time of ICE-GLS is linear in sample size N.

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 $A_{a_{jk}}(y_1^2)$ and $B_{b_{jk}}(y_1^2)$ 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 other two. The first one with parameters estimated by ICE-LS and data restored by exact restoration considering Gaussian margins and copulas, and the second one, identification and restoration through CGOMSM-ABF proposed in [20, 21]. 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 follow the simpler Gaussian CGOMSM, which is a particular case of CB-CMSHLM, GICE-LS based filtering and smoothing provide a result comparable to those obtained with ICE-LS and CGOMSM-ABF.

In the second series, data are sampled with respect to a CGOMSM. The aim is to verify that when data follow the simpler Gaussian CGOMSM, which is a particular case of CB-CMSHLM, GICE-LS based filtering and smoothing provide 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 R_1^N is stationary and has $K = 2$ jumps;
- The margins are of six possible forms (see Appendix for details):

$$H = \{H_1,...,H_6\} = \{\text{Gamma, Fisk, Gaussian, Laplace, Beta, Beta prime}\}, \tag{4.1}$$

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

$$G = \{G_1,...,G_7\} = \{\text{Gumble, Gaussian, Clayton, FGM, Arch12, Arch14, Product}\} \tag{4.2}$$

- 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 $F_1 \in H_1$, …, $F_L \in H_L$. The Kolmogorov distance between two CDFs F, F' is defined as:

$$d(F, F') = \sup_{y \in \mathrm{R}} \left| F(y) - F'(y) \right|, \tag{4.3}$$

and thus for a sample $u_1^Q = (u_1,...,u_Q)$ the chosen CDF $D^1(u_1^Q)$ among candidates $F_1 \in H_1$, …, $F_L \in H_L$, is defined with:

$$D^1(u_1^Q) = \underset{l\in\{1,\dots,L\}}{\arg\inf}\,[d(F_l,\hat{F})]\,, \tag{4.4}$$

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

$$\hat{F}(u) = \frac{1}{Q}\sum_{n=1}^{Q} 1_{[u_n \le u]}\,; \tag{4.5}$$

- Estimators $\hat{\alpha}_{jk}$ are obtained with the method presented in [26]. For a sample $u_1^{2Q} = ((u_1,u_2),\dots,(u_{2Q-1},u_{2Q}))$, we have:

$$\hat{\alpha}(u_1^{2Q}) = \underset{\alpha}{\arg\max}[\sum_{n=1}^{N-1} Log(c_\alpha(\hat{F}(u_n),\hat{F}(u_{n+1})))]\,, \tag{4.6}$$

where $\hat{F}(u_n)$ and $\hat{F}(u_{n+1})$ are empirical CDFs calculated from $(u_1,\dots,u_{2Q-1})$ and $(u_2,\dots,u_{2Q})$ respectively. Let us remark that other copula estimation methods [3, 23, 25] could replace the applied ones.

- Finally, the rule D^2 is the maximum of pseudo-likelihood: for a sample $u_1^{2Q} = ((u_1,u_2),\dots,(u_{2Q-1},u_{2Q}))$, copula $\hat{c}$ related to each distribution $p(u_{2n-1},u_{2n})$ is chosen among candidates $c_1 \in G_1$, …, $c_M \in G_M$ with:

$$D^2(u_1^{2Q}) = \underset{m\in\{1\dots,M\}}{\arg\sup}\prod_{n=1}^{N-1} c_m([\hat{F}(u_n),\hat{F}(u_{n+1})])\,, \tag{4.7}$$

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

Series 1.

In both series the probabilities $p_{jk} = p(r_1 = j, r_2 = k)$ defining the distribution of stationary R_1^N are $p_{11} = p_{22} = 0.45$, $p_{12} = p_{21} = 0.05$. A set of $N = 5000$ simulated (x_n, 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 $p(r_1^N, y_1^N)$ are set as in Table 5.2. $p(x_{n+1} | x_n, r_n = j, r_{n+1} = k, y_n^{n+1})$ are Gaussian with means $a_{jk}x_n + B_{jk}(y_n^{n+1})$ - where $B_{jk}(y_n^{n+1}) = b_{jk}y_n y_{n+1} + d_{jk}$ are non-linear in y_n, y_{n+1}, and the variances σ_{jk}^2. Let us recall that variances σ_{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 three methods are indicated in Table 4.1. 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 $p(y_1^N | r_1^N)$ is Gaussian; the difference lies in the fact that CGOMSM-ABF also assumes $p(x_n^{n+1} | r_n^{n+1}, y_n^{n+1})$ Gaussian, while ICE-LS limits the Gaussian assumption to $p(x_{n+1} | x_n, r_n^{n+1}, y_n^{n+1})$. Thus, the non-linearity of $B_{jk}(y_n^{n+1}) = b_{jk} y_n y_{n+1} + d_{jk}$ is better taken into account by ICE-LS;

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

4. The estimates from GICE-LS are quite close to the true ones as listed in Table 4.2. The average of estimated joint probabilities $p_{jk} = p(r_1 = j, r_2 = k)$ from GICE are $\hat{p}_{11} = 0.474$, $\hat{p}_{22} = 0.445$, $p_{12} = p_{21} = 0.040$;

5. According to Figure 4.1, where the error ratio of unsupervised switches estimation is concerned, GICE is much more effective than ICE;

6. A trajectory example displayed in Figure 4.2 clearly illustrates the superiority of GICE-LS over the other methods on the restoration of general CB-CMSHLM data considered;

7. According to Table 4.3 estimates of a_{jk}, b_{jk}, and d_{jk} are quite correct.

		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 4.1 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)

Margins and parameters	$f_1(\theta_1)$ (Gamma)	$f_2(\theta_2)$ (Fisk)	Copulas and parameters	$c_{11}(\alpha_{11})$ (Gaussian)	$c_{22}(\alpha_{22})$ (Clayton)	$c_{12}(\alpha_{12}) = c_{21}(\alpha_{21})$ (Gaussian)
True θ_i	16.00	4.00	True α_{ij}	0.45	4.67	0.45
Estimated θ_i	13.72	3.93	Estimated α_{ij}	0.46	4.46	0.46

Table 4.2. True margins, copulas, and their estimates (extracted from cases in which true copulas and margins are perfectly found).

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

Table 4.3. True a_{jk}, b_{jk}, d_{jk} and their estimates (extracted from cases in which true copulas and margins are perfectly found).

	Gamma	Fisk	Gaussian	Laplace	Beta	Beta Prime
Identified f_1	87%	12%	-	1%	-	-
Identified f_2	1%	99%	-	-	-	-

Table 4.4. Margins identification error ratio. f_1 is Gamma, and f_2 is Fisk, – see Table 4.2.

	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 4.5. Copulas identification error ratio. c_{11} is Gumbel, $c_{12} = c_{21}$ are Gaussian, and c_{22} is Clayton – see Table 4.2.

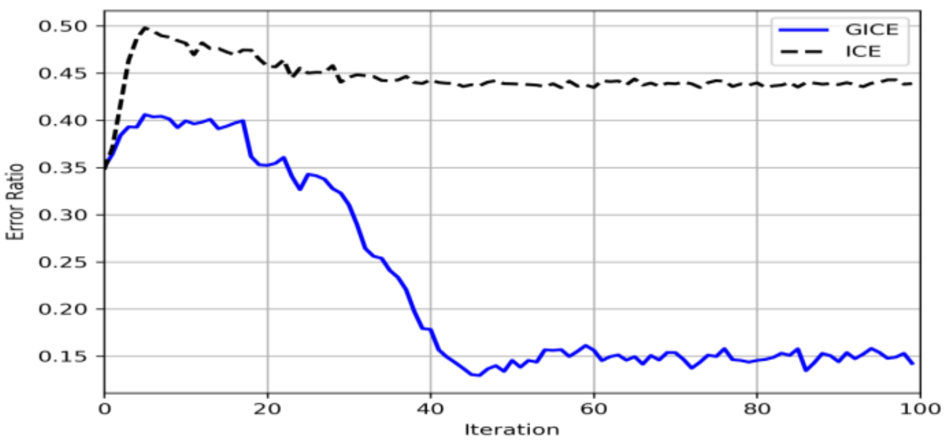

Figure 4.1. Error ratio tendency of estimated R_1^N according to GICE and ICE iterations in Series 1.

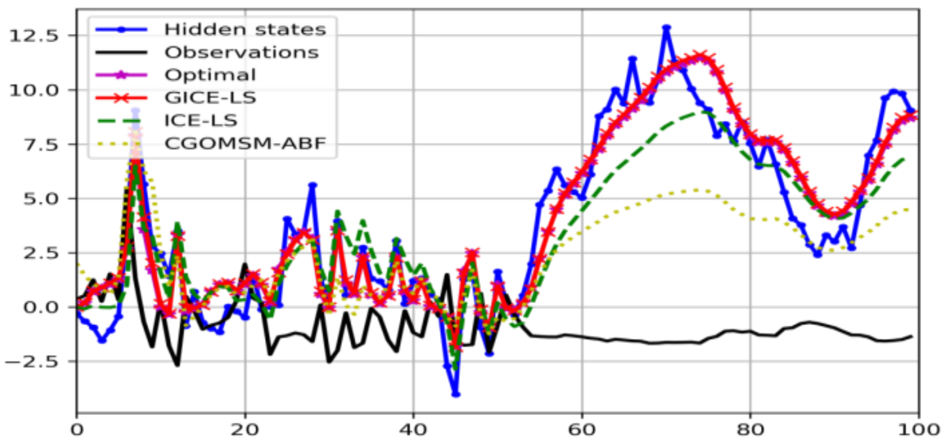

Figure 4.2. Trajectory example from Series 1 experiment (100 samples, smoothing).

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 $p(y_n^{n+1} | r_n^{n+1})$ and $p(x_{n+1} | x_n, r_n^{n+1} = (j,k), y_n^{n+1})$ are set to be Gaussian with $A_{jk}(y_n^{n+1}) = a_{jk}$ and $B_{jk}(y_n^{n+1}) = b_{jk} y_n + c_{jk} y_{n+1} + d_{jk}$, with a_{jk}, b_{jk}, c_{jk} and their estimates specified in Table 4.6. Estimated switching joint probabilities from GICE are $p_{11} = 0.485$, $p_{22} = 0.421$, $p_{21} = p_{12} = 0.047$; while from ICE, they are $p_{11} = 0.489$, $p_{22} = 0.419$, $p_{21} = p_{12} = 0.046$.

According to Table 4.8 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 (Table 4.9) and Gaussian copulas (Table 4.10). 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.

	a_{11}	a_{12}	a_{21}	a_{22}	b_{11}	b_{12}	b_{21}	b_{22}	c_{11}	c_{12}	c_{21}	c_{22}
True	0.30	0.50	0.50	0.70	0.61	0.05	0.25	-0.19	0.30	0.70	0.70	0.30
ICE-LS	0.30	0.52	0.48	0.69	0.60	0.03	0.25	-0.16	0.31	0.71	0.31	0.71
GICE-LS	0.34	0.56	0.47	0.67	0.50	0.05	0.20	-0.11	0.39	0.78	0.27	0.64

Table 4.6. True parameters and their estimates with ICE and GICE (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 θ_i	0.00	1.00	True α_{ij}	0.80	0.45	0.20
ICE	Estimated θ_i	0.01	1.00	Estimated α_{ij}	0.79	0.42	0.20
GICE	Estimated θ_i	-0.04	0.99	Estimated α_{ij}	0.78	0.49	0.20

Table 4.7. True margins, copulas (Gaussian) and their estimates (extracted from cases in which true copulas and margins are found).

MSE of observations 27.123		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

Table 4.8. 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 Table 5.4.

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

Table 4.9. Margins identification error ratio. f_1 and f_2 are Gaussian.

	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%	-

Table 4.10. Copulas identification error ratio. c_{11}, $c_{12} = c_{21}$, and c_{22} are Gaussian.

5. Conclusions and perspectives

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 $C_{n+1}(R_n^{n+1}, Y_n^{n+1})$ in (1.4), when dealing with the parameter estimations of $A_{n+1}(R_n^{n+1}, Y_n^{n+1})$ and $B_{n+1}(R_n^{n+1}, Y_n^{n+1})$ 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, while parametric or non-parametric methods [3, 25] are undoubtedly 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 Y_1^N and states X_1^N is required, while switches R_1^N are unknown. Extending the method to a fully unsupervised one, which would work from the 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 [47];

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 (1.4). Choosing the best model for a given concrete problem opens a huge field of perspectives. In particular, extending studies of stochastic volatility proposed in [21, 45] in the context of Gaussian models is an interesting perspective;

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 [8], 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 [19], or hidden Markov fields [27, 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 [5, 46], which results in as many possibilities of extensions of the proposed models.

Appendix

1. 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 $erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt$, $F(y)=\frac{1}{2}(1+erf(\frac{y}{\sqrt{2}}))$, and $f(y)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2})$;

- *Laplace*: $F(y)=\begin{cases}1-\frac{1}{2}\exp(-y) & if \quad y\geq 0\\ \frac{1}{2}\exp(y) & if \quad y<0\end{cases}$, $f(y)=\frac{1}{2}\exp(-|y|)$;

- *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): $F(y)=I(\frac{y}{1+y},\theta_1,\theta_2)$, and $f(y)=\frac{y^{\theta_1-1}(1+y)^{-\theta_1-\theta_2}}{B(\theta_1,\theta_2)}$ (for $\theta_1>0$, $\theta_2>0$).

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

- *Gumbel copula:* Setting $U_1=(-\ln(u_1))^{\alpha}$, $U_2=(-\ln(u_2))^{\alpha}$, CDF C and PDF c are (for $\alpha\in[1,+\infty[$),
$C(u_1,u_2)=\exp(-(U_1+U_2)^{1/\alpha}$;
$c(u_1,u_2)=\frac{U_1}{u_1\ln(u_1)}\frac{U_2}{u_2\ln(u_2)}(a-1+U_1+U_2)^{1/\alpha}(U_1+U_2)^{1/\alpha-2}\exp[-(U_1+U_2)^{1/\alpha}]$,

- *Gaussian copula.* Setting ϕ standard Gaussian PDF (mean 0 and variance 1), $\xi=\begin{bmatrix}\phi^{-1}(u_1)\\ \phi^{-1}(u_2)\end{bmatrix}$, $I=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$, and $\rho=\begin{bmatrix}1 & \alpha\\ \alpha & 1\end{bmatrix}$ (for $\alpha\in[-1,1]$): $C(u_1,u_2)=\int_0^{u_1}\phi(\frac{\phi^{-1}(u_2)-\alpha\phi^{-1}(u)}{\sqrt{1-\alpha^2}})du$,
$c(u_1,u_2)=\frac{1}{1-\alpha^2}\exp(-\frac{1}{2}\xi^T(\rho-I)\xi)$;

- *Clayton copula:* $C(u_1,u_2)=(u_1^{-\alpha}+u_2^{-\alpha}-1)^{1/\alpha}$, $c(u_1,u_2)=(1+\alpha)u_1^{-1-\alpha}u_2^{-1-\alpha}(u_1^{-\alpha}+u_2^{-\alpha}-1)^{-(1/\alpha)-2}$ (for $\alpha\in[0,+\infty[$);

- *FGM (Farlie-Gumbel-Morgenstern) copula.* $C(u_1,u_2)=u_1u_2(1+\alpha(1-u_1)(1-u_2))$, $c(u_1,u_2)=1+\alpha(1-2u_1)(1-2u_2))$.

- *Arc12 (Archimedean of order 12) copula*. Setting $U_1 = (\frac{1}{u_1} - 1)^{\alpha}$, $U_2 = (\frac{1}{u_2} - 1)^{\alpha}$ (for $\alpha \in [1, +\infty[$):

$$C(u_1, u_2) = (1 + (U_1 + U_2)^{1/\alpha})^{-1},$$

$$c(u_1, u_2) = \frac{U_1}{u_1(1-u_1)} \frac{U_2}{u_2(1-u_2)} \frac{[a - 1 + (1+\alpha)(U_1 + U_2)^{1/\alpha}](U_1 + U_2)^{(1/\alpha)-2}}{[1 + (U_1 + U_2)^{1/\alpha}]^3} .$$

- *Arc14 (Archimedean of order 14) copula*. Setting $U_1 = (u_1^{-1/\alpha})^{\alpha}$, $U_2 = (u_2^{-1/\alpha})^{\alpha}$ (for $\alpha \in [1, +\infty[$):

$$C(u_1, u_2) = (1 + (U_1 + U_2)^{1/\alpha})^{-\alpha},$$

$$c(u_1, u_2) = U_1 U_2 (U_1 + U_2)^{(1/\alpha)-2} [1 + (U_1 + U_2)^{1/\alpha}]^{-2-\alpha} \frac{[\alpha - 1 + 2\alpha(U_1 + U_2)^{1/\alpha}]}{\alpha u_1 u_2 (u_1^{1/\alpha} \text{-}1)(u_2^{1/\alpha} \text{-}1)} .$$

- *Product copula.* $C(u_1, u_2) = u_1 u_2$, $c(u_1, u_2) = 1$

References

[1] N. Abbassi, D. Benboudjema, and W. Pieczynski, Kalman filtering approximations in triplet Markov Gaussian switching models, IEEE Workshop on Statistical Signal Processing (SSP2011), Nice, France, June 28-30, 2011.

[2] N. Abbassi, D. Benboudjema, S. Derrode, and W. Pieczynski, Optimal filter approximations in conditionally Gaussian pairwise Markov switching models, *IEEE Trans. on Automatic Control* 60 (4) (2015) 1104-1109.

[3] A. K. Atiampo and G. L. Loum, Unsupervised image segmentation with pairwise Markov chains based on nonparametric estimation of copula using orthogonal polynomials, *International Journal of Image and Graphics* 16 (4) (2016) 1-15.

[4] A. Bjorck, *Numerical methods for least squares problems*, SIAM, 1996.

[5] Z. Bouyahia, S. Derrode, and W. Pieczynski, Filtering in Gaussian linear systems with fuzzy switches, *IEEE Trans. on Fuzzy Systems*, DOI: 10.1109/TFUZZ.2019.2921944, to appear.

[6] K. M. Brown and J. E. Dennis, A new algorithm for non-linear least-squares curve fitting, *Mathematical Software* (2014) 391-396.

[7] N. Brunel and W. Pieczynski, Unsupervised signal restoration using hidden Markov chains with copulas, Signal Processing 85 (12) (2005) 2304-2315.

[8] O. Cappe, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*, Springer Series in Statistics, Springer-Verlag, 2005.

[9] J. J. Dabrowski, J P de Villier, and C Beyers, Naive Bayes switching linear dynamical system: A model for dynamic system modelling, classification, and information fusion, *Information Fusion*, 42 (2018) 75-100.

[10] J. P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, *IEEE Trans. on Signal Processing* 45 (10) (1997) 2613-2615.

[11] S. Derrode, W. Pieczynski, Exact fast computation of optimal filter in Gaussian switching linear systems, *IEEE Signal Processing Letters* 20 (7) (2013) 701-704.

[12] S. Derrode and W. Pieczynski, Unsupervised data classification using pairwise Markov chains with automatic copulas selection, *Computational Statistics & Data Analysis* 63 (2013) 81-98.

[13] S. Derrode and W. Pieczynski, Unsupervised classification using hidden Markov chain with unknown noise copulas and margins, *Signal Processing* 128 (2016) 8-17.

[14] A. Doucet and A. M. Johansen, *A tutorial on particle filtering and smoothing: Fifteen years later*, Handbook of nonlinear filtering 12 (2009) 656-704.

[15] A. Doucet, N. J. Gordon, and V. Krishnamurthy, Particle filters for state estimation of jump Markov linear systems, *IEEE Trans. on Signal Processing* 49 (3) (2001) 613-624.

[16] A. Doucet and C. Andrieu, Iterative algorithms for state estimation of jump Markov linear systems, *IEEE Trans. on Signal Processing* 49 (2001) 1216-1227.

[17] P. Fearnhead and P. Clifford, On-line inference for hidden Markov models via particle filters, *Journal of Royal Statistical Society, Series B*, 65 (2003) 887-899.

[18] F. Flitti, C. Collet, and E. Slezak, Image fusion based on pyramidal multiband multiresolution markovian analysis, *Signal, Image, and Video Processing*, 3 (3) (2009) 275-289.

[19] V. Genon-Catalot, T. Jeantheau, and C. Laredo, Stochastic volatility models as hidden Markov models and statistical applications, *Bernoulli* 6 (6) (2000) 1051-1079.

[20] I. Gorynin, S. Derrode, E. Monfini, and W. Pieczynski, Fast smoothing in switching approximations of non-linear and non-Gaussian models, *Computational Statistics & Data Analysis* 114 (2017) 38-46.

[21] I. Gorynin, S. Derrode, E. Monfrini, W. Pieczynski, Fast filtering in switching approximations of nonlinear Markov systems with applications to stochastic volatility, *IEEE Trans. on Automatic Control* 62(2) (2017) 853-862.

[22] A. C. Harvey and N. Shephard, Estimation of an asymmetric stochastic volatility model for asset returns, *Journal of Business & Economic Statistics* 14 (4) (1996) 429-434.

[23] D. Huard, G. Evin, and A.-C. Favre, Bayesian copula selection, *Computational Statistics & Data Analysis* (2) (2006) 809-822.

[24] C. Kanzow, N. Yamashita, and M. Fukushima, Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints, *Journal of Computational and Applied Mathematics* 172 (2) (2004) 375-397.

[25] G. Kauermann, C. Schellhase, and D. Ruppert, Flexible copula density estimation with penalized hierarchical B-splines, *Scandinavian Journal of Statistics* 40 (4) (2013) 685-705.

[26] G. Kim, M. J. Silvapulle, and P. Silvapulle, Comparison of semi-parametric and parametric methods for estimating copulas, *Computational Statistics & Data Analysis* 51 (6) (2007) 2836-2850.

[27] V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, Supervised High-Resolution Dual-Polarization SAR Image Classification by Finite Mixtures and Copulas, *IEEE Journal of Selected Topics in Signal Processing*, 5 (3) (2011) 554-566.

[28] K. Levenberg, A method for the solution of certain non-linear problems in least squares, *Quarterly of Applied Mathematics* 2 (2) (1944) 164-168.

[29] W. Li and Y. Jia, Rao-Blackwellised particle filtering and smoothing for jump Markov non-linear systems with mode observation, *IET Signal Processing* 7 (4) (2013) 327-336.

[30] F. Li, Identifying asymmetric comovements of international stock market returns, *Journal of Financial Econometrics*, 12 (3) (2014) 507-543.

[31] F. Lindsten, P. Bunch, S. Sarkka, et al., Rao-Blackwellized Particle Smoothers for Conditionally Linear Gaussian Models, *IEEE Journal of Selected Topics in Signal Processing*, 10 (2) (2016), 353-365

[32] K. Madsen, H. B. Nielsen, O. Tingle, *Methods for Non-Linear LeastSquares Problems* (2nd ed.), Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2004.

[33] D. W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics* 11 (2) (1963) 431-441.

[34] S. Nadarajah, E. Afuecheta, and S. Chan, A compadium of copulas, *Statistica*, 77 (4) (2017) 279-328.

[35] T. Ngoc Minh Nguyen, S. Le Corff, and E. Moulines, On the two-filter approximations of marginal smoothing distribution in general state-soace models, *Advanced in Applied Probabilities*, 50 (1) (2018) 157-177.

[36] N. Ngoc Minh Nguyen, S. Le Corff, and E. Moulines, Particle rejuvenation of Rao-Blackwellized sequential Monte Carlo smoothers for conditionally linear and Gaussian models, *EURASIP Journal on Advances in Signal Processing*, 54 (2017).

[37] M. Osborne, Nonlinear least squares - the Levenberg algorithm revisited, *The ANZIAM Journal* 19 (3) (1976) 343-357.

[38] J. Olsson and J. Westerborn, Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm, *Bernoulli* 23 (3) (2017) 1951–1996.

[39] A. J. Patton, Modelling asymmetric exchange rate dependence, *International Economic Review* 47 (2) (2006) 527-556.

[40] Y. Petetin and F. Desbouvries, A class of fast exact Bayesian filters in dynamical models with jumps, *IEEE Trans. on Signal Processing*, 62 (14) (2014) 3643-53.

[41] W. Pieczynski, Exact filtering in conditionally Markov switching hidden linear models, *Comptes Rendus Mathematique*, 349 (9-10) (2011) 587-590.

[42] S. Sarkka, P. Bunch, and S. Godsill, A backward- simulation based Rao- Blackwellized particle smoother for conditionally linear Gaussian models, *IFAC Proceedings* 45 (16) (2012) 506-511.

[43] A. Sklar, Fonctions de repartition a N dimensions et leurs marges, Universite Paris 8, 1959 (URL https://books.google.fr/books?id=nreSmAEACAAJ).

[44] A. Voisin, V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, Supervised classification of multisensor and multiresolution remote sensing images with a hierarchical copula-based approach, *IEEE Trans. on Geoscience and Remote Sensing*, 52 (6) (2014) 3346-3358.

[45] G. L. C. Yap, Optimal Filter Approximations for Latent Long Memory Stochastic Volatility, *Computational Economics*, to appear, DOI: 10.1007/s10614-019-09933-8, to appear.

[46] M. Zhang, C. Shen, Z. G. Wu, and D. Zhang, Dissipative Filtering for Switched Fuzzy Systems With Missing Measurements, *IEEE Trans. on Cybernetics*, Doi: 10.1109/TCYB.2019.2908430, to appear.

[47] F. Zheng, S. Derrode, and W. Pieczynski, Parameter estimation in switching Markov systems and unsupervised smoothing, *IEEE Trans. on Automatic Control*, 64 (4), (2019) 1761-1767.