



HAL
open science

Bayesian Variable Selection in Markov Mixture Models

Roberta Paroli, Luigi Spezia

► **To cite this version:**

Roberta Paroli, Luigi Spezia. Bayesian Variable Selection in Markov Mixture Models. Communications in Statistics - Simulation and Computation, 2008, 37 (01), pp.25-47. 10.1080/03610910701459956 . hal-00514318

HAL Id: hal-00514318

<https://hal.science/hal-00514318>

Submitted on 2 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Bayesian Variable Selection in Markov Mixture Models

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2007-0008.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	14-May-2007
Complete List of Authors:	Paroli, Roberta; Università Cattolica S.C., Statistics Spezia, Luigi; Università Cà Foscari, Statistics
Keywords:	Gibbs variable selection, Kuo-Mallick method, Metropolized-Kuo-Mallick method, Stochastic search variable selection
Abstract:	Bayesian methods for variable selection have become increasingly popular in recent years, due to advances in MCMC computational algorithms. Several methods have been proposed in literature in the case of linear and generalized linear models. In this paper we adapt some of the most popular algorithms to a class of non-linear and non-Gaussian time series models, i.e. the Markov mixture models (MMM). We also propose the "Metropolization" of the algorithm of Kuo and Mallick (1998), in order to tackle variable selection efficiently. Numerical comparisons among the competing MCMC algorithms are also presented via simulation examples.
<p>Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.</p> <p>fig2-3.wmf fig2-4.wmf Table1.tex Table2.tex Table3.tex Table4.tex Table5.tex</p>	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

fig1-1.wmf
fig1-2.wmf
fig1-3.wmf
fig1-4.wmf
fig2-1.wmf
fig2-2.wmf
Paroli_Spezia2.tex



For Peer Review Only

BAYESIAN VARIABLE SELECTION IN MARKOV MIXTURE MODELS

ROBERTA PAROLI

LUIGI SPEZIA

*Dipartimento di Scienze Statistiche,
Università Cattolica S.C., Milano*

*Dipartimento di Statistica,
Università Ca' Foscari, Venezia*

May 2007

Abstract - Bayesian methods for variable selection and model choice have become increasingly popular in recent years, due to advances in Markov chain Monte Carlo (MCMC) computational algorithms. Several methods have been proposed in literature in the case of linear and generalized linear models. In this paper we adapt some of the most popular algorithms to a class of non-linear and non-Gaussian time series models, i.e. the Markov mixture models (MMM). We also propose the “Metropolization” of the algorithm of Kuo and Mallick (1998), in order to tackle variable selection efficiently, both when the complexity of the model is high, as in MMM, and when the exogenous variables are strongly correlated. Numerical comparisons among the competing MCMC algorithms are also presented via simulation examples.

Keywords - Gibbs variable selection; Kuo-Mallick method; Metropolized-Kuo-Mallick method; Stochastic search variable selection.

1 Introduction

In recent years, several methods have been proposed in Bayesian literature for selecting exogenous variables in regression models: they include Stochastic Search Variable Selection (SSVS) by George and McCulloch (1993), the unconditional priors Gibbs sampler (KM) by Kuo and Mallick (1998)

and Gibbs Variable Selection (GVS) by Dellaportas et al. (2000). These methods have been also extended to generalized linear models (George et al. 1996), log-linear models (Ntzoufras et al., 2000) and multivariate regression models (Brown et al., 1998). Other authors provide alternative solutions to the variable selection problem: see, among others, Carlin and Chib (1995), Denison et al. (1998), Kohn et al. (2001), Nott and Green (2004), Nott and Kohn (2005), Schneider and Corcoran (2004) for some different approaches and recent developments. We focus our attention on the first three methods because they are the most popular and the most referred in recent literature.

The basic idea of these approaches is that the promising predictors are identified by their highest posterior probability, that is the “best” subset of predictors is that with the most frequent appearance in the sequence of the Markov chain Monte Carlo (MCMC) algorithm developed to evaluate the posterior probability. To give a simple idea, we consider the multiple linear regression model defined by the usual equation

$$Y = \sum_{j=1}^q \beta_j X_j + E,$$

where Y is a dependent variable; X_1, X_2, \dots, X_q are the potential fixed exogenous variables or covariates, which are candidates for inclusion in the model, and β_1, \dots, β_q are the regression coefficients. As usual the r.v. E is Normal with zero mean and variance σ^2 ($\mathcal{N}(0, \sigma^2)$). There are 2^q possible submodels to select, i.e. there are 2^q possible subsets of predictors. The selection of a subset is also equivalent to setting to 0 those regression coefficients that are not included in the final model, so a binary indicator vector γ of dimension q is added in the specification of the model and it represents which of the q variables are included: any γ_j ($j = 1, \dots, q$) assumes the value 1 if the variable X_j is included, 0 if it is not included.

Let $f(\gamma)$ be the prior distribution of γ ; the “best” subset of regressors is identified by analysing the marginal posterior distribution of γ , $f(\gamma|y^n) \propto f(y^n|\gamma)f(\gamma)$, which contains the information relevant to variable selection supported by the data $y^n = (y_1, y_2, \dots, y_n)'$ and by the prior. Intractable integrals are involved in the computation of this posterior distribution and in the evaluation of its mode, so an MCMC sampling procedure can be developed.

The differences among the above mentioned methods of variable selection depend on how the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

vector γ is involved in the model. In particular, in the SSVS approach the indicator vector γ is involved through the prior of the regression coefficients; in the Kuo-Mallick approach γ is part of the regression equation, while the GVS approach is a “hybrid” of SSVS and KM.

Our aim is twofold: first, we extend these three approaches to Markov switching autoregressive models with exogeneous variables (MSARX) and non-homogeneous hidden Markov models (NHHMM), which belong to the class of Markov mixture models (MMM), as defined by Chib (1996); second, we propose the “Metropolization” of the algorithm of Kuo and Mallick (1998), giving rise to Metropolized-Kuo-Mallick (MKMK) method, which increases the performance of the algorithm both when the complexity of the model is high, as in MMM, and when exogenous variables are strongly correlated. The correct selection of exogenous variables is due to the acceptance in block of both the coefficients and the indicators of the exogenous variables, because blocking increases the precision of the estimators and improves the chain mixing.

The paper is organized as follows. The four techniques for variable selection, i.e. KM, SSVS, GVS and MKMK, are applied to MSARX and to NHHMM in Section 2, while, in Section 3, several simulation examples are performed to illustrate the methods and to compare numerically the competing algorithms.

2 Variable selection in Markov mixture models

In this section we apply KM, SSVS, GVS and MKMK to the Markov mixtures models. The general MMM is analysed by Chib (1996), who describes it in terms of finite mixture distributions in which the component populations are selected according to contemporary states of an unobserved Markov process: at each time t , a realization of the Markov chain, also called regime, occurs hiddenly. The conditional density of the current observation of the time series, given the previous ones and the previous regime, is a finite mixture of densities, whose mixing distribution is the row of the transition matrix corresponding to the previous regime.

Two special MMM are considered here: Markov switching autoregressive models and hidden Markov models. Suppose we have a couple of discrete-time stochastic processes, one observed and one unobserved, or hidden: the realizations of the latter can be observed only through the

1
2
3
4
5 realizations of the former one. Suppose also the dynamics of the unobserved process affects the
6
7 dynamics of the observed one. We model the hidden process as a finite-state Markov chain and
8
9 assume that the observed process, given the Markov chain, is a sequence of conditional independent
10
11 random variables, whose conditional distribution depends on the Markov chain only through its
12
13 contemporary state. The conditional distributions belong to a single parametric family, usually
14
15 Poisson or Normal. This class of models is called hidden Markov models (HMM). A hidden
16
17 Markov chain may be assumed because it is either suggested by the physical nature of the observed
18
19 phenomenon, such as in speech recognition (Elliott, Aggoun, Moore (1995)) and in biology for the
20
21 analysis of the architecture of proteins and nucleic acids (Koski (2001)), or just for convenience
22
23 in formulating the model, such as in overdispersed Poisson processes (MacDonald and Zucchini
24
25 (1997)).

26
27 By contrast, in Markov switching autoregressive (MSAR) models the conditional independence
28
29 condition is replaced by the autoregressive dependence condition and the conditional distributions
30
31 are always Normals. We have that different autoregressions, each one depending on a latent regime,
32
33 alternate according to the regime switching, which is driven by the Markov chain. These models
34
35 have been introduced in econometric literature to study economic and financial time series with
36
37 asymmetric cycles and changes in regime generated by a stochastic process (Hamilton (1993)).
38
39 Krolzig (1997), Kim and Nelson (1999) and Franses and van Dijk (2000) provide generalizations
40
41 and applications of this class of models to economic and financial time series, in which the hidden
42
43 states represent different macroeconomic regimes.

44 45 46 47 **2.1 Markov Switching Autoregressive models with exogenous variables**

48
49 A Markov switching autoregressive model of order m and p , i.e. $\text{MSAR}(m, p)$, is defined as a
50
51 bivariate discrete-time stochastic process $\{S_t; Y_t\}$, where $\{S_t\}$ is a latent, or hidden, finite-state
52
53 Markov chain and $\{Y_t\}$, given $\{S_t\}$, satisfies the order- p dependence and the contemporary depen-
54
55 dence conditions: we have a sequence of observed random variables $\{Y_t\}$ depending on the p past
56
57 observations, whose conditional distributions depend on $\{S_t\}$ only through the contemporary state
58
59 of the Markov chain.
60

Let $\{S_t\}$ be a discrete-time, first-order, homogeneous, ergodic Markov chain on a finite state-space with cardinality m . The transition matrix of the process is $P = [p_{i,j}]$, where $p_{i,j} = P(S_t = j \mid S_{t-1} = i)$, for any $i, j \in \{1, \dots, m\}$ and any $t = 2, \dots, n$. The sequence of the states of the Markov chain is denoted with $s^n = (s_1, \dots, s_n)'$ and, for any $t = 1, \dots, n$, any s_t has values in $\{1, \dots, m\}$.

Hence, given the order- p dependence and the contemporary dependence conditions, the equation describing MSAR(m, p) is

$$Y_{t(i)} = \mu_i + \sum_{\tau=1}^p \varphi_{\tau(i)} y_{t-\tau} + E_{t(i)},$$

where $Y_{t(i)}$ denotes the conditional variable Y_t when $S_t = i$, for any $1 \leq t \leq n$ and for any $i \in \{1, \dots, m\}$, and $E_{t(i)}$ denotes the Gaussian noise E_t when $S_t = i$, with zero mean and variance σ_i^2 ($E_{t(i)} \sim \mathcal{N}(0; \sigma_i^2)$), for any $i \in \{1, \dots, m\}$, with the discrete-time processes $\{E_t\}$, given $\{S_t\}$, satisfying the conditional independence and the contemporary dependence conditions. Any signal μ_i , any variance σ_i^2 and any autoregressive coefficient $\varphi_{\tau(i)}$, for any $\tau = 1, \dots, p$, depend on the current state i of the Markov chain, for any $i \in \{1, \dots, m\}$.

If we join to the MSAR(m, p) model q contemporary exogenous variables, we have the MSARX(m, p) model, defined by the equation

$$Y_{t(i)} = \mu_i + \sum_{\tau=1}^p \varphi_{\tau(i)} y_{t-\tau} + \sum_{j=1}^q \beta_{j(i)} X_{tj} + E_{t(i)}, \quad (1)$$

where $X_{t1}, X_{t2}, \dots, X_{tq}$ are the potential fixed exogenous variables which are candidates for inclusion in the final model and $\beta_i = (\beta_{1(i)}, \dots, \beta_{q(i)})'$ are their q coefficients when the hidden chain visits state i ; hence the observed process, at any time t , can be affected by different exogenous variables, according to the hidden state we assume in t . The matrix $n \times q$ of the observations of the exogenous variables will be denoted with $X = [x_{tj}]$, for $t = 1, \dots, n$ and $j = 1, \dots, q$. To automatically satisfy the constraint of stationarity of any AR subprocess on any vectors $\varphi_i = (\varphi_{1(i)}, \dots, \varphi_{p(i)})'$, we reparametrize the model in terms of the log transformation of the partial autocorrelations $r_i = (r_{1(i)}, \dots, r_{p(i)})'$, according to Mariott et al. (1996).

Identifiability of the model is ensured by imposing increasing signals μ_i , but the procedures we introduce can be easily adapted to models with any other type of constraint, provided the chosen constraint respects the geometry of the posterior distribution. An alternative option to

fixed constraints is postprocessing the MCMC outputs (Celeux et al. (2000)).

The unknown parameters of the model are $(\mu, \sigma^2, R, \beta, P)$ where μ is the vector of the m signals μ_i ; σ^2 is the vector of the m variances σ_i^2 ; R is the matrix of the m vectors R_i of the log transformations of the partial autocorrelations, i.e. $R = (R'_1, \dots, R'_i, \dots, R'_m)'$ with $R_i = (R_{1(i)}, \dots, R_{\tau(i)}, \dots, R_{p(i)})'$ and $R_{\tau(i)} = \ln\left(\frac{1+r_{\tau(i)}}{1-r_{\tau(i)}}\right)$; β is the matrix of the m vectors β_i , i.e. $\beta = (\beta'_1, \dots, \beta'_i, \dots, \beta'_m)'$; P is the matrix of the m^2 transition probabilities $p_{i,j}$; also the sequence of the hidden states s^n is unknown and it will be estimated. The priors are so specified:

- 1) any parameter μ_i is Normal with known μ_M and σ_M^2 ($\mu \sim \mathcal{N}(\mu_M, \sigma_M^2)$, for any $i = 1, \dots, m$);
- 2) each entry of matrix R is an independent Normal with known μ_R and σ_R^2 ($R_{\tau(i)} \sim \mathcal{N}(\mu_R, \sigma_R^2)$, for any $i = 1, \dots, m$ and for any $\tau = 1, \dots, p$);
- 3) each row of matrix β is an independent multivariate Normal of dimension q with known vector μ_B and matrix Σ_B ($\beta_i \sim \mathcal{N}_q(\mu_B, \Sigma_B)$, for any $i = 1, \dots, m$);
- 4) any variance σ_i^2 is an independent Inverse Gamma with known ν and η ($\sigma_i^2 \sim IG(\nu, \eta)$, for any $i = 1, \dots, m$);
- 5) any row of P , $P_{i\bullet} = (p_{i,1}, p_{i,2}, \dots, p_{i,m})$, is an independent Dirichlet with parameter $\omega = (\omega_1, \dots, \omega_m)$ ($P_{i\bullet} \sim \mathcal{D}(\omega)$, for any $i = 1, \dots, m$).

The following subsections will be devoted to develop the MCMC variables selection procedures of KM, SSVS, GVS and MKMK for this type of models. We assume that the observed process can be influenced by different exogenous variables, depending on the state visited by the hidden Markov chain; this means that we have a multiple selection problem, consisting in the estimation of m indicator vectors $\gamma_i = (\gamma_{1(i)}, \dots, \gamma_{j(i)}, \dots, \gamma_{q(i)})'$, with $i = 1, \dots, m$, which are entries of the indicator matrix $\gamma = (\gamma'_1, \dots, \gamma'_i, \dots, \gamma'_m)'$.

2.1.1 KM method

In the KM approach, we assume that binary indicators are associated to the parameters of any exogenous variables, for any state i , and that they are embedded in the model equation: model

(1) becomes

$$Y_{t(i)} = \mu_i + \sum_{\tau=1}^p \varphi_{\tau(i)} y_{t-\tau} + \sum_{j=1}^q \gamma_{j(i)} \beta_{j(i)} X_{tj} + E_{t(i)}.$$

The model also includes the prior distributions of the $\gamma_{j(i)}$'s that are independent Bernoulli with known probability ξ_j ($\gamma_{j(i)} \sim Be(\xi_j)$), for any i and for any $j = 1, \dots, q$, where $\xi_j = P(\gamma_{j(i)} = 1) = 1 - P(\gamma_{j(i)} = 0)$. If $\xi_j = 0.5$, for any $j = 1, \dots, q$, we have the uniform or indifference prior $f(\gamma) = 2^{-q}$, while if $\xi_j < 0.5$ large models are penalized in favour of the parsimonious ones.

The MCMC scheme used to generate, at any iteration k , the sequence of $s^{n(k)}$, $\mu^{(k)}$, $\sigma^{2(k)}$, $R^{(k)}$, $\beta^{(k)}$, $\gamma^{(k)}$, $P^{(k)}$ of the KM algorithm for this class of models is Metropolis-within-Gibbs. The algorithm proposed here provides the identifiability constraint on the signals ($\mu_i^{(k)} < \mu_j^{(k)}$, for any k and for any $i, j = 1, \dots, m$, so that $i < j$), but it can be easily rearranged when another type of constraint is specified.

[step 1] The sequence $s^{n(k)}$ of hidden states is generated by the well-known forward filtering-backward sampling (*ff-bs*) algorithm (Chib (1996)).

[step 2] The m parameters $\mu_i^{(k)}$, for any i , are independently generated from a Normal distribution with mean

$$\frac{\sigma_i^{-2(k-1)} \sum_{\{t \geq 1: s_t^{(k)} = i\}} \left(y_t - \sum_{\tau=1}^p \varphi_{\tau(i)}^{(k-1)} y_{t-\tau} - \sum_{j=1}^q \gamma_{j(i)}^{(k-1)} \beta_{j(i)}^{(k-1)} x_{t,j} \right) + \mu_M \sigma_M^{-2}}{N_i^{(k)} \sigma_i^{-2(k-1)} + \sigma_M^{-2}}$$

and variance

$$\left(N_i^{(k)} \sigma_i^{-2(k-1)} + \sigma_M^{-2} \right)^{-1},$$

where $N_i^{(k)}$ is the number of observations corresponding to the contemporary hidden state i in the sequence $s^{n(k)}$ generated at step [1]. The entries of the vector $\mu^{(k)}$ must be in increasing order to satisfy the identifiability constraint, so we apply the constrained permutation sampling algorithm (Frühwirth-Schnatter (2001)) to order them.

[step 3] The parameters $R_{\tau(i)}^{(k)}$, for any $\tau = 1, \dots, p$ and any i , are independently generated from the random walk $R_{\tau(i)}^{(k)} = R_{\tau(i)}^{(k-1)} + U_R$, where U_R is a univariate Gaussian noise with zero mean and constant variance, for any k . Then any vector $R_i^{(k)}$ is accepted with probability

$$A \left(R_i^{(k-1)}; R_i^{(k)} \right) = \min \left\{ 1; \frac{\pi \left(R_i^{(k)} \mid \mu^{(k)}, \sigma^{2(k-1)}, \beta^{(k-1)}, \gamma^{(k-1)}, s^{n(k)}, X, y^n \right)}{\pi \left(R_i^{(k-1)} \mid \mu^{(k)}, \sigma^{2(k-1)}, \beta^{(k-1)}, \gamma^{(k-1)}, s^{n(k)}, X, y^n \right)} \right\},$$

for any i , where the numerator of the acceptance ratio is proportional to

$$\exp \left\{ -\frac{1}{2} \sigma_i^{-2(k-1)} \sum_{\{t \geq 1: s_t^{(k)} = i\}} \left(y_t - \mu_i^{(k)} - \sum_{\tau=1}^p \varphi_{\tau(i)}^{(k)} y_{t-\tau} - \sum_{j=1}^q \gamma_{j(i)}^{(k-1)} \beta_{j(i)}^{(k-1)} x_{t,j} \right)^2 - \frac{1}{2} \sigma_R^{-2} \sum_{\tau=1}^p \left(R_{\tau(i)}^{(k)} - \mu_R \right)^2 \right\}$$

and the denominator is identical to the numerator except the introduction of $\varphi_{\tau(i)}^{(k-1)}$ and $R_i^{(k-1)}$.

Notice that once the $R_i^{(k)}$'s have been generated, the new vectors $\varphi_i^{(k)}$'s can be obtained by the inverse transformation.

[step 4] The parameters $\beta_i^{(k)}$, for any i , are independently generated from a Normal distribution of dimension q with mean vector

$$\left(X^{*T} Q_i^{(k)} X^* + \Sigma_B^{-1} \right)^{-1} \left(X^{*T} Q_i^{(k)} \widehat{y}_i^{T(k)} + \Sigma_B^{-1} \mu_B \right)$$

and covariance matrix

$$\left(X^{*T} Q_i^{(k)} X^* + \Sigma_B^{-1} \right)^{-1},$$

where X^* is the $n \times q$ matrix with elements $x_{tj}^* = \gamma_{j(i)}^{(k-1)} x_{tj}$, for any $t = 1, \dots, n$ and $j = 1, \dots, q$; $Q_i^{(k)}$ is a $n \times n$ diagonal matrix whose t -th term on the diagonal is either $\sigma_i^{-2(k-1)}$, if $s_t^{(k)} = i$, or zero, otherwise; \widehat{y}_i^n is a n -dimensional vector whose generic t -th element is either $y_t - \mu_i^{(k)} - \sum_{\tau=1}^p \varphi_{\tau(i)}^{(k)} y_{t-\tau}$, if $s_t^{(k)} = i$, or zero, otherwise.

[step 5] For any i -th row of the indicator matrix $\gamma^{(k)}$, every coefficient $\gamma_{j(i)}^{(k)}$ is independently generated from a Bernoulli distribution

$$\gamma_{j(i)}^{(k)} \sim f \left(\gamma_{j(i)} | \mu^{(k)}, \sigma^{2(k-1)}, \beta^{(k)}, \gamma_{\setminus j(i)}^{(k-1)}, s^{n(k)}, X, y^n \right) \sim Be \left(\pi_{j(i)}^{(k)} \right),$$

where $\gamma_{\setminus j(i)}^{(k-1)} = \left(\gamma_{1(i)}^{(k-1)}, \dots, \gamma_{j-1(i)}^{(k-1)}, \gamma_{j+1(i)}^{(k-1)}, \dots, \gamma_{q(i)}^{(k-1)} \right)'$ is the vector of the entries of $\gamma_{(i)}^{(k-1)}$ in which the j -th entry is suppressed and the first $(j-1)$ -th are updated by the new values. The parameter of the Bernoulli distribution is $\pi_{j(i)}^{(k)} = a_{j(i)}^{(k)} / \left(a_{j(i)}^{(k)} + b_{j(i)}^{(k)} \right)$, with

$$a_{j(i)}^{(k)} = f \left(y^n | \mu^{(k)}, \sigma^{2(k-1)}, \beta^{(k)}, s^{n(k)}, X, \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1 \right) f \left(\gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1 \right) \propto \exp \left\{ -\frac{1}{2} \sigma_i^{-2(k-1)} \sum_{\{t \geq 1: s_t^{(k)} = i\}} \left(y_t - \mu_i^{(k)} - \sum_{\tau=1}^p \varphi_{\tau(i)}^{(k)} y_{t-\tau} - \sum_{h=1}^q \theta_{h(i)}^{*(k)} x_{th} \right)^2 \right\} \xi_j$$

and

$$b_{j(i)}^{(k)} = f \left(y^n | \mu^{(k)}, \sigma^{2(k-1)}, \beta^{(k)}, s^{n(k)}, X, \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0 \right) f \left(\gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0 \right) \propto \exp \left\{ -\frac{1}{2} \sigma_i^{-2(k-1)} \sum_{\{t \geq 1: s_t^{(k)} = i\}} \left(y_t - \mu_i^{(k)} - \sum_{\tau=1}^p \varphi_{\tau(i)}^{(k)} y_{t-\tau} - \sum_{h=1}^q \theta_{h(i)}^{*(k)} x_{th} \right)^2 \right\} (1 - \xi_j),$$

where $f\left(y^n|\mu^{(k)}, \sigma^{2(k-1)}, \beta^{(k)}, s^{n(k)}, X, \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1\right)$ and $f\left(y^n|\mu^{(k)}, \sigma^{2(k-1)}, \beta^{(k)}, s^{n(k)}, X, \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0\right)$ are the likelihoods computed either with $\gamma_{j(i)}^{(k)} = 1$ or $\gamma_{j(i)}^{(k)} = 0$, with the $\theta_{h(i)}^{*(k)}$'s entries of the vector $\theta_{(i)}^{(k)} = \left(\gamma_{1(i)}^{(k)}\beta_{1(i)}^{(k)}, \dots, \gamma_{q(i)}^{(k-1)}\beta_{q(i)}^{(k)}\right)'$ in which the j -th entry is replaced by $\beta_{j(i)}^{(k)}$, i.e. $\theta_{(i)}^{*(k)} = \left(\gamma_{1(i)}^{(k)}\beta_{1(i)}^{(k)}, \dots, \gamma_{j-1(i)}^{(k)}\beta_{j-1(i)}^{(k)}, \beta_{j(i)}^{(k)}, \gamma_{j+1(i)}^{(k-1)}\beta_{j+1(i)}^{(k)}, \dots, \gamma_{q(i)}^{(k-1)}\beta_{q(i)}^{(k)}\right)'$, and the $\theta_{h(i)}^{**k}$'s entries of the vector $\theta_{(i)}^{(k)}$ in which the j -th entry is replaced by 0, i.e. $\theta_{(i)}^{**k} = \left(\gamma_{1(i)}^{(k)}\beta_{1(i)}^{(k)}, \dots, \gamma_{j-1(i)}^{(k)}\beta_{j-1(i)}^{(k)}, 0, \gamma_{j+1(i)}^{(k-1)}\beta_{j+1(i)}^{(k)}, \dots, \gamma_{q(i)}^{(k-1)}\beta_{q(i)}^{(k)}\right)'$.

[step 6] The m parameters $\sigma_i^{2(k)}$ are independently generated from an Inverse Gamma distribution with parameters

$$\frac{N_i^{(k)}}{2} + \nu \quad \text{and} \quad \frac{1}{2} \sum_{\{t \geq 1: s_t^{(k)} = i\}} \left(y_t - \mu_i^{(k)} - \sum_{\tau=1}^p \varphi_{\tau(i)}^{(k)} y_{t-\tau} - \sum_{j=1}^q \gamma_{j(i)}^{(k)} \beta_{j(i)}^{(k)} x_{t,j} \right)^2 + \eta.$$

[step 7] Each vector $P_{i\bullet}^{(k)}$ is independently generated from a Dirichlet distribution $\mathcal{D}\left(\omega + N_{i\bullet}^{(k)}\right)$, where $N_{i\bullet}^{(k)} = \left(N_{i,1}^{(k)}, \dots, N_{i,m}^{(k)}\right)$, with $N_{i,j}^{(k)}$ the number of couples of consecutive hidden states i, j in the sequence $s^{n(k)}$, for any i, j .

2.1.2 SSVS method

In the SSVS method, the indicator vectors γ_i are involved in the model through the prior of the coefficients $\beta_{j(i)}$, that are mixtures of two Normals with different variances

$$\beta_{j(i)}|\gamma_{j(i)} \sim \gamma_{j(i)}\mathcal{N}(\mu_{\beta_j}, c_j^2\tau_j^2) + (1 - \gamma_{j(i)})\mathcal{N}(\mu_{\beta_j}, \tau_j^2),$$

for specified c_j^2 and τ_j^2 and μ_{β_j} the j -th entry of μ_B . The prior distribution of any vector β_i , for any i , is a multivariate Normal of dimension q

$$\beta_i|\gamma_i \sim \mathcal{N}_q(\mu_B, D_{\gamma(i)}),$$

where $D_{\gamma(i)} = \text{diag}[(\delta_{1(i)}\tau_1)^2, \dots, (\delta_{q(i)}\tau_q)^2]$ with $\delta_{j(i)} = c_j$ if $\gamma_{j(i)} = 1$ and $\delta_{j(i)} = 1$ if $\gamma_{j(i)} = 0$.

The choice of the hyperparameters c_j^2 and τ_j^2 and the interpretation of this prior is discussed in George and McCulloch (1993) and Ntzoufras et al. (2000). The hyperparameters c_j and τ_j must be chosen carefully because the posterior probability is mostly dependent on them, given that the exogenous variables are never eliminated from the model space, but coefficients of the variables to be excluded are set close to zero with high probability. So it is recommended to run various

simulations with different values of c_j and τ_j and to analyse how the posterior probability of any vector γ_i is sensitive to the choice of these tuning factors.

The prior distributions of the remaining parameters are the same as for those in the KM method.

In the SSVS scheme used to generate the sequence of $s^{n(k)}, \mu^{(k)}, \sigma^{2(k)}, R^{(k)}, \beta^{(k)}, \gamma^{(k)}, P^{(k)}$, only the step [7] remains the same as for KM, since it does not depend on β_i and γ_i ; step [5] is completely different, while the remaining steps are identical to those in KM, except the presence of the $\gamma_{j(i)}$'s that multiply the $\beta_{j(i)}$'s. More in details, step [5] is so defined.

[step 5] For any i -th row of the indicator matrix $\gamma^{(k)}$, every coefficient $\gamma_{j(i)}^{(k)}$ is independently generated from a Bernoulli distribution $Be\left(\frac{a_{j(i)}^{(k)}}{a_{j(i)}^{(k)} + b_{j(i)}^{(k)}}\right)$, where

$$a_{j(i)}^{(k)} = f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1\right) f\left(\gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1\right) \propto f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1\right) \xi_j$$

and

$$b_{j(i)}^{(k)} = f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0\right) f\left(\gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0\right) \propto f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0\right) (1 - \xi_j),$$

with $f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1\right)$ or $f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0\right)$ the density functions of the prior distribution of $\beta_i | \gamma_i$ when $\gamma_{j(i)} = 1$ or $\gamma_{j(i)} = 0$, respectively. Their explicit forms are

$$f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1\right) \propto \exp\left\{-\frac{1}{2} \sum_{j \neq i} \left(\beta_i^{(k)} - \mu_B\right)' D_{\gamma^{(i)}}^{*(k)-1} \left(\beta_i^{(k)} - \mu_B\right)\right\}$$

and

$$f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0\right) \propto \exp\left\{-\frac{1}{2} \sum_{j \neq i} \left(\beta_i^{(k)} - \mu_B\right)' D_{\gamma^{(i)}}^{**(k)-1} \left(\beta_i^{(k)} - \mu_B\right)\right\},$$

where $D_{\gamma^{(i)}}^{*(k)}$ and $D_{\gamma^{(i)}}^{**(k)}$ are the covariance matrix of the prior of $\beta_i^{(k)}$ in which the j -th entries of $D_{\gamma^{(i)}}^{*(k)}$ are replaced by $\delta_{j(i)} = c_j$ (which corresponds to $\gamma_{j(i)}^{(k)} = 1$), while those of $D_{\gamma^{(i)}}^{**(k)}$ by $\delta_{j(i)} = 1$ (which corresponds to $\gamma_{j(i)}^{(k)} = 0$).

2.1.3 GVS method

GVS is a mixed approach between the two previous methods. Here the indicator variable influences both the prior distribution of the regression coefficients, as in SSVS, and the model equation, as in KM. The prior distributions are the same as for SSVS. In the MCMC scheme, steps [1], [2], [3], [6] and [7] are identical to those in KM, while steps [4] and [5] are a combination of KM and SSVS.

In details, the elements of the parameters of the Bernoulli distributions of any $\gamma_{j(i)}^{(k)}$ are

$$a_{j(i)}^{(k)} \propto f\left(y^n | \mu^{(k)}, \sigma^{2(k-1)}, \beta^{(k)}, s^{n(k)}, X, \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1\right) f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 1\right) \xi_j$$

and

$$b_{j(i)}^{(k)} \propto f\left(y^n | \mu^{(k)}, \sigma^{2(k-1)}, \beta^{(k)}, s^{n(k)}, X, \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0\right) f\left(\beta_i^{(k)} | \gamma_{\setminus j(i)}^{(k-1)}, \gamma_{j(i)}^{(k)} = 0\right) (1 - \xi_j).$$

Notice that both the likelihoods $f(y^n | \cdot)$, as in KM, and the conditional priors $f(\beta_i^{(k)} | \cdot)$, as in SSVS, are present in the expressions of $a_{j(i)}^{(k)}$ and $b_{j(i)}^{(k)}$.

2.1.4 MKMK method

When the complexity of the model is high, as for MSARX and NHHMM, or when exogenous variables are strongly correlated, the performances of KM, SSVS and GVS can be improved by introducing Metropolized-Kuo-MallicK method (MKMK), based on the acceptance in block of the coefficients β_i 's and the indicators of the exogenous variables γ_i 's: Liu, Wong, Kong (1994) showed that the precision of the estimators increases when the unknown parameters are gathered in separated blocks, while Gilks and Roberts (1996) showed that blocking improves the chain mixing.

In the basic MCMC algorithm only one step must be modified w.r.t. KM: for any state i , the coefficients of the exogenous variables and the associated dummy indicators are sampled jointly and updated in block within a Metropolis step. For MSARX models, the pair of vectors $(\beta_i^{(k)}; \gamma_i^{(k)})$ must be accepted in block, so steps [4] and [5] of Subsection 2.1.1 are joined in only one step and modified as follows. We assume a random walk proposal for any $\beta_{j(i)}^{(k)}$ and an independent Bernoulli proposal for any $\gamma_{j(i)}^{(k)}$.

For any i ($i = 1, \dots, m$), first, parameters $\beta_{j(i)}^{(k)}$, for any $j = 1, \dots, q$, are independently generated from the random walk $\beta_{j(i)}^{(k)} = \beta_{j(i)}^{(k-1)} + U_B$, where U_B is a Gaussian noise with zero mean and constant variance, for any k ; then, every parameter $\gamma_{j(i)}^{(k)}$, for any $j = 1, \dots, q$, is independently generated from a Bernoulli distribution with probability $\pi_{j(i)}^{(k)} = a_{j(i)}^{(k)} / (a_{j(i)}^{(k)} + b_{j(i)}^{(k)})$, where $a_{j(i)}^{(k)}$ and $b_{j(i)}^{(k)}$ are the same as in step [5] of Subsection 2.1.1; finally the pair of vectors $(\beta_i^{(k)}; \gamma_i^{(k)})$ is

accepted in block with probability

$$A \left[\left(\beta_i^{(k-1)}; \gamma_i^{(k-1)} \right); \left(\beta_i^{(k)}; \gamma_i^{(k)} \right) \right] = \min \left\{ 1; \frac{\pi \left(\beta_i^{(k)}; \gamma_i^{(k)} \right) q \left(\beta_i^{(k-1)}; \gamma_i^{(k-1)} \mid \beta_i^{(k)}; \gamma_i^{(k)} \right)}{\pi \left(\beta_i^{(k-1)}; \gamma_i^{(k-1)} \right) q \left(\beta_i^{(k)}; \gamma_i^{(k)} \mid \beta_i^{(k-1)}; \gamma_i^{(k-1)} \right)} \right\} =$$

$$= \min \left\{ 1; \frac{f \left(y^n \mid \beta_i^{(k)}; \gamma_i^{(k)} \right) p \left(\gamma_i^{(k)} \right) p \left(\beta_i^{(k)} \right) q \left(\gamma_i^{(k-1)} \mid \beta_i^{(k-1)}; \beta_i^{(k)}; \gamma_i^{(k)} \right)}{f \left(y^n \mid \beta_i^{(k-1)}; \gamma_i^{(k-1)} \right) p \left(\gamma_i^{(k-1)} \right) p \left(\beta_i^{(k-1)} \right) q \left(\gamma_i^{(k)} \mid \beta_i^{(k)}; \beta_i^{(k-1)}; \gamma_i^{(k-1)} \right)} \right\},$$

by the factorization of the proposal density, i.e.

$$q \left(\beta_i^{(k)}; \gamma_i^{(k)} \mid \beta_i^{(k-1)}; \gamma_i^{(k-1)} \right) = q \left(\gamma_i^{(k)} \mid \beta_i^{(k)}; \beta_i^{(k-1)}; \gamma_i^{(k-1)} \right) q \left(\beta_i^{(k)} \mid \beta_i^{(k-1)}; \gamma_i^{(k-1)} \right),$$

by the independence of $\beta_i^{(k)}$ on $\gamma_i^{(k-1)}$ and cancelling the ratio $q \left(\beta_i^{(k-1)} \mid \beta_i^{(k)} \right) / q \left(\beta_i^{(k)} \mid \beta_i^{(k-1)} \right)$

for the symmetry of the proposal distribution.

2.2 Non-homogeneous hidden Markov models

In the MSARX models of the previous section, the Markov chain of the latent process is supposed homogeneous, that is the transition probabilities do not depend on time t . In the most general case, the Markov chain can be non-homogeneous, that is the transition probabilities are time-varying, by assuming that their dynamics depend on exogenous variables. If we consider a HMM with a non-homogeneous Markov chain, we obtain a Non-Homogeneous Hidden Markov model (NHHMM). It is a discrete-time stochastic processes $\{S_t, Y_t\}$ where $\{S_t\}$ is now a non-homogeneous Markov chain whose transition probabilities $p_{i,j}^t = P(S_t = j \mid S_{t-1} = i)$ vary at any time t by depending on exogenous variables. We suppose that they can be expressed as logistic functions of q exogenous variables. Let X_t be a $(q+1)$ -dimensional vector of the observations of the exogenous deterministic variables at time t , $X_t = (1, x_{t,1}, \dots, x_{t,q})'$ and $\alpha_{i,j}$ be a $(q+1)$ -dimensional vector of parameters, $\alpha_{i,j} = (\alpha_{0(i,j)}, \alpha_{1(i,j)}, \dots, \alpha_{q(i,j)})'$, for any $i, j \in \{1, \dots, m\}$, if $i \neq j$, and a $(q+1)$ -dimensional vector of zeros, if $i = j$. For any $t = 2, \dots, n$, the transition probabilities can be expressed as

$$\text{logit}(p_{i,j}^t) = \ln \left(p_{i,j}^t / p_{i,i}^t \right) = \sum_{h=0}^q \alpha_{h(i,j)} x_{th} = X_t' \alpha_{i,j} \quad (2)$$

$$p_{i,j}^t = \left(\exp \left(X_t' \alpha_{i,j} \right) \right) / \left(1 + \sum_{j \neq i} \exp \left(X_t' \alpha_{i,j} \right) \right),$$

for any $i, j \in \{1, \dots, m\}$ and let $P_t = [p_{i,j}^t]$ be the $m \times m$ transition matrix of the chain, for any

$t = 2, \dots, n$.

Process $\{Y_t\}$, given process $\{S_t\}$, is an observed sequence of conditionally independent random variables, whose conditional distributions depend on $\{S_t\}$ only through the contemporary S_t . Here we assume the probability density function of any Y_t , given S_t , is Gaussian; hence we have Gaussian NHHMM. The model equation is

$$Y_{t(i)} = \mu_i + E_{t(i)},$$

where $E_{t(i)}$ denotes the Gaussian random variable E_t , when $S_t = i$, with zero mean and variance σ_i^2 ($E_{t(i)} \sim \mathcal{N}(0; \sigma_i^2)$), for any i , with the discrete-time process $\{E_t\}$, given $\{S_t\}$, satisfying the conditional independence and the contemporary dependence conditions; so $Y_{t(i)} \sim \mathcal{N}(\mu_i; \sigma_i^2)$, for any i . Notice that the inferential tools and the variable selection machinery we are going to describe can also be applied when another conditional distribution, both discrete and continuous, is hypothesized.

Identifiability of the model is again ensured by imposing increasing signals μ_i . The unknown parameters and the latent data of the NHHMM are $\mu, \sigma^2, \alpha, s^n$, where μ is the vector of the m signals μ_i , σ^2 is the vector of the m variances σ_i^2 , α is the $m \times m$ matrix of the vectors $\alpha_{i,j}$ and s^n is the sequence of the hidden states.

The following prior assumptions are stated in analogy with those in Subsection 2.1:

- 1) the prior distribution of any parameter μ_i is an independent Normal with known μ_M and σ_M^2 ($\mu_i \sim \mathcal{N}(\mu_M, \sigma_M^2)$, for any $i = 1, \dots, m$);
- 2) the prior distribution of any vector $\alpha_{i,j}$ is an independent multivariate Normal of dimension $(q+1)$ with known vector μ_A and matrix Σ_A ($\alpha_{i,j} \sim \mathcal{N}_{q+1}(\mu_A; \Sigma_A)$, for any $i, j = 1, \dots, m$, with $i \neq j$);
- 3) the prior distribution of any variance σ_i^2 is assumed to be an independent Inverse Gamma ($\sigma_i^2 \sim IG(\nu, \eta)$, for any $i = 1, \dots, m$).

2.2.1 KM method

Let $\alpha_i = (\alpha_{0(i)}, \alpha_{1(i)}, \dots, \alpha_{q(i)})'$, $i = 1, \dots, m$, be the i -th row of matrix α ; we associate to any entry of matrix α , i.e. the vector of parameters $\alpha_{i,j}$, a vector of binary indicators $\gamma_{(i)} = (1, \gamma_{1(i)}, \dots, \gamma_{q(i)})'$ of dimension $q+1$, which regulates the inclusion of the exogenous variables

when the hidden chain visits state i at time $t - 1$; finally, let γ be the $m \times (q + 1)$ matrix of any vector $\gamma_{(i)}$, i.e. $\gamma = (\gamma'_{(1)}, \dots, \gamma'_{(m)})'$. Expression (2) becomes

$$\begin{aligned} \text{logit}(p_{i,j}^t) &= \sum_{h=0}^q \gamma_{h(i)} \alpha_{h(i,j)} x_{th} = X_t' \theta_{(i,j)} \\ p_{i,j}^t &= \exp \left(\sum_{h=0}^q \gamma_{h(i)} \alpha_{h(i,j)} x_{th} \right) / \left(1 + \sum_{j \neq i} \exp \left(\sum_{h=0}^q \gamma_{h(i)} \alpha_{h(i,j)} x_{th} \right) \right) = \\ &= \exp \left(X_t' \theta_{(i,j)} \right) / \left(1 + \sum_{j \neq i} \exp \left(X_t' \theta_{(i,j)} \right) \right), \end{aligned} \quad (3)$$

with $\theta_{(i,j)} = (\alpha_{0(i,j)}, \gamma_{1(i)} \alpha_{1(i,j)}, \dots, \gamma_{q(i)} \alpha_{q(i,j)})'$.

Notice that we assume the transition from state i to any other state j is regulated by the same set of covariates, hence the same vector $\gamma_{(i)}$ is associated to any $(\alpha_{0(i,j)}, \alpha_{1(i,j)}, \dots, \alpha_{q(i,j)})'$ ($j = 1, \dots, m$) and consequently to $P_{i\bullet}$.

The prior distribution of any $\gamma_{h(i)}$ is an independent Bernoulli with known probability ξ_h ($\gamma_{h(i)} \sim Be(\xi_h)$, for any $i = 1, \dots, m$ and any $h = 1, \dots, q$).

The steps of the MCMC algorithm used to simulate the sequence of $s^{n(k)}, \mu^{(k)}, \alpha^{(k)}, \gamma^{(k)}, \sigma^{2(k)}$ are the following.

[step 1] The sequence of hidden states $s^{n(k)}$ is still generated by the *ff-bs* algorithm.

[step 2] The m parameters $\mu_i^{(k)}$, for any i , are independently generated from a Normal distribution with mean

$$\frac{\sigma_i^{-2(k-1)} \sum_{\{t \geq 1: s_t^{(k)} = i\}} y_t + \mu_M \sigma_M^{-2}}{N_i^{(k)} \sigma_i^{-2(k-1)} + \sigma_M^{-2}}$$

and variance

$$\left(N_i^{(k)} \sigma_i^{-2(k-1)} + \sigma_M^{-2} \right)^{-1},$$

with the entries of the vector $\mu^{(k)}$ in increasing order to satisfy the identifiability constraint, so we apply the constrained permutation sampling algorithm (Frühwirth-Schnatter (2001)) to order them.

[step 3] Any vector $\alpha_{i,j}^{(k)}$, for any i, j , with $i \neq j$, is independently generated from a random walk $\alpha_{i,j}^{(k)} = \alpha_{i,j}^{(k-1)} + U_A$, where U_A is a multivariate Gaussian noise with zero mean and constant

precision matrix, for any k . Then any matrix $\alpha_i = (\alpha'_{i,1}, \dots, \alpha'_{i,m})'$ is accepted with probability

$$A\left(\alpha_i^{(k-1)}; \alpha_i^{(k)}\right) = \min \left\{ 1; \frac{\pi\left(\alpha_i^{(k)} \mid s^{n(k)}, \gamma_{(i)}^{(k-1)}, X\right)}{\pi\left(\alpha_i^{(k-1)} \mid s^{n(k)}, \gamma_{(i)}^{(k-1)}, X\right)} \right\},$$

where the numerator of the acceptance ratio is proportional to

$$\prod_{\{t \geq 2: s_{t-1}^{(k)} = i\}} p_{s_{t-1}^{(k)}, s_t^{(k)}}^{t(k)} \exp \left\{ -\frac{1}{2} \sum_{j \neq i} \left(\theta_{(i,j)}^{(k)} - \mu_A \right)' \Sigma_A^{-1} \left(\theta_{(i,j)}^{(k)} - \mu_A \right) \right\}$$

and the denominator is identical to the numerator except the introduction of $\alpha_i^{(k-1)}$ and $p_{i,j}^{t(k-1)}$, with $p_{i,j}^{t(k)}$ and $p_{i,j}^{t(k-1)}$ functions of $\alpha_i^{(k)}$ and $\alpha_i^{(k-1)}$, respectively, and $X = (X'_1, \dots, X'_t, \dots, X'_n)'$.

[step 4] Any coefficient $\gamma_{h(i)}^{(k)}$ ($h = 1, \dots, q; i = 1, \dots, m$) is independently generated from a Bernoulli distribution $Be\left(a_{j(i)}^{(k)} / (a_{j(i)}^{(k)} + b_{j(i)}^{(k)})\right)$, with

$$a_{h(i)}^{(k)} \propto \prod_{\{t \geq 2: s_{t-1}^{(k)} = i\}} \frac{\exp\left(X'_t \theta_{(i,s_t^{(k)})}^{*(k)}\right)}{1 + \sum_{j \neq i} \exp\left(X'_t \theta_{(i,j)}^{*(k)}\right)} \xi_h$$

and

$$b_{h(i)}^{(k)} \propto \prod_{\{t \geq 2: s_{t-1}^{(k)} = i\}} \frac{\exp\left(X'_t \theta_{(i,s_t^{(k)})}^{** (k)}\right)}{1 + \sum_{j \neq i} \exp\left(X'_t \theta_{(i,j)}^{** (k)}\right)} (1 - \xi_h),$$

where $\theta_{(i,j)}^{*(k)}$ is the vector $\theta_{(i,j)}^{(k)}$ in which the h -th entry is replaced by $\alpha_{h(i,j)}^{(k)}$, i.e. $\theta_{(i,j)}^{*(k)} = \left(\alpha_{0(i,j)}^{(k)}, \dots, \gamma_{h-1(i)}^{(k)} \alpha_{h-1(i,j)}^{(k)}, \alpha_{h(i,j)}^{(k)}, \gamma_{h+1(i)}^{(k-1)} \alpha_{h+1(i,j)}^{(k)}, \dots, \gamma_{q(i)}^{(k-1)} \alpha_{q(i,j)}^{(k)}\right)'$, while in $\theta_{(i,j)}^{** (k)}$ the h -th entry of $\theta_{(i,j)}^{(k)}$ is replaced by 0, i.e. $\theta_{(i,j)}^{** (k)} = \left(\alpha_{0(i,j)}^{(k)}, \dots, \gamma_{h-1(i)}^{(k)} \alpha_{h-1(i,j)}^{(k)}, 0, \gamma_{h+1(i)}^{(k-1)} \alpha_{h+1(i,j)}^{(k)}, \dots, \gamma_{q(i)}^{(k-1)} \alpha_{q(i,j)}^{(k)}\right)'$.

[step 5] The m parameters $\sigma_i^{2(k)}$ are generated from independent Inverse Gamma distributions with parameters

$$\frac{N_i^{(k)} + \nu}{2} \quad \text{and} \quad \frac{1}{2} \sum_{\{t \geq 1: s_t^{(k)} = i\}} \left(y_t - \mu_i^{(k)} \right)^2 + \eta$$

2.2.2 SSVS method

In the SSVS method, the indicator variable is involved in the model through the conditional priors of the coefficients $\alpha_{i,j}$ that are mixtures of two Normals with different variances: i.e. any vector $\alpha_{i,j}$, given $\gamma_{(i)}$, is an independent multivariate Normal of dimension $(q + 1)$

$$\alpha_{i,j} | \gamma_{(i)} \sim \mathcal{N}_{q+1}(\mu_A, D_{\gamma_{(i)}}),$$

where $D_{\gamma(i)} = \text{diag}[1, (\delta_{1(i)}\tau_1)^2, \dots, (\delta_{q(i)}\tau_q)^2]$ with $\delta_{h(i)} = c_h$ if $\gamma_{h(i)} = 1$ and $\delta_{h(i)} = 1$ if $\gamma_{h(i)} = 0$, for specified c_h and τ_h .

The prior distributions of the remaining parameters are the same of those elicited for the KM method; the expression of the logit transformation of the $p_{i,j}^t$'s is still the (2).

In the SSVS scheme used to generate the sequence $s^{n(k)}, \mu^{(k)}, \sigma^{2(k)}, \alpha^{(k)}, \gamma^{(k)}$, steps [2] and [5] remain the same as for KM, since they do not depend on $\alpha_{(i)}$ and $\gamma_{(i)}$, steps [3] and [4] change totally and step [1] is identical to that in KM except the presence of the $\gamma_{(i)}$'s that multiply the $\alpha_{i,j}$'s.

[step 3] Any vector $\alpha_{i,j}^{(k)}$, for any i, j , with $i \neq j$, is independently generated from a multivariate random walk $\alpha_{i,j}^{(k)} = \alpha_{i,j}^{(k-1)} + U_A$. Then any matrix $\alpha_i^{(k)} = \left(\alpha_{i,1}^{(k)}, \dots, \alpha_{i,m}^{(k)} \right)'$ is accepted with probability

$$A \left(\alpha_i^{(k-1)}; \alpha_i^{(k)} \right) = \min \left\{ 1; \frac{\pi \left(\alpha_i^{(k)} \mid s^{n(k)}, X \right)}{\pi \left(\alpha_i^{(k-1)} \mid s^{n(k)}, X \right)} \right\},$$

where the numerator of the acceptance ratio is proportional to

$$\prod_{\{t \geq 2: s_{t-1}^{(k)} = i\}} p_{s_{t-1}^{(k)}, s_t^{(k)}}^{t(k)} \cdot \exp \left\{ -\frac{1}{2} \sum_{j \neq i} \left(\alpha_{i,j}^{(k)} - \mu_A \right)' D_{\gamma(i)}^{(k)-1} \left(\alpha_{i,j}^{(k)} - \mu_A \right) \right\}$$

and the denominator is identical to the numerator except the introduction of $\alpha_i^{(k-1)}$ and $p_{i,j}^{t(k-1)}$, with $p_{i,j}^{t(k)}$ and $p_{i,j}^{t(k-1)}$ functions of $\alpha_i^{(k)}$ and $\alpha_i^{(k-1)}$, respectively, and $X = (X'_1, \dots, X'_t, \dots, X'_n)'$.

[step 4] For every hidden state i , any coefficient $\gamma_{h(i)}^{(k)}$ is independently generated, for $h = 1, \dots, q$, from a Bernoulli distribution $Be \left(a_{j(i)}^{(k)} / \left(a_{j(i)}^{(k)} + b_{j(i)}^{(k)} \right) \right)$, with

$$a_{h(i)}^{(k)} \propto f \left(\alpha_i^{(k)} \mid \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 1 \right) \xi_h \quad \text{and} \quad b_{h(i)}^{(k)} \propto f \left(\alpha_i^{(k)} \mid \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 0 \right) (1 - \xi_h),$$

where $f \left(\alpha_i^{(k)} \mid \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 1 \right)$ or $f \left(\alpha_i^{(k)} \mid \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 0 \right)$ are the density functions of the $(q+1)$ -variate Normal prior distribution of $\alpha_{i,j} \mid \gamma_{(i)}$ when $\gamma_{h(i)} = 1$ or $\gamma_{h(i)} = 0$, respectively, that is

$$f \left(\alpha_i^{(k)} \mid \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 1 \right) \propto \exp \left\{ -\frac{1}{2} \sum_{j \neq i} \left(\alpha_{i,j}^{(k)} - \mu_A \right)' D_{\gamma(i)}^{*(k)-1} \left(\alpha_{i,j}^{(k)} - \mu_A \right) \right\}$$

and

$$f \left(\alpha_i^{(k)} \mid \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 0 \right) \propto \exp \left\{ -\frac{1}{2} \sum_{j \neq i} \left(\alpha_{i,j}^{(k)} - \mu_A \right)' D_{\gamma(i)}^{** (k)-1} \left(\alpha_{i,j}^{(k)} - \mu_A \right) \right\},$$

where $D_{\gamma^{(i)}}^{*(k)}$ and $D_{\gamma^{(i)}}^{***(k)}$ are the covariance matrix of the prior of $\alpha_{i,j}^{(k)}$ in which the h -th entries $\delta_{h(i)}$ are replaced by c_h (which corresponds to $\gamma_{h(i)}^{(k)} = 1$), and by 1 (which corresponds to $\gamma_{h(i)}^{(k)} = 0$), respectively.

2.2.3 GVS method

As in the previous sections, the GVS method is a mixed approach between the two previous methods: the indicator variable influences both the prior distributions of the coefficients of the exogenous variables, as in SSVS, and the equation of the model (3), as in KM.

The prior distributions are the same of those in SSVS; in the MCMC scheme, steps [1], [2], [5] are identical to those in KM, step [3] is identical to that in SSVS, while step [4] is a combination of KM and SSVS. In details, the elements of the parameters of the Bernoulli distributions of any $\gamma_{j(i)}^{(k)}$ are

$$a_{h(i)}^{(k)} \propto f\left(s^n | \alpha_i^{(k)}, X, \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 1\right) f\left(\alpha_i^{(k)} | \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 1\right) \xi_h$$

and

$$b_{h(i)}^{(k)} \propto f\left(s^n | \alpha_i^{(k)}, X, \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 0\right) f\left(\alpha_i^{(k)} | \gamma_{\setminus h(i)}^{(k-1)}, \gamma_{h(i)}^{(k)} = 0\right) (1 - \xi_h).$$

Notice that both the densities $f(s^n | \cdot)$, as in KM, and the conditional priors $f(\alpha_i^{(k)} | \cdot)$, as in SSVS, are present in the expressions of $a_{h(i)}^{(k)}$ and $b_{h(i)}^{(k)}$.

2.2.4 MKMK method

For NHHMM the pair of vectors $(\alpha_i^{(k)}; \gamma_{(i)}^{(k)})$ must be accepted in block, so steps [3] and [4] of Subsection 2.2.1 are joined in only one step and modified as follows. We assume the following proposals: a multivariate random walk for any $\alpha_{i,j}^{(k)}$ and an independent Bernoulli for any entry of the vector $\gamma_{(i)}^{(k)}$.

For any i , first, parameters $\alpha_{i,j}^{(k)}$ ($j = 1, \dots, m$, with $i \neq j$) are independently generated from the multivariate random walk $\alpha_{i,j}^{(k)} = \alpha_{i,j}^{(k-1)} + U_A$, then, every parameter $\gamma_{h(i)}^{(k)}$, for any $h = 1, \dots, q$, is independently generated from a Bernoulli distribution with probability $\pi_{h(i)}^{(k)} = a_{h(i)}^{(k)} / (a_{h(i)}^{(k)} + b_{h(i)}^{(k)})$, where $a_{h(i)}^{(k)}$ and $b_{h(i)}^{(k)}$ are the same as in step [4] of Subsection 2.2.1; finally

the pair of vectors $(\alpha_i^{(k)}; \gamma_{(i)}^{(k)})$ is accepted in block with probability

$$A \left[(\alpha_i^{(k-1)}; \gamma_{(i)}^{(k-1)}); (\alpha_i^{(k)}; \gamma_{(i)}^{(k)}) \right] = \min \left\{ 1; \frac{\pi(\alpha_i^{(k)}; \gamma_{(i)}^{(k)}) q(\alpha_i^{(k-1)}; \gamma_{(i)}^{(k-1)} | \alpha_i^{(k)}; \gamma_{(i)}^{(k)})}{\pi(\alpha_i^{(k-1)}; \gamma_{(i)}^{(k-1)}) q(\alpha_i^{(k)}; \gamma_{(i)}^{(k)} | \alpha_i^{(k-1)}; \gamma_{(i)}^{(k-1)})} \right\} =$$

$$= \min \left\{ 1; \frac{f(y^n | \alpha_i^{(k)}; \gamma_{(i)}^{(k)}) p(\gamma_{(i)}^{(k)}) p(\alpha_i^{(k)}) q(\gamma_{(i)}^{(k-1)} | \alpha_i^{(k-1)}; \alpha_i^{(k)}; \gamma_{(i)}^{(k)})}{f(y^n | \alpha_i^{(k-1)}; \gamma_{(i)}^{(k-1)}) p(\gamma_{(i)}^{(k-1)}) p(\alpha_i^{(k-1)}) q(\gamma_{(i)}^{(k)} | \alpha_i^{(k)}; \alpha_i^{(k-1)}; \gamma_{(i)}^{(k-1)})} \right\}$$

by the factorization of the proposal density, by the independence of $\alpha_i^{(k)}$ on $\gamma_{(i)}^{(k-1)}$ and cancelling the ratio $q(\alpha_i^{(k-1)} | \alpha_i^{(k)}) / q(\alpha_i^{(k)} | \alpha_i^{(k-1)})$ for the symmetry of the proposal distribution.

3 Simulation examples

3.1 MSARX models

We generate time series of length $n = 500$, for each pair (m, p) with $m = 2, 3$ and $p = 1, 2, 3$ and different coefficient vectors β_i and γ_i for any states $i = 1, \dots, m$. The data generation details are the following: we obtain the exogenous variables X_1, \dots, X_5 as a multivariate Normal, with independent components or with low or high correlations between them. Specifically, for the case of correlated variables, we set to non zero the correlations between variables (X_2, X_4) and (X_1, X_5) , while the correlations between the other variables are random values in modulo less then the previous ones. With these values we explore several cases of correlation between variables which are included in or excluded from the true model. For any state $i = 1, \dots, m$ and $\tau = 1, \dots, p$, the values $r_{\tau(i)}$'s are randomly generated from uniform distributions between -1 and $+1$ and then transformed to obtain the autoregressive coefficients $\varphi_{\tau(i)}$'s. The remaining parameters are set as follows:

- for $m = 2$, we suppose that the true model contains the covariates X_4 and X_5 in state 1, and X_2, X_3, X_4 in state 2. The corresponding coefficients and the values of the other parameters are:

state	μ_i	σ_i^2	$\beta_{j(i)}$					$\gamma_{j(i)}$					P	
1	1	1	0	0	0	1	1.2	0	0	0	1	1	0.6	0.4
2	3	0.8	0	1	0.8	1	0	0	1	1	1	0	0.2	0.8

- for $m = 3$, we suppose that in states 1 and 2 there are the same covariates of the case $m = 2$, while in state 3 the covariates in the model are X_2, X_4 . The corresponding coefficients and the

true values of the parameters are:

state	μ_i	σ_i^2	$\beta_{j(i)}$					$\gamma_{j(i)}$				P			
1	1	1	0	0	0	1	1.2	0	0	0	1	1	0.6	0.3	0.1
2	3	0.8	0	1	0.8	1	0	0	1	1	1	0	0.05	0.8	0.15
3	5	1.25	0	1.2	0	0.8	0	0	1	0	1	0	0.1	0.2	0.7

Notice that the signals μ_i are in increasing order to ensure the identifiability.

For all the analysed models, we use independent $\mathcal{N}(0, 10)$ priors for any parameter μ_i and $R_{\tau(i)}$ (for any $i = 1, \dots, m$, and $\tau = 1, \dots, p$); independent $IG(0.5, 0.5)$ for any variances σ_i^2 and indifference priors for any $\gamma_{j(i)}$ with $\xi_j = 0.5$. The priors of any row of the transition matrix $P_{i\bullet}$ are independent Dirichlet with parameter $\omega = (\omega_1, \dots, \omega_k, \dots, \omega_m)$, and $\omega_k = m \cdot I(i = k) + 0.6 \cdot I(i \neq k)$, for any $i, k = 1, \dots, m$, where $I(\cdot)$ is the indicator function. By this prior the probability of persistence is greater than the probability of transition: the probability of persistence is about 0.7 and it slowly decreases as the number of states increases. Finally, in KM the priors of any $\beta_{j(i)}$ are independent $\mathcal{N}(0, 10)$, for $i = 1, \dots, m$ and $j = 1, \dots, q$, while in SSVS and GVS the conditional priors of $\beta_i | \gamma_i$ are $\mathcal{N}_5(0, D_{\gamma(i)})$, where $D_{\gamma(i)}$ is defined through the choice of the tuning parameters c_j and τ_j : by the results of several simulations of the full models and following the suggestions by George and McCulloch (1993), we use $c_j = 10$ and $\tau_j = 0.3$, for any $j = 1, \dots, q$.

Tables 1, 2, 3, 4 show the results of the posterior model probabilities, estimated through the relative frequencies, of the highest frequency models, for all the variable selection methods we analysed, both in the case of independent covariates and in some special cases of correlation (i.e. 0.3, 0.7, 0.9). The results are based on 300,000 iterations for all the algorithms, after a burn-in period of 100,000 iterations. All the Markov chain are initialized at the full model, while the starting points of the parameters are generated at random from the priors.

TABLES 1, 2, 3, 4 ABOUT HERE

The results show that in general all the four methods predict the right MSARX model, choosing that model which includes the right covariates. In the case of independent covariates the model probability is high for each pair of m and p , while as the correlation between covariates increases the

1
2
3
4
5 model probability decreases till to assume very small values (i.e. 7%, for the case of MSARX(2,1),
6
7 in Table 4, for the SSVS method).

8
9 It can be pointed out also that all the sampling schemes explore local modes of the posterior
10
11 distribution because different results are obtained in different runs of the algorithms, although all
12
13 runs are able to identify the true model.

14
15 It can be seen that SSVS generally performs worse than the other methods, but reasonably
16
17 well, and only in the case of independent covariates the model probabilities are higher than 60%,
18
19 while in the case of high correlation they are always less than 30%.

20
21 On the other hand, GVS seems to be the best method, since in all the analysed cases, the true
22
23 model is identified with probability greater than 60%, sometimes exceeding the value 90%, even
24
25 if the choice of the tuning factors can represent a serious drawback of this method, as in SSVS.
26
27 Finally, the performances of KM and MKMK are almost equivalent and their model probabilities,
28
29 in some cases, tend to the values reached by GVS.

30
31 In Figure 1 the MCMC evolution of the corresponding ergodic posterior probabilities are plotted
32
33 in some special cases, at every 50-*th* iteration after the burn-in. **We dealt with the convergence**
34
35 **of the MCMC algorithms by using many plots of output values and formal convergence**
36
37 **tests (i.e. Gamerman (1997)).**

38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
FIGURE 1 ABOUT HERE

We compared also the performances of all methods in problems with more data, for example
with $n = 1000$, or with a large number of potential covariates, $q = 30$ or 50 . As we would expect,
the results are similar to those presented here and, again, GVS seems to perform better than other
methods. Moreover, as the number of data increases the values of the highest posterior model
probability increases and the convergence is reached more fastly.

3.2 NHHMM

The simulation examples for evaluating the performance of the previous methods are based on
time series of length $n = 500$, generated from NHHMM(m), for each values of $m = 2, 3$, with

$q = 5$ potential covariates and different coefficients $\alpha_{h(i,j)}$ of the logistic transformation in (2), for $i = 1, \dots, m; j = 1, \dots, m$ and $h = 0, 1, \dots, q$.

The exogenous variables X_1, \dots, X_5 are obtained again as multivariate Normal; for the cases of correlated covariates we suppose that the pairs of variables (X_2, X_4) and (X_1, X_5) have a non-zero fixed correlation, while the correlations between other variables are zero or random values, in modulo less than the previous ones. The parameters values are taken as follows:

- for $m = 2$, we suppose that the true model contains covariates X_2, X_3 , in state 1 and X_2, X_3, X_4 , in state 2; the values of the parameters are:

state	μ_i	σ_i^2	$\alpha_{h(i,j)}$						$\gamma_{h(i)}$						
1	1	1	$\alpha_{(1,2)}$	2	0	-2	2	0	0	1	0	1	1	0	0
2	3	0.8	$\alpha_{(2,1)}$	2	0	-2	2	2	0	1	0	1	1	1	0

- for $m = 3$, we suppose in states 1 and 2 the model contains the same covariates as in case $m = 2$, while in the state 3 the covariates in the model are X_1 and X_3 ; the values of the parameters are:

state	μ_i	σ_i^2	$\alpha_{h(i,j)}$						$\gamma_{h(i)}$						
1	1	1	$\alpha_{(1,2)}$	2	0	-2	2	0	0	1	0	1	1	0	0
2	3	0.8	$\alpha_{(1,3)}$	0.5	0	1	0.5	0	0	1	0	1	1	0	0
3	5	1.25	$\alpha_{(2,1)}$	2	0	-2	2	2	0	1	0	1	1	1	0
			$\alpha_{(2,3)}$	1	0	0.5	-0.5	1	0	1	0	1	1	1	0
			$\alpha_{(3,1)}$	2	-2	0	2	0	0	1	1	0	1	0	0
			$\alpha_{(3,2)}$	2	-2	0	2	0	0	1	1	0	1	0	0

We use independent $\mathcal{N}(0, 10)$ priors for any parameter μ_i ; independent $IG(0.5, 0.5)$ for any variance σ_i^2 and indifference priors for any $\gamma_{h(i)}$, with $\xi_h = 0.5$, for any $h = 1, \dots, q$. For KM the prior of any vector $\alpha_{i,j}$ (for any $i = 1, \dots, m$, and $j = 1, \dots, m$) is a multivariate Normal $N_6(0, \Sigma_A)$, with $\Sigma_A = 10 * I$ (where I is the identity matrix), while, for SSVS and GVS, the conditional priors of vectors $\alpha_{i,j} | \gamma_{h(i)}$ are $\mathcal{N}_6(0, D_{\gamma(i)})$, where $D_{\gamma(i)}$ is defined through the choice of the tuning parameters c_h and τ_h , which are still set $c_h = 10$ and $\tau_h = 0.3$, for any $i = 1, \dots, m$ and $h = 1, \dots, 5$. Again, all the Markov chain are initialized at the full model and the starting values of the parameters are generated at random from the priors.

1
2
3
4
5 In Table 5 the highest posterior model probabilities are displayed, both in the case of independent
6 covariates and in some special cases of correlation (i.e. 0.3, 0.7, 0.9). The results are based on
7
8 300,000 iterations, after a burn-in period of 100,000 iterations.
9

10
11 TABLE 5 ABOUT HERE
12

13
14 For KM, SSVS, GVS, the results are very similar to those described for the MSARX models,
15 even if the values of the posterior model probabilities are much smaller: this means that the
16 performance of the methods decreases when the complexity of the model increases ($m > 2$),
17 especially also when exogenous variables are strongly correlated. The algorithms in fact visit
18 a larger set of promising models that incorrectly include other variables or exclude the correct
19 variables: we can see that, in the case of independence, the true models have posterior model
20 probabilities less than 30% or, when high correlation exists (0.7 or 0.9), less than 10%. In these
21 last cases (marked with (*) in Table 5) the algorithms do not identify the correct model and the
22 associated probabilities are not the highest.
23
24

25
26 By contrast, MKMK works extremely well and always identify the right set of covariates with
27 high probabilities, both when the correlations we fixed are null and when they are positive.
28

29
30 In Figure 2 the MCMC evolution of the corresponding ergodic posterior probabilities are plotted
31 in some special cases, at every 50-th iteration after the burn-in. **The convergence of the sam-**
32 **plers has been checked through plots and traditional convergence tests (i.e. Gaman-**
33 **man (1997)).**
34
35

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
FIGURE 2 ABOUT HERE

The comparison among the results of Table 6 and those of the previous tables shows that our method is very satisfactory: the values of the posterior model probabilities are always very high, also in the critical cases of NHHMM.

4 Conclusions

Bayesian variable selection is an active research area in recent years. In this paper we developed some alternative Bayesian procedures, based on Markov chain Monte Carlo algorithms, for variable

1
2
3
4
5 selection in special Markov mixture models, i.e. the Markov switching autoregressive models with
6
7 exogenous variables (MSARX) and the non-homogeneous hidden Markov models (NHHMM). We
8
9 focused our attention on three methods based on the analysis of the highest model posterior
10
11 probability: the Stochastic Search Variable Selection (SSVS) by George and McCulloch (1993),
12
13 the unconditional priors Gibbs sampler (KM) by Kuo and Mallick (1998) and the Gibbs Variable
14
15 selection (GVS) by Dellaportas et al. (2000). These methods have been initially proposed for
16
17 linear models and so we decided to explore their performances when they are applied to highly
18
19 complex stochastic processes.

20
21 Simulation results show that GVS selects the true model with an excellent relative frequency,
22
23 both when the exogenous variables are independent and when they are correlated. By contrast,
24
25 SSVS and KM visit the right models with a much lower probability and sometimes they are not
26
27 able to choose the best ones.

28
29 The selection procedures of SSVS and GVS provides the employment of some tuning factors
30
31 which, in our modelling, are hard to identify; hence, our goal was the improvement of the perfor-
32
33 mances of KM, which is extremely straightforward, since it requires only to specify the prior and
34
35 no pilot runs are required to define the hyperparameters.

36
37 Our method is based on the scheme of KM, which is improved by accepting in block, through a
38
39 Metropolis step, both the coefficients and the indicators of the exogenous variables; so we decided
40
41 to call it Metropolized-Kuo-MallicK (MKMK). MKMK works very well and better than the other
42
43 methods, especially in the NHHMM case, because always selects the right model with a high model
44
45 posterior probability. The correct selection of exogenous variables is due to the acceptance in block
46
47 of both the coefficients and the indicators of the exogenous variables, because blocking increases
48
49 the precision of the estimators and improves the chain mixing.

50
51 In Paroli and Spezia (2005), we successfully applied MKMK to a real data problem in the context
52
53 of non-homogeneous Markov mixtures of periodic autoregressions, that is MSARX models with
54
55 a Markov-dependent periodic component, all driven by NHHMM. We observed a series of hourly
56
57 mean concentrations of sulphur dioxide simultaneously to six meteorological variables (wind speed,
58
59 temperature, rain, solar radiation, relative humidity, pressure) and MKMK allows us to select the
60

1
2
3
4 exogenous variables that, for any state of the hidden Markov chain, can influence the dynamics of
5
6 the observed process and/or the latent one.
7
8
9

10 11 References

12
13 Brown, P.J. Vannucci, M., Fearn, T. (1998) Multivariate Bayesian variable selection and pre-
14 diction. *Journal of the Royal Statistical Society, Series B*, 60, 627-641.

15
16 Carlin, B.P. and Chib, S. (1995) Bayesian model choice via Markov chain Monte Carlo. *Journal*
17 *of the Royal Statistical Society, Series B*, 57, 473-484.

18
19 Celeux G., Hurn M., Robert C.P. (2000) Computational and Inferential Difficulties With Mix-
20 ture Posterior Distributions. *Journal of the American Statistical Association*, 95, 957-970.

21
22 Chib, S. (1996) Calculating posterior distributions and modal estimates in Markov mixture
23 models. *Journal of Econometrics*, 75, 79-97.

24
25 Dellaportas, P., Forster, J.J., Ntzoufras, I. (2000) Bayesian Variable Selection Using the Gibbs
26 Sampler. In *Generalized Linear Models: A Bayesian Perspective* (eds D.K. Dey, S.K. Ghosh, B.K.
27 Mallick), 273-286. Marcel Dekker, New York.

28
29 Denison, D.G.T., Mallick, B.K., Smith, A.F.M. (1998) Automatic Bayesian curve fitting. *Jour-*
30 *nal of the Royal Statistical Society, Series B*, 60, 333-350.

31
32 Elliott R.J., Aggoun L., Moore J.B. (1995) *Hidden Markov Models: Estimation and Control*.
33 Springer, New York.

34
35 Franses P.H. and van Dijk D. (2000) *Nonlinear Time Series Models in Empirical Finance*.
36 Cambridge University Press, Cambridge.

37
38 Frühwirth-Schnatter, S. (2001) Markov Chain Monte Carlo Estimation of Classical and Dynamic
39 Switching and Mixture Models. *Journal of the American Statistical Association*, 96, 194-209.

40
41 Gamerman, D. (1997) *Markov Chain Monte Carlo: Stochastic simulation for Bayesian infer-*
42 *ence*. Chapman & Hall, London.

43
44 George, E.I. and McCulloch, R.E. (1993) Variables selection via Gibbs-Sampling. *Journal of*
45 *the American Statistical Association*, 88, 881-889.

46
47 George, E.I., McCulloch, R.E., Tsay, R. (1996) Two Approaches to Bayesian Model Selection
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 with Applications. In *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of*
5
6 *Arnold Zellner* (eds D.A. Berry, K.M. Chaloner, J.K. Geweke), 339-348. John Wiley & Sons, New
7
8 York.
9

10
11 Gilks W.R. and Roberts G.O. (1996) Strategies for improving MCMC. In *Markov Chain Monte*
12
13 *Carlo in Practice* (eds W.R. Gilks, S. Richardson, D.J. Spiegelhalter), 89-114. Chapman & Hall,
14
15 London.

16
17 Hamilton J.D. (1993) Estimation, Inference and Forecasting of Time Series Subject to Changes
18
19 in Regime. In *Handbook of Statistics, vol. 11* (eds G.S. Maddala, C.R. Rao, H.D. Vinod), 231-259.
20
21 North-Holland, Amsterdam.

22
23 Kim C.J. and Nelson C.R. (1999) *State-Space Models with Regime Switching: Classical and*
24
25 *Gibbs-Sampling Approaches with Applications*. The MIT Press, Cambridge.

26
27 Kohn, R., Smith, M., Chan, D. (2001) Nonparametric regression using linear combinations of
28
29 basis functions. *Statistics and Computing*, 11, 313-322.

30
31 Koski T. (2001) *Hidden Markov Models for Bioinformatics*. Kluwer Academic Publishers, Dor-
32
33 drecht.

34
35 Krolzig H.-M. (1997) *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference*
36
37 *and Applications to Business Cycle Analysis*. Springer, Berlin.

38
39 Kuo, L. and Mallick, B. (1998) Variable Selection for Regression Models. *Sankhya, Series B*,
40
41 60, 65-81.

42
43 Liu J.S., Wong W.H., Kong A. (1994) Covariance structure of the Gibbs sampler with applica-
44
45 tions to the comparisons of the estimators and augmentation scheme. *Biometrika*, 81, 27-40.

46
47 MacDonald I.L. and Zucchini W. (1997) *Hidden Markov and Other Models for Discrete-valued*
48
49 *Time Series*. Chapman & Hall, London.

50
51 Mariott, J., Ravishanker, N., Gelfand, A., Pai, J. (1996) Bayesian Analysis of ARMA Processes:
52
53 Complete Sampling-Based Inference under Exact Likelihoods. In *Bayesian Analysis in Statistics*
54
55 *and Econometrics: Essays in Honor of Arnold Zellner* (eds D.A. Berry, K.M. Chaloner, J.K.
56
57 Geweke), 243-256. John Wiley & Sons, New York.

58
59
60 Nott, D.J. and Green, P.J. (2004) Bayesian variable selection and the Swendsen-Wang Algo-

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

rithm. *Journal of Computational and Graphical Statistics*, 13, 1-17.

Nott, D.J. and Kohn, R. (2005) Adaptive sampling for Bayesian variable selection. *Biometrika*, 92, 747-763.

Ntzoufras, I., Forster, J.J., Dellaportas, P. (2000) Stochastic Search Variable Selection for Log-linear Models. *Journal of Statistical Computation and Simulation*, 68, 23-37.

Paroli, R. and Spezia, L. (2005) *Non-Homogeneous Markov Mixture of Periodic Autoregressions and Sulphur Dioxide Concentrations*. Serie E.P. n. 126, Istituto di Statistica, Università Cattolica S.C., Milano.

Schneider, U. and Corcoran, J.N. (2004) Perfect Sampling for Bayesian Variable Selection in a Linear Regression Model. *Journal of Statistical Planning and Inference*, 126, 153-171.

		independent						
		m=2		m=3				
		state	1	2	state	1	2	3
		model var.	4,5	2,3,4	model var.	4,5	2,3,4	2,4
p=1	KM		0.71	0.89	KM	0.61	0.83	0.48
	SSVS		0.53	0.42	SSVS	0.42	0.35	0.28
	GVS		0.92	0.91	GVS	0.91	0.94	0.85
	MKMK		0.59	0.84	MKMK	0.52	0.78	0.64
p=2	KM		0.60	0.78	KM	0.51	0.37	0.75
	SSVS		0.56	0.28	SSVS	0.42	0.35	0.36
	GVS		0.92	0.95	GVS	0.63	0.95	0.91
	MKMK		0.52	0.83	MKMK	0.42	0.82	0.52
p=3	KM		0.72	0.84	KM	0.65	0.86	0.65
	SSVS		0.64	0.42	SSVS	0.63	0.29	0.30
	GVS		0.92	0.95	GVS	0.77	0.94	0.80
	MKMK		0.74	0.72	MKMK	0.79	0.59	0.68

Table 1: Highest posterior model probabilities for independent covariates

		corr=0.3						
		m=2		m=3				
		state	1	2	state	1	2	3
		model var.	4,5	2,3,4	model var.	4,5	2,3,4	2,4
p=1	KM		0.61	0.76	KM	0.66	0.73	0.60
	SSVS		0.53	0.35	SSVS	0.55	0.46	0.24
	GVS		0.92	0.95	GVS	0.77	0.94	0.75
	MKMK		0.78	0.82	MKMK	0.46	0.73	0.61
p=2	KM		0.62	0.77	KM	0.75	0.78	0.54
	SSVS		0.53	0.44	SSVS	0.33	0.38	0.20
	GVS		0.89	0.94	GVS	0.85	0.93	0.91
	MKMK		0.65	0.60	MKMK	0.33	0.72	0.55
p=3	KM		0.75	0.87	KM	0.60	0.64	0.64
	SSVS		0.52	0.39	SSVS	0.21	0.20	0.22
	GVS		0.66	0.92	GVS	0.83	0.93	0.83
	MKMK		0.60	0.75	MKMK	0.54	0.81	0.55

Table 2: Highest posterior model probabilities for correlated covariates: $\text{corr}(X_2, X_4) = \text{corr}(X_1, X_5) = 0.3$

corr=0.7							
m=2				m=3			
state	1	2		state	1	2	3
model var.	4,5	2,3,4		model var.	4,5	2,3,4	2,4
p=1	KM	0.68	0.74	KM	0.71	0.76	0.33
	SSVS	0.36	0.31	SSVS	0.37	0.11	0.25
	GVS	0.67	0.89	GVS	0.74	0.78	0.43
	MKMK	0.68	0.74	MKMK	0.13	0.80	0.56
p=2	KM	0.63	0.81	KM	0.36	0.60	0.70
	SSVS	0.45	0.30	SSVS	0.28	0.10	0.11
	GVS	0.77	0.88	GVS	0.86	0.87	0.83
	MKMK	0.60	0.85	MKMK	0.40	0.58	0.71
p=3	KM	0.41	0.59	KM	0.58	0.41	0.61
	SSVS	0.36	0.23	SSVS	0.26	0.07	0.18
	GVS	0.87	0.85	GVS	0.81	0.82	0.84
	MKMK	0.55	0.61	MKMK	0.61	0.62	0.60

Table 3: Highest posterior model probabilities for correlated covariates: $\text{corr}(X_2, X_4) = \text{corr}(X_1, X_5) = 0.7$

		corr=0.9						
		m=2		m=3				
		state	1	2	state	1	2	3
		model var.	4,5	2,3,4	model var.	4,5	2,3,4	2,4
p=1	KM		0.65	0.25	KM	0.32	0.33	0.42
	SSVS		0.18	0.07	SSVS	0.27	0.04	0.26
	GVS		0.70	0.91	GVS	0.86	0.80	0.70
	MKMK		0.52	0.45	MKMK	0.87	0.71	0.45
p=2	KM		0.60	0.21	KM	0.35	0.31	0.34
	SSVS		0.26	0.31	SSVS	0.24	0.19	0.20
	GVS		0.87	0.89	GVS	0.83	0.61	0.91
	MKMK		0.40	0.42	MKMK	0.19	0.38	0.27
p=3	KM		0.36	0.46	KM	0.37	0.41	0.50
	SSVS		0.24	0.18	SSVS	0.18	0.17	0.23
	GVS		0.68	0.69	GVS	0.85	0.75	0.89
	MKMK		0.43	0.35	MKMK	0.33	0.51	0.39

Table 4: Highest posterior model probabilities for correlated covariates: $\text{corr}(X_2, X_4) = \text{corr}(X_1, X_5) = 0.9$

		independent				corr=0.3			
	model var.	KM	SSVS	GVS	MKMK	KM	SSVS	GVS	MKMK
m=2									
state 1	2,3	0.25	0.05	0.11	0.75	0.46	0.05	0.16	0.71
state 2	2,3,4	0.19	0.06	0.27	0.85	0.30	0.04	0.26	0.88
m=3									
state 1	2,3	0.09	0.03	0.11	0.87	0.08	0.03	0.09	0.90
state 2	2,3,4	0.08	0.04	0.10	0.79	0.10	0.04	0.26	0.95
state 3	1,3	0.05	0.03	0.03	0.63	0.04	0.03	0.04	0.70
		corr=0.7				corr=0.9			
	model var.	KM	SSVS	GVS	MKMK	KM	SSVS	GVS	MKMK
m=2									
state 1	2,3	0.41	0.04	0.18	0.74	0.49	0.04	0.08	0.68
state 2	2,3,4	0.34	0.06	0.25	0.85	0.48	0.04	0.17	0.67
m=3									
state 1	2,3	0.07(*)	0.04(*)	0.13	0.64	0.02(*)	0.02(*)	0.05	0.27
state 2	2,3,4	0.08(*)	0.04(*)	0.12	0.44	0.01(*)	0.04(*)	0.09	0.81
state 3	1,3	0.05(*)	0.03(*)	0.10	0.70	0.03(*)	0.02(*)	0.02	0.20

Table 5: Highest posterior model probabilities for NHHMM(m) with independent or correlated covariates - (*) = not the highest probability

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

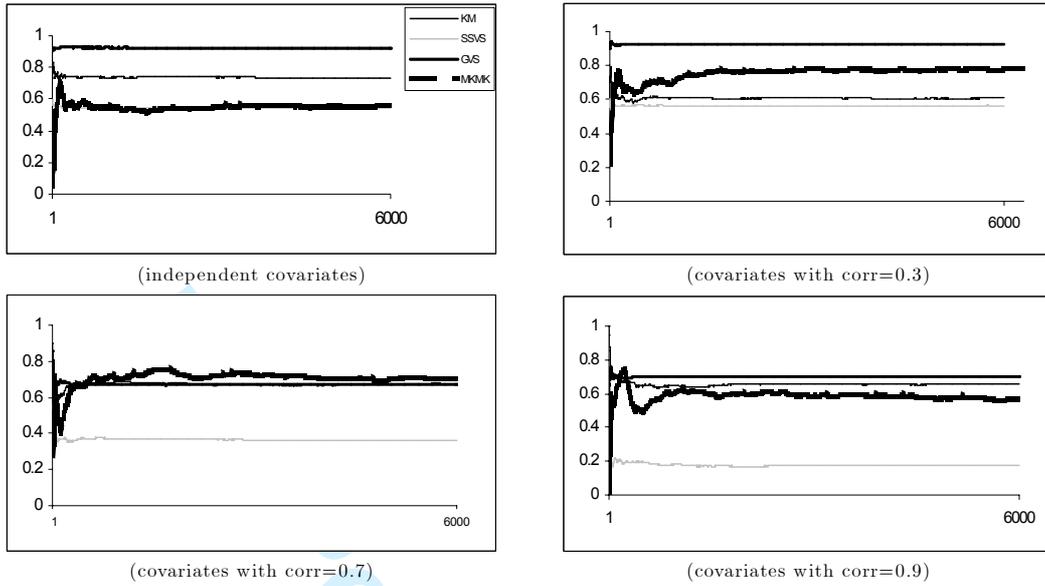


Figure 1: Ergodic highest posterior model probabilities for the simulated datasets of MSARX(2,1), in state 1, for KM (black line); SSVS (grey line); GVS (bold line); MKMK(double black line)

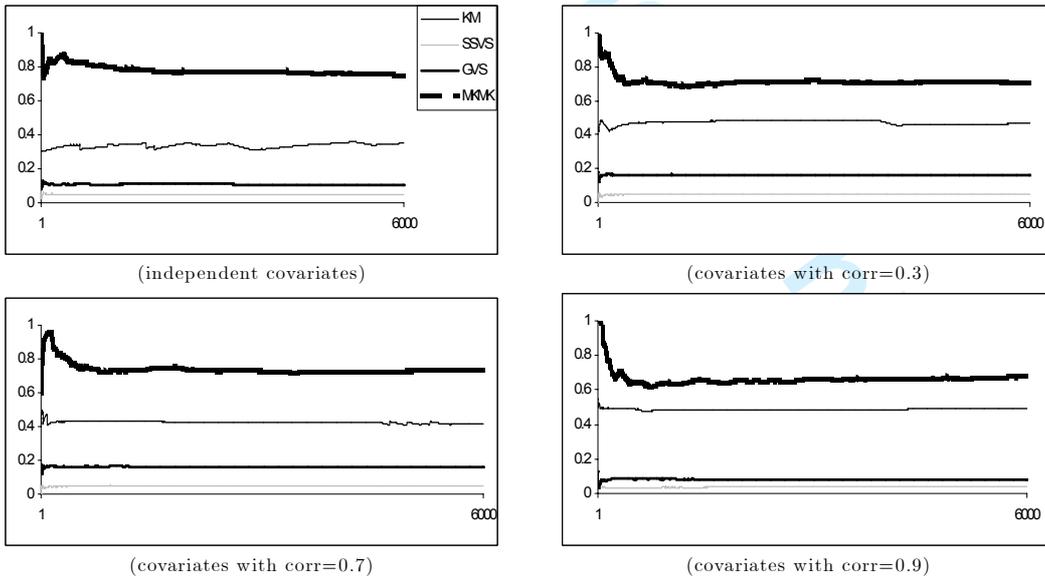


Figure 2: Ergodic highest posterior model probabilities for the simulated datasets of NHHMM(2), in state 1, for KM (black line); SSVS (grey line); GVS (bold line); MKMK (double black line)