Abstract - We consider Markov switching autoregressive models to tackle the prediction problem of nonlinear time series with missing data. We propose an iterative procedure based on the EM algorithm to obtain the maximum likelihood estimators of the parameters of the model. We use a Monte Carlo procedure to reconstruct the sequence of the hidden states and then we restore the missing observations; this methodology allows us to forecast the future hidden states and the future values of the observed variable, too. This method is applied to analyse a time series of daily mean concentrations of sulphur dioxide; the different levels of the pollution are described by the dynamics of the hidden states.

Keywords - nonlinear time series, hidden Markov chain, EM algorithm, missing data.
Introduction

Markov Switching AutoRegressive (MSAR) models make up a class of models for nonlinear time series that present different behaviours according to the dynamics of hidden (or latent) random variables, usually known as state variables or regime variables. The MSAR models assume the dynamics of the regime variables is described by an unobservable Markov chain and the dynamics of the observed time series is modelled by an AR process, whose parameters depend on the states of the Markov chain. These models have been introduced in econometric literature by Hamilton (see for example Hamilton (1994), ch. 22 and the references therein) to study economic and financial time series with asymmetric cycles and changes in regime generated by a stochastic process (see Krolzig (1997), Kim and Nelson (1999), Franses and van Dijk (2000) for many applications and generalizations). These models are also an extension of hidden Markov models (MacDonald and Zucchini (1997)).

MSAR models of order \((m,p)\) are discrete-time stochastic processes \(\{Y_t; X_t\}\) such that \(\{X_t\}\) is an unobservable discrete Markov chain with a finite number of states, \(m\), while \(\{Y_t\}\), given \(\{X_t\}\), is an observed autoregressive process of order \(p\), with the conditional distribution of \(Y_t\) depending on \(\{X_t\}\) only through the contemporary \(X_t\).

Optimal inference and forecasting method for MSAR models have been discussed by Hamilton (1994), but without analysing the sequence of the hidden states. In this paper we propose an iterative procedure to reconstruct the hidden states and they will be used in making inference and forecasting. According to the optimal selected model, given the maximum likelihood estimators of the parameters, the sequence of the hidden states of the Markov chain is reconstructed by means of a Monte Carlo procedure; after that the missing observations occurring within the series can be restored computing the expected value of the conditional distribution of \(Y_t\) given the \(p\) past observations and the contemporary reconstructed state. The same procedure can be adopted to forecast the values of the observed variable.

The basic model used to study univariate stationary nonlinear time series will be introduced in Section 1; in Section 2, by means of the EM algorithm, the maximum likelihood estimators of the unknown parameters will be obtained and they will be used in Section 3 to reconstruct the hidden chain, to restore the missing observations and to forecast the future hidden states and the future values of the observed process. Finally, in Section 4, an environmental application will be shown: we will examine a data set about the daily mean concentrations of sulphur dioxide, recorded by an air pollution testing station in Bergamo, where we have 21 missing values within a sequence of 1064 observations.

1. The MSAR model

Let \(\{X_t\}_{t \in (1, \ldots, T)}\) be a discrete, homogeneous, aperiodic, irreducible Markov chain on a finite state-space \(S_X = \{1, 2, \ldots, m\}\). The transition probabilities matrix is \(\Gamma = [\gamma_{ij}]\), with \(\gamma_{ij} = P(X_t = j \mid X_{t-1} = i)\), for any \(i, j \in S_X\), and the initial distribution is \(\delta = (\delta_1, \delta_2, \ldots, \delta_m)\), where \(\delta_i = P(X_1 = i)\), for any \(i \in S_X\). Since \(\{X_t\}\) is a homogeneous, irreducible Markov chain, defined on a finite state-space, it has an initial distribution \(\delta\) which is stationary, that is \(\delta_t = P(X_t = i)\) for any \(t = 1, \ldots, T\); hence the equality \(\delta = \delta \Gamma\).

\(^1\) Work partially supported by MURST 2000 grant “Statistical Methods for Environmental Analysis”.
holds. Finally, the hypothesis characterizing this class of models is the unobservability of the Markov chain \( \{X_t\} \): the sequence of states of the Markov chain is hidden in the observations. Notice that the states of the chain can have either a convenient interpretation suggested by the nature of the observed phenomenon, or be supposed only for convenience in formulating the model.

Let \( \{Y_t; t \in \{1, \ldots, T\} \subseteq \mathbb{N} \) be some discrete stochastic process, on a continuous space \( S_Y = \mathbb{R} \). The process \( \{Y_t\} \) must satisfy two conditions: (i) \textit{contemporary dependence condition} - every observation depends on the contemporary state of the chain; (ii) \textit{order-p dependence condition} - the present observation depends on the \( p \) past observations. Hence the equation describing MSAR\((m, p)\) models is

\[
Y_{t(i)} = \mu_{t(i)} + \theta_{1(i)} Y_{t-1(i-1)} + \theta_{2(i)} Y_{t-2(i-2)} + \cdots + \theta_{p(i)} Y_{t-p(i-p)} + E_{t(i)},
\]

where \( Y_{t(i)} \) denotes the generic variable \( Y_t \) when \( X_t = i \), for any \( 1 \leq t \leq T \), and the autoregressive coefficients \( \theta_{h(i)} \), for any \( h = 1, \ldots, p \), depend on the current state \( i \) of the Markov chain. \( E_{t(i)} \) denotes the gaussian random variable \( E_t \) when \( X_t = i \), with zero mean and variance \( \sigma_i^2 \) \( (E_{t(i)} \sim \mathcal{N}(0, \sigma_i^2)) \), for any \( i \in S_X \), with the discrete process \( \{E_t\} \), given \( \{X_t\} \), satisfying the conditional independence and the contemporary dependence conditions.

As special cases of MSAR models, we have linear autoregressive models of order \( p \), AR\((p)\), when \( m = 1 \), and gaussian hidden Markov models, when \( p = 0 \).

From equation (1), the generic distribution of \( Y_{t(i)} \), given the \( p \) past observations and the current latent state, is gaussian with mean \( \mu_t + \sum_{h=1}^{p} \theta_{h(i)} y_{t-h} \) and variance \( \sigma_i^2 \),

\[
(Y_t \mid y_{t-1}, \ldots, y_{t-p}, i) \sim \mathcal{N}\left(\mu_t + \sum_{h=1}^{p} \theta_{h(i)} y_{t-h}; \sigma_i^2\right),
\]

with probability density function (pdf)

\[
f(y_t \mid y_{t-1}, \ldots, y_{t-p}, i) = \frac{1}{\sqrt{2\pi \sigma_i}} \exp\left[-\frac{1}{2} \left(\frac{y_t - \mu_t - \sum_{h=1}^{p} \theta_{h(i)} y_{t-h}}{\sigma_i}\right)^2\right],
\]

for any \( t = 1, \ldots, T \).

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

Hence the so-defined MSAR models \( \{Y_t; X_t \in \{1, \ldots, m\} \subseteq \mathbb{N}\) are characterized by the stationary initial distribution \( \delta \), by the transition probabilities matrix \( \Gamma \) and by the stationary pdf's \( f(y_t \mid y_{t-1}, \ldots, y_{t-p}, i) \).

### 2. Parameter estimation

In MSAR\((m, p)\) models the parameters to be estimated are the \( m^2 - m \) transition probabilities \( \gamma_{ij} \) for any \( i = 1, \ldots, m; j = 1, \ldots, m - 1 \) (the entries of the \( m \)th column of \( \Gamma \) are obtained by difference, given that each row of \( \Gamma \) sums up to one), the \( m \) entries of the vector \( \delta \), the \( m \) parameters \( \mu_\alpha \), entries of the vector \( \mu = (\mu_1, \ldots, \mu_m)' \), the \( m \) parameters \( \sigma_\alpha^2 \), entries of the vector \( \sigma^2 = (\sigma_1^2, \ldots, \sigma_m^2)' \) and the \( mp \) autoregressive coefficients \( \theta_{h(i)} \),
entries of the vector \( \theta = (\theta_{1(1)}, \ldots, \theta_{1(m)}, \ldots, \theta_{p(m)})^T \). The initial distribution \( \delta \) will be estimated by the equality \( \delta' = \delta' \), after the estimation of the matrix \( \Gamma \). Hence the vector of the \( m(m + p + 1) \) unknown parameters is

\[
\phi = \left( \gamma_{1,1}, \ldots, \gamma_{1,m-1}, \ldots, \gamma_{m,1}, \ldots, \gamma_{m,m-1}, \mu^T, \sigma^2, \theta' \right)^T,
\]

belonging to the parameter space \( \Phi \). The estimator of the vector \( \phi \) will be obtained by the maximum likelihood (ML) method.

Let \( y_T = (y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T)^T \) be the vector of the observed data, that is the sequence of the realizations of the stochastic process \( \{Y_t\} \), where \( (y_{-p+1}, \ldots, y_0) \) are initial values, fixed for the order \( p \) dependence condition; the data vector is incomplete because the sequence \( x_T = (i_1, \ldots, i_T)^T \) of the states of the chain \( \{X_t\} \) is unobserved; the complete data vector will be \( (y_T^T, x_T^T) = (y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T, i_1, \ldots, i_T)^T \).

Moreover let \( L_T(\phi) \) be the likelihood function of the complete data and \( L_T(\phi) \) be that of the observed data, given the \( p \) initial values:

\[
L_T(\phi) = f(y_1, \ldots, y_T, i_1, \ldots, i_T \mid y_{-p}, \ldots, y_{-p+1}) = \\
= \delta_{i_1} f(y_1 \mid y_{-p}, \ldots, y_{-p+1}, i_1) \prod_{t=2}^T \gamma_{i_{t-1},i_t} f(y_t \mid y_{t-1}, \ldots, y_{t-p}, i_t);
\]

\[
L_T(\phi) = f(y_1, \ldots, y_T \mid y_{-p}, \ldots, y_{-p+1}) = \\
= \sum_{i_1 \in S_X} \ldots \sum_{i_T \in S_X} \delta_{i_1} \left[ \prod_{t=1}^{T-1} \gamma_{i_{t-1},i_t} \prod_{t=1}^T f(y_t \mid y_{t-1}, \ldots, y_{t-p}, i_t) \right].
\]

The maximization of the incomplete log-likelihood function, \( \ln L_T(\phi) \), is a non-linear optimization problem, so it requires numerical techniques. For the MSAR models, the most popular numerical method to obtain the ML estimators is the Expectation-Maximization (EM) algorithm (Dempster, Laird, Rubin (1977)) in conjunction with the iterative algorithms of Hamilton (1994) and Kim (1993) for computing the filtered and the smoothed probabilities of the hidden states given the observations.

The EM algorithm is an iterative ML estimation technique, designed for a more general class of data-incomplete problems. The iterative scheme is the following: let \( \phi^{(0)} \) be some starting value of \( \phi \); at the first iteration, the E step requires the computation of the conditional expectation of the complete data log-likelihood, given the observed data, in \( \phi = \phi^{(0)} \) and the M step provides the search for that special value \( \phi^{(1)} \) which maximizes the previous function. In general, at the \( (k + 1) \)th iteration, the E and M steps are so defined:

**E step** - given \( \phi^{(k)} \), compute \( Q(\phi; \phi^{(k)}) = E_{\phi^{(k)}} (\ln L_T(\phi) \mid y_T^T) \);

**M step** - search for that \( \phi^{(k+1)} \) which maximizes \( Q(\phi; \phi^{(k)}) \), for any \( \phi \in \Phi \).

The E and M steps must be repeated in an alternating way until we have the convergence of the sequence of the log-likelihood values \( \left\{ \ln L_T(\phi^{(k)}) \right\} \). If the convergence is reached at the \( (k + 1) \)th iteration, the vector \( \phi^{(k+1)} = \hat{\phi} \) is then the ML estimator of \( \phi \). The main property of the EM algorithm is the monotonicity of the log-likelihood function for incomplete data, with respect to the iterations of the algorithm (Dempster, Laird, Rubin (1977), Theorem 1).

Nevertheless, if the likelihood surface is multimodal, the convergence of the EM algorithm to the global maximum depends on the starting value \( \phi^{(0)} \). To avoid the convergence
to a stationary point which is not a global maximum, the best strategy is to start the algorithm from several different, possibly random, points in $\Phi$ and to compare the stationary points obtained at each run.

As shown in Hamilton (1994), to obtain the ML estimators of the parameters of MSAR models, the EM algorithm can be used in conjunction with the filtering and the smoothing algorithms. The basic filtering algorithm is an iterative procedure to compute the estimates of the probabilities of being in regime $i$ at time $t$ or $t + 1$, given the informations available up to time $t$, for any $i \in S_X$. Let $y^t = (y_1, \ldots, y_t)$ be the vector of the observations up to time $t$, $\xi_{t+1|t}$ and $\xi_{t|t}$ respectively be the $m$-dimensional vectors of the conditional probabilities $P(X_{t+1} = i \mid y^t; \phi)$, $i = 1, \ldots, m$, $t = 1, \ldots, T - 1$, and $P(X_t = i \mid y^t; \phi)$, $i = 1, \ldots, m$, $t = 1, \ldots, T$. These vectors are recursively computed starting from $\xi_{1|0} = \delta' = \delta' \Gamma$:

$$\xi_{t|t} = \frac{\xi_{t|t-1} \odot F_t}{V_{(m)} (\xi_{t|t-1} \odot F_t)}$$

(3)

$$\xi_{t+1|t} = \Gamma' \xi_{t|t}$$

(4)

where $F_t$ denotes the $m$-dimensional vector containing the conditional pdf's (2),

$$F_t = (f(y_t \mid y_{t-1}, \ldots, y_{t-p}, i_t = 1), \ldots, f(y_t \mid y_{t-1}, \ldots, y_{t-p}, i_t = m)),$$

$1_{(m)}$ is the $m$-dimensional vector of ones and the symbol $\odot$ indicates the Hadamard product (Hamilton (1994)).

The smoothing algorithm allows us to compute the estimates of the probabilities of being in regime $i$ at time $t$, given all the available informations, $y^T$, for any $i \in S_X$. Let $\xi_{i|T}$ be the $m$-dimensional vector of the conditional probabilities $P(X_t = i \mid y^T; \phi)$, $i = 1, \ldots, m$, $t = 1, \ldots, T$; these vectors are recursively computed backward starting from $\xi_{T|T}$, obtained by (3):

$$\xi_{i|T} = \xi_{i|t} \odot \left[ \Gamma (\xi_{i+1|T} \div \xi_{i+1|t}) \right],$$

(5)

for $t = T - 1, T - 2, \ldots, 1$, where the symbol $\div$ indicates element-by-element division (Kim (1993)).

Finally, the incomplete log-likelihood function can be computed:

$$\ln L_T(\phi) = \sum_{t=1}^{T} \ln f(y_t, y_2, \ldots, y_T \mid y_0, \ldots, y_{t-p}) = \sum_{t=1}^{T} \ln (1_{(m)} (\xi_{t|t-1} \odot F_t))$$

(Hamilton (1994)).

Now the two steps of the EM algorithm at the $(k + 1)^{th}$ iteration are analysed in details, remembering that at the $k^{th}$ iteration the vector of estimates $\phi^{(k)}$ has been obtained,

$$\phi^{(k)} = \left( \gamma_1^{(k)}, \ldots, \gamma_{l-1}^{(k)}; \gamma_{l}^{(k)}; \ldots, \gamma_{m-1}^{(k)}, \gamma_{m}^{(k)}; \gamma_{m}^{(k)}; \mu^{(k)}; \sigma^{(k)}; \rho^{(k)} \right)^T,$$

and the filtered and smoothed probabilities have been computed. Henceforth the superscript $(k)$ will denote a quantity obtained at the $k^{th}$ iteration as a function of the vector $\phi^{(k)}$; notice that $\delta^{(k)}$ is the left eigenvector of the matrix $\Gamma^{(k)} = \left[ \gamma^{(k)}_{t|t} \right]$, associated with the eigenvalue one, such that $\delta^{(k)} = \delta^{(k)} \Gamma^{(k)}$. 

5
At the E step of the \((k+1)th\) iteration, the analytic closed-form of the function 

\[ Q(\phi; \phi^{(k)}) \]

is

\[
Q(\phi; \phi^{(k)}) = E_{\phi^{(k)}} \{ \ln L_{\phi}(\phi) \mid y^T \} = \sum_{i} \frac{\ell_{(i)}^{(k)}(y_{i+1} \ldots Y_{T})}{\ell_{(i)}^{(k)}(y_{1} \ldots i)} \ln \delta_{i} + \\
+ \sum_{i} \sum_{j} \left[ \frac{\sum_{t=1}^{T-1} \ell_{(i)}^{(k)}(y_{i+1} \ldots Y_{t})}{\ell_{(i)}^{(k)}(y_{1} \ldots i)} \ln \gamma_{i,j} \right] + \\
+ \sum_{i} \left[ \frac{\sum_{t=1}^{T-1} \ell_{(i)}^{(k)}(y_{i+1} \ldots Y_{t})}{\ell_{(i)}^{(k)}(y_{1} \ldots i)} \ln f(y_{t} \mid y_{t-1}, \ldots, y_{t-p}, i) \right].
\]

Using the filtered and smoothed probabilities, computed at the \(k^{th}\) iteration, we obtain

\[
Q(\phi; \phi^{(k)}) = \sum_{i} \ell_{(i)}^{(k)}(i) \ln \delta_{i} + \sum_{i} \sum_{j} \left[ \frac{\sum_{t=1}^{T-1} \ell_{(i)}^{(k)}(y_{i+1} \ldots Y_{t})}{\ell_{(i)}^{(k)}(y_{1} \ldots i)} \right] \ln \gamma_{i,j} + \\
+ \sum_{i} \sum_{t=1}^{T-1} \ell_{(i)}^{(k)}(i) \ln f(y_{t} \mid y_{t-1}, \ldots, y_{t-p}), i).
\]

At the M step of the \((k+1)th\) iteration, to obtain \(\phi^{(k+1)}\), the function \(Q(\phi; \phi^{(k)})\) must be maximizing with respect to the \(m^{2} - m\) parameters \(\gamma_{i,j}\), for any \(i = 1, \ldots, m; j = 1, \ldots, m - 1\), the \(m\) parameters \(\mu_{i}\), the \(m\) parameters \(\sigma_{i}^{2}\), for any \(i \in S_{X}\), and the \(mp\) parameters \(\theta_{h}(\phi)\), for any \(i = 1, \ldots, m; h = 1, \ldots, p\). We shall not maximize \(Q(\phi; \phi^{(k)})\) with respect to the \(m\) parameters \(\delta_{i}\), for any \(i \in S_{X}\), because, as we said previously, the initial distribution \(\delta\) will be estimated by the equality \(\delta = \delta^{T}\), after the estimation of the matrix \(\Gamma\). But, by the stationarity assumption, \(\delta\) contains informations about the transition probabilities matrix \(\Gamma\), since \(\delta_{j} = \sum_{i \in S_{X}} \delta_{i} \gamma_{i,j}\), for any \(j \in S_{X}\). However, for large \(T\), the effect of \(\delta\) is negligible; so the first term of the function \(Q(\phi; \phi^{(k)})\) can be ignored searching for the maximum likelihood estimator of \(\gamma_{i,j}\), for any \(i, j\) (Basawa and Prakasa Rao (1980), pp. 53-54).

In practice the M step can be performed in two ways: numerically, using a standard numerical optimization procedure (like Newton-Raphson method) or, when possible, analytically, obtaining the closed-form solutions of the parameter estimators.

Notice that the analytic expression of \(Q(\phi; \phi^{(k)})\) is the sum of three terms: the first two are functions only of the parameters of the Markov chain, while the third is a function only of the parameters of the pdfs (2). This separation of parameters makes the maximization problem into a simple closed-form, that gives the explicit solutions of the estimators of \(\gamma_{i,j}, \sigma_{i}^{2}\), while for the other estimators, the entries of the vectors \(\mu\) and \(\theta\), the solutions must be obtained using the Newton-Raphson method. The explicit expression of the estimators of the parameters \(\gamma_{i,j}\), \(i = 1, \ldots, m; j = 1, \ldots, m - 1\), at the \((k+1)th\) iteration is

\[
\gamma_{i,j}^{(k+1)} = \frac{\sum_{t=1}^{T-1} \ell_{(i)}^{(k)}(y_{i+1} \ldots Y_{t}) \xi_{i,t}^{(k)(i)}}{\sum_{t=1}^{T-1} \ell_{(i)}^{(k)(i)}}
\]
while that of the estimators of the parameters $\sigma_i^2$, $i = 1, \ldots, m$, is

$$\sigma_i^{(k+1)} = \frac{\sum_{t=1}^{T} \xi_{IT}^{(k)}(i) \left( y_t - \mu_i^{(k+1)} - \sum_{h=1}^{p} \theta_{h(i)}^{(k+1)} y_{t-h} \right)^2}{\sum_{t=1}^{T} \xi_{IT}^{(k)}(i)}$$

performed after having numerically obtained the values of $\mu_i^{(k+1)}$ and $\theta_{h(i)}^{(k+1)}$, for $i = 1, \ldots, m$ and $h = 1, \ldots, p$.

In the case of convergence of the algorithm at the $(k+1)$th iteration, $(\phi^{(k+1)}; \ln L_T (\phi^{(k+1)}))$ is a stationary point of $\ln L_T (\phi)$, and therefore $\phi^{(k+1)}$ is the ML estimator of the unknown parameter $\phi$. Assuming $\sigma_i \in [\varepsilon, 1/\varepsilon]$, for any $i \in S_X$ and for some small $\varepsilon > 0$, this fundamental result holds because the four regularity conditions on the convergence of the EM algorithm to a stationary point (Wu (1983), conditions (5), (6), (7), p. 96; (10), p. 98) are satisfied.

As Douc, Moulines, Rydén (2001) proved, the maximum likelihood estimator of $\phi$ is strongly consistent and asymptotically normal.

3. Missing restoration and forecasting

One of the major aims in time series analysis is forecasting the future values of an observed process. Computing forecasts for nonlinear time series is generally more complex than for the linear ones, because the non linearity of the function that defines the model makes the optimal predictor not to be linear, but the case of one-step-ahead forecasting. Hence we perform the $h$-step-ahead ($h > 0$) forecasting in the "naive" way, iterating $h$ times the one-step-ahead forecasting and using every time the forecasted value as the last observation.

For the MSAR models the one-step-ahead forecast of the future values of the observed series can be decomposed into the forecast of the conditional observed variable given the $p$ past observations and the current state of the Markov chain and the estimated probabilities of occurrence of each state. Let $\hat{y}_{T+1}$ be the one-step-ahead forecasting value:

$$\hat{y}_{T+1} = E(Y_{T+1} \mid y_T, \ldots, y_{T-p+1}) = \sum_{i=1}^{m} E(Y_{T+1} \mid y_T, \ldots, y_{T-p+1}, X_{T+1} = i) P(X_{T+1} = i \mid y_T),$$

where the forecast of the conditional variable $Y_{T+1}$ given the $p$ past observations and the current state of the Markov chain is the expected value

$$E(Y_{T+1} \mid y_T, \ldots, y_{T-p+1}, X_{T+1} = i) = \mu_i + \sum_{h=1}^{p} \theta_{h(i)} y_{T-h+1},$$

whereas $P(X_{T+1} = i \mid y_T)$, for any $i \in S_X$, are given by the filtered probabilities (Hamilton (1994)).

We now introduce an alternative forecasting method based on the reconstruction of the sequence of the hidden states. The sequence $x^T = (i_1, \ldots, i_T)$ of the hidden states can be
generated by Monte Carlo simulations through the forward filtering-backward sampling (FF-bs) algorithm (Frühwirth-Schnatter (1994)), based on the equality

\[
P(X_1 = i_1, \ldots, X_T = i_T \mid y^T) = P(X_T = i_T \mid y^T) \prod_{t=1}^{T-1} P(X_t = i_t \mid X_{t+1} = i_{t+1}, y^T).
\]

Given \( \hat{\phi} \), the vector of the ML estimators of the parameters, the iterative scheme of the algorithm is the following:

(i) for \( t = 1, \ldots, T \), compute the vectors of the filtered probabilities \( \hat{\xi}_{it} = P(X_t = i \mid y^t, \hat{\phi}) \),

for any \( i \in S_X \);

(ii) compute \( i_T \) as the mode of a sufficiently large number of values generated from \( \hat{\xi}_{iT} \);

(iii) for \( t = T - 1, \ldots, 1 \), compute the vector

\[
\hat{\xi}_t = \frac{\hat{\xi}_{it} \odot \hat{\Gamma}_{i_{t+1}}}{\Pi_{(m)} \left( \hat{\xi}_{it} \odot \hat{\Gamma}_{i_{t+1}} \right)}
\]

and \( i_t \) as the mode of a sufficiently large number of values generated from \( \hat{\xi}_{it} \), where \( \hat{\Gamma}_{i_{t+1}} \) represents the column of \( \hat{\Gamma} \) corresponding to the state previously generated.

To obtain the state of the hidden chain for the future time \( T + 1 \) and consequently to forecast the future observation \( y_{T+1} \), we run the FF-bs algorithm up to \( T + 1 \), with the same iterative scheme, replacing the vector \( F_{T+1} \) in (3) with the unit vector. If we repeat this procedure \( h \) times, using the previously obtained one-step-ahead forecast as the last observation, we can have the \( h \)-step-ahead ‘naive’ forecast.

Now we introduce the method we use to analyse Markov switching time series with missing observations; it is based on the iteration of three consecutive procedures: EM algorithm for the estimation of the parameters, reconstruction of the hidden chain, restoration of the missing observations as forecasting based on the best model estimated through the subsseries bounded by the missing values. Consider for example that two observations, \( y_{i1} \) and \( y_{i2} \), are missing at times \( t_1 \) and \( t_2 \) (\( 1 < t_1 < t_2 < T \)). First we consider the subsseries \( (y_1, \ldots, y_{i1-1}) \); the parameters vector \( \hat{\phi} \) is estimated by means of the EM algorithm; the sequence of hidden states is reconstructed up to \( t_1 \) by the FF-bs algorithm; \( \hat{y}_{i1} \) is forecasted given \( i_{t1} \):

\[
\hat{y}_{i1} = E(Y_{i1} \mid y_{i1-1}, \ldots, y_{i1-p}, X_{i1} = i_{t1}) = \hat{\rho}_{i1} + \sum_{h=1}^{p} \hat{\theta}_{h(i1)} y_{ih} - h.
\]  

Then we consider the subsseries \( (y_1, \ldots, y_{i1-1}) \) in which the missing value \( y_{i1} \) is filled with \( \hat{y}_{i1} \); the parameters vector \( \hat{\phi} \) is estimated; the sequence of hidden states is reconstructed up to \( t_2 \); \( y_{i1} \) is updated and finally \( \hat{y}_{i2} \) is computed, according to (8). Now the iterative procedure can restart using all the series \( (y_1, \ldots, y_T) \), to find the final estimates of \( \phi \) and to make forecasts.

In the forthcoming section we will apply this methodology to a real data set with missing observations that will be restored. Then using the restored sequence of observations we will make forecasting.
4. Application to air pollution data

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\(^2\). It records seven types of pollutants and each hour we have the mean concentration of every pollutant. We are interested in the analysis of daily mean concentrations of sulphur dioxide (SO\(_2\)). SO\(_2\), measured in µg/m\(^3\), is a gas with a characteristic pungent and choking smell; it is produced by the combustion of substances containing sulphur (coal, fuel oil, diesel oil). The principal sources of SO\(_2\) are the industries that need a lot of energy ( refineries, steelworks, thermoelectric power stations), the domestic heating (even if the use of methane drastically reduced the SO\(_2\) emissions), the traffic of the heavy transport (diesel vehicular traffic). SO\(_2\) is responsible, with the nitrogen oxides (NO\(_x\)), for the acid rains. It can cause the onset and the worsening of the respiratory tract diseases (persistent cough, bronchitis, sinusitis); very high concentrations can cause tissues destruction, giving rise to emphysema.

We analysed the time series of the daily mean concentrations of SO\(_2\) recorded by the air pollution testing station placed in Via San Giorgio, Bergamo, a medium density of population and intense traffic area, during the period 1.1.1997-30.11.1999.

Figure 1 shows the behaviours of the series of 1064 observations. A seasonal pattern is evident: it depends on the unobserved levels representing the different climates along the year. As in finite mixture analysis, plotting a histogram of the values in Figure 1, we have a rough indication of the number of the different unobserved levels: in Figure 2 we can guess the presence of some hidden states, noticing a multi-peaks distribution.

\[\text{Figure 1: Series of the daily mean concentrations of SO}_2\text{, recorded by the air pollution testing station placed in Via San Giorgio, Bergamo, from 1.1.1997 to 30.11.1999}\]

\(^2\)The data set has been provided by the Bergamo ARPA.
Notice that in Figure 1 some observations are missing, either because the station must be stopped for automatic calibration, because of occasional mechanical failure, ordinary maintenance, or because some data are clearly anomalous and so are removed by technicians. There are 21 missing values within the series: the observation $y_{147}$ and the groups of consecutive observations $(y_{156}, \ldots, y_{159})$, $(y_{312}, y_{313})$, $(y_{500}, \ldots, y_{502})$. So the estimation procedure we propose will be iteratively repeated for the four subseries $y^{146} = (y_1, \ldots, y_{146})$, $y^{156} = (y_{156}, \ldots, y_{159})$, $y^{311} = (y_{311}, \ldots, y_{313})$, $y^{503} = (y_{503}, \ldots, y_{502})$ and for the whole series $y^{1064} = (y_1, \ldots, y_{1064})$.

In the following analysis the sample period 1.1.1997-23.11.1999 ($y^{1067}$) has been used as the fit period for the model selection, while the observations of the last week, from 24.11.1999 to 30.11.1999 ($y_{1068}, \ldots, y_{1064}$), have been left to evaluate and compare the one-step-ahead “naive” forecasts performance.

The procedures introduced in Sections 2 and 3 have been implemented in a GAUSS code to analyze several models, for a range of values of the model order $m$, the number of states of the Markov chain, and $p$, the autoregressive order. For every analyzed subseries, we select the optimal model for $m = 1, 2, 3, 4$ and $p = 0, 1, \ldots, 5$, according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) or the minimum Mean Squared Error (MSE).

To avoid reporting local maxima of the log-likelihood surface, the EM algorithm runs several times with different initial values. Among the acceptable solutions of the parameters (i.e., those such that the Markov chain is irreducible and the autoregressive coefficients satisfy the stationarity condition), we choose the stationary point which corresponds to the largest log-likelihood value.

The variance-covariance matrix of the parameters estimates are obtained from the inverse of a numerical approximation of the Hessian matrix with reverse sign.

The results of model selection for the five steps of the iterative procedure are summarized in Tables A1-A5 in Appendix, where the values of the maximized log-likelihood and the corresponding AIC, BIC and MSE are given. The values obtained by $m = 4$ and $p = 5$ are not included in the tables because they give rise to reducible hidden chains or nonstationary models.

First we consider the subseries $y^{146} = (y_1, \ldots, y_{146})$. We choose as best model the MSAR(3,4), as in Table A1: three states of the hidden chain are selected by two criteria, while four autoregressive components are selected by all the criteria. The estimates of the
parameters of the hidden Markov chain are (standard errors in brackets)

\[
\Gamma^{(92)} = \begin{bmatrix}
0.9401 & 0.0001 & 0.0597 \\
(0.1261) & (0.1270) & (0.0354) \\
0.0001 & 0.9119 & 0.0880 \\
(0.0692) & (0.0071) & (0.0334) \\
0.2794 & 0.6473 & 0.0733 \\
(0.1424) & (0.1560) & (0.0928)
\end{bmatrix},
\]

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

\[
\delta^{(92)} = (0.3590; 0.5643; 0.0767)',
\]

while those of the parameters of the three gaussian pdfs are

<table>
<thead>
<tr>
<th>(i)</th>
<th>(\mu_i^{(92)})</th>
<th>(\sigma_i^{(92)})</th>
<th>(\theta_i^{(92)})</th>
<th>(\theta_i^{(92)})</th>
<th>(\theta_i^{(92)})</th>
<th>(\theta_i^{(92)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.6002</td>
<td>23.3265</td>
<td>0.4066</td>
<td>0.1650</td>
<td>0.1240</td>
<td>-0.0706</td>
</tr>
<tr>
<td></td>
<td>(2.3561)</td>
<td>(4.1969)</td>
<td>(0.1222)</td>
<td>(0.1262)</td>
<td>(0.1257)</td>
<td>(0.1181)</td>
</tr>
<tr>
<td>2</td>
<td>9.4890</td>
<td>3.2689</td>
<td>0.4536</td>
<td>-0.5276</td>
<td>0.0316</td>
<td>-0.0822</td>
</tr>
<tr>
<td></td>
<td>(0.9340)</td>
<td>(0.5479)</td>
<td>(0.0924)</td>
<td>(0.0975)</td>
<td>(0.0989)</td>
<td>(0.0681)</td>
</tr>
<tr>
<td>3</td>
<td>10.5535</td>
<td>0.0047</td>
<td>0.6268</td>
<td>-0.3155</td>
<td>0.1389</td>
<td>-0.1738</td>
</tr>
<tr>
<td></td>
<td>(0.0437)</td>
<td>(0.0020)</td>
<td>(0.0046)</td>
<td>(0.0116)</td>
<td>(0.0081)</td>
<td>(0.0093)</td>
</tr>
</tbody>
</table>

The missing observation \(y_{147}\) is now restored by means of the procedure introduced in the previous section and its value results 8.4321.

Then we consider the second subseries \(y^{(95)} = (y_1, \ldots, y_{156})\). We choose as best model the MSAR(3,4), as in Table A2: three states of the hidden chain are selected by two criteria, while four autoregressive components are selected by all the criteria. The estimates of the parameters of the hidden Markov chain are

\[
\Gamma^{(187)} = \begin{bmatrix}
0.9403 & 0.0001 & 0.0596 \\
(0.1216) & (0.1230) & (0.0358) \\
0.0001 & 0.9239 & 0.0760 \\
(0.0000) & (0.0439) & (0.0296) \\
0.2818 & 0.6444 & 0.0738 \\
(0.1498) & (0.1624) & (0.0938)
\end{bmatrix},
\]

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

\[
\delta^{(187)} = (0.3326; 0.5970; 0.0704)',
\]

while those of the parameters of the three gaussian pdfs are

<table>
<thead>
<tr>
<th>(i)</th>
<th>(\mu_i^{(187)})</th>
<th>(\sigma_i^{(187)})</th>
<th>(\theta_i^{(187)})</th>
<th>(\theta_i^{(187)})</th>
<th>(\theta_i^{(187)})</th>
<th>(\theta_i^{(187)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.5960</td>
<td>23.3232</td>
<td>0.4066</td>
<td>0.1651</td>
<td>0.1240</td>
<td>-0.0705</td>
</tr>
<tr>
<td></td>
<td>(2.3504)</td>
<td>(4.1962)</td>
<td>(0.1223)</td>
<td>(0.1262)</td>
<td>(0.1258)</td>
<td>(0.0186)</td>
</tr>
<tr>
<td>2</td>
<td>9.6846</td>
<td>3.3015</td>
<td>0.4218</td>
<td>-0.5024</td>
<td>0.0332</td>
<td>-0.0982</td>
</tr>
<tr>
<td></td>
<td>(0.9133)</td>
<td>(0.5206)</td>
<td>(0.0904)</td>
<td>(0.0933)</td>
<td>(0.0950)</td>
<td>(0.0665)</td>
</tr>
<tr>
<td>3</td>
<td>10.5535</td>
<td>0.0047</td>
<td>0.6266</td>
<td>-0.3154</td>
<td>0.1387</td>
<td>-0.1737</td>
</tr>
<tr>
<td></td>
<td>(0.0439)</td>
<td>(0.0021)</td>
<td>(0.0045)</td>
<td>(0.0118)</td>
<td>(0.0082)</td>
<td>(0.0101)</td>
</tr>
</tbody>
</table>
Given that we have more than one missing value, i.e., \( y_{156}, \ldots, y_{159} \), the procedure used to restore the missing values, described at Sections 2 and 3, must be modified in the following way:

(i) compute \( \hat{\xi}_{1|0} = \delta' = \delta' \Gamma \);  
(ii) compute 

\[
\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \odot F_t}{1'_{(3)} \left( \hat{\xi}_{t|t-1} \odot F_t \right)}
\]

and 

\[
\hat{\xi}_{t+1|t} = \Gamma \hat{\xi}_{t|t},
\]

for any \( t = 1, \ldots, 155 \);  
(iii) compute 

\[
\hat{\xi}_{156|155} = \frac{\hat{\xi}_{156|155} \odot 1_{(3)}}{1'_{(3)} \left( \hat{\xi}_{156|155} \odot 1_{(3)} \right)} = \frac{\hat{\xi}_{156|155}}{1'_{(3)} \hat{\xi}_{156|155}}
\]

(iv) compute \( i_{156} \) as the mode of a sufficiently large number of values generated from \( \hat{\xi}_{156|155} \);  
(v) compute \( \hat{\gamma}_{156} \), according to (8);  
(vi) for any \( t = 155, 154, \ldots \), compute 

\[
\hat{\xi}_t = \frac{\hat{\xi}_{t|t} \odot \Gamma_{i_{t+1}}}{1'_{(3)} \left( \hat{\xi}_{t|t} \odot \Gamma_{i_{t+1}} \right)}
\]

and \( i_t \) as the mode of a sufficiently large number of values generated from \( \hat{\xi}_t \). This sixth step runs up to the special \( t \) such that the generated state is equal to the previously reconstructed state;  
(vii) compute (ii) for \( t = 156 \);  
(viii) to obtain \( i_{157} \) and \( \hat{\gamma}_{157} \), repeat (iii), (iv), (v), (vi) for \( t = 157 \). Notice that formula (8) must be reapplied to all the restored values corresponding to the states that are different from those previously generated;  
(ix) to obtain \( i_{158} \) and \( \hat{\gamma}_{158} \), repeat (vii) for \( t = 157 \) and (viii) for \( t = 158 \);  
(x) to obtain \( i_{159} \) and \( \hat{\gamma}_{159} \), repeat (vii) for \( t = 158 \) and (viii) for \( t = 159 \).

By this procedure the restored missing values are \((7.7807, 9.1870, 9.0142, 8.4585)\), while the missing value \( y_{147} \), restored by the new set of parameters, is 8.4609.

For the third subsample \( y_{311} = (y_1, \ldots, y_{311}) \) we choose as best model the MSAR(3,4) again, as in Table A3: three states of the hidden chain and four autoregressive components are selected by two criteria. The estimates of the parameters of the hidden Markov chain are 

\[
\Gamma^{(337)} = \begin{bmatrix}
0.9645 & 0.0001 & 0.0354 \\
0.0184 & 0.0269 & 0.0256 \\
0.0001 & 0.9731 & 0.0268 \\
0.0000 & 0.0181 & 0.0196 \\
0.0573 & 0.0374 & 0.9053 \\
0.0324 & 0.0345 & 0.0471
\end{bmatrix},
\]

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

\[
\delta^{(337)} = (0.4034; 0.3471; 0.2495)\]
while those of the parameters of the three gaussian pdfs are

\[
\begin{array}{ccccccc}
   i & \mu_i^{(337)} & \sigma_i^{(337)} & \phi_i^{(337)}(1) & \phi_i^{(337)}(2) & \phi_i^{(337)}(3) & \phi_i^{(337)}(4) \\
1 & 1.0046 & 1.9815 & 0.5001 & -0.0681 & 0.0060 & 0.0636 \\
   & (0.3455) & (0.3091) & (0.027) & (0.1036) & (0.1250) & (0.0840) \\
2 & 4.4789 & 21.7986 & 0.5016 & 0.0722 & 0.1062 & -0.0371 \\
   & (1.8748) & (3.2731) & (0.1015) & (0.1049) & (0.1084) & (0.1073) \\
3 & 11.3052 & 6.4493 & 0.2191 & -0.2645 & -0.1091 & -0.1089 \\
   & (1.3834) & (2.1350) & (0.1141) & (0.1223) & (0.1434) & (0.0883) \\
\end{array}
\]

The missing observations \((y_{112}, y_{313})\) are now restored by means of the procedure used for the second subsamples, updated for \(t = 312, 313\). We obtain \((11.8267, 11.7593)\), while the missing values \(y_{147}, (y_{150}, \ldots, y_{159})\), restored by the new set of parameters, are \(6.2452, (6.2912, 7.2307, 8.4254, 8.0725)\).

For the fourth subsamples \(y^{(38)} = (y_{1}, \ldots, y_{38})\) we choose as best model the MSAR(3,3), as in Table A4: three states of the hidden chain and three autoregressive components are selected by two criteria. The estimates of the parameters of the hidden Markov chain are

\[
\Gamma^{(119)} = \begin{bmatrix}
0.9536 & 0.0463 & 0.0001 \\
(0.0290) & (0.0238) & (0.0325) \\
0.0302 & 0.9619 & 0.0079 \\
(0.0163) & (0.0180) & (0.0082) \\
0.0001 & 0.0097 & 0.9902 \\
(0.0031) & (0.0082) & (0.0062) \\
\end{bmatrix},
\]

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

\[
\delta^{(119)} = (0.2645; 0.4061; 0.3294)',
\]

while those of the parameters of the three gaussian pdfs are

\[
\begin{array}{ccccccc}
   i & \mu_i^{(119)} & \sigma_i^{(119)} & \phi_i^{(119)}(1) & \phi_i^{(119)}(2) & \phi_i^{(119)}(3) & \phi_i^{(119)}(4) \\
1 & 0.9955 & 1.6283 & 0.8624 & -0.1522 & 0.0864 & 0.0850 \\
   & (0.3586) & (0.2374) & (0.0942) & (0.1143) & (0.0825) & (0.0739) \\
2 & 4.9824 & 11.5426 & 0.2751 & 0.0563 & 0.0650 & 0.0650 \\
   & (0.8519) & (1.3748) & (0.0811) & (0.0800) & (0.0739) & (0.0666) \\
3 & 5.3720 & 27.8386 & 0.5377 & 0.1368 & -0.0109 & -0.0109 \\
   & (1.0616) & (2.6125) & (0.0651) & (0.0727) & (0.0666) & (0.0666) \\
\end{array}
\]

The missing observations \((y_{139}, \ldots, y_{552})\) are now restored by the usual procedure for \(t = 539, \ldots, 552\). Their values are \((6.5991, 7.4203, 7.5984, 7.9194, 8.0711, 8.1425, 8.1916, 8.2190, 8.2339, 8.2427, 8.2478, 8.2506, 8.2753, 8.2532)\), while the missing values \(y_{447}, (y_{550}, \ldots, y_{559})\), restored by the new set of parameters, are \(6.1673, (6.1121, 6.1406, 6.8641, 7.7046), (12.1249, 13.2061)\).

Finally, for the whole series \(y^{(067)} = (y_{1}, \ldots, y_{567})\) the final best model is MSAR(3,4) as in Table A5: three states of the hidden Markov chain and four autoregressive components are selected by two criteria. The estimates of the parameters of the hidden Markov chain
are

$$\Gamma^{(76)} = \begin{bmatrix} 0.9633 & 0.0366 & 0.0001 \\ 0.0221 & 0.0139 & 0.0090 \\ 0.0705 & 0.9181 & 0.0114 \\ 0.0033 & 0.0326 & 0.0093 \\ 0.0001 & 0.0090 & 0.9909 \\ 0.0000 & 0.0054 & 0.0000 \end{bmatrix},$$

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

$$\delta^{(76)} = (0.4598; 0.2388; 0.3014)'$$

while those of the parameters of the three gaussian pdfs are

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\mu_i^{(76)}$</th>
<th>$\sigma_i^{2(76)}$</th>
<th>$\theta_{1i}^{(76)}$</th>
<th>$\theta_{2i}^{(76)}$</th>
<th>$\theta_{3i}^{(76)}$</th>
<th>$\theta_{4i}^{(76)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7266</td>
<td>2.2774</td>
<td>0.6404</td>
<td>0.0047</td>
<td>0.0890</td>
<td>0.1190</td>
</tr>
<tr>
<td></td>
<td>(0.2182)</td>
<td>(0.2448)</td>
<td>(0.0595)</td>
<td>(0.0746)</td>
<td>(0.0661)</td>
<td>(0.0495)</td>
</tr>
<tr>
<td>2</td>
<td>5.3030</td>
<td>11.3926</td>
<td>0.2234</td>
<td>0.2090</td>
<td>0.0100</td>
<td>0.0962</td>
</tr>
<tr>
<td></td>
<td>(0.9970)</td>
<td>(1.3905)</td>
<td>(0.0870)</td>
<td>(0.0770)</td>
<td>(0.0612)</td>
<td>(0.0728)</td>
</tr>
<tr>
<td>3</td>
<td>6.4221</td>
<td>26.3704</td>
<td>0.5733</td>
<td>0.0427</td>
<td>-0.0263</td>
<td>0.0044</td>
</tr>
<tr>
<td></td>
<td>(0.9274)</td>
<td>(1.9306)</td>
<td>(0.0522)</td>
<td>(0.0588)</td>
<td>(0.0735)</td>
<td>(0.0711)</td>
</tr>
</tbody>
</table>

The missing observations $y_{147}$, $(y_{158}, \ldots, y_{169})$, $(y_{312}, y_{313})$, $(y_{539}, \ldots, y_{562})$, restored by the new set of parameters, are 6.0081, (6.4333, 6.1257, 6.4197, 6.5651), (12.9059, 13.6257), (4.5367, 6.2918, 6.6621, 7.1060, 7.9268, 8.5677, 9.0788, 9.5212, 9.9262, 10.2802, 10.5886, 10.8584, 11.0056, 11.3035).

The one-step-ahead “naive” forecasts for the period from 24.11.99 to 30.11.99 are now obtained by means of the usual procedure for $t = 1058, \ldots, 1064$. Their values are 9.7455, 8.3064, 7.9564, 8.6002, 8.4116, 8.6410, 8.1725 and they are plotted in Figure 3 with the series of the fitted values by the best model, for $t = 5, \ldots, 1057$. Furthermore Figure 4 shows the final sequence of the reconstructed states of the hidden Markov chain, representing the three different levels of pollution occurred during the analysed period.

![Figure 3: Series of the fitted values from time 5 to time 1057 and the forecasted values from time 1058 to time 1064.](image-url)
Figure 4: Reconstructed series of the hidden states.

To evaluate the forecasting performance of our method, the one-step-ahead "naive" forecasts are compared with the actual values through some descriptive statistics: the mean squared error (MSE), the mean absolute error (MAE) and the correlation coefficient (CORR) between actual and forecasted values. The values of these statistics for the forecasts are reported in Table 1, left column. They are also compared with those computed by means of Hamilton’s method (Formula 7), right column, which is now applied to the complete series, where the missing observations have been restored through the procedure based on the $ff-bs$ algorithm.

<table>
<thead>
<tr>
<th></th>
<th>Predictions statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>53.0423</td>
</tr>
<tr>
<td>MAE</td>
<td>6.9492</td>
</tr>
<tr>
<td>CORR</td>
<td>-0.0589</td>
</tr>
</tbody>
</table>

The results show that the two methods approximatively have the same forecasting ability, but it must be noticed that the advantage in using the $ff-bs$ algorithm is that we can reconstruct the sequence of the hidden states.

Conclusions

We recurred to Markov Switching AutoRegressive (MSAR) models to analyse the series of 1064 daily mean concentrations of sulphur dioxide, recorded by an air pollution testing station, placed in Bergamo. MSAR models are sequences of observed variables depending on the past observations and on the realization of a contemporary unobserved variable which is the current state of a hidden Markov chain. We performed the EM algorithm, jointly with the filtering and the smoothing algorithms, to obtain the maximum likelihood estimators of the parameters of the model; we also reconstructed the sequence of the hidden state through the forward filtering-backward sampling algorithm, restored the missing observations and forecasted the future values. We compared the different models, varying for both the order of the autoregressive component and the cardinality of the state-space of the hidden Markov chain. The comparisons have been made by means of the Akaike Information Criterion, the Bayesian Information Criterion and the minimum Mean Squared Error.
Appendix

| Table A1 - Results of the model selection for the first subseries $y_{146}=(y_1,...,y_{145})$. |
|---|---|---|---|---|
| $m$ | $p$ | 0 | 2 | 3 | 4 |
| 1 | ln $L_T(\phi)$ | -469.309 | -425.321 | -421.856 | -414.583 | -411.543 |
| | BIC | -474.492 | -432.786 | -431.796 | -426.990 | -426.411 |
| 2 | ln $L_T(\phi)$ | -418.194 | -402.117 | -392.923 | -386.024 | -381.062 |
| | AIC | -424.194 | -410.117 | -402.923 | -398.024 | -395.062 |
| | BIC | -433.144 | -422.024 | -417.772 | -415.801 | -415.753 |
| 3 | ln $L_T(\phi)$ | -407.963 | -399.165 | -379.618 | -369.619 | -356.322 |
| | AIC | -419.963 | -414.165 | -397.618 | -390.619 | -380.322 |
| | BIC | -437.864 | -436.490 | -424.346 | -421.729 | -415.792 |

| Table A2 - Results of the model selection for the second subseries $y_{155}=(y_{15},...,y_{155})$. |
|---|---|---|---|---|
| $m$ | $p$ | 0 | 2 | 3 | 4 |
| 1 | ln $L_T(\phi)$ | -496.274 | -448.437 | -444.881 | -437.324 | -434.306 |
| | BIC | -501.317 | -455.992 | -454.942 | -449.884 | -449.358 |
| | MSE | 35.592 | 19.934 | 19.733 | 18.596 | 18.560 |
| 2 | ln $L_T(\phi)$ | -437.482 | -420.328 | -412.037 | -405.003 | -399.394 |
| | AIC | -443.482 | -428.328 | -422.037 | -417.003 | -413.394 |
| | BIC | -452.613 | -440.476 | -437.189 | -435.146 | -434.515 |
| 3 | ln $L_T(\phi)$ | -426.158 | -417.546 | -402.005 | -388.582 | -375.469 |
| | AIC | -438.158 | -432.546 | -420.005 | -409.582 | -399.469 |
| | BIC | -456.418 | -455.323 | -447.279 | -441.332 | -435.677 |
| | MSE | 14.113 | 18.254 | 12.521 | 11.9142 | 11.596 |
Table A3 - Results of the model selection for the third subseries \( y_{31}^3 = (y_{11}, y_{31}) \).

<table>
<thead>
<tr>
<th>( m )</th>
<th>( p )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \ln L_T(\phi) )</td>
<td>-988.320</td>
<td>-853.005</td>
<td>-846.448</td>
<td>-834.305</td>
<td>-831.873</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>-900.320</td>
<td>-856.005</td>
<td>-850.448</td>
<td>-839.305</td>
<td>-837.873</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-994.090</td>
<td>-861.610</td>
<td>-857.914</td>
<td>-848.630</td>
<td>-849.054</td>
</tr>
<tr>
<td></td>
<td>( \ln L_T(\phi) )</td>
<td>-865.225</td>
<td>-784.129</td>
<td>-779.005</td>
<td>-771.723</td>
<td>-767.321</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>-871.225</td>
<td>-792.129</td>
<td>-789.005</td>
<td>-783.723</td>
<td>-781.321</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-882.444</td>
<td>-807.076</td>
<td>-807.672</td>
<td>-806.103</td>
<td>-807.409</td>
</tr>
<tr>
<td></td>
<td>( \ln L_T(\phi) )</td>
<td>-812.631</td>
<td>-765.842</td>
<td>-761.823</td>
<td>-755.093</td>
<td>-751.251</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>-824.631</td>
<td>-780.842</td>
<td>-779.823</td>
<td>-776.093</td>
<td>-775.251</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-847.070</td>
<td>-808.866</td>
<td>-813.423</td>
<td>-815.259</td>
<td>-819.973</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>13.963</td>
<td>11.093</td>
<td>12.099</td>
<td>10.144</td>
<td>10.109</td>
</tr>
</tbody>
</table>

Table A4 - Results of the model selection for the fourth subseries \( y_{33}^{33} = (y_{13}, y_{33}) \).

<table>
<thead>
<tr>
<th>( m )</th>
<th>( p )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \ln L_T(\phi) )</td>
<td>-1794.574</td>
<td>-1565.986</td>
<td>-1549.280</td>
<td>-1539.886</td>
<td>-1535.870</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>-1796.574</td>
<td>-1568.986</td>
<td>-1553.280</td>
<td>-1544.886</td>
<td>-1541.870</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-1800.867</td>
<td>-1575.415</td>
<td>-1561.848</td>
<td>-1555.592</td>
<td>-1554.711</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>46.306</td>
<td>19.974</td>
<td>18.971</td>
<td>18.517</td>
<td>18.434</td>
</tr>
<tr>
<td></td>
<td>( \ln L_T(\phi) )</td>
<td>-1506.147</td>
<td>-1482.186</td>
<td>-1471.348</td>
<td>-1464.332</td>
<td>-1459.814</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>-1602.147</td>
<td>-1490.186</td>
<td>-1481.348</td>
<td>-1476.332</td>
<td>-1473.814</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-1615.011</td>
<td>-1507.330</td>
<td>-1502.768</td>
<td>-1502.026</td>
<td>-1503.777</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>25.601</td>
<td>19.134</td>
<td>18.400</td>
<td>18.084</td>
<td>18.051</td>
</tr>
<tr>
<td></td>
<td>( \ln L_T(\phi) )</td>
<td>-1522.272</td>
<td>-1455.469</td>
<td>-1448.867</td>
<td>-1442.825</td>
<td>-1440.650</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>-1534.272</td>
<td>-1470.469</td>
<td>-1466.867</td>
<td>-1463.825</td>
<td>-1464.650</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-1559.999</td>
<td>-1502.615</td>
<td>-1505.425</td>
<td>-1508.789</td>
<td>-1516.015</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>15.906</td>
<td>17.647</td>
<td>17.108</td>
<td>16.885</td>
<td>16.588</td>
</tr>
</tbody>
</table>

Table A5 - Results of the model selection for the whole series \( y_{13}^{33} = (y_{11}, y_{33}) \).

<table>
<thead>
<tr>
<th>( m )</th>
<th>( p )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \ln L_T(\phi) )</td>
<td>-3476.812</td>
<td>-2957.361</td>
<td>-2933.961</td>
<td>-2920.827</td>
<td>-2911.997</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>-3478.812</td>
<td>-2960.361</td>
<td>-2939.961</td>
<td>-2925.827</td>
<td>-2917.997</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-3483.775</td>
<td>-2967.805</td>
<td>-2949.884</td>
<td>-2938.228</td>
<td>-2932.875</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>42.170</td>
<td>15.850</td>
<td>15.315</td>
<td>14.961</td>
<td>14.789</td>
</tr>
<tr>
<td></td>
<td>( \ln L_T(\phi) )</td>
<td>-2978.014</td>
<td>-2716.584</td>
<td>-2705.131</td>
<td>-2693.166</td>
<td>-2683.910</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>-2984.014</td>
<td>-2724.584</td>
<td>-2715.131</td>
<td>-2705.166</td>
<td>-2697.910</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-2998.903</td>
<td>-2744.433</td>
<td>-2739.937</td>
<td>-2734.928</td>
<td>-2732.626</td>
</tr>
<tr>
<td></td>
<td>( \ln L_T(\phi) )</td>
<td>-2819.432</td>
<td>-2677.750</td>
<td>-2672.033</td>
<td>-2661.983</td>
<td>-2654.901</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>-2831.432</td>
<td>-2692.750</td>
<td>-2690.033</td>
<td>-2682.983</td>
<td>-2678.901</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>-2861.211</td>
<td>-2729.967</td>
<td>-2734.684</td>
<td>-2735.066</td>
<td>-2738.414</td>
</tr>
</tbody>
</table>
References


