NON-HOMOGENEOUS MARKOV MIXTURES
OF PERIODIC AUTOREGRESSIONS
AND SULPHUR DIOXIDE CONCENTRATIONS
NON-HOMOGENEOUS MARKOV MIXTURES OF PERIODIC AUTOREGRESSIONS AND SULPHUR DIOXIDE CONCENTRATIONS

ROBERTA PAROLI
Istituto di Statistica,
Università Cattolica S.C., Milano

LUIGI SPEZIA
Dipartimento di Statistica,
Università Ca’ Foscari, Venezia

Abstract - We propose a non-homogeneous Markov mixture of periodic autoregressions for relating the hourly mean concentrations of sulphur dioxide with six meteorological variables, recorded for three years by an air pollution testing station located in the lagoon of Venice (Italy). Bayesian analysis is performed within a Markov chain Monte Carlo framework, along four consecutive steps: the specification of the identifiability constraint; the choice of the cardinality of the hidden Markov chain state-space and the autoregressive order; the selection of the meteorological variables which influence the observed process and the time-varying transition probabilities of the hidden Markov chain; the estimation of the parameters. The main goal of selecting exogenous variables is performed in the case of correlation among the variables, by means of a procedure based on the “Metropolization” of the approach proposed by Kuo and Mallick (1998). The reconstruction of the sequence of the hidden states, the restoration of the missing values occurring within the observed series, the description of the periodic component are also given.

Keywords - Hidden Markov chain; marginal likelihood; Metropolized-Kuo-Mallick; permutation sampling; regime switching.
1 Introduction

Public authorities are requested to control air quality by measuring through air pollution testing stations the concentrations of some pollutants. In this paper, we analyse a time series of hourly mean concentrations of sulphur dioxide (SO$_2$), recorded in Venice (Italy) in years from 2001 to 2003. In Venice there are ten air pollution testing stations; nine of them are on the dry land and one in the lagoon, on the Isle of Giudecca. This last station is exactly that in which we are interested, because it recorded concentrations of SO$_2$ much higher than those recorded on the dry land. So, now, it is of great importance to the public authorities to understand both the relationships between SO$_2$ and the meteorological variables and which of them have more influence on the dynamics of SO$_2$, to be able to predict, in the future, when the alarm level can be exceeded.

The main features of SO$_2$ time series, which will be illustrated in more details in Section 2, are: i) non-normality, ii) daily periodicity, iii) occurrence of missing values within the series, iv) presence of six meteorological variables which could influence the dynamics of the data, v) alternation of different unobserved regimes which influence the observed levels of pollution. So, a non-homogeneous Markov mixture of periodic autoregressions can be an efficient tool to model all the above-mentioned characteristics of the series.

Markov mixture of autoregressions have been introduced in econometric literature by Hamilton under the name of time series subject to changes in regime (Hamilton (1990)) or Markov-switching time-series models (Hamilton (1996)), because we have that different autoregressions, each one depending on a latent regime, alternate according to the regime switching, which is driven by an unobserved homogeneous Markov chain. Krozlig (1997), Kim and Nelson (1999), Franses and van Dijk (2000) provide generalizations and applications for a wide range of economic and financial time series. Chib (1995) proposed to call this class of models Markov mixtures, because the conditional density of the current observation, 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. Following Chib (1995), Hurn et al (2003) and Wong and Li (2000), we call them Markov mixture of autoregressions (MMAR).

We generalize MMAR by assuming that the unobserved Markov chain is non-homogeneous and by joining a periodic component to the autoregressive processes, giving rise to non-homogeneous Markov mixtures of periodic autoregressions (NHMM-PAR) and showing that this class of model is an appropriate instrument not only for econometricians, but for environmental statisticians, too. In this paper we also increase the complexity of the stochastic environmental systems drawn by us in some earlier works, i.e. Paroli and Spezia (2003); Spezia et al (2004); Spezia (2005).

The data are described in Section 2. In Section 3 we introduce the model, by using the Bayesian approach. Section 4 presents the Markov chain Monte Carlo (MCMC) algorithm we use for the estimation procedure. Sections 5, 6 and 7 describe how the basic MCMC algorithm must be adapted to perform variable selection, model choice and identifiability constraint specification, respectively. In Section 8 we provide the results of our analysis.
2 The data

ARPA Veneto, one of the twenty regional agencies for environmental prevention and protection in Italy, placed a series of SO$_2$ concentrations to our disposal. The data have been recorded any hour, from January 1st, 2001 to September 30th, 2003 (24072 observations), by the air pollution testing station located on the Isle of Giudecca, in the lagoon of Venice. This series is reproduced in Figure 1a (the horizontal line represents the alarm level, 400 $\mu g/m^3$, which has been exceeded twice), showing higher levels of concentrations in the colder days and lower levels in the warmer ones.

In this paper, we analyse the series of the natural logarithms of the SO$_2$ concentrations (Figure 1b), thereafter called “observations”. We take the logarithm of the data both to reduce the variability of the series and to reverse the peaks and the spikes which are always hard to fit.

By the continuous approximation of the histogram of the observations (Figure 1c), we can notice the non-normality of the series, that has been confirmed by suitable normality tests. Moreover, by the autocorrelogram of the observations (Figure 1d), the daily periodicity is evident.

About the 7.5% of observations is missing: 1785 missing values occur within the series and they are handled as unknown parameters and estimated.

![Diagram](a)
![Diagram](b)
![Diagram](c)
![Diagram](d)

Figure 1: Series of the SO$_2$ hourly mean concentrations with the attention-level (a); series of the natural logarithms of the SO$_2$ concentrations (b); continuous approximation of the histogram of ln(SO$_2$) (c); 240 hours autocorrelations (d)

Six meteorological variables had been recorded simultaneously to the pollutants and they have been made available to us for verifying their possible influences on the SO$_2$ dynamics. They are: wind speed, temperature, rain, solar radiation, relative humidity, pressure.
3 The model

Let

\[ Y_{t(i)} = \mu_i + \sum_{\tau=1}^{p} \varphi_{\tau(i)} y_{t-\tau} + E_{t(i)}, \quad E_{t(i)} \sim \mathcal{N} \left( 0; \sigma_i^2 \right), \quad i = 1, \ldots, m \tag{1} \]

be \( m \) autoregressions of order \( p \), which differ for the value \( i \). At any time \( t \), we have the autoregression \( Y_{t(i)} \), if the discrete unobserved variable \( X_t^* \) takes the value of \( i \), for any \( i = 1, \ldots, m \) and for any \( t = 1, \ldots, T \). Any autoregression \( Y_{t(i)} \) is characterized by a signal \( \mu_i \), a variance \( \sigma_i^2 \) and \( p \) autoregressive coefficients \( \varphi_i = (\varphi_{1(i)}, \ldots, \varphi_{p(i)})' \) and depends on a normal noise \( E_{t(i)} \).

A linear combination of exogenous variables and a daily periodic component are joined to model (1):

\[ Y_{t(i)} = \mu_i + \sum_{\tau=1}^{p} \varphi_{\tau(i)} y_{t-\tau} + \sum_{j=1}^{q_i} \theta_{j(i)} w_{t,j}^i + \beta_{t(i)} + E_{t(i)}, \tag{2} \]

where \( w_{t,j}^i = (w_{t,j_1}^i, \ldots, w_{t,j_q}^i)' \) are \( q_i \) exogenous, deterministic variables observed at any time \( t = 1, \ldots, T \), possibly different for any \( i = 1, \ldots, m \); \( \theta_i = (\theta_{1(i)}, \ldots, \theta_{q_i(i)})' \) are the \( q_i \) coefficients of the exogenous variables, for any \( i = 1, \ldots, m \);

\[ \beta_{t(i)} = \sum_{j=1}^{v} \left( \beta_{1,j(i)} \cos \left( \pi j t / 12 \right) + \beta_{2,j(i)} \sin \left( \pi j t / 12 \right) \right); \]

\( v \) is the number of significant harmonics in the periodic component \( \beta_{t(i)} \), whose period is 24 \((v \leq 24)\); \( \beta_i = (\beta_{1,1(i)}, \beta_{2,1(i)}, \ldots, \beta_{1,v(i)}, \beta_{2,v(i)})' \) are the \( 2v \) coefficients of any \( \beta_{t(i)} \), for any \( i = 1, \ldots, m \).

In order to let the daily harmonic component depend on the latent variables, we need that the hidden regime \( i \) persists for all the 24 hours of the day. Hence it can serve our purpose to replace the time \( t \) subscript with the day \( d \) and hour \( h \) subscripts, so that \( t = (d-1)24 + h \), where \( d = 1, \ldots, D = T/24 \) and \( h = 1, \ldots, 24 \), and to assume \( X_{t(d-1)24+h}^* = \ldots = X_{t(d-1)24+h}^* = \ldots = X_{24d}^* = X_d \) for any \( d = 1, \ldots, D \). As a consequence, we can have \( d = \lceil t/24 \rceil + 1 \), where \( \lceil \cdot \rceil \) denotes the integer part of a real number.

The sequence of the unobserved variables \( X_d \), for any \( d = 1, \ldots, D \), is modelled here as a hidden non-homogeneous Markov chain, defined on a finite state-space \( S_X = \{1, \ldots, m\} \). The time-varying transition probabilities are \( \gamma_{a,b}^d = P(X_d = b \mid X_{d-1} = a) \), for any \( a, b \in S_X \) and for any \( d = 2, \ldots, D \).

We assume that, at any time \( d = 2, \ldots, D \), the transition probabilities \( \gamma_{a,i}^d \) can be obtained as logistic functions of \( z_d^a \alpha_{a,i} \), for any \( a, i \in S_X \):

\[ \logit(\gamma_{a,i}^d) = \log \left( \frac{\gamma_{a,i}^d}{\gamma_{a,a}^d} \right) = z_d^a \alpha_{a,i} \quad \text{for any } a, i \in S_X. \]

\[ \gamma_{a,i}^d = \frac{\exp(z_d^a \alpha_{a,i})}{\left( 1 + \sum_{i \neq a} \exp(z_d^a \alpha_{a,i}) \right)} \quad \text{for any } a, i \in S_X, \]
where \( \alpha_{a,i} \), when \( a \neq i \), is a vector of \( n_i \) coefficients, \( \alpha_{a,i} = (\alpha_{a,i,0}, \alpha_{a,i,1}, \ldots, \alpha_{a,i,n_i-1})' \), while, when \( a = i \), it is a vector of \( n_i \) zeros; \( z_d^i \) is a vector of \( n_i \) exogenous, deterministic variables, \( z_d^i = (1, z_{d,1}^i, \ldots, z_{d,n_i-1}^i) \), possibly different for any \( i = 1, \ldots, m \). Notice that when the last \( n_i - 1 \) entries of \( z_d^i \) are equal to zero for any \( d \) and for any \( i \), the Markov chain is homogeneous. The initial distribution of the Markov chain is vector \( \delta = (\delta_1, \ldots, \delta_m)' \), where \( \delta_i = P(X_1 = i) \), for any \( i \in S_X \); \( x^D = (x_1, \ldots, x_D)' \) is the sequence of the states of the non-homogeneous Markov chain and, for any \( d = 1, \ldots, D \), \( x_d \) has values in \( S_X \).

We also assume that the \( m \) autoregressive processes in (1) are stationary, that is, for any \( i \in S_X \), the roots of the auxiliary equations \( \lambda^p - \varphi_{1(i)} \lambda^{p-1} - \cdots - \varphi_{p(i)} = 0 \), where \( \lambda \) is a complex variable, are all inside the unit circle. To automatically satisfy the constraint on any \( \varphi_i = (\varphi_{1(i)}, \ldots, \varphi_{p(i)})' \), we reparametrize \( \varphi_i \) in terms of the partial autocorrelations \( r_i = (r_{1(i)}, \ldots, r_{p(i)})' \) of any sub-process, for any \( i \in S_X \), according to Barndorff-Nielsen and Schou (1973) and Jones (1987).

The functional relation between \( r_i \) and \( \varphi_i \), for any \( i \in S_X \), is recursively defined, for \( J = 1, \ldots, p \):

\[
\begin{align*}
&g_{1(i)}^1 = r_{1(i)} \\
&g_{K(i)}^j = g_{K(i)}^{j-1} - r_{J(i)} g_J^{j-1}, \quad \text{for any } K = 1, \ldots, J - 1 \\
&g_{J(i)}^j = r_{J(i)}, \\
&\varphi_i = (g_{1(i)}^p, g_{2(i)}^p, \ldots, g_{p(i)}^p)'.
\end{align*}
\]

Our inference will be based on a logarithmic transformation of any \( r_{j(i)} \), denoted by \( R_{j(i)} \), which maps partial autocorrelations \( r_{j(i)} \)'s from \((-1;1)\) to \( \mathbb{R} \), for any \( j = 1, \ldots, p \) and any \( i \in S_X \):

\[
R_{j(i)} = \ln \left( \frac{1 + r_{j(i)}}{1 - r_{j(i)}} \right).
\]

Identifiability of model (2) is ensured by imposing increasing variances \( (\sigma_i^2 < \sigma_j^2 \) for any \( i, j \in S_X \) so that \( i < j \)), but the procedures we shall introduce can be easily adapted to models with any other type of constraint. Notice that the identifiability constraint is chosen through a data driven procedure in order to respect the geometry and the shape of the unconstrained posterior distribution (Section 7); so, different identifiability constraints can be derived by different data sets.

Let \( \psi \) be the vector of the unknown parameters and latent data of the NHMMPAR to be estimated,

\[
\psi = (\mu, \sigma^{-2}, R, \theta, \beta, \alpha, x^D, y^*)',
\]

where \( \mu \) is the vector of the \( m \) signals \( \mu_i \); \( \sigma^{-2} \) is the vector of the \( m \) reciprocals of variances \( \sigma_i^{-2} \); \( R \) is the matrix of the \( m \) vectors \( R_i \), i.e. \( R = (R_1', \ldots, R_p', \ldots, R_m')' \) with \( R_i = (R_{1(i)}', \ldots, R_{j(i)}', \ldots, R_{p(i)}')' \); \( \theta \) is the matrix of the \( m \) vectors \( \theta_i \), i.e. \( \theta = (\theta_1', \ldots, \theta_j', \ldots, \theta_m')' \); \( \beta \) is the matrix of the \( m \) vectors of seasonal coefficients \( \beta_j \), i.e. \( \beta = (\beta_1', \ldots, \beta_j', \ldots, \beta_m')' \); \( \alpha \) is the matrix of the \( m^2 \) vectors \( \alpha_{a,i} \), \( x^D = (x_1, \ldots, x_d, \ldots, x_D)' \) is the sequence of the hidden states and \( y^* \) is the vector of all the missing observations \( y^*_r \) occuring within the sequence \( y^T = (y_1, \ldots, y_i, \ldots, y_T)' \). All the parameters and the latent data will be estimated numerically, by performing
an MCMC algorithm. For concreteness, we place the initial distribution as a vector of equal probabilities: \( \delta = (1/m, 1/m, \ldots, 1/m)' \).

For our Bayesian inference, we place the following priors:

- independent normal priors with known \( \mu_M \) and \( \sigma_M^2 \) on each entry of vector \( \mu \) (\( \mu_i \sim \mathcal{N}(\mu_M; \sigma_M^2) \), for any \( i \in S_X \));
- independent gamma priors with known \( \alpha_\Sigma \) and \( \beta_\Sigma \) on each entry of vector \( \sigma^{-2} \), under the identifiability constraint (\( \sigma_i^{-2} \sim \mathcal{G}(\alpha_\Sigma; \beta_\Sigma) \), for any \( i \in S_X \));
- independent normal priors with known \( \mu_R \) and \( \sigma_R^2 \) on each entry of matrix \( R \) (\( R_{ji} \sim \mathcal{N}(\mu_R; \sigma^2_R) \), for any \( i \in S_X \) and for any \( j = 1, \ldots, p \));
- independent multivariate normal priors of dimension \( q_i \) with known vector \( \mu_i \) and matrix \( \Sigma_i \) on each row of matrix \( \Theta \) (\( \theta_i \sim \mathcal{N}(\mu_i; \Sigma) \), for any \( i \in S_X \));
- independent multivariate normal priors of dimension \( 2v \) with known vector \( \mu_B \) and matrix \( \Sigma_B \) on each row of matrix \( \beta \) (\( \beta_i \sim \mathcal{N}(\mu_B; \Sigma_B) \), for any \( i \in S_X \));
- independent multivariate normal priors of dimension \( n_i \) with known vector \( \mu_A \) and matrix \( \Sigma_A \) on each off-diagonal entry of matrix \( \alpha \) (\( \alpha_{a,i} \sim \mathcal{N}(\mu_A; \Sigma_A) \), for any \( a; i \in S_X \), with \( a \neq i \)).

The posterior distribution of \( \psi \) is

\[
\pi(\psi | y^T, y^0, W, V, Z, \delta) = f(\mu, \sigma^{-2}, R, \theta, \beta, x^D, y^* | y^T, y^0, W, V, Z, \delta) \propto f(y^T | \mu, \sigma^{-2}, R, \theta, \beta, x^D, y^*, W, V, y^0) f(x^D | \alpha, Z, \delta) p(\mu)p(\sigma^{-2})p(R)p(\theta)p(\beta)p(\alpha),
\]

where \( y^0 = (y_{-p+1}, \ldots, y_0)' \) are the initial values fixed for the \( p \)-dependence condition; \( W \) is an \( m \)-block matrix, \( W = (W_1, \ldots, W_m) \), whose generic \( i \)-th block \( W_i \) is a \((T \times q_i) \) matrix, whose generic \( t \)-th row is \( w_{it} \); \( V \) is \((T \times 2v) \) matrix, whose generic element on the \( t \)-th row of the \( j \)-th odd column is \( \cos(\pi jt/12) \), while the generic element on the \( t \)-th row of the \( j \)-th even column is \( \sin(\pi jt/12) \), for any \( j = 1, 2, \ldots, v \); \( Z \) is an \( m \)-block matrix, \( Z = (Z_1, \ldots, Z_i, \ldots, Z_m) \), whose generic \( i \)-th block \( Z_i \) is a \((D \times n_i) \) matrix, whose generic \( d \)-th row is \( z_{id} \);

\[
f(y^T | \mu, \sigma^{-2}, R, \theta, \beta, x^D, y^*, W, V, y^0) = \prod_{t=1}^T f(y_t | y_{t-1}, \ldots, y_{t-p}, \mu, \sigma^{-2}, R, \theta, \beta, x_{[t/24]+1}, y^*, W, V, y^0),
\]

with

\[
f(y_t | y_{t-1}, \ldots, y_{t-p}, \mu, \sigma^{-2}, R, \theta, \beta, x_{[t/24]+1}, y^*, W, V, y^0) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left\{- \frac{1}{2\sigma_i^2} \left( y_t - \mu_{t-1} - \sum_{\tau=1}^{p} \varphi_{\tau(i)} y_{t-\tau} - \sum_{j=1}^{q_i} \theta_{j(i)} w_{ij} - \beta_{i(i)} \right)^2 \right\},
\]

whenever \( x_{[t/24]+1} = i \);

\[
f(x^D | \alpha, Z, \delta) = \delta_{x^D} \prod_{d=2}^{D} x^d_{x_{d-1}, x_d},
\]

with \( \logit(\gamma_{x_{d-1},x_d}) = z_{d,x_{d-1},x_d} \).
4 Parameter estimation

The basic MCMC algorithm for the estimation of the parameters of a NHMMPAR is now introduced, by showing all the steps of the generic \( k \)-th iteration of the sampler, given vector \( \psi^{(k-1)} = \left( \mu^{(k-1)}, \sigma^{-2(k-1)}, R^{(k-1)}, \beta^{(k-1)}, \alpha^{(k-1)}, \pi^{(k-1)} \right)' \), generated at the previous iteration. The following scheme provides identifiability constraint on the variances \( \sigma_i^{(2(k))} < \sigma_j^{(2(k))} \), for any \( k \) and for any \( i, j \in S_X \), so that \( i < j \), but it can be easily rearranged when another type of constraint is specified.

1) The sequence \( x^{D(k)} \) of hidden states is generated by the forward filtering-backward sampling (ff-bs) algorithm (Carter and Kohn (1994); Frühwirth-Schnatter (1994); Chib (1996)), which is so called because first the filtered probabilities of the hidden states are computed going forwards; then the conditional probabilities of the hidden states are computed going backwards, sampling the states from the full conditional,

\[
\pi \left( x^D \mid y^T \right) = \pi \left( x^D \mid y^T \right) \prod_{d=1}^{D-1} \pi \left( x_d \mid x_{d+1}, y^{24d} \right),
\]

suppressing all the conditioning on \( \mu, \sigma^{-2}, R, \theta, \beta, \alpha, \delta, W, V \) and \( Z \).

Let \( \xi_{d+1|d} \) be the \( m \)-dimensional vector whose generic entry is \( P \left( X_{d+1} = i \mid y^{24d} \right) \), for any \( i = 1, \ldots, m \); \( \xi_{d|d} \) be the \( m \)-dimensional vector whose generic entry is \( P \left( X_d = i \mid y^{24d} \right) \), for any \( i = 1, \ldots, m \); \( \xi_{d} \) be the \( m \)-dimensional vector whose generic entry is \( P \left( X_d = i \mid X_{d+1} = x_{d+1}, y^{24d} \right) \), for any \( i = 1, \ldots, m \). The iterative scheme of the ff-bs algorithm is the following.

1.1) Place

\( \xi^{(k)}_{1|0} = \delta' \).

1.2) Compute

\[
\xi^{(k)}_{d+1|d} = \frac{\xi^{(k)}_{d|d-1} F^{(k-1)}_d}{\Gamma^{d(k-1)'} \xi^{(k)}_{d|d}},
\]

for any \( d = 1, \ldots, D-1 \), where \( F^{(k-1)}_d = \text{diag} \left[ \prod_{h=1}^{24} f \left( y_{(d-1)24+h} \mid y_{(d-1)24+h-1}, \ldots, y_{(d-1)24+h-p}, x_d^{(k-1)} = 1 \right), \ldots, \prod_{h=1}^{24} f \left( y_{(d-1)24+h} \mid y_{(d-1)24+h-1}, \ldots, y_{(d-1)24+h-p}, x_d^{(k-1)} = m \right) \right], \Gamma^{d(k-1)} = \left[ \gamma^{d(k-1)}_{a,i} \right] \), with \( \gamma^{d(k-1)}_{a,i} = \left( \exp \left( z_d^{a,i}(k-1) \right) \right) \right) \left( 1 + \sum_{i \neq a} \exp \left( z_d^{a,i}(k-1) \right) \right) \), for any \( a, i \in S_X \), and \( 1_{(m)} \) is the \( m \)-dimensional vector of ones.

1.3) Compute

\[
\xi^{(k)}_{D|D} = \frac{\xi^{(k)}_{D|D-1} F^{(k-1)}_D}{\Gamma^{D(k-1)'} \xi^{(k)}_{D|D-1}},
\]

1.4) Generate \( x_{D}^{(k)} \) from \( \xi^{(k)}_{D|D} \).
1.5) Compute

\[
\xi_d^{(k)} = \frac{\xi_d^{(k)} \Gamma_{d+1}^{(k-1)}}{1'} \left( \xi_t^{(k)} \Gamma_{d+1}^{(k-1)} \right)
\]

and generate \( x_d^{(k)} \) from \( \xi_d^{(k)} \), for any \( d = D - 1, \ldots, 1 \). \( \Gamma_{d+1}^{(k-1)} \) represents the column of \( \Gamma^{(k-1)} \) corresponding to the state generated previously.

2) The parameters \( \sigma_i^{-2(k)} \), for any \( i \in S_X \), are independently generated from a gamma distribution with parameters

\[
\frac{D_i^{(k)}}{2} + \alpha \Sigma
\]

and

\[
\frac{1}{2} \sum_{t \geq 1; x_t^{(k)} \in S_{D(i)}} \left( y_t - \mu_i^{(k-1)} - \sum_{\tau=1}^p \varphi_{\tau(i)}^{(k-1)} y_{t-\tau} - \sum_{j=1}^{q_i} \theta_{j(i)}^{(k-1)} w_{ij}^i - \beta_{i}^{(k-1)} \right)^2 + \beta_{\Sigma},
\]

where \( D_i^{(k)} \) is the number of observations corresponding to the contemporary hidden state \( i \) in the sequence \( x^{D(k)} \) generated at the previous step. The entries of the vector \( \sigma^{-2(k)} \) must be in decreasing order to satisfy the identifiability constraint: \( \sigma_i^{2(k)} < \sigma_j^{2(k)} \), for any \( i, j \in S_X \), so that \( i < j \). Whenever the entries of \( \sigma^{-2(k)} \) are not ordered, we apply the constrained permutation sampling algorithm (Frühwirth-Schnatter (2001)): given \( m \) pairs \( (i, \sigma_i^{-2(k)}) \), a special permutation \( \rho(\cdot) \) is selected to order the variances; consequently also the corresponding \( i \)'s are permuted, \( \rho(S_X) = \{ \rho(1), \ldots, \rho(m) \} \); finally the permutation is extended to the generated sequence of states \( x^{D(k)} \), \( \rho(x^{D(k)}) = \left( \rho(x_1^{(k)}), \ldots, \rho(x_d^{(k)}), \ldots, \rho(x_D^{(k)}) \right)' \), and to the switching-parameters generated previously, \( \rho(\mu^{(k-1)}) \), \( \rho(R^{(k-1)}) \), \( \rho(\varphi^{(k-1)}) \), \( \rho(\theta^{(k-1)}) \) and \( \rho(\alpha^{(k-1)}) \), where \( \varphi^{(k-1)} \) is obtained from \( R^{(k-1)} \) by means of (3).

3) The parameters \( \mu_i^{(k)} \), for any \( i \in S_X \), are independently generated from a normal distribution with mean

\[
\sigma_i^{-2(k)} \sum_{t \geq 1; x_t^{(k)} \in S_{D(i)}} \left( y_t - \sum_{\tau=1}^p \varphi_{\tau(i)}^{(k-1)} y_{t-\tau} - \sum_{j=1}^{q_i} \theta_{j(i)}^{(k-1)} w_{ij}^i - \beta_{i}^{(k-1)} \right) + \mu_M \sigma_M^{-2}
\]

and variance

\[
\left( \frac{D_i^{(k)}}{\sigma_i^{-2(k)} + \sigma_M^{-2}} \right)^{-1}.
\]

4) The parameters \( R_{j(i)}^{(k)} \), for any \( j = 1, \ldots, p \) and any \( i \in S_X \), are independently generated from the random walk \( R_{j(i)}^{(k)} = R_{j(i)}^{(k-1)} + U_R \), where \( U_R \) is a univariate
Gaussian noise with zero mean and constant variance $\sigma_\mathcal{U_h}^2$, for any $k$. Then any vector $R_i^{(k)}$ is accepted with probability

$$A\left(R_i^{(k-1)}; R_i^{(k)}\right) = \min\left\{ \frac{\pi\left(R_i^{(k)} \mid \mu^{(k)}, \sigma^{-2(k)}, \theta^{(k-1)}, \beta^{(k-1)}, x^{D(k)}, y^{s(k-1)}, W, V, y^0, y^T\right)}{\pi\left(R_i^{(k-1)} \mid \mu^{(k)}, \sigma^{-2(k)}, \theta^{(k-1)}, \beta^{(k-1)}, x^{D(k)}, y^{s(k-1)}, W, V, y^0, y^T\right)}, 1 \right\},$$

for any $i \in S_X$, where

$$\pi\left(R_i^{(k)} \mid \mu^{(k)}, \sigma^{-2(k)}, \theta^{(k-1)}, \beta^{(k-1)}, x^{D(k)}, y^{s(k-1)}, W, V, y^0, y^T\right) \propto \exp\left\{ -\frac{1}{2}\sigma_i^{-2(k)} \sum_{t \geq x[i/24]+1} \left(y_t - \mu_i^{(k)} - \sum_{\tau=1}^{p} \varphi_{\tau(i)}^{(k)} y_{t-\tau} + \sum_{j=1}^{q_i} \theta_{\ell(j)}^{(k-1)} u_{i,j} - \beta_{\ell(i)}^{(k-1)} \right)^2 - \frac{1}{2}\sigma_R^{-2} \sum_{j=1}^{p} \left(R_j^{(k)} - \mu_R\right)^2 \right\}$$

(the same applies to the denominator in the r.h.s. of (4), replacing $R_j^{(k)}$ with $R_j^{(k-1)}$ and $\varphi_{\tau(i)}^{(k)}$ with $\varphi_{\tau(i)}^{(k-1)}$).

5) The parameters $\theta_i^{(k)}$, for any $i \in S_X$, are independently generated from a normal distribution of dimension $q_i$ with mean vector

$$\left(W_i'Q_i^{(k)}W_i + \Sigma_{\Theta}^{-1}\right)^{-1} \left(W_i'Q_i^{(k)}\tilde{y}_i^{(k)} + \Sigma_{\Theta}^{-1}\mu_{\Theta}\right)$$

and covariance matrix

$$\left(W_i'Q_i^{(k)}W_i + \Sigma_{\Theta}^{-1}\right)^{-1},$$

where $Q_i^{(k)}$ is a $(T \times T)$ diagonal matrix whose $t$-th term on the diagonal is either $\sigma_i^{-2(k)}$, if $x^{(k)}_{i[t/24]+1}$ is $i$, or zero, otherwise; $\tilde{y}_i^{(k)}$ is a $T$-dimensional vector whose generic $t$-th element is either $y_t - \mu_i^{(k)} - \sum_{\tau=1}^{p} \varphi_{\tau(i)}^{(k)} y_{t-\tau} - \beta_{\ell(i)}^{(k-1)}$, if $x^{(k)}_{i[t/24]+1}$ is $i$, or zero, otherwise.

6) The parameters $\beta_i^{(k)}$, for any $i \in S_X$, are independently generated from a normal distribution of dimension $2\nu$ with mean vector

$$\left(V'Q_i^{(k)}V + \Sigma_{\beta}^{-1}\right)^{-1} \left(V'Q_i^{(k)}\tilde{y}_i^{(k)} + \Sigma_{\beta}^{-1}\mu_{\beta}\right)$$

and covariance matrix

$$\left(V'Q_i^{(k)}V + \Sigma_{\beta}^{-1}\right)^{-1},$$

where $\tilde{y}_i^{(k)}$ is a $T$-dimensional vector whose generic $t$-th element is either $y_t - \mu_i^{(k)} - \sum_{\tau=1}^{p} \varphi_{\tau(i)}^{(k)} y_{t-\tau} - \sum_{j=1}^{q_i} \theta_{\ell(j)}^{(k)} w_{i,j}$, if $x^{(k)}_{i[t/24]+1}$ is $i$, or zero, otherwise.

7) The parameters $\alpha_{a,b}^{(k)}$, for any $a, b \in S_X$, with $a \neq b$, are independently generated from the random walk $\alpha_{a,b}^{(k)} = \alpha_{a,b}^{(k-1)} + U_A$, where $U_A$ is a multivariate Gaussian noise
with zero mean and constant covariance matrix $\Sigma_U$, for any $k$. Then any vector $\alpha_i^{(k)} = (\alpha_{i,1}, \ldots, \alpha_{i,m})^T$ is accepted with probability

$$
A\left(\alpha_i^{(k-1)}; \alpha_i^{(k)}\right) = \min \left\{ 1; \frac{\pi\left(\alpha_i^{(k)} | x^{D(k)}, Z, \delta\right)}{\pi\left(\alpha_i^{(k-1)} | x^{D(k)}, Z, \delta\right)} \right\},
$$

for any $i \in S_X$, where

$$
\pi\left(\alpha_i^{(k)} | x^{D(k)}, Z, \delta\right) \propto \delta_{x_i^{(k)}} \prod_{d \geq 2; x_{d-1}^{(k)} = i} \gamma_{i,b}^{d(k)} \exp \left\{ -\frac{1}{2} \sum_{b \neq i} (\alpha_{i,b}^{(k)} - \mu_A)^T \Sigma_A^{-1} (\alpha_{i,b}^{(k)} - \mu_A) \right\},
$$

with $\gamma_{i,b}^{d(k)}$ function of $\alpha_{i,b}^{(k)}$ (the same applies to the denominator in the r.h.s. of (5)).

8) Every missing observation $y_t^i$, given the hidden state $x_{[t/24]+1}^{(k)} = i$, is generated from the normal distribution

$$
\mathcal{N} \left( \mu_i^{(k)} + \sum_{\tau = 1}^{p} \phi^{(k)}(\tau) y_{t-\tau} + \sum_{j = 1}^{q} \varphi_{j(i)}^{(k)} w_{t,j}^{i} + \phi_{(i)}^{(k)}; \sigma_i^{(k)} \right),
$$

for any $i \in S_X$, and this concludes the $k$-th iteration.

The procedure for the estimation of the parameter vector $\psi$ has been illustrated within a framework in which the exogenous variables (matrices $W$ and $Z$) are given, the number of hidden states ($m$) and the autoregressive order ($p$) are fixed, and the special identifiability constraint is imposed. In the following three sections, the procedure to select the entries of $W$ and $Z$, to choose $m$ and $p$ and to adopt increasing variances will be examined.

## 5 Multiple selection of correlated exogenous variables

In NHMMPAR (2), at any time $t = (d-1)24 + h$, we have that, when the hidden Markov chain visits state $i$, the observed process is influenced both directly by $q_i$ exogenous variables $w_t^i$ and indirectly by the $m$ vectors of $n_i - 1$ exogenous variables $z_d^i$, for any $i = 1, \ldots, m$. So, for any time $t$, and $d$, and for any $i \in S_X$, the $q_i$ variables $w_t^i$ and $(n_i - 1)$ variables $z_d^i$ are selected within larger sets of covariates, of cardinality $Q$ and $N-1$ respectively ($q_i \leq Q, n_i \leq N$, for any $i = 1, \ldots, m$). Hence, the selection of the variables is multiple, because $2m$ procedures of selection are contemporary performed. Notice that the exogenous variables $w_t^i$’s in $W$ and $z_d^i$’s in $Z$ can be different, or be partially coincident, or be coincident in their whole.

In this section, we see how multiple variable selection can be performed, in the non-trivial case of correlated covariates, by means of a procedure based on the “Metropolization” of the approach proposed by Kuo and Mallick (1998), when $m$ and $p$ are fixed and the constraint on the variances is imposed.
For choosing the variables \( w_t^i \) and \( z_{t,j}^i \) to be included in model (2), a new setting must be introduced. Let

\[
Y_{t(i)} = \mu_i + \sum_{t=1}^{p} \varphi_{t(i)} Y_{t-\tau} + \sum_{j=1}^{Q} \theta_{j(i)} \bar{w}_{t,j} + \beta_{t(i)} + \varepsilon_{t(i)}
\]

be the NHMMPAR including, for any state \( i \), \( Q \) exogenous variables \( \bar{w}_{t,j} \), that is \( \bar{w}_t = (\bar{w}_{t,1}, \ldots, \bar{w}_{t,j}, \ldots, \bar{w}_{t,Q})' \) for any \( t = 1, \ldots, T \), and \( N - 1 \) exogenous variables \( z_{d,j} \), that is \( z_d = (1, z_{d,1}, \ldots, z_{d,j}, \ldots, z_{d,N-1})' \) for any \( d = 1, \ldots, D \), so that

\[
\logit(\gamma_{t,b}^d) = \bar{z}_d \alpha_{t,b}.
\]

Then, let \( \theta_i = (\theta_{1(i)}, \ldots, \theta_{j(i)}, \ldots, \theta_{Q(i)})' := (\lambda_1(i) \bar{\theta}_{1(i)}, \ldots, \lambda_j(i) \bar{\theta}_{j(i)}, \ldots, \lambda_{Q(i)}(i) \bar{\theta}_{Q(i)})' \) be the vector of coefficients of \( \bar{w}_t \), for any \( t \) and any \( i \in S_X \), where \( \bar{\theta}_{j(i)}' \)'s are real coefficients and \( \lambda_{j(i)}'s \) are binary coefficients, which can exclude variables \( \bar{w}_{t,j} \)'s when they assume value of zero or include them when they assume value of one, for any \( j = 1, \ldots, Q \).

After that, let \( \alpha_{t,b} = (\alpha_{t,b,0}, \alpha_{t,b,1}, \ldots, \alpha_{t,b,N-1})' := (\alpha_{t,b,0}, \eta_{1(i)} \bar{\alpha}_{t,b,1}, \ldots, \eta_{N-1(i)} \bar{\alpha}_{t,b,N-1})' \) be the vector of coefficients of \( \bar{z}_d \), for any \( d \) and any \( i \in S_X \), where \( \bar{\alpha}_{t,b,\ell} \)'s \( (\ell = 0, \ldots, N - 1) \) are real coefficients and \( \eta_{i(i)}'s \) are binary coefficients regulating the selection of the variables.

Finally, let \( \lambda \) and \( \eta \) be the vectors of binary coefficients, i.e. \( \lambda = (\lambda_1', \ldots, \lambda_j', \ldots, \lambda_m')' \) with \( \lambda_i = (\lambda_1(i), \ldots, \lambda_j(i), \ldots, \lambda_{Q(i)})' \) and \( \eta = (\eta_1', \ldots, \eta_{j'}, \ldots, \eta_{m'})' \) with \( \eta_i = (1, \eta_{1(i)}, \ldots, \eta_{N-1(i)})' \); \( \bar{\theta} \) be the vector of real coefficients, i.e. \( \bar{\theta} = (\bar{\theta}_1', \ldots, \bar{\theta}_j', \ldots, \bar{\theta}_m')' \) with \( \bar{\theta}_i = (\bar{\theta}_{1(i)}, \ldots, \bar{\theta}_{j(i)}, \ldots, \bar{\theta}_{Q(i)})' \) and \( \bar{\pi} \) be the matrix of real coefficients \( \bar{\pi}_{i,b} \) with \( \bar{\pi}_{i,b} = (\bar{\pi}_{i,b,0}, \bar{\pi}_{i,b,1}, \ldots, \bar{\pi}_{i,b,N-1})' \) and \( \bar{\pi}_i = (\bar{\pi}_{i,1}', \ldots, \bar{\pi}_{i,m}')' \).

The following priors are specified:

- independent multivariate normal priors of dimension \( Q \) with known vector \( \mu_{\bar{\pi}} \) and matrix \( \Sigma_{\bar{\pi}} \) on each row of matrix \( \bar{\theta} \) \( (\bar{\theta}_i \sim \mathcal{N}(\mu_{\bar{\pi}}, \Sigma_{\bar{\pi}}), \text{ for any } i \in S_X) \);
- independent multivariate normal priors of dimension \( N \) with known vector \( \mu_{\bar{\pi}} \) and matrix \( \Sigma_{\bar{\pi}} \) on each off-diagonal entry of matrix \( \bar{\pi} \) \( (\bar{\pi}_{i,b} \sim \mathcal{N}(\mu_{\bar{\pi}}, \Sigma_{\bar{\pi}}), \text{ for any } i; b \in S_X, \text{ with } i \neq b) \);
- independent Bernoulli with known probability \( p_A \) on each entry of matrix \( \lambda \) \( (\lambda_j(i) \sim B(\alpha_i), \text{ for any } i \in S_X \text{ and for any } j = 1, \ldots, Q) \);
- independent Bernoulli with known probability \( p_B \) on each entry of matrix \( \eta \) \( (\eta_{j(i)} \sim B(\alpha_i), \text{ for any } i \in S_X \text{ and for any } j = 1, \ldots, N - 1) \).

To perform variable selection, steps five and seven in the basic MCMC algorithm (Section 4) must be modified. This version of the algorithm updates both \( (\bar{\theta}_i; \lambda_i) \) and \( (\bar{\pi}_i; \eta_i) \) in block, for any \( i \in S_X \). Let \( (\bar{\theta}_i^{(k-1)}; \lambda_i^{(k-1)}) \) be the current state of the MCMC chain and \( (\bar{\theta}_i^{(k)}; \lambda_i^{(k)}) \) the candidate drawn from the proposal density \( q(\bar{\theta}_i^{(k)}; \lambda_i^{(k)} | \bar{\theta}_i^{(k-1)}; \lambda_i^{(k-1)}) \); \( (\bar{\theta}_i^{(k)}; \lambda_i^{(k)}) \) is accepted with probability

\[
A \left [ \left ( \bar{\theta}_i^{(k-1)}; \lambda_i^{(k-1)} \right ); \left ( \bar{\theta}_i^{(k)}; \lambda_i^{(k)} \right ) \right ] = \min \left \{ 1; \frac{\pi(\bar{\theta}_i^{(k)}; \lambda_i^{(k)}) q(\bar{\theta}_i^{(k)}; \lambda_i^{(k)} | \bar{\theta}_i^{(k-1)}; \lambda_i^{(k-1)})}{\pi(\bar{\theta}_i^{(k-1)}; \lambda_i^{(k-1)}) q(\bar{\theta}_i^{(k)}; \lambda_i^{(k)} | \bar{\theta}_i^{(k-1)}; \lambda_i^{(k-1)})} \right \}.
\]

(7)
By the factorization of the proposal density,
\[
q(\theta^{(k)}_i; \lambda^{(k)}_i | \theta^{(k-1)}_i; \lambda^{(k-1)}_i) = q(\lambda^{(k)}_i | \theta^{(k)}_i; \theta^{(k-1)}_i; \lambda^{(k-1)}_i) q(\theta^{(k)}_i | \theta^{(k-1)}_i; \lambda^{(k-1)}_i),
\]
and by the independence of \( \theta^{(k)}_i \) on \( \lambda^{(k-1)}_i \), the acceptance ratio in (7) becomes
\[
\frac{f(y_i^T | \theta^{(k)}_i; \lambda^{(k)}_i) p(\lambda^{(k)}_i) p(\theta^{(k)}_i) q(\lambda^{(k-1)}_i | \theta^{(k)}_i; \theta^{(k-1)}_i; \lambda^{(k-1)}_i) q(\theta^{(k-1)}_i | \theta^{(k)}_i; \lambda^{(k)}_i)}{f(y_i^T | \theta^{(k-1)}_i; \lambda^{(k-1)}_i) p(\lambda^{(k-1)}_i) p(\theta^{(k-1)}_i) q(\lambda^{(k)}_i | \theta^{(k-1)}_i; \theta^{(k)}_i; \lambda^{(k-1)}_i) q(\theta^{(k)}_i | \theta^{(k-1)}_i; \lambda^{(k-1)}_i)},
\]
suppressing all the conditioning on the remaining parameters and the exogenous variables.

The same holds for \((\alpha^{(k)}_i; \eta^{(k)}_i)\) and \((\alpha^{(k)}_i; \eta^{(k)}_i)\): the acceptance probability is
\[
A \left[ \left( \alpha^{(k-1)}_i; \eta^{(k-1)}_i \right); \left( \alpha^{(k)}_i; \eta^{(k)}_i \right) \right] = \min \left\{ 1; \frac{f(x^T | \alpha^{(k)}_i; \eta^{(k)}_i) p(\alpha^{(k)}_i) p(\eta^{(k)}_i) q(\eta^{(k-1)}_i | \alpha^{(k)}_i; \eta^{(k)}_i) q(\alpha^{(k-1)}_i | \alpha^{(k)}_i; \eta^{(k)}_i)}{f(x^T | \alpha^{(k-1)}_i; \eta^{(k-1)}_i) p(\alpha^{(k-1)}_i) p(\eta^{(k-1)}_i) q(\eta^{(k)}_i | \alpha^{(k-1)}_i; \eta^{(k-1)}_i) q(\alpha^{(k)}_i | \alpha^{(k-1)}_i; \eta^{(k-1)}_i)} \right\}.
\]

We assume the following proposal distributions: multivariate random walks for any \( \theta^{(k)}_i \) and any \( \alpha^{(k)}_{i,b} \) and products of independent Bernoulli for any \( \lambda^{(k)}_i \) and any \( \eta^{(k)}_i \).

Steps five and seven become the following.

5) The parameters \( \theta^{(k)}_i \), for any \( i \in S_X \), are independently generated from the random walk \( \theta^{(k)}_i = \theta^{(k-1)}_i + U_\Sigma \), where \( U_\Sigma \) is a multivariate Gaussian noise with zero mean and constant covariance matrix \( \Sigma \), for any \( k \). Then, every parameter \( \lambda^{(k)}_j \), for any \( j = 1, \ldots, Q \) and for any \( i \in S_X \), is independently generated from a Bernoulli distribution with probability \( p_j^{(k)} = c_j^{(k)}(n_{j(i)}) \), where, suppressing the conditioning on \( \mu^{(k)}, \sigma^{(k)}, \theta^{(k)}, \alpha^{(k-1)}, x^{(k)}, y^{(k-1)}, y^0, W, V \),
\[
c_j^{(k)} = p_\lambda \prod_{\{t=1, \ldots, 24+1\}} f(y_t | y^t_i, \theta^{(k)}_i; \lambda^{(k)}_i)
\]
and
\[
d_j^{(k)} = (1 - p_\lambda) \prod_{\{t=1, \ldots, 24+1\}} f(y_t | y^t_i, \theta^{(k)}_i, \lambda^{(k)}_i ; \lambda^{(k)}_i)
\]
with \( \theta^{(k)}_{ij} = (\theta^{(k)}_{ij(1)}, \ldots, \theta^{(k)}_{ij(1+i)}, \ldots, \theta^{(k)}_{ij(Q)}), \), \( \lambda^{(k)}_{ij} = (\lambda^{(k)}_{ij(1)}, \ldots, \lambda^{(k)}_{ij(1+i)}, \ldots, \lambda^{(k)}_{ij(Q)}) \), and \( \lambda^{(k)}_{ij} = (\lambda^{(k)}_{ij(1)}, \ldots, \lambda^{(k)}_{ij(1+i)}, \ldots, \lambda^{(k)}_{ij(Q)}) \). After that, any pair of vectors \( (\theta^{(k)}_i; \lambda^{(k)}_i), \) for any \( i \in S_X \), is accepted in block with probability (8), cancelling the ratio \( q(\theta^{(k-1)}_i | \theta^{(k)}_i) / q(\theta^{(k)}_i | \theta^{(k-1)}_i) \) for the symmetry of the proposal distribution.

7) The parameters \( \alpha^{(k)}_{a,b} \), for any \( a, b \in S_X \), with \( a \neq b \), are independently generated from the random walk \( \alpha^{(k)}_{a,b} = \alpha^{(k-1)}_{a,b} + U_\Sigma \), where \( U_\Sigma \) is a multivariate Gaussian
noise with zero mean and constant covariance matrix $\Sigma_{ll'}$ for any $k$. Then, every parameter $\eta^{(k)}_{(i)}$, for any $l = 1, \ldots, N - 1$ and for any $i \in S_X$, is independently generated from a Bernoulli distribution with probability $P^{(k)}_{(i)} = c^{(k)}_{(i)} / \left( c^{(k)}_{(i)} + d^{(k)}_{(i)} \right)$, where

$$c^{(k)}_{(i)} = p_H \delta_x \prod_{\{d>2x_{d-1}^i=1\}} \exp \left( \frac{\tilde{\alpha}^{(k)}_{(i)} \tilde{\alpha}^{(k)}_{(d)}}{\tilde{\alpha}^{(k)}_{(i)}} \right) \frac{1}{1 + \sum_{b \neq i} \exp \left( \frac{\tilde{\alpha}^{(k)}_{(i)} \tilde{\alpha}^{(k)}_{(b)}}{\tilde{\alpha}^{(k)}_{(i)}} \right)}$$

and

$$d^{(k)}_{(i)} = (1 - p_H) \delta_x \prod_{\{d>2x_{d-1}^i=1\}} \exp \left( \frac{\tilde{\alpha}^{(k)}_{(i)} \tilde{\alpha}^{(k)}_{(d)}}{\tilde{\alpha}^{(k)}_{(i)}} \right) \frac{1}{1 + \sum_{b \neq i} \exp \left( \frac{\tilde{\alpha}^{(k)}_{(i)} \tilde{\alpha}^{(k)}_{(b)}}{\tilde{\alpha}^{(k)}_{(i)}} \right)},$$

with $\tilde{\alpha}^{(k)}_{(i,b,l)} = (\tilde{\alpha}^{(k)}_{(i,b,0)}, \ldots, \tilde{\alpha}^{(k)}_{(i,b,k-1)}, \tilde{\alpha}^{(k)}_{(i,b,k+1)}, \ldots, \tilde{\alpha}^{(k)}_{(i,b,N-1)})'$, $\eta^{(k)}_{(i)} = \text{diag} \left( 1, \eta^{(k)}_{(i)}, \ldots, \eta^{(k)}_{(i)}, 1, \eta^{(k)}_{(i)}, \ldots, \eta^{(k)}_{(i)} \right)$, and $\eta^{(k)}_{(i)} = \text{diag} \left( 1, \eta^{(k)}_{(i)}, \ldots, \eta^{(k)}_{(i)}, 0, \eta^{(k)}_{(i)}, \ldots, \eta^{(k)}_{(i)} \right)$. After that, any pair of vectors $(\tilde{\alpha}^{(k)}_{(i)}; \tilde{\alpha}^{(k)}_{(i)})$, for any $i \in S_X$, is accepted in block with probability (9), cancelling the ratio $q \left( \tilde{\alpha}^{(k)}_{(i)} \left| \tilde{\alpha}^{(k)}_{(i)} \right. \right) / q \left( \tilde{\alpha}^{(k)}_{(i)} \left| \tilde{\alpha}^{(k)}_{(i)} \right. \right)$ for the symmetry of the proposal distribution.

At the end of the iterations of the MCMC algorithm, vectors $\lambda_i$ and $\eta_i$ are estimated as the maximizers of their respective posterior distributions and then

$$q_i = \sum_{j=1}^{Q} \lambda_{j(i)} \quad n_i = \sum_{j=0}^{N-1} \eta_{j(i)}$$

for any $i = 1, \ldots, m$. Hence, for any $i \in S_X$, vectors $w^t_i$ and $z^t_i$ are made up by the entries of vectors $(\tilde{\alpha}^{(k)}_{(i)}, \ldots, \tilde{\alpha}^{(k)}_{(i)}, \tilde{\alpha}^{(k)}_{(i)}, \ldots, \tilde{\alpha}^{(k)}_{(i)})' \tilde{\alpha}^{(k)}_{(i)} \tilde{\alpha}^{(k)}_{(i)} = (1, \eta^{(k)}_{(i)}, \ldots, \eta^{(k)}_{(i)})'$, corresponding to entries equal to one in vectors $\lambda_i$ and $\eta_i$, respectively.

### 6 Model choice

The dimensions of the NHMMPAR, i.e. $m$ and $p$, are chosen by comparing pairs of competing models (with different $m$ and/or $p$) through Bayes factors, in which the marginal likelihoods are computed numerically by an MCMC algorithm, based on the techniques proposed by Chib (1995) and Chib and Jeliazkov (2001).

Here, any block $W_t$ is a $(T \times Q)$ matrix, whose $t$-th row is $\tilde{W}_t$, and any block $Z_t$ is a $(D \times N)$ matrix, whose $d$-th row is $\tilde{\omega}_t$; the constraint on the variance is imposed.

The natural logarithm of the marginal likelihood, $\ln f (y^T \mid y^0, W, V, Z, \delta)$, is estimated in a special point $(\mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, \alpha^*)'$, the posterior mode of $(\mu, \sigma^{-2}, R, \theta, \beta, \alpha)'$; its estimate, denoted by $\ln \hat{f} (y^T \mid y^0, W, V, Z, \delta)$, is given by

$$\ln \hat{f} (y^T \mid y^0, W, V, Z, \delta) = \ln f (y^T \mid \mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, \alpha^*; y^0, W, V, Z, \delta) + \ln p (\mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, \alpha^*) - \ln \pi (\mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, \alpha^* \mid y^T, y^0, W, V, Z, \delta),$$

(10)
where \( \hat{\pi} \left( \cdot \mid y^T, y^0, W, V, Z, \delta \right) \) is an estimate of the value that the posterior assumes in the mode.

The algorithm introduced in Section 4 is identically used to estimate the various marginal likelihoods. It is run for \( N \) iterations to obtain the posterior mode \((\mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, \alpha^*)\) and to fill the series of the observations with the estimates of the missing values; then, at the end of the \( N \) iterations, the first two addenda in the r.h.s. of (10) are computed; finally, it is run for other \( 8N \) iterations to estimate the eight addenda in which \( \ln \hat{\pi} \left( \cdot \mid y^T, y^0, W, V, Z, \delta \right) \) can be decomposed, as we see.

The first expression in the r.h.s. of (10) is

\[
\ln f(y^T \mid \mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, \alpha^*, y^0, W, V, Z, \delta) = \\
= \sum_{d=1}^{24} \sum_{h=1}^{m} \ln \left( \sum_{i=1}^{m} f \left( y_{(d-1)24+h} \mid y_{(d-1)24+h-1}, \mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, x_d = i, y^0, W, V \right) \cdot \\
\cdot P \left( X_d = i \mid y^{24(d-1)}, \alpha^*, \mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, y^0, W, V, Z, \delta \right) \right),
\]

where \( P \left( X_d = i \mid y^{24(d-1)}, \alpha^*, \mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, y^0, W, V, Z, \delta \right) \), for any \( d = 1, \ldots, D \) and for any \( i = 1, \ldots, m \), is the filtered probability (see Section 4). The second expression of r.h.s. of (10) becomes

\[
\ln p(\mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, \alpha^*) = \sum_{i=1}^{m} \ln p(\mu^*_i) + \sum_{i=1}^{m} \ln p(\sigma^{-2*}_i) + \\
+ \sum_{i=1}^{m} \ln p(R^*_i) + \sum_{i=1}^{m} \ln p(\theta^*_i) + \sum_{i=1}^{m} \ln p(\beta^*_i) + \sum_{i=1}^{m} \ln p(\alpha^*_i).
\]

The third expression in the r.h.s. of (10) is

\[
\ln \hat{\pi} \left( \mu^*, \sigma^{-2*}, R^*, \theta^*, \beta^*, \alpha^* \mid y^T, y^0, W, V \right) = \\
= \ln \left[ \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{m} A \left( R^*_i ; R^*_i \right) \prod_{j=1}^{m} q \left( R^*_j \mid R^*_j \right) \right] - \ln \left[ \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{m} A \left( R^*_i ; R^*_i \right) \right] + \\
+ \ln \left[ \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{m} A \left( \alpha^*_i ; \alpha^*_i \right) \prod_{b \neq i} \prod_{b=1}^{m} q \left( \alpha^*_b \mid \alpha^*_b \right) \right] - \ln \left[ \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{m} A \left( \alpha^*_i ; \alpha^*_i \right) \right] + \\
+ \ln \left[ \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{m} \pi \left( \mu^*_i \mid y^T, y^0, W, V, \sigma^{-2(k)}, R^*, \beta^{(k)}, \theta^{(k)}, \alpha^*, x^{D(k)} \right) \right] + \\
+ \ln \left[ \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{m} \pi \left( \beta^*_i \mid y^T, y^0, W, V, \mu^*, \sigma^{-2(k)}, R^*, \beta^{(k)}, \alpha^*, x^{D(k)} \right) \right] + \\
+ \ln \left[ \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{m} \pi \left( \theta^*_i \mid y^T, y^0, W, V, \mu^*, \sigma^{-2(k)}, R^*, \beta^{(k)}, \alpha^*, x^{D(k)} \right) \right] + \\
+ \ln \left[ \frac{1}{N} \sum_{k=1}^{N} \prod_{i=1}^{m} \pi \left( \sigma^{-2*}_i \mid y^T, y^0, W, V, \mu^*, R^*, \beta^*, \theta^*, \alpha^*, x^{D(k)} \right) \right],
\]

and any addendum in the r.h.s. of (11) is computed by using \( N \) extra-iterations, labelled by \( k \), of the MCMC sampler, where \( q \left( R^*_j \mid R^*_j \right) \) is the probability density function (pdf) of the univariate normal with mean \( R^*_j \) and variance \( \sigma^2_{R^*_j} \) evaluated.
in $R_{j(i)}^*$;

$$A \left( R_i^{(k)}; R_i^* \right) = \min \left\{ \frac{\pi \left( R_i^* \mid \mu^{(k)}, \sigma^{(k)} \right)}{\pi \left( R_i \mid \mu^{(k)}, \sigma^{(k)} \right)}, 1; \frac{\pi \left( R_i \mid \mu^{(k)}, \sigma^{(k)} \right)}{\pi \left( R_i^* \mid \mu^{(k)}, \sigma^{(k)} \right)} \right\};$$

the acceptance ratio in $A \left( R_i^{(k)}; R_i^* \right)$ is the reciprocal of that in $A \left( R_i^*; R_i^{(k)} \right)$;

$$q \left( \alpha_{i,b}^* \mid \alpha_{i,b}^{(k)} \right)$$

is the multivariate normal pdf evaluated at $\alpha_{i,b}^*$, with mean $\alpha_{i,b}^{(k)}$ and covariance matrix $\Sigma_U$;

$$A \left( \alpha_{i}^{(k)}; \alpha_i^* \right) = \min \left\{ 1; \frac{\pi \left( \alpha_i^* \mid x_{D(k)}, Z, \delta \right)}{\pi \left( \alpha_i \mid x_{D(k)}, Z, \delta \right)} \right\};$$

the acceptance ratio in $A \left( \alpha_{i}^{(k)}; \alpha_i^* \right)$ is the reciprocal of that in $A \left( \alpha_{i}^*; \alpha_i^{(k)} \right)$.

Notice that all the values, labelled by $k$, are drawn from the posterior distribution, except those used to build $A \left( R_i^{(k)}; R_i^* \right)$ which are drawn from the univariate normals $R_{j(i)}^{(k)} = R_{j(i)}^* + U_R$, for any $j = 1, \ldots, p$ and $i = 1, \ldots, m$, and those used to build $A \left( \alpha_{i}^{(k)}; \alpha_i^* \right)$ which are drawn from the multivariate normals $\alpha_{i,b}^{(k)} = \alpha_{i,b}^* + U_A$, for any $i, b \in S_X$, with $i \neq b$.

### 7 Identifiability constraint specification

At the beginning of our analysis, we have to investigate the consistency of the cardinality of the state-space of the hidden Markov chain and to select some suitable identifiability constraint, which respects the geometry of the posterior distribution.

So, the MCMC algorithm described at Section 4, with matrices $W$ and $Z$ designed as in the previous section, is run unconstrained, that is, at any iteration, the variances are not reordered; by contrast, before generating the missing values, all the parameters are reordered according to a random permutation of hidden states (Frühwirth-Schnatter (2001)). This strategy allows to explore the whole set of possible labelings and to find a data-driven identifiability constraint that follows the shape of the posterior distribution: the pairs of signals and reciprocals of variances generated at any sweep of the sampler are plotted and then, by looking at the graphs, we check if there are as many groups as the hidden states and if these groups can suggest some special ordering in the parameters.

### 8 Results

The series of the natural logarithms of the SO$_2$ concentrations described in Section 2 is now analysed by means of the methodologies introduced in Sections 3-7, placing $T = 24072$, $D = 1003$ and $v = 1$, given that there is only one significant harmonic in the periodic component $\beta_{i(i)}$ (the autocorrelation function presents only one peak per period, as it is shown in Figure 1d).
Six exogenous meteorological variables, recorded any hour simultaneously to the SO₂ concentrations, are available (wind speed, temperature, rain, solar radiation, relative humidity, pressure) and their centered and scaled values are included, for any $t = 1, \ldots, T$, in $\mathbf{\tau}_t$, which is a vector of dimension $Q = 6$ (relative humidities have been standardized after the logit transformation). Moreover, for any $d = 1, \ldots, D$, in $\mathbf{\tau}_d$, which is a vector of dimension $N = 7$, we include, besides a first entry equal to one, the standardized values of the six meteorological variables, recorded in the hour in which the SO₂ has its daily maximum.

Figure 2: Signals and reciprocals of variances generated by the unconstrained MCMC algorithm with random permutations for $p=1$ and $m=2$ (a), $m=3$ (b), $m=4$ (c), $m=5$ (d), $m=6$ (e)

The following hyperparameters have been chosen for all the models and used in all the four steps of our empirical analysis, if not specified differently:

- $\mu_M = \ln(200/2)$ and $\sigma^2_M = 2.5$, i.e. the concentrations of SO₂ are in the middle of the tolerance interval defined by the attention-level (200 $\mu g/m^3$);
- $\alpha_{\Sigma} = \beta_{\Sigma} = 0.5$, i.e. each reciprocal of variances is assumed a priori to follow a gamma with mean 1 and variance 2, leading to low variability within each state;
- $\mu_R = 0$ and $\sigma^2_R = 10$, i.e the prior information on any $R_{ja(i)}$ is quite vague;
- $\mu_\Theta = 0_{(6)}$ and $\Sigma_\Theta = 10 \cdot I_{(6)}$, where $0_{(6)}$ is a 6-dimensional zero vector and $I_{(6)}$ is a 6-dimensional identity matrix, i.e the prior information on any $\theta_i$ is quite vague;
- $\mu_B = 0_{(2)}$ and $\Sigma_B = 10 \cdot I_{(2)}$, i.e the prior information on any $\beta_i$ is quite vague;
The first step of our analysis is the investigation of the consistency of the cardinality of the state-space of the hidden Markov chain and the specification of a data driven identifiability constraint. We start by considering eighteen competing models, which differ for the cardinality of the state-space of the hidden Markov chain \((m = 1, \ldots, 6, \text{where } m = 1 \text{ implies that no hidden chain underlies})\) and for the order of the autoregressive process \((p = 0, 1, 2; p > 2 \text{ is not consistent with physical theories on SO}_2 \text{ dynamics})\). By the graphical analysis of the pairs of signals and reciprocals of variances generated by the unconstrained MCMC algorithm, first we can notice that, for any \(p, m = 6\) is not consistent with the data we are studying, because six groups do not emerge in any plot (Figures 2e); hence we develop our analysis for \(m = 1, \ldots, 5\) only. Moreover, the constraint on the variances \((\sigma_i^2 < \sigma_j^2, \text{for any } i, j \in S_X, \text{ so that } i < j)\) is specified, because the increasing ordering is evident in any graph (Figures 2a-d).

Then we can compute the marginal likelihoods of the fifteen data-consistent competing models. By Table 1 (in Appendix) the model with four hidden states and autoregressions of the first order is chosen as the best among all the competing models.

After that we can select the exogenous variables which influence the observations and the time-varying transition probabilities of the chosen model. In this part of the analysis, we fix the hyperparameters \(\mu_{\Theta}, \Sigma_{\Theta}, \mu_{\Lambda}, \Sigma_{\Lambda}, p_{\Lambda} \text{ and } p_{H}\) at:

- \(\mu_{\Theta} = 0_{(6)}\) and \(\Sigma_{\Theta} = 10 \cdot I_{(6)}\), i.e. the prior information on any \(\theta_i\) is quite vague;
- \(\mu_{\Lambda} = 0_{(7)}\) and \(\Sigma_{\Lambda} = 10 \cdot I_{(7)}\), i.e the prior information on any \(\alpha_{a,i}\) is quite vague;
- \(p_{\Lambda} = p_{H} = 0.5\), i.e the inclusion and the exclusion of any exogenous variable are a priori equiprobable.

Exogenous variables are selected within a set made up by wind speed, temperature, rain, solar radiation, relative humidity and pressure. By means of Metropolized-Kuo-Mallick procedure, the meteorological variables included in the model for the last step of our analysis are:

- variables influencing the observations (W) - state 1: wind speed - state 2: temperature - state 3: wind speed - state 4: wind speed, temperature
- variables influencing the transition probabilities (Z) - state 1: temperature - state 2: temperature - state 3: solar radiation - state 4: temperature, pressure, solar radiation

We can notice that for any state \(i = 1, 2, 3\) only one covariate is selected in matrices \(W_i\)'s and \(Z_i\)'s, while two or three covariates are selected in matrices \(W_4\) and \(Z_4\), respectively. Remember that the meteorological variables influencing the observations are recorded any hour, while those influencing the transition probabilities are recorded in that hour of the day in which \(\text{SO}_2\) has its maximum daily concentration.

Finally, the parameters of the NHMMPAR with four hidden states and autoregressions of the first order are estimated by running the MCMC algorithm on the
constrained subspace, selected by $\sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$. Hyperparameters of vague priors $p(\theta_i)$ and $p(\alpha_a,i)$ have been fixed at $\mu_\Theta = 0_{(q_i)}$, $\Sigma_\Theta = 10 \cdot I_{(q_i)}$, $\mu_A = 0_{(n_i)}$, $\Sigma_A = 10 \cdot I_{(n_i)}$, where $q_i$ is 1 for states 1, 2, 3 and 2 for state 4, while $n_i$ is 2 for states 1, 2, 3 and 4 for state 4, respectively.

The estimates are reported in Table 2 (in Appendix). We can see the variances are ordered and increase as the hidden states increase, while the corresponding signals do not follow the same order. The daily dynamics of the series, described by the $\beta_t(i)$'s, presents a maximum at eight a.m. and a minimum at eight p.m., for states 1, 3 and 4, and a maximum at nine a.m. and a minimum at nine p.m., for state 2; moreover, the amplitudes of the harmonic components decrease as the hidden states increase (Figure 4a).

![Figure 3: The sequence of the fitted values](image)

The fitting ability of the model is satisfactory, as shown by the fitted series in Figure 3a, compared with the the observations in Figure 1b, and by the matching of three subseries of actual and fitted values (Figures 3b-3d).

The dynamics of the hidden states can be observed in Figure 4b: state 1 underlies the group of observations with the lowest variability, while state 4 underlies that with the highest variability. Hidden states are estimated by the modes of the corresponding smoothed probabilities, that is the probabilities of any state, at any time, given all the observations and the parameters. Smoothed probabilities are computed backwards, by using the algorithm of Kim (1993):

$$P(X_d = i \mid y^T) = P(X_d = i \mid y^{2d}) \sum_{j=1}^{m} \gamma_{i,j} P(X_{d+1} = j \mid y^{2d}) \frac{P(X_{d+1} = j \mid y^T)}{P(X_{d+1} = j \mid y^{2d})},$$
for any \( d = D - 1, \ldots, 1 \) and any \( i = 1, \ldots, m \), suppressing all the conditioning on the estimates of \( \mu, \sigma^{-2}, R, \theta, \beta, \alpha, y^* \) and on \( y^0, \delta, W, V, Z \). The algorithm is started from \( P(X_D = i \mid y^T) \), computed by means of the filtered probabilities (Section 4).

![Figure 4](image1)

(a) The daily component [solid line=state 1 - dashes=state 2 - dots=state 3 - dots and dashes=state 4] (a); the sequence of the hidden states (b)

![Figure 5](image2)

(a) Actual (triangles) and fitted (circles) values of days 200 (a) and 1001 (b)

Within the sequence of observations, we have 1785 missing values which can be grouped in three sets: 279 missing observations gathered in 143 small blocks of 9 data at maximum, 740 missing observations gathered in 24 medium blocks whose number of elements is between 11 and 59, 766 missing observations gathered in 6 huge blocks whose number of elements is between 83 and 158. Missing observations are simulated according to (6): we can see in Figures 5a and 5b that these simulated values correctly fill the series according to the dynamics of the twenty-four hours.

9 Conclusions

Non-homogeneous Markov mixtures of periodic autoregressions have been used to analyse a series of hourly mean concentrations of sulphur dioxide and to investigate what meteorological variables, among wind speed, temperature, rain, solar
radiation, relative humidity and pressure, influence the dynamics of the observed pollutant. This investigation has been developed by performing a multiple selection procedure for correlated exogenous variables, we called “Metropolized-Kuo-Mallick”, given that a Metropolis-step is introduced in the procedure proposed by Kuo and Mallick (1998). Alternative methods can be obtained by Metropolizing the techniques of George and McCullogh (1993) and those of Dellaportas et al (2000), even if both methods need pre-simulations for choosing the tuning factors that specify some hyperparameters.

The public authorities which placed the data to our disposal were prevalently interested in the selection of the meteorological variables, but we also performed the reconstruction of the sequence of the hidden states, the restoration of the missing values occurring within the observed series and the description of the periodic component.

Finally, this class of models can also be efficiently applied both in forecasting pollutant concentrations, by considering future values as missing, and in synthesizing air quality indices, by explaining the hidden states of the Markov chain as unobserved levels of pollution. Further extensions in the modelling can be developed by considering multivariate observations possibly recorded by a multisite monitoring net.

**Acknowledgements**

The data set analysed in the paper has been provided by “Laboratorio ARPAV Provinciale di Venezia”.

**References**


Association, 96, 194-209.

Appendix

<table>
<thead>
<tr>
<th>p \ 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>-37576.591</td>
<td>-32422.943</td>
<td>-30014.406</td>
<td>-28526.868</td>
<td>-27331.534</td>
</tr>
<tr>
<td>1</td>
<td>-23419.559</td>
<td>-18275.672</td>
<td>-17400.587</td>
<td><strong>17202.763</strong></td>
<td>-17434.625</td>
</tr>
<tr>
<td>2</td>
<td>-23115.978</td>
<td>-18058.984</td>
<td>-17341.055</td>
<td>-17401.190</td>
<td>-17587.905</td>
</tr>
</tbody>
</table>

Table 1: Marginal likelihoods of the competing NHMMPAR
$\mu_1 = 0.182$, $\mu_2 = 0.325$, $\mu_3 = 0.409$, $\mu_4 = 0.215$,
$\sigma_1^2 = 13.382$, $\sigma_2^2 = 4.618$, $\sigma_3^2 = 1.634$, $\sigma_4^2 = 0.581$
$(\sigma_1^2 = 0.075$, $\sigma_2^2 = 0.217$, $\sigma_3^2 = 0.612$, $\sigma_4^2 = 1.721)$,
$R_{1(1)} = 3.381$, $R_{1(2)} = 2.609$, $R_{1(3)} = 2.090$, $R_{1(4)} = 0.937$
($\varphi_{1(1)} = 0.934$, $\varphi_{1(2)} = 0.862$, $\varphi_{1(3)} = 0.780$, $\varphi_{1(4)} = 0.437$),
$\beta_{1,1(1)} = -0.034$, $\beta_{2,1(1)} = 0.046$, $\beta_{1,1(2)} = -0.063$, $\beta_{2,1(2)} = 0.081$,
$\beta_{1,1(3)} = -0.048$, $\beta_{2,1(3)} = 0.101$, $\beta_{1,1(4)} = -0.113$, $\beta_{2,1(4)} = 0.209$,
$\theta_{ws,1} = 0.004$, $\theta_{t,2} = -0.005$, $\theta_{ws,3} = -0.010$, $\theta_{ws,4} = -0.106$, $\theta_{t,4} = 0.022$,
$\alpha_{1,2,0} = -0.176$, $\alpha_{1,2,t} = 0.3682$, $\alpha_{1,3,0} = -1.758$, $\alpha_{1,3,sr} = -0.507$, $\alpha_{1,4,0} = -3.624$,
$\alpha_{1,4,t} = 1.065$, $\alpha_{1,4,p} = -1.310$, $\alpha_{1,4,sr} = -0.711$, $\alpha_{2,1,0} = -0.679$, $\alpha_{2,1,t} = -0.368$,
$\alpha_{2,3,0} = -0.593$, $\alpha_{2,3,sr} = -0.488$, $\alpha_{2,4,0} = -2.649$, $\alpha_{2,4,t} = 0.846$, $\alpha_{2,4,p} = -0.186$,
$\alpha_{2,4,sr} = -0.815$, $\alpha_{3,1,0} = -1.510$, $\alpha_{3,1,t} = -0.948$, $\alpha_{3,2,0} = 0.523$, $\alpha_{3,2,t} = -0.283$,
$\alpha_{3,4,0} = -1.099$, $\alpha_{3,4,t} = 0.404$, $\alpha_{3,4,p} = -0.408$, $\alpha_{3,4,sr} = -0.414$, $\alpha_{4,1,0} = -4.955$,
$\alpha_{4,1,t} = -0.135$, $\alpha_{4,2,0} = -0.651$, $\alpha_{4,2,t} = -0.322$, $\alpha_{4,3,0} = -0.196$, $\alpha_{4,3,sr} = 0.097$

Table 2: Estimates of the parameters of the NHMPAR with m=4 and p=1, where