Periodic Markov switching autoregressive models for Bayesian analysis and forecasting of air pollution
Luigi Spezia, Roberta Paroli and Petros Dellaportas
Statistical Modelling 2004; 4; 19
DOI: 10.1191/1471082X04st062oa

The online version of this article can be found at:
http://smj.sagepub.com/cgi/content/abstract/4/1/19

Published by:
© SAGE Publications
http://www.sagepublications.com

Additional services and information for Statistical Modelling can be found at:
Email Alerts: http://smj.sagepub.com/cgi/alerts
Subscriptions: http://smj.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 11 articles hosted on the SAGE Journals Online and HighWire Press platforms):
http://smj.sagepub.com/cgi/content/abstract/4/1/19#BIBL
Periodic Markov switching autoregressive models for Bayesian analysis and forecasting of air pollution

Luigi Spezia\textsuperscript{1}, Roberta Paroli\textsuperscript{2} and Petros Dellaportas\textsuperscript{1}
\textsuperscript{1}Department of Statistics, Athens University of Economics and Business, Athens, Greece
\textsuperscript{2}Istituto di Statistica, Università Cattolica S.C. di Milano, Milan, Italy

Abstract: Markov switching autoregressive models (MSARMs) are efficient tools to analyse nonlinear and non-Gaussian time series. A special MSARM with two harmonic components is proposed to analyse periodic time series. We present a full Bayesian analysis based on a Gibbs sampling algorithm for model choice and the estimations of the unknown parameters, missing data and predictive distributions. The implementation and modelling steps are developed by tackling the problem of the hidden states labeling by means of random permutation sampling and constrained permutation sampling. We apply MSARMs to study a data set about air pollution that presents periodicities since the hourly mean concentration of carbon monoxide varies according to the dynamics of the 24 day-hours and of the year. Hence, we introduce in the model both a hidden state-dependent daily component and a state-independent yearly component, giving rise to periodic MSARMs.

Key words: hidden Markov chain; harmonic components; Gibbs sampling; label switching; permutation sampling; carbon monoxide

Data and Programs available from: http://stat.uibk.ac.at/SMIJ
Received January 2003; revised September 2003; accepted October 2003

1 Introduction

Suppose we have a couple of discrete-time stochastic processes, \{X_t\} and \{Y_t\}, with the dynamics of \{X_t\} affecting the dynamics of \{Y_t\}. We can only observe the process \{Y_t\}, while the process \{X_t\} is unobservable; it is hidden in the observations. We model \{X_t\} as a finite-state Markov chain and assume that \{Y_t\}, given \{X_t\}, is an observed autoregressive process with the conditional distribution of \(Y_t\) depending on \{X_t\} only through the contemporary state of the Markov chain. This class of models is called Markov switching autoregressive models (MSARMs). A hidden Markov chain may be assumed because it is either suggested by the physical nature of the observed phenomenon or just for convenience in formulating the model.

Such MSARMs were originally studied by Hamilton (see, for example, Hamilton, 1994, Chapter 22 and references therein) to model economic and financial nonlinear
and non-normal time series (Franses and van Dijk, 2000; Kim and Nelson, 1999; Krozlig, 1997; for many applications and generalizations). Bayesian analysis of MSARMs has been developed, among others, by McCulloch and Tsay (1994), Chib (1996), Billio et al. (1999), Frühwirth-Schnatter (1999, 2001).

In this paper we propose MSARMs to analyse a data set about the hourly mean concentration of carbon monoxide (CO), recorded by an air pollution testing station located in Bergamo, Italy. The hidden states of the Markov chain represent unobserved levels of the observed process, depending on the weather conditions, since we have higher levels of CO in the colder periods of the year. The data set is also characterized by periodicities since the hourly mean concentrations vary according to both the yearly and daily dynamics. Hence, we introduce in the model both a yearly and a state-dependent daily component, giving rise to periodic MSARMs (PMSARMs).

Our inferential tool is Bayesian, so we introduce 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.

The paper makes the following distinct contributions. First, we introduce two harmonics components, one depending on the hidden chain and the other state-independent. We develop a block Gibbs sampling procedure to estimate the hidden chain that needs some extra care due to the identifiability constraints caused by the state-dependent harmonic component. We treat the number of the hidden states and the autoregressive order as unknown and we use Chib’s (1995) marginal likelihood approach together with the random permutation sampling of Frühwirth-Schnatter (2001) to make inference about the posterior probabilities of the number of states and the autoregressive order. We also tackle 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.

This paper is organized as follows. Some useful notes on the data will be given in Section 2; PMSARMs will be described in Section 3; Section 4 will be devoted to model choice, selection of the identifiability constraints and parameter estimation; finally, in Section 5 our methodology will be illustrated by analysing the hourly mean concentrations of CO.

2 Carbon monoxide

Air quality control includes the study of data sets recorded by air pollution testing stations. We consider one of the five stations located in Bergamo (a town in northern Italy, with 116 000 inhabitants). It records seven types of pollutants providing hourly mean concentrations of every pollutant. We are interested in the analysis of the dynamics of hourly mean concentrations of carbon monoxide (CO).

Carbon monoxide is produced in the combustion processes that take place in those situations with oxygen deficiency. Unlike other air pollutants, CO is mainly produced by petrol vehicular traffic and so it is released at ground level. It can reach very high concentrations in urban areas where traffic jams occur. Other sources of CO are domestic and industrial heating and industrial productive processes. It is noxious
because it permanently settles in blood haemoglobin, preventing it from its usual function of carrying oxygen; the central nervous system and the sense organs are the first to be damaged. The CO concentration is measured in milligrams per cubic meter (mg/m³) and the alarm and attention levels have been defined to be 30 and 15mg/m³, respectively. It is of great importance to the public authorities to be able to predict, as quickly as possible (for example, one day in advance), when these levels will be exceeded.

Figure 1(a) shows the series of hourly mean concentrations of CO recorded by the air pollution testing station located in Via San Giorgio, Bergamo, from 1 January 1997, 1 a.m., to 30 November 1999, 12 p.m. (25536 observations), whereas Figure 1(b) shows the daily mean concentrations obtained by the same data set (1064 observations). A yearly periodic component is evident in both figures and it is confirmed by the 19200 hours (that is, 800 days) autocorrelations (Figure 1(c)). As in finite mixture analysis, plotting a histogram of the values in Figure 1(b), we have a rough indication of the number of different unobserved levels. In Figure 1(d) we can guess the presence of hidden states by noticing an asymmetric and multipeak distribution.

By just looking at two generic subseries of observations, for example, the first 10 days of December 1998 (Figure 2(a)) and June 1999 (Figure 2(b)), we notice a daily periodicity with two peaks in rush hours, even if the two subseries present different levels of pollution due to the different climatic situations occurring in June and December. The daily periodicity is confirmed by the correlograms of five days of the two observed subseries of December 1998 and June 1999 (Figures 2(c) and 2(d)).

Notice that in Figures 1(a), 1(b), 2(a), 2(b) some observations are missing. This happens either because the station must be stopped every 72 hours for automatic

---

![Figure 1](http://smj.sagepub.com)

**Figure 1** Series of CO hourly mean concentrations (a); series of the CO daily mean concentrations (b); the 800 days autocorrelation, obtained by data plotted in Figure 1(a) (c); histogram of the CO hourly mean concentrations (d)
Figure 2  Subseries of the hourly mean concentrations of CO, recorded in the first 10 days of December 1998 (a) and June 1999 (b); five days autocorrelations of the same subseries recorded in December (c) and June (d)

calibration or because of occasional mechanical failure, ordinary maintenance, or data quality inspections.

The main characteristics of the series that must be modelled are:

1) serially correlated non-Gaussian data;
2) different unobserved levels of the hourly mean concentration of CO described by the hidden states of the Markov chain;
3) yearly and daily periodicities;
4) missing observations.

PMSARMs are appropriate model candidates since they incorporate the necessary flexibility to accommodate these characteristics.

The observations of the first two years will be used for our inference; the remaining eleven months data will be used to evaluate the forecasting performance of our model.

3  Periodic Markov switching autoregressive models for air pollution analysis

3.1  Model formulation

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; that is, 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 \( m \) (\( S_X = \{1, \ldots, m\} \)). Let also \( \Gamma = [\gamma_{ij}] \) be the \((m \times m)\) transition matrix, where \( \gamma_{ij} = P(X_t = j | X_{t-1} = i) \), for any \( i, j \in S_X \), and \( \delta = (\delta_1, \ldots, \delta_m)' \) is the stationary distribution, so that \( \delta' = \delta \Gamma \). Finally, let \( x^T = (x_1, \ldots, x_T)' \) be the sequence of the states of the Markov chain and, for any \( t = 1, \ldots, T \), \( x_t \) has values in \( S_X \).

Hence, given the order-\( p \) dependence and the contemporary dependence conditions, the equation describing PMSARMs of order \((m; p)\), henceforth PMSAR\((m; p)\), is

\[
Y_{t(x_i)} = \mu_{x_t} + \sum_{\tau=1}^{p} \phi_{\tau(x_i)} y_{t-\tau} + \eta_t + \beta_{\tau(x_i)} + E_{t(x_i)}
\]

(3.1)

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 \( \phi_{\tau(i)} \), 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 \( 2s \),

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

where \( s^* \) is the number of significant harmonics \((s^* \leq s)\); \( \beta_{\tau(x_i)} \) is a harmonic component of periodicity \( 2q \), depending on the current state \( i \) of the Markov chain,

\[
\beta_{\tau(i)} = \sum_{j=1}^{q^*} \left( \beta_{1,j(i)} \cos \frac{\pi jt}{q} + \beta_{2,j(i)} \sin \frac{\pi jt}{q} \right)
\]

where \( q^* \) is the number of significant harmonics \((q^* \leq q)\). \( E_{t(i)} \) denotes the Gaussian random variable \( E_t \) when \( X_t = i \), with zero mean and precision \( \lambda_i \), for any \( 1 \leq t \leq T \) and 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. From (3.1), the generic distribution of \( Y_{t(i)} \), given the \( p \) past observations and the current hidden state \( x_t = i \), is Gaussian with mean \( \mu_t + \sum_{\tau=1}^{p} \phi_{\tau(i)} y_{t-\tau} + \eta_t + \beta_{\tau(i)} \) and precision \( \lambda_i \). In our CO application we will fix \( s = 182 \) and \( s^* = 1 \) for the yearly harmonic component, which has one peak only (Figure 1(c)); we will fix \( q = 12 \) and \( q^* = 3 \) for the daily harmonic components, which has three peaks (Figures 2(c) and 2(d)).

Notice that the yearly harmonic component does not depend on the hidden Markov chain for identifiability reasons. If it did, in order to have an identified model, we would have assumed the same hidden state all along the period \( 2s \). By contrast, in order to let the daily harmonic component depend on the latent variables, we need to assume the same hidden state for all 24 hours of the day. Hence it serves 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 \( h = 1, \ldots, 24 \):

\[
Y_{[(d-1)24+b](x_d)} = \mu_{x_d} + \sum_{\tau=1}^{p} \phi_{\tau}(x_d)Y_{[(d-1)24+b-\tau]} + \eta_{(d-1)24+b} + \beta_{[(d-1)24+b](x_d)} + E_{[(d-1)24+b](x_d)}
\] (3.2)

A sufficient condition for the stationarity of the process (3.2) is that all the \( m \) subprocesses generated by the \( m \) states of the chain are stationary, that is, for any \( i \in S_X \), the roots of the auxiliary equations \( z^p - \phi_1(i)z^{p-1} - \cdots - \phi_p(i) = 0 \), where \( z \) is a complex variable, are all inside the unit circle.

The parameters to be estimated are the transition matrix \( \Gamma \), the stationary distribution \( \delta \), the vector \( \mu \) of the \( m \) parameters \( \mu_j \), the vector \( \lambda \) of the \( m \) parameters \( \lambda_j \), the matrix \( \phi = (\phi_1, \ldots, \phi_i, \ldots, \phi_m)' \) of the \( m \) autoregressive coefficients vectors \( \phi_i = (\phi_1(i), \ldots, \phi_i(i), \ldots, \phi_p(i))' \), the vector \( \eta = (\eta_{1,1}, \eta_{2,1}, \ldots, \eta_{1,s^+}, \eta_{2,s^+})' \) of the \( 2 \cdot s^+ \) state-independent harmonic coefficients and the matrix \( \beta = (\beta_1, \ldots, \beta_i, \ldots, \beta_m)' \) of the \( m \) state-dependent harmonic coefficient vectors \( \beta_i = (\beta_{1,1}(i), \beta_{2,1}(i), \ldots, \beta_{1,s^+}(i), \beta_{2,s^+}(i))' \). We also want to estimate the sequence of hidden states \( x^D = (x_1, \ldots, x_d, \ldots, x_D)' \), and all the missing observations \( y^v_{[(d-1)24+b]} \), collected in a vector \( y^v \). Using Tanner and Wong (1987) terminology, \( y^{24D} = (y_1, \ldots, y_{(d-1)24+b}, \ldots, y_{24b})' \) are 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; \( z = (x^D, y^v)' \) are the latent data and \( (y^{24D}, y^v, z)' \) 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 \( \delta = \delta \Gamma \)).

3.2 Prior specification

In this class of models we can place conditional independent conjugate priors on the unknown parameters. Notice that we need priors that are invariant to the relabeling of the states. For fixed \( m \) there are \( m! \) ways of labeling the states, which can alternate during the Gibbs sampling iterations, giving rise to the well-known label switching problem (Celeux et al., 2000; Frühwirth-Schnatter, 2001; Richardson and Green, 1997; Stephens, 2000). We tackle label switching by sampling from the posterior distribution, which must be invariant to the relabeling of the hidden states. Thus our prior distributions must also be relabeling-invariant.

Let \( D(\cdot) \) stand for the Dirichlet distribution, \( N(\mu; \lambda) \) stand for the normal distribution with precision \( \lambda \), \( G(\kappa; \beta) \) stand for the gamma distribution with mean \( \kappa / \beta \), and \( I(A) \) be the indicator function that takes the value of 1 if \( A \) is true and the value of 0 otherwise. We can elicit the following relabeling-invariant prior distributions.

Let \( \Gamma_i = (\gamma_{i,1}, \gamma_{i,2}, \ldots, \gamma_{i,m})' \) be the \( i \)th row of \( \Gamma \); \( \Gamma_i \sim D(\omega_{1,1}, \omega_{1,2}, \ldots, \omega_{1,m}) \), with \( \omega_{i,j} = m \cdot I(i = j) + 0.6 \cdot I(i \neq j) \), for any \( i, j = 1, \ldots, m \), that is, the probability of persistence is greater than the probability of transition; the probability of persistence is about 0.7 and it slowly decreases as the number of states increases (for \( m = 2, 3, 4, 5 \), \( E(\gamma_{i,i}) \)
are 0.77, 0.72, 0.70, 0.68, respectively, while \( E(\gamma_{1,y}) \) are 0.23, 0.14, 0.10, 0.08, respectively; \( \mu_i \sim \mathcal{N}(\mu_M; \lambda_M) \), for any \( i = 1, \ldots, m \), with \( \mu_M = \ln(15/2) \) and \( \lambda_M = 0.3 \), that is, the concentrations of CO are quite close to the attention level (15 mg/m³), because CO reaches high concentrations in the urban areas where traffic jams occur and the air pollution testing station from which our data have been recorded was placed in a heavy traffic road; \( \lambda_i \sim \mathcal{G}(\alpha_\Lambda; \beta_\Lambda) \), for any \( i = 1, \ldots, m \), with \( \alpha_\Lambda = \beta_\Lambda = 0.5 \), that is, each precision is assumed a priori to follow a gamma with mean 1 and variance 2, leading to low variability within each state; \( \phi_i \sim \mathcal{N}(\mu_\phi; \Lambda_\phi) \), for any \( i = 1, \ldots, m \), with \( \mu_\phi = 0(p) \), 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, that is, the autoregressive coefficients belong to a space circumscribing the stationarity region; \( \eta \sim \mathcal{N}(\mu_H; \Lambda_H) \) with \( \mu_H = 0(2r^s) \), \( \Lambda_H = 0.1 \cdot I(2r^s) \), that is, the prior information on \( \eta \) is quite vague; \( \beta_i \sim \mathcal{N}(\mu_B; \Lambda_B) \), for any \( i = 1, \ldots, m \), with \( \mu_B = 0(2q^r) \), \( \Lambda_B = 0.1 \cdot I(2q^r) \), that is, the prior information on \( \beta_i \) is quite vague.

### 3.3 Posterior distribution

Let \( \theta \) be the vector of both the unknown parameters and latent data of the PMSARM:

\[
\theta = (\Gamma, \mu, \lambda, \phi, \eta, \beta, x^D, y^*)^\prime
\]

The posterior distribution of \( \theta \) is

\[
\pi(\theta|y^{24D}, y^0, W, V) = f(\Gamma, \mu, \lambda, \phi, \eta, \beta, x^D, y^*|y^{24D}, y^0, W, V) \propto \\
f(y^{24D}, y^*|\mu, \lambda, \phi, \eta, \beta, x^D, W, V, y^0) f(x^D|\Gamma)p(\Gamma)p(\mu)p(\lambda)p(\phi)p(\eta)p(\beta),
\]

where \( W \) is a \((T \times 2s^*)\) matrix whose generic element on the \( t \)th row of the \( j \)th odd column is \( \cos(\pi jt/s) \), while the generic element on the \( t \)th row of the \( j \)th even column is \( \sin(\pi jt/s) \), for any \( j = 1, 2, \ldots, s^* \); \( V \) is an \( m \)-block matrix, \( V = (V_1, \ldots, V_i, \ldots, V_m) \), whose generic \( i \)th block \( V_i \) is a \((T \times 2q^r)\) matrix whose generic element on the \( t \)th row of the \( j \)th odd column is \( \cos(\pi jt/q) \), while the generic element on the \( t \)th row of the \( j \)th even column is \( \sin(\pi jt/q) \), for any \( j = 1, 2, \ldots, q^r \) and any \( i = 1, \ldots, m \).

The density of the observed and missing data is given by

\[
f(y^{24D}, |\mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, \Gamma^*, y^0, W, V) = \\
= \prod_{d=1}^{D} \prod_{b=1}^{24} \sum_{i=1}^{m} f(y_{(d-1)24+b}^{d-1} y_{(d-1)24+b-1}^{d-1}, \ldots, y_{(d-1)24+b-p}^{d-1}, \mu^*, \lambda^*, \phi^*, \eta^*_p, \beta^*, W, V) P(X_d = i|y_{d-1}^{d-1}, \Gamma^*, \mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, y^0, W, V),
\]

\[
(3.3)
\]
by the order-$p$ dependence and the contemporary dependence conditions, for any $d = 1, \ldots, D$ and any $h = 1, \ldots, 24$. Finally notice that

$$f(x^D | \Gamma) = \delta_{x_1} \prod_{d=2}^{D} \gamma_{x_{d-1}, x_d} = \delta_{x_1} \prod_{i=1}^{m} \prod_{j=1}^{m} \gamma_{ij}^D$$

by the Markov dependence condition, where $D_{ij}$ is the number of couples of consecutive hidden states $i, j$. In the right-hand side of (3.3) all missing observations have been replaced by the corresponding simulated value $y_{(d-1)24+h}^*$.

## 4 Bayesian inference

Our inference is developed along three consecutive steps: model choice, constraint identification, parameter estimation. These steps will be executed by means of a complete Gibbs sampling algorithm.

### 4.1 Model choice

The PMSAR model choice is based on the selection of the autoregressive order and of the cardinality of the hidden Markov chain state space. The selection of the best model is limited to $m = 1, \ldots, 4$ and $p = 0, 1, 2$: when $m$ is greater than four, the computing time becomes prohibitive given that the number of iterations is proportional to $m!$, while when $p$ is greater than two, the normalizing constants of the full conditionals generating the $\phi$, values are very complicated and cumbersome to code. Moreover, as we explained in Section 3.1, we fix $s^* = 1$ and $q^* = 3$.

#### 4.1.1 Chib’s marginal likelihood

Model choice is performed by means of Bayes factors (Kass and Raftery, 1995) in which the marginal likelihoods are computed according to Chib (1995) and corrected by the relabeling of the hidden states (Neal, 1999), through permutation sampling (Frühwirth-Schnatter, 2001).

Chib’s marginal likelihood method proceeds as follows. Assume that we have obtained a Gibbs sampler output of size $N$; the natural logarithm of the marginal likelihood at a point $(\mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, \Gamma^*)$, taken here to be the posterior mode of $(\mu, \lambda, \phi, \eta, \beta, \Gamma)$, can be estimated as $\ln \hat{f}(y^{24D} | y^0, W, V)$ by

$$\ln \hat{f}(y^{24D} | y^0, W, V) = \ln f(y^{24D} | \mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, \Gamma^*, y^0, W, V) + \ln p(\mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, \Gamma^*)$$

$$- \ln \hat{p}(\mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, \Gamma^* | y^{24D}, y^0, W, V)$$

(4.1)
The first expression in the right-hand sides of (4.1) is obtained by marginalizing over the sequence of the hidden states:

\[
f(y^{24D}, y^* | \mu, \lambda, \phi, \eta, \beta, \mathbf{x}, \mathbf{W}, \mathbf{V}, y^0) = \prod_{d=1}^{D-1} \prod_{b=1}^{24} f(y_{(d-1)24+b} | y_{(d-1)24+b-1}, \ldots, y_{(d-1)24+b-p}, \mu, \lambda, \phi, \eta_{(d-1)24+b}, \beta_{(d-1)24+b-1} x_d, y^0),
\]

where \( P(X_d = i | y^{d-1}, \Gamma^*, \mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, y^0, \mathbf{W}, \mathbf{V}) \), for any \( d = 1, \ldots, D \) and for any \( i = 1, \ldots, m \), is the filtered probability (see Appendix and Hamilton, 1994, pp. 692–93, for its computation). The second expression of the right-hand side of (4.1) becomes

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

and the third can be decomposed as

\[
\ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi(\mu^* | y^{24D}, y^0, \mathbf{W}, \mathbf{V}, \lambda^{(k)}, \phi^{(k)}, \eta^{(k)}, \beta^{(k)}, \Gamma^{(k)}, z^{(k)}) \right)
+ \ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi(\lambda^* | y^{24D}, y^0, \mathbf{W}, \mathbf{V}, \mu^*, \phi^{(k)}, \eta^{(k)}, \beta^{(k)}, \Gamma^{(k)}, z^{(k)}) \right)
+ \ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi(\phi^* | y^{24D}, y^0, \mathbf{W}, \mathbf{V}, \mu^*, \lambda^*, \eta^{(k)}, \beta^{(k)}, \Gamma^{(k)}, z^{(k)}) \right)
+ \ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi(\eta^* | y^{24D}, y^0, \mathbf{W}, \mathbf{V}, \mu^*, \lambda^*, \phi^*, \beta^{(k)}, \Gamma^{(k)}, z^{(k)}) \right)
+ \ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi(\beta^* | y^{24D}, y^0, \mathbf{W}, \mathbf{V}, \mu^*, \lambda^*, \phi^*, \eta^*, \Gamma^{(k)}, z^{(k)}) \right)
+ \ln \left( \frac{1}{N} \sum_{k=1}^{N} \pi(\Gamma^* | y^{24D}, y^0, \mathbf{W}, \mathbf{V}, \mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, z^{(k)}) \right)
\]

(4.3)

and estimated using \( 6 \cdot N \) extra-iterations, labeled by \( k \), of Gibbs sampling; for details, see Chib (1995).

When a current observation \( y^*_{(d-1)24+b} \) is missing, its corresponding probability density functions (PDFs) in (4.2) are replaced with 1, for any \( i = 1, \ldots, m \). Hence, we introduce in the recursive computations of the filtered probabilities (see Appendix and Hamilton, 1994, pp. 692–93) the powers of the transition matrix, which, by the Chapman–Kolmogorov equations, represent a skip of the missing observation. By
contrast, when missing observations occur among the \( p \) past observations, they are replaced by the expected values

\[
E(Y_{(d-1)24+b}|y^{d-1}, \Gamma^*, \mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, y^0, W, V) = \sum_{i=1}^{m} E(Y_{(d-1)24+b}|y^{d-1}, \Gamma^*, \mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, y^0, W, V, x_d = i) \times P(X_d = i|y^{d-1}, \Gamma^*, \mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, y^0, W, V)
\]

for any \( i = 1, \ldots, m \), where \( P(X_d = i|y^{d-1}, \Gamma^*, \mu^*, \lambda^*, \phi^*, \eta^*, \beta^*, y^0, W, V) \) is the filtered probability.

This completes the specificities of calculating the marginal likelihood.

### 4.1.2 Random permutation sampling

When the hidden chain is defined on a state space with cardinality \( m \), we have \( m! \) ways of labeling the states and hence the marginal likelihood is defined on a space that contains \( m! \) subspaces. In estimating marginal likelihood, Gibbs sampling with permutation sampling explores the whole support of the posterior distribution, improving the mixing property of the sampler. The Markov chain is free to move through the different subspaces, while permutation sampling encourages the moves between subspaces.

Notice that although the posterior density is invariant to the relabeling of the states, parameter estimation must be based on a special labeling. Chib (1995), in a finite mixture example, assumed that this ordering is a priori known (see also Neal, 1999). However, this is not the case here, so we use the ideas of Frühwirth-Schnatter (2001) as follows.

At the \( k \)th iteration of the Gibbs sampling we draw the sequence of the hidden states \((x_{1}^{(k)}, \ldots, x_{D}^{(k)})\)' and the parameters \( \Gamma^{(k)}, \mu^{(k)}, \lambda^{(k)}, \phi^{(k)}, \eta^{(k)}, \beta^{(k)} \). Then we randomly select a permutation \( (\rho(1), \ldots, \rho(m)) \) of the current labeling \((1, \ldots, m)\) and relabel the hidden states, \( (\rho(x_{1}^{(k)}), \ldots, \rho(x_{D}^{(k)})) \), and the switching parameters, \( \rho(\Gamma^{(k)}), \rho(\mu^{(k)}), \rho(\lambda^{(k)}), \rho(\phi^{(k)}), \rho(\beta^{(k)}) \). Finally, we generate the missing observations. During the Gibbs iterations needed for the calculation of the marginal likelihood, the entries of the posterior mode involved in the computation of (4.3) are permuted before the generation of the missing observations.

### 4.2 Constraint identification

We follow closely Frühwirth-Schnatter (2001) and provide guidelines on how to select one special labeling to perform our parameter estimation. This tackles the so-called label switching problem. The identifiability constraint must prevent label switching and so it must respect the geometry and the shape of the posterior distribution. The choice is made by first collecting all the draws of the switching parameters obtained during the \( N \) iterations of Gibbs sampling with random permutation. We can then derive a data-driven identifiability constraint by looking at the graphs of the output of the unconstrained Gibbs sampling performed associated with random permutation.
sampling: we plot couples of the generated parameters and then we check if there are
groups corresponding to the different states and if these groups can suggest special
ordering in the labeling. It is sufficient to plot the values corresponding to the first label
only, because it represents all the states, given the continuous jumping among all the
possible labels.

4.3 Parameter estimation and constrained permutation sampling

We estimate the unknown parameters of PMSARMs via a Gibbs sampling pro-
dure. The Appendix contains the details of the full conditionals; we just give a brief
discussion here.

Following Section 4.2, the Gibbs sampler is based on some identifiability constraint.
Without loss of generality, and since for our data set this constraint is based on the
precisions, see Section 5.2, we discuss our methodology assuming that the entries of $\lambda^{(k)}$
must be in increasing order. If $\lambda^{(k)}$ is not ordered, instead of rejecting the vector and
going on sampling until ordered vector is obtained, we adopt the constrained permutation sampling algorithm (Frühwirth-Schnatter, 2001). After the generation of
the sequence of the hidden states, we generate the vector of the precisions; so we have $m$
couples $(i, \lambda_i^{(k)})$. If the $\lambda_i^{(k)}$ values are unordered, we apply a permutation $\rho(\cdot)$ to order them; consequently also the corresponding is must be permuted according to the
permutation $\rho(\cdot), \rho(S_X) = \{\rho(1), \ldots, \rho(m)\}$; then the permutation $\rho(S_X)$ is extended to
the sequence of states $x^{D(k)}$ just generated, $\rho(x^{D(k)}) = \{\rho(x_1^{(k)}), \ldots, \rho(x_d^{(k)}), \ldots, \rho(x_D^{(k)})\}$,
and to the switching parameters generated in the previous iteration, $\rho(\Gamma^{(k-1)}), \rho(\mu^{(k-1)})$,
$\rho(\Phi^{(k-1)}), \rho(\beta^{(k-1)})$; finally all the other parameters and the missing observations are
generated as described in the Appendix.

5 Results

The series of the natural logarithms of the hourly mean concentrations of CO is
analysed here and the results we obtained performing model choice, constraint
identification, parameter estimation and forecasting are introduced. All the inference
is based on the data collected in the first two years of the series, while the predictions
made are referring to and checked against the last 11 months data. The code has been
implemented in Fortran, by using a sample of $5000 \cdot m$! successive values generated
after a $500 \cdot m$! burn-in period, except for predictions that have been obtained by means
of $1000 + 10\,000$ iterations.

5.1 Model choice

For each of the 12 PMSARMs considered here, their natural logarithms of the marginal
likelihoods are computed. Their values are presented in Table 1 and, using the
guidelines of Kass and Raftery (1995), the PMSAR(3;1) (shown in bold) is the best
among all the competing models.
Table 1  Natural logarithms of the marginal likelihoods

<table>
<thead>
<tr>
<th>ρ \ m</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-7085.156</td>
<td>-6765.429</td>
<td>-5738.714</td>
<td>-5400.767</td>
</tr>
<tr>
<td>1</td>
<td>-3578.737</td>
<td>-2175.933</td>
<td>-988.718</td>
<td>-1053.620</td>
</tr>
<tr>
<td>2</td>
<td>-4252.607</td>
<td>-3135.293</td>
<td>-3173.314</td>
<td>-2675.988</td>
</tr>
</tbody>
</table>

5.2 Constraint identification

Graphically analysing the outputs of the unconstrained PMSAR(3; 1), we can choose the constraint on the precision, λ₁ < λ₂ < λ₃, while no ordering is evident on the diagonal elements of Γ and on the entries of μ, φ and β (Figure 3). Increasing precisions is a reasonable constraint because it means that when the low hidden states occur, the variability of CO data depending on it is high and the level of pollution is also high (Figure 2(a)). By contrast, when the high hidden states occur, the variability of CO data depending on it is low and the level of pollution is also low (Figure 2(b)).

5.3 Parameter estimation

The estimates of the parameters of the hidden Markov chain are

\[
\Gamma = \begin{bmatrix}
0.647 & 0.224 & 0.129 \\
0.025 & 0.623 & 0.352 \\
0.021 & 0.132 & 0.847
\end{bmatrix}
\]

from which we have the estimate of the stationary initial distribution, \( \delta = (0.059; 0.271; 0.670)' \), so that \( \delta' = \delta' \Gamma \), while those of the parameters of the Gaussian PDFs are given in Table 2.

Figure 3  Some outputs of unconstrained Gibbs sampling with random permutations
Table 2  Posterior means of the PMSAR(3;1) model

<table>
<thead>
<tr>
<th>( i )</th>
<th>( \mu_i )</th>
<th>( \lambda_i )</th>
<th>( \phi_i )</th>
<th>( \beta_i )</th>
<th>( \eta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.497</td>
<td>2.293</td>
<td>0.162</td>
<td>-0.273</td>
<td>-0.251</td>
</tr>
<tr>
<td>2</td>
<td>0.117</td>
<td>7.073</td>
<td>0.795</td>
<td>-0.131</td>
<td>-0.078</td>
</tr>
<tr>
<td>3</td>
<td>0.185</td>
<td>16.157</td>
<td>0.776</td>
<td>-0.091</td>
<td>-0.086</td>
</tr>
</tbody>
</table>

Notice the (imposed) precision ordering and the fact that the variability within each state decreases as the unobserved level of pollution increases. Moreover, the CO yearly dynamics, described by \( \eta_{(d-1)24+b} \), respects the climatic conditions since higher levels of CO are obtained in the colder periods of the year (Figure 4(e)). The hourly dynamics of the series, described by \( \beta_{[(d-1)24+b](j)} \), respects rush hours obtaining two peaks at 8 a.m. and at 5 or 6 p.m., respectively (Figure 4(f)). In Figure 4(f) only two peaks appear,
possibly suggesting that the number of significant harmonics of the hourly component can be reduced to two. We feel that our model performs extremely well in terms of fitting (Figures 4(a)–(c)) for comparisons of two subseries of five days of actual and fitted values.

The dynamics of the hidden states, representing the three different levels of pollution occurring during the analysed period, can be observed in Figure 4(d), where we have depicted the sequence of the posterior modes of all generated states. State 1 underlies the observations with the highest level of pollution, while state 3 underlies those with the lowest level of pollution.

From the diagonal entries of the transition matrix, it is also possible to compute the time spent in each state $i$ of the Markov chain upon each return to it, which has a geometric distribution with mean $1/(1 - \gamma_{ii})$. The expected times spent in states 1–3 are 2.833, 2.659 and 6.536 days, respectively.

Within the sequence of 17520 observations, we have 1053 missing values, which can be grouped in four sets: 245 single missing observations, 57 missing observations gathered in 21 small blocks of almost 10 data points, 420 missing observations gathered in 11 medium blocks with number of points between 11 and 71, and two large blocks of 132 and 199 observations, respectively. Missing observations are simulated as extra latent variables; see Appendix for details. Figure 5 shows how simulated values fill the series according to the dynamics of two observed data subsets.

An interesting indicator of the fit of our model, which is strictly related to the real data problem we dealt with, is the ability to fit the points in which the attention level is exceeded. In the observed series there are 44 values greater than the attention level, split in 28 days. PMSAR (3; 1) captures one day out of 19 in which the threshold is exceeded once; two days out of three in which the threshold is exceeded twice; four days out of five in which the threshold is exceeded three times; one day out of one in which the threshold is exceeded four times.

Residual analysis has been graphically performed through histograms and QQ plots (Figure 6), showing that for any state the normality of the residuals is respected. The residual normality becomes more evident as the unobserved level rises, because the number of visits to the corresponding hidden state also increases.

![Figure 5](http://smj.sagepub.com) Actual (triangles) and fitted (circles) values of days 161 and 659
5.4 Forecasting

A principal aim of our analysis is to obtain the one-day ahead hidden state predictions and the $k$-hour ahead pollutant predictions ($1 \leq k \leq 24$). This problem is tackled through the usual Gibbs sampling machinery by considering future values as missing.

To evaluate the forecasting ability of our model we use the last 11 months of CO series to compare it with all the $k$-hour ahead predictions. In Figure 7(b) we can see that the dynamics of the 24-hour ahead forecasting respects the dynamics of the real data plotted in Figure 7(a) (see also Figures 7(c) and 7(d) for the comparisons of two subseries of five days of the data); this is true for all $k$-ahead predictions we tried.

We are also interested in the ability of the model to forecast the exceeding of the attention level (15 mg/m$^3$), given that the alarm level has never been reached. In the real data the attention level was exceeded in seven different days (736, 749, 750, 751, 752, 753, 755) with frequencies 1, 1, 4, 3, 2, 1, 2 hours per day, but the model fails to predict the spikes, which we suppose to depend on state 1, because, on the contrary, state 2 is estimated to underly these exceedings. Hence, to induce the hidden chain to visit state 1, when high concentrations must be predicted, it will be suitable in the future to introduce transition probabilities which depend on covariates.
6 Conclusions

The previously described empirical studies about air pollution show that Markov switching autoregressive models with harmonic components analyse and predict well periodic time series whose dynamics depend nonlinearly on latent variables. Model choice, inference and forecasting can be performed through Gibbs sampling, considering the label switching problem, which has been efficiently tackled by permutation sampling.

The models we considered can be extended in many directions, such as time-varying transition matrices, multivariate pollutants and multisites recording analysis. These extensions concern current research activities from the authors.

Acknowledgements

The research of the first author has been supported by the European Science Foundation Program on Highly Structured Stochastic System (HSSS), by the European Union ‘Training and Mobility of Researchers’ (TMR) network ERB-FMRX-CT96-0095 on ‘Statistical and Computational Methods for the Analysis of Spatial Data’ and by a Training Abroad Fellowship of Milan Catholic University. The authors are thankful to Leonhard Knorr-Held for his invaluable suggestions and to the associate editor and two referees for their useful comments, which led to a substantial improvement of the first version of this article.
References


Appendix

The complete expressions of the full conditionals are described here. Assume that we are in the kth iteration of Gibbs sampler and we have obtained the vector

\[ \theta^{(k-1)} = (\Gamma^{(k-1)}, \mu^{(k-1)}, \lambda^{(k-1)}, \phi^{(k-1)}, \eta^{(k-1)}, \beta^{(k-1)}, x^{D(k-1)}, y^{(k-1)}) \]

at the (k – 1)th iteration.

1) The sequence \( x^{D(k)} \) of hidden states is block-generated from the full conditional \( \pi(x^D | y^{24D}, \Gamma^{(k-1)}, \mu^{(k-1)}, \lambda^{(k-1)}, \phi^{(k-1)}, \eta^{(k-1)}, \beta^{(k-1)}, y^{(k-1)}, W, V, y^0) \), 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 forwards and then
the conditional probabilities of the hidden states are computed going backwards. The sampling of the states is based on the full conditionals derived according to Lemma 2.1 of Carter and Kohn (1994), based on Markov property and equality 22.A.13 by Hamilton (1994):

$$\pi(x^D|y^{24D}, \Gamma, \mu, \lambda, \phi, \eta, \beta, y^*, W, V, y^0) = \pi(x_D|y^{24D}, \Gamma, \mu, \lambda, \phi, \eta, \beta, y^*, W, V, y^0) \times \prod_{d=1}^{D-1} \pi(x_d|x_{d+1}, y^{24d}, \Gamma, \mu, \lambda, \phi, \eta, \beta, y^*, W, V, y^0)$$

Let $\xi_{d+1|d}$ be the $m$-dimensional vector whose generic entry is $P(X_{d+1} = i|y^{24d}, \Gamma, \mu, \lambda, \phi, \eta, \beta, y^*, W, V, y^0)$ for any $i = 1, \ldots, m$, $\xi_{d|d}$ be the $m$-dimensional vector whose generic entry is $P(X_{d} = i|y^{24d}, \Gamma, \mu, \lambda, \phi, \eta, \beta, y^*, W, V, y^0)$, for any $i = 1, \ldots, m$, and $\xi_d$ be the $m$-dimensional vector whose generic entry is $P(X_{d} = i|X_{d+1} = x_{d+1}, y^{24d}, \Gamma, \mu, \lambda, \phi, \eta, \beta, y^*, W, V, y^0)$, for any $i = 1, \ldots, m$. The iterative scheme of the FF-BS 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 $\Gamma^{(k-1)} = [\gamma_{ij}^{(k-1)}]$, associated with the eigenvalue equal to one.

1.2) Compute

$$\xi^{(k)}_{d|d} = \frac{\xi^{(k)}_{d|d-1}' F^{k-1}_d}{1_{(m)}(\xi^{(k)}_{d|d-1}' F^{k-1}_d)} \quad \text{and} \quad \xi^{(k)}_{d+1|d} = \Gamma^{(k-1)} \xi^{(k)}_{d|d}$$

for any $d = 1, \ldots, D-1$, where $F^{k-1}_d = \text{diag}[\prod_{b=1}^{24} f(y_{(d-1)24+b}, y_{(d-1)24+b-1}, \ldots, y_{(d-1)24+b-p}, \mu_{(k-1)}, \lambda_{(k-1)}, \phi_{(k-1)}, \eta_{(k-1)}, \beta_{(k-1)}, W, V, y^0, x_{(k-1)}^d = 1), \ldots, \prod_{b=1}^{24} f(y_{(d-1)24+b}, y_{(d-1)24+b-1}, \ldots, y_{(d-1)24+b-p}, \mu_{(k-1)}, \lambda_{(k-1)}, \phi_{(k-1)}, \eta_{(k-1)}, \beta_{(k-1)}, \mu_{(k-1)}, \lambda_{(k-1)}, \phi_{(k-1)}, \eta_{(k-1)}, \beta_{(k-1)}, W, V, y^0, x_{(k-1)}^d = m)]$ and $1_{(m)}$ is the $m$-dimensional vector of ones.

1.3) Compute

$$\xi^{(k)}_{D|D} = \frac{\xi^{(k)}_{D|D-1}' F^{k-1}_D}{1_{(m)}(\xi^{(k)}_{D|D-1}' F^{k-1}_D)}$$

(for details on the derivation of formulae at steps 1.2 and 1.3, see Hamilton (1994), pp. 692–93).

1.4) Generate $x^{(k)}_D$ from $\xi^{(k)}_{D|D}$. Let $\Xi^{(k)}_{D|D}(i)$ be the $i$th entry of vector $\Xi^{(k)}_{D|D}$, so that $\Xi^{(k)}_{D|D}(i) = \sum_{j=1}^{i} P(X_D = j|y^{24D}, \Gamma, \mu, \lambda, \phi, \eta, \beta, y^*, W, V, y^0)$, for any $i = 1, \ldots, m$. Then generate a random quantity $\nu$ from a uniform distribution on the unit interval $[0; 1]$: if $\Xi^{(k)}_{D|D}(i) < \nu \leq \Xi^{(k)}_{D|D}(i + 1)$, we obtain $x^{(k)}_D = i$. 

Downloaded from http://smj.sagepub.com at UCLA COLLEGE SERIALS/YRL on May 13, 2007 © 2004 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.
1.5) Compute

\[ \xi_{d}^{(k)} = \frac{\xi_{d|d}^{(k)} \Gamma_{d+1}^{(k-1)}}{1^{(m)} \xi_{d|d}^{(k)} \Gamma_{d+1}^{(k-1)}} \]

and generate \( x_{d}^{(k)} \) from \( \xi_{d}^{(k)} \), with the same procedure described at step 1.4, for any \( d = D - 1, \ldots, 1 \). \( \Gamma_{x_{d+1}}^{(k-1)} \) represents the column of \( \Gamma^{(k-1)} \) corresponding to the state previously generated.

2) Let \( \Gamma_{i} = (\gamma_{i,1}, \gamma_{i,2}, \ldots, \gamma_{i,m}) \), be the \( i \)th row of \( \Gamma \). Placing a Dirichlet prior with parameter \( \omega = (\omega_1, \ldots, \omega_m) \) on \( \Gamma_{i} \), each row \( \Gamma_{i}^{(k)} \), for any \( i \in S_{X} \), is independently generated from a Dirichlet \( \mathcal{D}(\omega + D_{l_{i}}^{(k)}) \), where \( D_{l_{i}}^{(k)} = (D_{l_{i,1}}^{(k)}, \ldots, D_{l_{i,m}}^{(k)}) \).

3) Placing a normal prior \( \mathcal{N}(\mu_{M}^{i}, \lambda_{M}^{i}) \) on any \( \mu_{i} \), the parameters \( \mu_{i}^{(k)} \), for any \( i \in S_{X} \), are independently generated from a normal distribution with mean

\[ \hat{\lambda}_{i}^{(k-1)} = \sum_{d} \sum_{\{ d : x_{d}^{(k)} = i \}}^{24} \left( y_{(d-1)24+b} - \sum_{\tau=1}^{p} \varphi_{\tau(i)}^{(k-1)} y_{(d-1)24+b-\tau} - \eta_{(d-1)24+b}^{(k-1)} - \beta_{[(d-1)24+b]i}^{(k-1)} + \mu_{M}^{i} \right) \]

and precision

\[ D_{l_{i}}^{(k)} \hat{\lambda}_{i}^{(k-1)} + \lambda_{M} \]

where \( D_{l_{i}}^{(k)} \) is the number of observations corresponding to the contemporary hidden state \( i \).

4) Placing a gamma prior \( \mathcal{G}(\alpha_{\Lambda}; \beta_{\Lambda}) \) on any \( \hat{\lambda}_{i} \), the parameters \( \lambda_{i}^{(k)} \), for any \( i \in S_{X} \), are independently generated from a gamma distribution with parameters

\[ \frac{D_{l_{i}}^{(k)}}{2} + \alpha_{\Lambda} \]

and

\[ \frac{1}{2} \sum_{\{ d : x_{d}^{(k)} = i \}}^{24} \left( y_{(d-1)24+b} - \mu_{i}^{(k)} - \sum_{\tau=1}^{p} \varphi_{\tau(i)}^{(k-1)} y_{(d-1)24+b-\tau} - \eta_{(d-1)24+b}^{(k-1)} - \beta_{[(d-1)24+b]i}^{(k-1)} \right)^2 + \beta_{\Lambda} \]

5) Placing a truncated multivariate normal prior of dimension \( p \) \( \mathcal{N}(\mu_{\phi}; \Lambda_{\phi}) \cdot I(\phi) \) on any \( \phi_{i} \), where \( I(\phi) \) is an indicator function so that

\[ I(\phi) = \begin{cases} 1 & \text{if the roots of the auxiliary equation are all inside the unit circle} \\ 0 & \text{otherwise} \end{cases} \]
the parameter vectors $\phi_{i}^{(k)}$, for any $i \in S_{X}$, are generated from a multivariate normal distribution of dimension $p$, under the stationarity constraint, with mean vector

$$[\lambda_{i}^{(k)}Z'Q_{i}^{(k)}Z + \Lambda_{\phi}]^{-1}[Z'Q_{i}^{(k)}(y^{24D} - \mu_{i}^{(k)}1_{T} - W\eta^{(k-1)} - V\beta_{i}^{(k-1)})\lambda_{i}^{(k)} + \Lambda_{\phi}\mu_{\phi}]$$

and precision matrix

$$\lambda_{i}^{(k)}Z'Q_{i}^{(k)}Z + \Lambda_{\phi}$$

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 $Q_{i}^{(k)}$ is a $(T \times T)$ diagonal matrix whose $t$th term is one if $x_{d}^{(k)}$ is $i$ or zero if it is not $i$, with $t = (d - 1)24 + b$ and $T = 24D$.

6) Placing a multivariate normal prior of dimension $2s^{*}N(\mu_{H}; \Lambda_{H})$ on $\eta$, the parameter $\eta^{(k)}$ is generated from a multivariate normal distribution of dimension $2s^{*}$ with mean vector

$$(W'\Lambda^{(k)}W + \Lambda_{H})^{-1}(W'\Lambda^{(k)}\tilde{y}^{24D(k)} + \Lambda_{H}\mu_{H})$$

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 $\lambda_{x_{d}(k)}$, $\tilde{y}^{24D}$ is a $24D$-dimensional vector whose generic $[(d - 1)24 + b]$-th element is $y_{(d-1)24+b} - \mu_{x_{d}(k)} - \sum_{t=1}^{p}\phi_{(x_{d}(k))}^{(k)}y_{(d-1)24+b-\tau} - \beta_{[(d-1)24+b](x_{d}(k))}^{(k-1)}$.

7) Placing a multivariate normal prior of dimension $2q^{*}N(\mu_{B}; \Lambda_{B})$ on any $\beta_{i}$, the parameters $\beta_{i}^{(k)}$, for any $i \in S_{X}$, are independently generated from a normal distribution of dimension $2q^{*}$ with mean vector

$$(V'\Lambda^{(k)}Q_{i}^{(k)}V + \Lambda_{B})^{-1}(V'\Lambda^{(k)}Q_{i}^{(k)}\tilde{y}^{24D} + \Lambda_{B}\mu_{B})$$

and precision matrix

$$V'\Lambda^{(k)}Q_{i}^{(k)}V + \Lambda_{B}$$

where $\tilde{y}^{24D}$ is a $24D$-dimensional vector whose generic $[(d - 1)24 + b]$-th element is $y_{(d-1)24+b} - \mu_{x_{d}(k)} - \sum_{t=1}^{p}\phi_{(x_{d}(k))}^{(k)}y_{(d-1)24+b-\tau} - \eta_{(d-1)24+b}^{(k)}$.

8) Every missing observation $y_{(d-1)24+b}^{*}$ is generated from the normal distribution

$$\mathcal{N}\left(\mu_{x_{d}(k)} + \sum_{\tau=1}^{p}\phi_{(x_{d}(k))}^{(k)}y_{(d-1)24+b-\tau} + \eta_{(d-1)24+b}^{(k)} + \beta_{[(d-1)24+b](x_{d}(k))}^{(k-1)} + \lambda_{x_{d}(k)}^{(k)} ; \lambda_{x_{d}(k)}^{(k)} \right) .$$