Abstract - Markov switching autoregressive models (MSARMs) are efficient tools to analyse non linear and non Gaussian time series. A special MSARM with a harmonic component in the Bayesian framework is here proposed to analyse periodic time series. We present a complete Gibbs sampling algorithm for model choice (the selection of the autoregressive order and of the cardinality of the hidden Markov chain state-space), for constraint identification (the research of the identifiability constraints which respect the geometry and the shape of the posterior distribution) and for the estimation of the unknown parameters and the latent data. These three consecutive steps are developed tackling the problem of the hidden states labeling by means of random permutation sampling and constrained permutation sampling. We illustrate our methodology with two examples about the dynamics of air pollutants.

Keywords - Bayesian model choice and inference, Gibbs sampling, hidden Markov chain, label switching, random and constrained permutation sampling.
Introduction

Markov switching autoregressive models (MSARMs) make up a class of models for non linear time series that present different behaviours according to the dynamics of hidden (or latent) random variables, usually known as state variables or regime variables. The MSARMs assume the dynamics of the regime variables is described by an unobservable Markov chain and the dynamics of the observed time series is modelled by an autoregressive process, whose parameters depend on the state of the Markov chain. These models have been introduced in the econometric literature by Hamilton to study economic and financial time series with asymmetric cycles and changes in regime generated by a stochastic process (Hamilton (1989), (1990), (1993)). See also Krozlig (1997) and Franses and van Dijk (2000) for many applications and generalizations of this class of models. Bayesian analysis of MSARMs has been developed together with others by McCulloch and Tsay (1993), Chib (1996), Billio, Monfort, Robert (1999), Frühwirth-Schnatter (1999), (2001a).

As special cases of MSARMs we have linear autoregressive models (e.g. Hamilton (1994), ch. 3), when no Markov chain underlies, or Gaussian hidden Markov models (e.g. Robert, Rydén, Titterington (2000)), when the observed variables, given the states of the Markov chain, are conditionally independent, i.e. when we have an autoregressive process of order zero, or independent finite mixture models (e.g. Titterington, Smith, Makov (1985)), when all the rows of the transition matrix of the Markov chain are equal. Spreading the terms used by Hurn, Justel, Robert (2000), MSARMs can be seen as Markov mixtures of autoregressions: the conditional distribution of any observation, given the previous ones, is a mixture of normal distributions weighed by the stationary distribution of the hidden Markov chain.

A special MSARM with a harmonic component is here proposed in the Bayesian framework, giving rise to Harmonic MSARMs (HMSARMs). We shall study the parameter estimation of the models through a Markov chain Monte Carlo method, also tackling the problem of missing data within the series, that is the sequence of the hidden states and all the missing observations are handled as unknown parameters. Parameter estimation must consider the identifiability constraints within the model: instead of rejecting the values not satisfying the constraints, we apply a thrifty technique called constrained permutation sampling (Frühwirth-Schnatter (2001a)). Before performing parameter estimation we need to choose the best model and to select its identifiability constraints. MSAR model choice is the investigation of the number of hidden states and the order the autoregressive process, through Bayes factor (Kass and Raftery (1995)), computing the marginal likelihoods by the Chib-Neal method (Chib (1995), Neal (1999)). The selection of the identifiability constraints is done exploiting the mixing properties of the random permutation sampling algorithm (Frühwirth-Schnatter (2001a)).

The paper is organized as follows. HMSARMs will be described in Section 1; Bayesian estimates of the parameters of HMSARMs will be obtained in Section 2 performing Gibbs sampling associated with constrained permutation sampling; finally in Section 3 two applications of HMSARMs will be shown and two environmental time series with different periodicities will be examined; Section 3 is also devoted to choose the best model and to detect its identifiability constraints by means of random permutation sampling.

1. Harmonic Markov switching autoregressive models

Markov switching autoregressive models are discrete-time stochastic processes \{Y_t; X_t\}, so that \{X_t\} is a latent finite-state Markov chain and \{Y_t\}, given \{X_t\}, satisfies the order-p dependence and the contemporary dependence conditions: we have a sequence of observed random variables \{Y_t\} depending on the p past observations, whose conditional distributions depend on \{X_t\} only through the contemporary \(X_t\).

Let \{X_t\} be a discrete-time, first-order, homogeneous, ergodic Markov chain on a finite state-space \(S_X\) with cardinality \(m (S_X = \{1, \ldots, m\})\). \(\Gamma = [\gamma_{i,j}]\) is the \((m \times m)\) transition matrix, where \(\gamma_{i,j} = P(X_t = j | X_{t-1} = i)\), for any \(i,j \in S_X\), and \(\delta = (\delta_1, \ldots, \delta_m)\) is the stationary distribution, so that \(\delta' = \delta' \Gamma; x_T' = (x_1, \ldots, x_T)\) is the sequence of the states of the Markov chain and, for any \(t = 1, \ldots, T\), \(x_t\) assumes values in \(S_X\).

Hence, given the order-p dependence and the contemporary dependence conditions, the equation
describing HMSARMs is

\[ Y_t(x_i) = \mu_{x_i} + \sum_{\tau=1}^{p} \varphi_{\tau(i)} Y_{t-\tau(x_i)} + \eta_t + E_t(x_i), \]  

(1)

where \( Y_t(i) \) denotes the generic variable \( Y_t \) when \( X_t = i \), for any \( 1 \leq t \leq T \) and for any \( i \in S_X \); the autoregressive coefficients \( \varphi_{\tau(i)} \), for any \( \tau = 1, \ldots, p \) and for any \( i \in S_X \), depend on the current state \( i \) of the Markov chain; \( \eta_t \) is a harmonic component of periodicity 2s,

\[ \eta_t = \sum_{j=1}^{s^*} \left( \eta_{1,j} \cos(\pi j t/s) + \eta_{2,j} \sin(\pi j t/s) \right), \]

where \( s^* \) is the number of significant harmonics \( (s^* \leq s); \) \( E_t(i) \) denotes the Gaussian random variable \( E_t \) when \( X_t = i \), with zero mean and precision \( \lambda_i \) \( (E_t(i) \sim \mathcal{N}(0; \lambda_i)) \), for any \( i \in S_X \), with the discrete-time process \( \{E_t\} \), given \( \{X_t\} \), satisfying the conditional independence and the contemporary dependence conditions. Notice that the harmonic component does not depend on the hidden Markov chain for identifiability reasons: if it depended, to have an identified model, we would assume the same hidden state all along the period 2s. From equation (1), the generic distribution of \( Y_t(i) \), given the \( p \) past observations and the current hidden state, is Gaussian with mean \( \mu_i + \sum_{\tau=1}^{p} \varphi_{\tau(i)} y_{t-\tau} + \eta_t \) and precision \( \lambda_i \).

A sufficient condition for the stationarity of the process (1) is that all the \( m \) sub-processes generated by the \( m \) states of the chain are stationary, that is, for any \( i \in S_X \), the roots of the auxiliary equations \( z^m - \varphi_{1(i)} z^{m-1} - \cdots - \varphi_{p(i)} = 0 \), where \( z \) is a complex variable, are all inside the unit circle.

The labels of the states and the sub-models, given a state, are interchangeable: the model (1) is unidentifiable in data fitting. This is the so-called label switching problem and it can be overcome placing suitable identifiability constraints on some parameters, i.e. \( \mu_i < \mu_j \) or \( \lambda_i < \lambda_j \) or \( \gamma_{i,i} < \gamma_{j,j} \), for any \( i, j \in S_X \) so that \( i < j \). In this paper the special HMSARM with the constraint on the precisions is analysed, but the procedures we shall introduce can be easily adapted to any other type of constraints. In Section 3 we shall see how and why we derive this type of constraint by a data-driven procedure, based on random permutation sampling algorithm (Fruwirth-Schnatter (2001a)). At this point it is important only to notice that the constraint is chosen ex post after simulations so as to respect the geometry and the shape of the unconstrained posterior distribution, that is different identifiability constraints can be derived by different data sets.

The parameters to be estimated are the transition matrix \( \Gamma \), the stationary distribution \( \delta \), the vector \( \mu \) of the \( m \) parameters \( \mu_i \), the vector \( \lambda \) of the \( m \) parameters \( \lambda_i \), the matrix \( \varphi \) of the \( m \) autoregressive coefficients vectors \( \varphi_i \), i.e. \( \varphi = (\varphi_1', \ldots, \varphi_i', \ldots, \varphi_m')' \), where \( \varphi_i = (\varphi_{1(i)}', \ldots, \varphi_{\tau(i)}', \ldots, \varphi_{p(i)}')' \) and the vector \( \eta \) of the harmonic coefficients, i.e. \( \eta = (\eta_{1,1}, \eta_{2,1}, \ldots, \eta_{1,s^*}, \eta_{2,s^*})' \). We also want to estimate the sequence of hidden states \( x^T = (x_1, \ldots, x_t, \ldots, x_T)' \), and all the missing observations \( y^*_t \), collected in a vector \( y^* \). Using Tanner and Wong (1987) terminology, \( y^T \) are the observed data, \( z = (x^T, y^*)' \) are the latent data and \( (y^T, z)' \) are the augmented data. All the parameters and the latent data will be estimated by simulation, performing Gibbs sampling (except for the stationary distribution \( \delta \) that will be estimated by the equality \( \delta' = \delta' \Gamma \)). As usual for this class of models we place conditional independent conjugate priors: independent Dirichlet priors on each row of \( \Gamma \); independent normal priors on each entry of vector \( \mu = (\mu_1, \ldots, \mu_m)' \); independent gamma priors on each entry of vector \( \lambda = (\lambda_1, \ldots, \lambda_m)' \), under the identifiability constraint; independent multivariate normal priors of dimension \( p \) on each row of the matrix \( \varphi = (\varphi_1', \ldots, \varphi_i', \ldots, \varphi_m')' \), under the stationarity constraint; a multivariate normal prior of dimension \( 2s^* \) on the vector \( \eta = (\eta_{1,1}, \eta_{2,1}, \ldots, \eta_{1,s^*}, \eta_{2,s^*})' \). Finally, let \( \theta \) be the vector of the unknown parameters and latent data of the HMSARM to be estimated through Gibbs sampling,

\[ \theta = (\Gamma, \mu, \lambda, \varphi, \eta, x^T, y^*)'. \]
The posterior distribution of $\theta$ is

$$
\pi (\theta \mid y^T, y^0, W) = f(\Gamma, \mu, \lambda, \varphi, \eta, x^T, y^* \mid y^T, y^0, W) \propto
$$

$$
f (y^T, y^* \mid \mu, \lambda, \varphi, \eta, x^T, W, y^0) f (x^T \mid \Gamma) p(\Gamma) p(\mu)p(\lambda)p(\varphi)p(\eta),
$$

where $y^T = (y_1, \ldots, y_T)'$ is the vector of the observed data, that is the sequence of the realizations of the stochastic process $\{Y_t\}$, with the initial values $y^0 = (y_{-p+1}, \ldots, y_0)'$ fixed for the $p$-dependence condition; $W$ is a $(T \times 2s^*)$ matrix whose generic element on the $i$-th row of the $j$-th odd column is $\cos(\pi j t/s)$, while the generic element on the $t$-th row of the $j$-th even column is $\sin(\pi j t/s)$, for any $j = 1, 2, \ldots, s^*$, and

$$
f (y^T, y^* \mid \mu, \lambda, \varphi, \eta, x^T, W, y^0) = \prod_{t=1}^T f (y_t \mid y_{t-1}, \ldots, y_{t-p}, \mu, \lambda, \varphi, \eta, x_t, y^0),
$$

(2)

by the order-$p$ dependence and the contemporary dependence conditions, for any $t = 1, \ldots, T$, and

$$
f (x^T \mid \Gamma) = \delta_{x_1} \prod_{t=2}^T \gamma_{x_{t-1}, x_t} = \delta_{x_1} \prod_{i=1}^m \prod_{j=1}^m \gamma_{i,j},
$$

by the Markov dependence condition, where $T_{i,j}$ is the number of couples of consecutive hidden states $i,j$. Notice that in the right side of equality (2) there are no missing observations: if one or more missing observations occur within $y^T$, any missing observation will be replaced by the corresponding simulated value $y^*_t$.

Now we can apply the Gibbs sampler to HMSARMs to estimate the vector $\theta$.

2. Parameter estimation of HMSARMs

Gibbs sampling is a simulation scheme, via Monte Carlo methods, from a posterior distribution, which is the stationary distribution of a Markov chain, that can be adopted when the posterior distribution in closed form is unavailable, but the full conditionals are available. We shall not describe in detail the iterative scheme of Gibbs sampling, that can be seen for example in Gamerman (1997), Chapter 5, to which we refer.

The Gibbs sampling procedure associated with the constrained permutation sampling algorithm is now developed for the special HMSARM with identifiability constraint on the precision, noticing this scheme it can be easily rearranged whenever another type of constraint is imposed. To be able to perform permutation sampling, we need all the priors to be invariant to relabelling the states, i.e. their hyperparameters must not depend on the hidden states.

Here we can analyse the generic $k$-th iteration of Gibbs sampling only, remembering that at the $(k-1)$-th iteration the vector $\theta^{(k-1)}$ has been generated,

$$
\theta^{(k-1)} = \left(\Gamma^{(k-1)}, \mu^{(k-1)}, \lambda^{(k-1)}, \varphi^{(k-1)}, \eta^{(k-1)}, x^{(k-1)}, y^{(k-1)}\right),
$$

and the identifiability constraint on the precision has been chosen, $\lambda_i^{(k-1)} < \lambda_j^{(k-1)}$, for any $i,j \in S_X$ so that $i < j$.

1) The sequence $x^{T(k)}$ of hidden states is generated in block from the full conditional $\pi (x^T \mid y^T, \Gamma^{(k-1)}, \mu^{(k-1)}, \lambda^{(k-1)}, \varphi^{(k-1)}, y^{(k-1)}, W, y^0)$, by means of the procedure proposed by Chib (1996), based on the forward filtering-backward sampling ($ff-bs$) algorithm by Carter and Kohn (1994) and Frühwirth-Schnatter (1994) for state-space models. The $ff-bs$ algorithm is so called because first the filtered probabilities of the hidden states are computed going forward; then the conditional probabilities of the hidden states are computed going backward, sampling the states from the full
conditional, according to Lemma 2.1 of Carter and Kohn (1994), based on Markov property and equality 22.A.13 by Hamilton (1994),

\[
\pi (x^T, y^T, \Gamma, \mu, \lambda, \varphi, \eta, y^*, W, y^0) = \\
= \pi (x_T, y_T, \Gamma, \mu, \lambda, \varphi, \eta, y^*, W, y^0) \prod_{t=1}^{T-1} \pi (x_t, x_{t+1}, y^t, \Gamma, \mu, \lambda, \varphi, \eta, y^*, W, y^0).
\]

Let \( \xi_{t+1} \) be the \( m \)-dimensional vector whose generic entry is \( P (X_{t+1} = i | y^t, \Gamma, \mu, \lambda, \varphi, \eta, y^*, W, y^0) \), for any \( i = 1, \ldots, m \); \( \xi_t \) be the \( m \)-dimensional vector whose generic entry is \( P (X_t = i | y^t, \Gamma, \mu, \lambda, \varphi, \eta, y^*, W, y^0) \), for any \( i = 1, \ldots, m \); \( \xi_t \) be the \( m \)-dimensional vector whose generic entry is \( P (X_t = i | X_{t+1} = x_{t+1}, y^t, \Gamma, \mu, \lambda, \varphi, \eta, y^*, W, y^0) \), for any \( i = 1, \ldots, m \). The iterative scheme of the \( \phi \)-hbs algorithm is the following.

1.1) Compute

\[
\xi^{(k)}_{1|0} = \delta^{(k-1)} = \delta^{(k-1)} \Gamma^{(k-1)},
\]

that is \( \delta^{(k-1)} \) is the left eigenvector of the matrix \( \Gamma^{(k-1)} = \left[ \gamma_{i,j}^{(k-1)} \right] \), associated with the eigenvalue one.

1.2) Compute

\[
\xi^{(k)}_{t|t-1} = \frac{\xi^{(k)}_{t|t-1} \Gamma^{(k-1)}_t}{1^{(m)} (\xi^{(k)}_{t|t-1} \Gamma^{(k-1)}_t)} \quad \text{and} \quad \xi^{(k)}_{t+1|t} = \Gamma^{(k-1)} \xi^{(k)}_{t|t-1},
\]

for any \( t = 1, \ldots, T-1 \), where \( \Gamma^{(k-1)}_t = \text{diag} \left[ f \left( y_t | y_{t-1}, \ldots, y_{t-p}, \mu^{(k-1)}, \lambda^{(k-1)}, \varphi^{(k-1)}, \eta^{(k-1)} \right), \ldots, f \left( y_t | y_{t}, \ldots, y_{t-p}, y_{t-1}, \ldots, y_{t-p}, \mu^{(k-1)}, \lambda^{(k-1)}, \varphi^{(k-1)}, \eta^{(k-1)} \right), W, y^0, x^{(k-1)}_t = 1 \right] \) and \( 1^{(m)} \) is the \( m \)-dimensional vector of ones.

1.3) Compute

\[
\xi^{(k)}_{T|T} = \frac{\xi^{(k)}_{T|T-1} \Gamma^{(k-1)}_T}{1^{(m)} (\xi^{(k)}_{T|T-1} \Gamma^{(k-1)}_T)}
\]

(for details on the derivation of formulae at steps 1.2 and 1.3, see Hamilton (1994), pp. 692-693).

1.4) Generate \( x^{(k)}_T \) from \( \xi^{(k)}_{T|T} \).

1.5) Compute

\[
\xi^{(k)}_{t|t} = \frac{\xi^{(k)}_{t|t} \Gamma^{(k-1)} \mathbf{w}^{(k)}}{1^{(m)} (\xi^{(k)}_{t|t} \Gamma^{(k-1)} \mathbf{w}^{(k)})}
\]

and generate \( x^{(k)}_t \) from \( \xi^{(k)}_{t|t} \), for any \( t = T-1, \ldots, 1 \). \( \Gamma^{(k-1)} \mathbf{w}^{(k)} \) represents the column of \( \Gamma^{(k-1)} \) corresponding to the state previously generated.

2) Placing a gamma prior \( G (x; \alpha, \beta) \) on any \( \lambda \), the parameters \( \lambda_i^{(k)} \), for any \( i \in S_X \), are independently generated from a gamma distribution with parameters

\[
\frac{T_i^{(k)}}{2} + \alpha_i \]
and

\[
\frac{1}{2} \sum_{t \geq 1; i \neq k} \left( y_t - \mu_i^{(k-1)} - \sum_{\tau=1}^p \psi_{\tau(i)}^{(k-1)} y_{t-\tau} - \eta_t^{(k-1)} \right)^2 + \beta_\lambda,
\]

where \( T_i^{(k)} \) is the number of observations corresponding to the contemporary hidden state \( i \) in the sequence \( xT(k) \) generated at step 1.

The entries of the vector \( \lambda_i^{(k)} \) must be in increasing order to satisfy the identifiability constraint: \( \lambda_i^{(k)} < \lambda_j^{(k)} \), for any \( i, j \in S_X \), so that \( i < j \). If \( \lambda^{(k)} \) is not ordered, instead of rejecting the vector and going on sampling till we have an ordered one, we introduce the constrained permutation sampling algorithm (Früwirth-Schnatter (2001a)): we have \( m \) couples \((i, \lambda_i^{(k)})\); if the \( \lambda_i^{(k)} \)'s are unordered, we apply a permutation \( \rho(\cdot) \) to order them; consequently also the corresponding \( i \)'s must be permuted according to the permutation \( \rho(\cdot) \), \( \rho(S_X) = \{\rho(1), \ldots, \rho(m)\} \); finally the permutation \( \rho(S_X) \) is extended to the generated sequence of states \( xT(k) \), \( \rho(xT(k)) = (\rho(x_1^{(k)}), \ldots, \rho(x_t^{(k)}), \ldots, \rho(x_T^{(k)}))^T \), and to the switching-parameters previously generated, \( \rho(\Gamma^{(k-1)}) \), \( \rho(\mu^{(k-1)}) \), \( \rho(\varphi^{(k-1)}) \).

Notice that if we had had a different constraint, either on the means or on the diagonal entries of transition matrix, the corresponding full-conditional would have been placed at this step.

3) Placing a normal prior \( \mathcal{N}(\mu_M; \lambda_M) \) on any \( \mu_i \), the parameters \( \mu_i^{(k)} \), for any \( i \in S_X \), are independently generated from a normal distribution with mean

\[
\rho \left( \lambda_i^{(k)} \right) \sum_{t \geq 1; \rho(\tau^i) = i} \left( y_t - \sum_{\tau=1}^p \psi_{\tau(i)}^{(k-1)} y_{t-\tau} - \eta_t^{(k-1)} \right) + \mu_M \lambda_M
\]

and precision

\[
\rho \left( T_i^{(k)} \right) \rho \left( \lambda_i^{(k)} \right) + \lambda_M,
\]

where \( \rho \left( T_i^{(k)} \right) \) is the number of observations corresponding to the contemporary hidden state \( i \) in the permuted sequence \( xT(k) \).

4) Placing a truncated multivariate normal prior of dimension \( p \mathcal{N}(\mu_\Phi; \Lambda_\Phi) \cdot I(\varphi) \) on any \( \varphi_i \), where \( I(\varphi) \) is an indicator function so that

\[
I(\varphi) = \begin{cases} 
1 & \text{if the roots of the auxiliary equation are all inside the unit circle} \\
0 & \text{otherwise}
\end{cases}
\]

the parameter vectors \( \varphi_i^{(k)} \), for any \( i \in S_X \), are generated from a multivariate normal distribution of order \( p \), under the stationarity constraint, with mean vector

\[
\left[ \rho \left( \lambda_i^{(k)} \right) \right] Z' Q_{\rho(i)}^{(k)} Z + \Lambda_\Phi
\]

and precision matrix

\[
\rho \left( \lambda_i^{(k)} \right) Z' Q_{\rho(i)}^{(k)} Z + \Lambda_\Phi,
\]

where \( Z \) is a \((T \times p)\) matrix whose generic element on the \( t \)-th row and the \( j \)-th column is \( y_{t-j} \) (\( t = 1, \ldots, T \) and \( j = 1, \ldots, p \)) and \( Q_{\rho(i)}^{(k)} \) is a \((T \times T)\) diagonal matrix whose \( t \)-th term is one if \( \rho \left( x_t^{(k)} \right) = i \) or zero if it is not \( i \).
5) Placing a multivariate normal prior of dimension $2s^*$ $\mathcal{N}(\mu_H; \Lambda_H)$ on $\eta$, the parameter $\eta^{(k)}$ is generated from a multivariate normal distribution of dimension $2s^*$ with mean vector
\[
\left(W'\Lambda^{(k)}W + \Lambda_H\right)^{-1}\left(W'\Lambda^{(k)}\tilde{y}^{(k)} + \Lambda_H\mu_H\right)
\]
and precision matrix
\[
W'\Lambda^{(k)}W + \Lambda_H,
\]
where $\Lambda^{(k)}$ is a $(T \times T)$ diagonal matrix whose generic $t$-th element of the diagonal is $\rho\left(\lambda^{(k)}_{1|1} \right)$; $\tilde{y}^T$ is a $T$-dimensional vector whose generic $t$-th element is $y_t - \rho\left(\mu^{(k)}_{1|1}\right) - \sum_{\tau=1}^p y_{t-\tau}$.

6) Let $\Gamma_{i\bullet} = (\gamma_{i,1}, \gamma_{i,2}, \ldots, \gamma_{i,m})$, be the $i$-th row of $\Gamma$. Placing a Dirichlet prior with parameter $\omega = (\omega_1, \ldots, \omega_m)$ on $\Gamma_{i\bullet}$, each row $\Gamma^{(k)}_{i\bullet}$, for any $i \in S_X$, is independently generated from a Dirichlet $\mathcal{D}(\omega + \rho(T^{(k)}_{i\bullet}))$, where $\rho(T^{(k)}_{i\bullet}) = \left(\rho(T^{(k)}_{1i}), \ldots, \rho(T^{(k)}_{mi})\right)$.

7) Every missing observation $y^*_t$ is generated from the normal distribution:
\[
\mathcal{N}\left(\mu^{(k)}_{x_{1|1}}, \sum_{\tau=1}^p \rho\left(\varphi^{(k)}_{\tau|1}\right) y_{t-\tau} + \eta^{(k)}_t; \rho\left(\lambda^{(k)}_{1|1}\right)\right).
\]

Now, at the end of the $k$-th iteration of the Gibbs sampling, the vector $\theta^{(k)}$ has been simulated from $\pi(\theta \mid y^T, \rho^0, W)$, if $k$ is large enough. We shall repeat these steps till we have an $N$-dimensional sample. This sample will be used to estimate each entry of $\theta$ by means of posterior means, but the sequence of states, estimated through the posterior modes.

3. Empirical studies

Two applications of HMSARMs to real data will be studied in the following and the two time series, the first is about the daily mean concentrations of sulphur dioxide (SO2) and the second is about the hourly mean concentrations of carbon monoxide (CO), will be analysed in detail.

In each application we shall compare fifteen competing models which differ for the cardinality of the states-space of the hidden Markov chain ($m = 1, \ldots, 5$) and for the order of the autoregressive process ($p = 0, 1, 2$), henceforth said HMSAR($m$;$p$). We shall go on three consecutive steps: i) model selection, ii) constraint identification, iii) parameter estimation.

All the codes used for the applications have been written in Fortran.

3.1 Application to daily mean concentrations of sulphur dioxide

The first periodic time series we consider is about the daily mean concentrations of SO2, in micrograms per cubic meter ($\mu$g/$m^3$), recorded by the air pollution testing station placed in Via Goisis, Bergamo (Italy) from 13th of September, 1996, to 25th of November, 1999 (1169 observations). The data were collected and provided by “Assessorato all’Ambiente della Provincia di Bergamo”.

In the series of SO2 (Figure 1a) a yearly periodicity is evident and it is confirmed by the correlogram of 800 days (Figure 1b). The yearly periodicity 2s is 365 and the number $s^*$ of the harmonics is one, as it can be deduced by Figure 1b; so the harmonic component is
\[
\eta_t = \eta_1 \cos \left(\pi t/182\right) + \eta_2 \sin \left(\pi t/182\right).
\]

The analysed data are the natural logarithms of SO2 concentrations, while the sequence of the hidden chain makes up the latent data $\left(z \equiv x^T\right)$, because no missing value occurs within the series.
3.1.1 Model selection

Model selection will be performed by means of Bayes factors (Kass and Raftery (1995)) in which the marginal likelihoods, i.e. the normalizing constants of the posterior densities, are computed according to Chib (1995), corrected by the relabeling of the hidden states (Neal (1999)): by Chib’s method the Gibbs sampling chain visits only one of the m! possible labelings of the hidden states; hence we consider one labeling only and multiply the likelihood by m!. An alternative method to Chib’s to compute marginal likelihood from the output of the Gibbs sampler is bridge sampling (Meng and Wong (1996)); Frühwirth-Schnatter (1999), (2001b) deal with applications of bridge sampling to mixture models and non linear time series.

The natural logarithm of the marginal likelihood, \( \ln f(y^T \mid y^0, W) \), is estimated in a special point \( (\mu^*, \lambda^*, \varphi^*, \eta^*, \Gamma^*)' \), the posterior mode of \( (\mu, \lambda, \varphi, \eta, \Gamma)' \), and we obtain the estimate \( \ln \hat{f}(y^T \mid y^0, W) \):}

\[
\ln \hat{f}(y^T \mid y^0, W) = \ln m! + \ln f(y^T \mid \mu^*, \lambda^*, \varphi^*, \eta^*, \Gamma^*, y^0, W) + \ln p(\mu^*, \lambda^*, \varphi^*, \eta^*, \Gamma^*) - \ln \hat{\pi}(\mu^*, \lambda^*, \varphi^*, \eta^*, \Gamma^* \mid y^T, y^0, W). \tag{4}
\]

Computing the exp(\( \cdot \)), we have the marginal likelihood we need to compute Bayes factor.

The second addendum in Equality (4) is

\[
f(y^T \mid \mu^*, \lambda^*, \varphi^*, \eta^*, \Gamma^*, y^0, W) = \sum_{x_1 \in S_X} \sum_{x_2 \in S_X} \cdots \sum_{x_T \in S_X} \prod_{t=2}^{T} \gamma_{x_{t-1},x_{t}}^{*}(y_t \mid y_{t-1}, \ldots, y_{t-p}, \mu^*, \lambda_{x_t}^*, \varphi_{x_t}^*, \eta^*, W, x_t), \tag{5}
\]

the third becomes

\[
\ln p(\mu^*) + \ln p(\lambda^*) + \ln p(\varphi^*) + \ln p(\eta^*) + \ln p(\Gamma^*)
\]

and the fourth can be decomposed as

\[
\ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi \left( \mu^* \mid y^T, y^0, W, \lambda^{(k)}, \varphi^{(k)}, \eta^{(k)}, \Gamma^{(k)}, z^{(k)} \right) \right) + \ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi \left( \lambda^* \mid y^T, y^0, W, \mu^*, \varphi^{(k)}, \eta^{(k)}, \Gamma^{(k)}, z^{(k)} \right) \right) + \ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi \left( \varphi^* \mid y^T, y^0, W, \mu^*, \lambda^*, \eta^{(k)}, \Gamma^{(k)}, z^{(k)} \right) \right) + \ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi \left( \eta^* \mid y^T, y^0, W, \mu^*, \lambda^*, \varphi^*, \Gamma^{(k)}, z^{(k)} \right) \right) + \ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi \left( \Gamma^* \mid y^T, y^0, W, \mu^*, \lambda^*, \varphi^*, \eta^*, z^{(k)} \right) \right),
\]

and estimated using 5 \( \cdot \) \( N \) extra-iterations, labelled by \( k \), of the Gibbs sampling (Chib (1995)).
The iterative procedure for the selection of the best HMSARM is limited to \( m = 1, \ldots, 5 \) and \( p = 0, 1, 2 \): when \( p \) is greater than two, the normalizing constants of the full-conditionals generating the \( \varphi_j \)'s are very complicated and uncodable. Moreover we do not select \( s^* \) via Bayes factors, but we fix it \textit{a priori}: if \( s^* \) was free to vary from 1 to \( s \), we should have a considerable proliferation in the number of competing models to check, i.e. \( m \cdot p \cdot s \). For each of the fifteen HMSARMs here considered, the marginal likelihood is computed using a sample of 10000 successive values generated after a 1000 iterations burn-in period.

The following hyperparameters have been chosen for all the models:

- \( \omega_j = 1 \), for any \( j = 1, \ldots, m \), i.e. each row of \( \Gamma \) is assumed \textit{a priori} to be a multivariate uniform on the unit hypercube;
- \( \mu_M = \ln(125)/8, \lambda_M = 0.3 \), i.e. we model our prior belief that the concentrations of SO2 are much less than the attention level \((125 \, \mu g/m^3)\), both because the use of methane drastically reduce the SO2 emissions and because the air pollution testing station which recorded our data set is placed in a park;
- \( \alpha_{\Lambda} = \beta_{\Lambda} = 0.5 \), i.e. each precision is assumed \textit{a priori} to be a gamma with mean 1 and variance 2, leading to low variability within each state;
- \( \mu_\Phi = 0_{(p)}, \) where \( 0_{(p)} \) is a \( p \)-dimensional zero vector, \( \Lambda_\Phi = 2.75 \cdot I_{(p)} \), where \( I_{(p)} \) is a \( p \)-dimensional identity matrix, i.e the autoregressive coefficients belong to a space circumscribing the stationarity region;
- \( \mu_H = 0_{(2)}, \Lambda_H = 0.1 \cdot I_{(2)} \), i.e the prior information on \( \eta \) is quite vague.

We notice that the HMSAR(3:1) is the best among all the competing models (see Appendix for all the values of the marginal likelihoods) hence we shall estimate its unknown parameters.

3.1.2 Constraint identification

We have just seen that we can run unconstrained Gibbs sampling to select the number of the hidden states and then we can run it constrained for parameter estimation. Between these two steps, we have to identify carefully the constraint which must respect the geometry and the shape of the unconstrained posterior distribution. We derive a data-driven identifiability constraint looking at the graphs of the output of the unconstrained Gibbs sampling performed associated with \textit{random permutation sampling} \cite{fruhwirth-schnatter2001a}: we plot couples of outputs of the estimates of the parameters obtained via unconstrained Gibbs sampling with random permutations of the hidden states; after that we check if there are any groups corresponding to the different states and if these groups can suggest special ordering in the labeling.

Figure 2: Output of unconstrained Gibbs sampling with random permutations
Random permutation sampling is an easy adjustment we introduce in the procedure described in Section 3: at any iteration all the steps of Gibbs sampling run unconstrained; then we randomly generate one of \( m! \) ways of labelling the states and consequently update the sequence of the hidden states and any switching-parameter according to the selected permutation of the states. Random permutation sampling allows us to explore the whole support of the posterior distribution, improving the mixing property of the sampler because the chain is free to move through the different subspaces, and encourages the moves from the current subspace to one of the other \((m! - 1)\).

Graphically analysing the outputs of the unconstrained HMSAR(3;1) model, we choose the constraint on the precisions: \( \lambda_1 < \lambda_2 < \lambda_3 \), while no ordering is evident on the diagonal elements of \( \Gamma \), on \( \mu \) and on \( \varphi \), even if they are grouped (Figure 2). Notice that it is sufficient to plot the values corresponding to the first label only, because they represent all the three states, given the continuous jumps among all the possible labeling.

### 3.1.3 Parameter estimation

Running unconstrained Gibbs sampling with random permutations, it has been possible to choose HMSAR(3;1) as the best model and to detect its identifiability constraints. Now we can run constrained permutation Gibbs sampling for the HMSAR(3;1) model to estimate its parameters. The estimates are computed using a sample of 10000 successive values generated after a 1000 iterations burn-in period. The hyperparameters are listed in Section 3.1.1.

The estimates of the parameters of the hidden Markov chain are

\[
\Gamma = \begin{bmatrix}
0.500 & 0.472 & 0.028 \\
0.107 & 0.640 & 0.253 \\
0.044 & 0.081 & 0.875 \\
\end{bmatrix},
\]

from which we have the estimate of the stationary initial distribution,

\[
\delta = (0.114; 0.284; 0.602)',
\]

while those of the parameters of the three Gaussian probability density functions (pdfs) are

<table>
<thead>
<tr>
<th>( i )</th>
<th>( \mu_i )</th>
<th>( \lambda_i )</th>
<th>( \varphi_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.531</td>
<td>0.867</td>
<td>0.550</td>
</tr>
<tr>
<td>2</td>
<td>0.958</td>
<td>2.857</td>
<td>0.226</td>
</tr>
<tr>
<td>3</td>
<td>0.546</td>
<td>6.724</td>
<td>0.681</td>
</tr>
</tbody>
</table>
The SO2 yearly dynamics, described by the $\eta_t$’s, respects the climatic conditions: higher levels of SO2 in the colder periods of the year and lower levels in the warmer ones (Figure 3d).

To evaluate the fitting performance of the HMSAR(3;1) model, the fitted and actual values are analysed through some descriptive statistics: the Root Mean Squared Error (RMSE), the mean absolute error (MAE) and the correlation coefficient (CORR) between actual and fitted values. We obtain $\text{RMSE} = 0.485$, $\text{MAE} = 0.364$, $\text{CORR} = 0.896$ and, by these values, the fitting ability of the model sounds good. Moreover in Figure 3 we can see the dynamics of the fitted values (Figure 3b) respects the dynamics of the real data (Figure 3a).

Dealing with the hidden states, from the diagonal entries of the transition matrix, it is also possible to compute the time spent in state $i$ of the Markov chain upon each return to it, which has a geometric distribution with mean $1/(1 - \gamma_{i,i})$; hence the expected time spent in state $i$, is

<table>
<thead>
<tr>
<th>$i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>days</td>
<td>2.000</td>
<td>1.894</td>
<td>8.000</td>
</tr>
</tbody>
</table>

Finally we are interested in the dynamics of the hidden states, representing the three different levels of pollution occurred during the analysed period, which we can observe in Figure 3c, where we have the sequence of the posterior modes of any generated state $x_t$, for any $t = 1, \ldots, T$.

3.2 Application to hourly mean concentrations of carbon monoxide

The second periodic time series we consider is about the hourly mean concentrations of CO, in milligrams per cubic meter (mg/m$^3$), recorded by the air pollution testing station placed in Via San Giorgio, Bergamo (Italy) from 20th of October, 1998, 1 a.m., to 8th of December, 1998, 12 p.m. (1200 observations). The data were collected by “Assessorato all’Ambiente della Provincia di Bergamo” and provided by the local “Agenzia Regionale Protezione Ambiente”.

In the series of CO (Figure 4a) a daily periodicity is evident and it is confirmed by the correlogram of 120 hours (Figure 4b). The daily periodicity $2s$ is 24 and the number $s^*$ of the harmonics is three, as it can be deduced by Figure 4b, so the harmonic component is

$$\eta_t = \sum_{j=1}^{3} \left( \eta_{1,j} \cos \left( \pi j t / 12 \right) + \eta_{2,j} \sin \left( \pi j t / 12 \right) \right).$$

The analysed data are the natural logarithms of CO concentrations, while the latent data are made up both by the hidden states and by the 25 missing values occurring within the observed series.

3.2.1 Model selection

Model selection is based on the marginal likelihoods, whose computation is described in Subsection 3.1.1. Given that we have missing values within the sequence of observations, Formula (5) must
be updated. When a current observation $y^*_t$ is missing, the pdf in (5) must be replaced with 1, for any $x_t \in S_X$ (Paroli and Spezia (2002)), while, when a missing value occurs among the $p$ past observations, it must be replaced by the expected value

$$E\left(y_t \mid y^{t-1}, \mu^*, \lambda^*, \varphi^*, \eta^*, \Gamma^*, W, y^0\right) =$$

$$= \sum_{i=1}^{m} E\left(y_t \mid y^{t-1}, \mu^*, \lambda^*, \varphi^*, \eta^*, \Gamma^*, W, y^0, x_t = i\right) P\left(X_t = i \mid y^{t-1}, \mu^*, \lambda^*, \varphi^*, \eta^*, \Gamma^*, W, y^0\right),$$

where $P\left(X_t = i \mid y^{t-1}, \mu^*, \lambda^*, \varphi^*, \eta^*, \Gamma^*, W, y^0\right)$ is the filtered probability.

For each of the fifteen HMSARMs here considered, the marginal likelihood is computed using a sample of 10000 successive values generated after a 1000 iterations burn-in period. The following hyperparameters have been chosen for all the models:

- $\omega_j = 1$, for any $j = 1, \ldots, m$;
- $\mu_M = \ln(15)/2$, $\lambda_M = 0.3$, i.e. we model our prior belief that the concentrations of CO are quite close to the attention level $\left(15 \text{ mg/m}^3\right)$, because CO reaches high concentrations in the urban areas where the traffic jams occur and the air pollution testing station which recorded our data set is placed in a heavy traffic road;
- $\alpha_\Lambda = \beta_\Lambda = 0.5$;
- $\mu_\Phi = 0(p)$, $\Lambda_\Phi = 2.75 \cdot I(p)$;
- $\mu_H = 0(6)$, $\Lambda_H = 0.1 \cdot I(6)$.

We notice that the HMSAR(2;1) model is the best among all the competing models (see Appendix for all the values of the marginal likelihoods); hence we shall estimate its unknown parameters.

### 3.2.2 Constraint identification

Graphically analysing the outputs of the unconstrained HMSAR(2;1), we choose the constraint on the precisions again: $\lambda_1 < \lambda_2$ (Figure 5).

### 3.2.3 Parameter estimation

Now we run constrained permutation Gibbs sampling for the best model for CO analysis, i.e. HMSAR(2;1), to estimate its parameters, using the usual 1000 + 10000 iterations sampler, running with the same hyperparameters used in model choice and constraint identification.

The estimates of the parameters of the hidden Markov chain are

$$\Gamma = \begin{bmatrix} 0.907 & 0.093 \\ 0.012 & 0.988 \end{bmatrix},$$

from which we have the estimate of the stationary initial distribution,

$$\delta = (0.117; 0.883)^t,$$

while those of the parameters of the two Gaussian pdfs are

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\mu_i$</th>
<th>$\lambda_i$</th>
<th>$\varphi_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.275</td>
<td>3.077</td>
<td>0.542</td>
</tr>
<tr>
<td>2</td>
<td>0.319</td>
<td>11.969</td>
<td>0.687</td>
</tr>
</tbody>
</table>

and

$$\eta = (-0.106; -0.173; -0.154; -0.152; 0.141; 0.050)^t.$$

The CO daily dynamics, $\eta_t$'s, respects rush hours, in fact we have the peaks at eight a.m. and five p.m.; the presence in the plot (Figure 6d) of two peaks only seems to suggest that the number of harmonics can be reduced to two.
Figure 5: Output of unconstrained Gibbs sampling with random permutations

Figure 6: Actual (a) and fitted (b) values of CO series; the sequence of the hidden states (c) and the harmonic component (d)

Figure 7: Actual (triangles) and fitted (circles) values of days 2 (a) and 14 (b)
The fitting performance of the model is evaluated again through descriptive statistics (i.e. RMSE = 0.316, MAE = 0.242, CORR = 0.884) and plots of actual and fitted values (Figures 6a and 6b): we have the fitted series well describes the observed phenomenon.

Within the sequence of 1200 observations, we have 25 missing values which can be grouped in three sets: 18 single missing observations, 2 couples of missing observations and 1 block of 3 missing observations. Missing observations are simulated according to Formula (3): we can see from Figures 7a and 7b these simulated values correctly fill the series according to the dynamics of the twenty-four hours.

By the Markov chain side, the mean time spent in state $i$ is

<table>
<thead>
<tr>
<th>$i$</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>hours</td>
<td>10.753</td>
<td>83.333</td>
</tr>
</tbody>
</table>

and the estimated sequence of hidden states is plotted in Figure 6c.

Conclusions

The previously described empirical studies about air pollution show that Markov switching autoregressive models with a harmonic component well analyse periodic time series whose dynamics non linearly depend on latent variables. Model choice and inference have been performed through Gibbs sampling, considering the label switching problem, which has been efficiently tackled by permutation sampling. For alternatives to random permutation sampling, see Celeux, Hurn, Robert (2000) and Stephens (2000). Permutation sampling first allows to select identifiability constraints which rightly accords with the posterior distribution; then it improves the efficiency of Gibbs sampler removing the step of the rejections of the generated values which do not satisfy the identifiability constraint. Model choice is limited to the models whose autoregressive order is not greater than two, because of coding complications, which could be overcome reparameterising the model in terms of the reciprocal roots of the stationary autoregressive processes and imposing on them independent Beta priors, rescaled between -1 and +1 (Ehlers and Brooks (2002)).

The models we considered can be extended in many ways (i.e. time-varying transition matrices, multivariate pollutants and multisites recording analysis) to apply them more extensively to air quality control, focusing attention on the prediction of the future values of the pollutants; these extensions are concern future researches.

Acknowledgements

The authors are thankful to Petros Dellaportas for a number of useful remarks. The research of the second author has been supported by a UCSC fellowship.

References


**Appendix**

Natural logarithms of the marginal likelihoods of the models considered in the SO2 application, for any couple \((m; p)\)

<table>
<thead>
<tr>
<th>(p\backslash m)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>−1494.925</td>
<td>−1274.362</td>
<td>−1145.464</td>
<td>−1150.332</td>
<td>−1091.001</td>
</tr>
<tr>
<td>1</td>
<td>−1243.110</td>
<td>−1090.657</td>
<td><strong>−1087.239</strong></td>
<td>−1109.067</td>
<td>−1270.041</td>
</tr>
<tr>
<td>2</td>
<td>−1937.514</td>
<td>−1777.069</td>
<td>−1764.091</td>
<td>−1731.080</td>
<td>−1558.150</td>
</tr>
</tbody>
</table>

Natural logarithms of the marginal likelihoods of the models considered in the CO application, for any couple \((m; p)\)

<table>
<thead>
<tr>
<th>(p\backslash m)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>−816.881</td>
<td>−629.002</td>
<td>−562.667</td>
<td>−564.813</td>
<td>−616.034</td>
</tr>
<tr>
<td>1</td>
<td>−412.073</td>
<td><strong>−387.865</strong></td>
<td>−400.554</td>
<td>−440.949</td>
<td>−491.150</td>
</tr>
<tr>
<td>2</td>
<td>−1107.716</td>
<td>−1071.565</td>
<td>−1084.328</td>
<td>−803.991</td>
<td>−1159.014</td>
</tr>
</tbody>
</table>