

Seasonal Autoregressions with Regime Switching

Autoregressioni Stagionali con Cambiamenti di Regime

Roberta Paroli

Istituto di Statistica
Univ. Cattolica del S.C., Milano
roberta.paroli@unicatt.it

Luigi Spezia

Dip. Sc. Econ. e Metodi Quantitativi
Univ. del Piemonte Orientale, Novara
luigi.spezia@eco.unipmn.it

Riassunto: I modelli autoregressivi con cambiamenti di regime costituiscono un valido strumento nell'analisi di serie temporali non lineari e non normali. In questo lavoro un particolare modello autoregressivo, con cambiamenti di regime di natura markoviana e con una componente stagionale dipendente dallo stato della catena di Markov latente, viene studiato per l'analisi di serie temporali periodiche. La selezione dei vincoli di identificabilità, la scelta del modello migliore, la stima dei parametri e dei dati latenti e la previsione dei dati futuri sono affrontati utilizzando algoritmi di tipo Metropolis-within-Gibbs, che tengono conto anche dei differenti ordinamenti possibili degli stati latenti. La metodologia proposta viene illustrata con l'analisi della dinamica di un inquinante dell'aria.

Keywords: Hidden Markov chain, label switching, permutation sampling, Metropolis-within-Gibbs algorithm

1. Introduction

Autoregressive processes can be extended to model non-linear and non-normal time series by assuming that different autoregressions, each one depending on a latent regime, alternate according to the regime switching, which is driven by an unobserved Markov chain. These models, widely known as Markov switching autoregressive models (MSARMs), have been introduced in the econometric literature by Hamilton to study economic and financial time series (see Hamilton (1994), ch. 22, and references therein).

MSARMs are discrete-time stochastic processes $\{Y_t; X_t\}$, so that $\{X_t\}$ is a latent, or hidden, finite-state Markov chain and $\{Y_t\}$, given $\{X_t\}$, satisfies the order- p dependence and the contemporary dependence conditions: we have a sequence of observed random variables $\{Y_t\}$ depending on the p previous observations, whose conditional distributions depend on $\{X_t\}$ only through the contemporary state, or regime, of the Markov chain. A special MSARM with a seasonal component, called Seasonal MSARM (SMSARM), is proposed here in the Bayesian framework.

2. Seasonal Markov switching autoregressive models

Let $\{X_t\}$ be a discrete-time, first-order, homogeneous, ergodic Markov chain on a finite state-space S_X with cardinality m ($S_X = \{1, \dots, m\}$). $\Gamma = [\gamma_{i,j}]$ is the $(m \times m)$ transition matrix, where $\gamma_{i,j} = P(X_t = j | X_{t-1} = i)$, for any $i, j \in S_X$ and any $t = 2, \dots, T$; $\delta = (\delta_1, \dots, \delta_m)'$ is the

stationary distribution, so that $\delta' = \delta' \Gamma$; $\mathbf{x}^T = (x_1, \dots, x_T)'$ is the sequence of the states of the Markov chain and, for any $t = 1, \dots, T$, x_t has values in S_X .

Our aim is the analysis of seasonal time series characterized by periodic variations with period s and we assume that different seasonalities occur, according to the regime switching. Seasonal components, depending on the hidden states, are modelled by parameters $\beta_{h(i)}$, for any $i \in S_X$ and for any $h = 1, \dots, s$, under the constraint $\sum_{h=1}^s \beta_{h(i)} = 0$, with $\beta_i = (\beta_{1(i)}, \dots, \beta_{s(i)})'$, for any $i \in S_X$. Hence the equation describing SMSARMs is

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

where $Y_{t(i)}$ denotes the conditional variable Y_t when $X_t = i$, for any $1 \leq t \leq T$ and for any $i \in S_X$, and $E_{t(i)}$ denotes the Gaussian noise E_t when $X_t = i$, with zero mean and precision λ_i ($E_{t(i)} \sim \mathcal{N}(0; \lambda_i)$), for any $i \in S_X$, with the discrete-time processes $\{E_t\}$, given $\{X_t\}$, satisfying the conditional independence and the contemporary dependence conditions. To automatically satisfy the constraint on any $\phi_i = (\phi_{1(i)}, \dots, \phi_{p(i)})'$ for the stationarity of the process, we reparametrize ϕ_i in terms of the partial autocorrelations $\mathbf{r}_i = (r_{1(i)}, \dots, r_{p(i)})'$ of any sub-process, for any $i \in S_X$, according to Barndorff-Nielsen and Schou (1973) and Jones (1987). Model (1) is unidentifiable in data fitting: when we have m states, we have $m!$ ways to label them; so different models are interchangeable by permuting their labeling. This is the so-called label switching problem and it can be overcome by placing some parameters in increasing or decreasing order. In this paper the special SMSARM with decreasing precisions is analysed, but the procedures we shall introduce can be easily adapted to any other type of constraint. We can derive suitable constraints by a data-driven procedure, based on random permutation sampling algorithm (Frühwirth-Schnatter (2001)). Here it is important only to notice that the constraint is chosen ex post after simulations so as to respect the geometry and the shape of the unconstrained posterior distribution, that is different identifiability constraints can be derived by different data sets. Furthermore to be able to estimate the state-dependent seasonal component, given that $\sum_{h=1}^s \beta_{h(i)} = 0$ for any $i \in S_X$, we also need that the hidden chain visits the same hidden state i for all the s times of any sub-period. To estimate the parameters of the model, the hidden states and the missing observations the Metropolis-within-Gibbs procedures associated with the constrained permutation sampling algorithm have been developed for the special SMSARM with decreasing precisions.

3. Application to Ozone data

An application of SMSARMs to real data will be studied in the following: we shall analyse the time series of the hourly mean concentrations of Ozone (O3), in micrograms per cubic meter ($\mu\text{g}/\text{m}^3$), recorded by the air pollution testing station placed in Via San Giorgio, Bergamo (Italy), from April 1st, 1998, 1 a.m., to September 30th, 1998, 12 p.m. (4392 observations). The analysed data are the natural logarithms of O3 concentrations, while the latent data are made up both by the sequence of the hidden states and by 549 missing values occurring within the observed series. In the series (Figure 1a) a daily periodicity ($s=24$) is evident and it is confirmed by the correlogram.

Our data set is collected in the six warmest months, because O₃ is a substance which is not directly emitted in the air, but it forms owing to complex chemical reactions in the presence of solar radiation and high temperature. We shall develop our empirical analysis in four steps: *i*) constraint identification, *ii*) model selection, *iii*) parameter estimation, *iv*) forecasting.

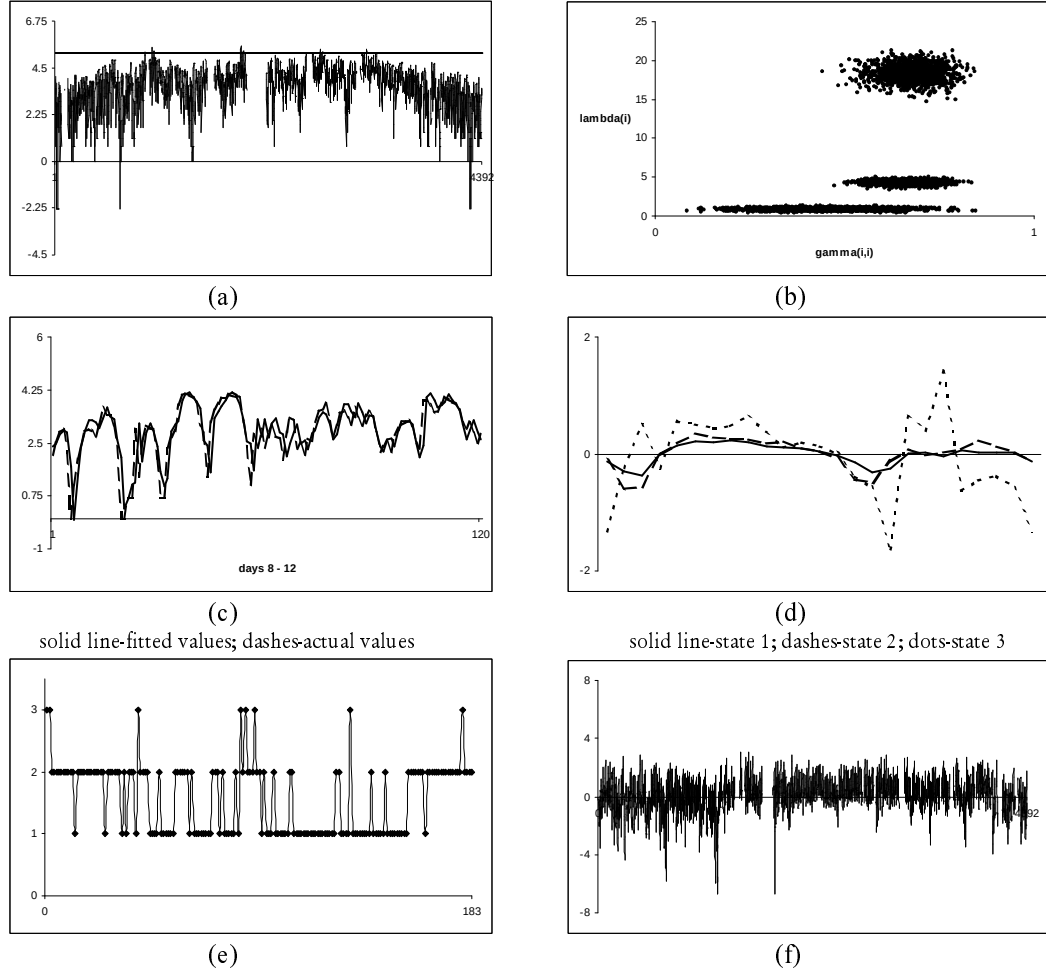
i) Constraint identification. By plotting the couples of the outputs of the estimates, obtained via unconstrained Metropolis-within-Gibbs algorithm performed associated with random permutation sampling, we can check if there are as many groups as the hidden states and if these groups can suggest special ordering in their labeling. We can compare the outputs of thirty competing models, henceforth said SMSAR($m;p$), which differ for the cardinality of the state-space of the hidden Markov chain ($m=1, \dots, 5$) and for the order of the autoregressive process ($p=0, \dots, 5$). By graphically analysing the outputs of the unconstrained SMSAR($m;p$), first we can notice that, for any p , $m=4$ and $m=5$ are not consistent with the data we are studying, because four or five groups do not emerge in any plot, hence we shall develop our analysis for $m=1;2;3$ only. After that we can select the constraint on the precisions because the decreasing ordering is evident in any graph (see for example Figure 1b). Decreasing precisions is a reasonable constraint, because when the low hidden state occurs, the variability of O₃ data depending on it is low and the concentrations of pollution are also low; by contrast when the high hidden state occurs, the variability of O₃ data depending on it is high and the concentrations of pollution are also high.

ii) Model choice. Model choice will be performed by means of Bayes factors in which the marginal likelihoods are computed according to Chib (1995) and Chib and Jeliazkov (2001), through the relabeling of the hidden states by means of constrained permutation sampling. From the values of the marginal likelihoods, we choose the SMSAR(3;2) as the best among all the competing models.

iii) Parameter estimation. From the results of the estimates we can see the variability within each state increases as the unobserved level of pollution increases. Moreover the O₃ hourly dynamics of the series, described by the β_i 's, presents peaks which become more evident as the unobserved level of pollution increases (Figure 1d). The dynamics of the hidden states can be observed in Figure 1e, where we have the sequence of the posterior modes of any generated state: state 1 underlies the observations with the lowest level of pollution, while state 3 underlies those with the highest level of pollution. The probability of persistence in states 1 and 2 is greater than the probability of transition; by contrast the probability of persistence in state 3 is less than the probability of transition, that is the mean time of persistence in the dangerous state is the lowest one. Another interesting result of our analysis is the comparison between observed and fitted values (Figure 1c) with regards to the concentrations of O₃ exceeding the attention-level ($180 \mu\text{g}/\text{m}^3$). In the observed series there are 37 values greater than the attention-level, split in 10 days, and SMSAR(3;2) captures 5 day out of 10 in which the threshold is exceeded.

iv) Forecasting. The model we fitted by the series of O₃ concentrations recorded from April 1st, 1998, 1 a.m., to September 30th, 1998, 12 p.m., is now used to obtain the one-day ahead hidden state predictions and the k -hour ahead pollutant predictions ($1 \leq k \leq 24$), for the data from April 1st, 1999, 1 a.m., to September 30th, 1999, 12 p.m.; this problem is tackled considering future values as missing. To evaluate the forecasting ability of our model we used the O₃ spring and summer 1999 data to compare them with all the k -hour ahead predictions (see Figure 1f for the sequence of the residuals of the case $k=6$).

Figure 1: (a) Series of the O3 hourly mean log-concentrations with the log of the attention-level; (b) output of unconstrained Metropolis-within-Gibbs algorithm with random permutations for $p=2$ and $m=3$; (c) actual and fitted values of the sub-series of days 8 - 12; (d) the daily component from 5 a.m. to 5 a.m.; (e) the sequence of the hidden states; (f) sequence of the predictive residuals for $k=6$



References

- Barndorff-Nielsen O., Schou G. (1973). On the Parametrization of Autoregressive Models by Partial Autocorrelations, *Journal of Multivariate Analysis*, 3, 408-419.
- Chib S. (1995). Marginal Likelihood From the Gibbs Output, *Journal of the American Statistical Association*, 90, 1313-1321.
- Chib S., Jeliazkov I. (2001). Marginal Likelihood From the Metropolis-Hastings Output, *Journal of the American Statistical Association*, 96, 270-281.
- Frühwirth-Schnatter S. (2001). Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models, *Journal of the American Statistical Association*, 96, 194-209.
- Hamilton J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.
- Jones M. C. (1987). Randomly Choosing Parameters from the Stationarity and Invertible Region of Autoregressive-Moving Average Models, *Applied Statistics*, 36, 134-138.