

# Parsimonious periodic autoregressive models for time series with evolving trend and seasonality

Francesco Battaglia · Domenico Cucina · Manuel Rizzo

the date of receipt and acceptance should be inserted later

**Abstract** This paper proposes an extension of Periodic AutoRegressive (PAR) modelling for time series with evolving features. The large scale of modern datasets, in fact, implies that the time span may subtend several evolving patterns of the underlying series, affecting also seasonality. The proposed model allows several regimes in time and a possibly different PAR process with a trend term in each regime. The means, autocorrelations and residual variances may change both with the regime and the season, resulting in a very large number of parameters. Therefore as a second step we propose a grouping procedure on the PAR parameters, in order to obtain a more parsimonious and concise model. The model selection procedure is a complex combinatorial problem, and it is solved basing on Genetic Algorithms that optimize an information criterion. The model is tested in both simulation studies and real data analysis from different fields, proving to be effective for a wide range of series with evolving features, and competitive with respect to more specific models.

**Keywords** Genetic Algorithms · Multiregime Models · Information Criteria · Structural Change

---

F. Battaglia, M. Rizzo  
Department of Statistical Sciences, University La Sapienza,  
Rome, Italy  
E-mail: {francesco.battaglia manuel.rizzo}@uniroma1.it

D. Cucina  
Department of Economics and Statistics, University of  
Salerno, Italy  
E-mail: dcucina@unisa.it

## 1 Introduction

In modern times an increasing amount of data is available. For time series this implies that both the observation period increases, and the time interval between observations decreases. But series observed over a very large time span are usually subject to changes of their structure and features, concerning both the first order (means) and the second order (variance and autocorrelations) properties. Moreover, due to the usually infra-annual observation rate, seasonality is present and cannot be disregarded. To analyse such a complex scenario, sufficiently simple models are needed for describing, and following in their evolution, both first and second order behaviour of the time series.

The present research is concerned with time series with trend and seasonality, and an evolving structure. We shall consider a framework for which the seasonality has a complex structure, and has an effect on means, variances and autocorrelations of the underlying process. We shall base our procedures on periodic modelling (Gladyshev 1961). Such methods have been introduced because standard linear filtering techniques have been proven to be not adequate for dealing with seasonality in the autocorrelation structure (see Franses 1995), so periodic modelling is suitable for our goals. We shall focus on Periodic AutoRegressive (PAR) models, for which each season of the year follows a possibly different AR process, with season-varying means and residual variances (for an account see Franses and Paap 2004; Hipel and McLeod 1994).

The evolving structure will be analysed in the framework of structural changes specifications (Bai and Perron 1998; Zeileis et al. 2003; Davis et al. 2006; Caporale et al. 2012). Therefore, we will assume the possible presence of several regimes in time, where the trend

slope, the means, the autocorrelations and the residual variances can have a different behaviour. The resulting model will naturally require a very large number of parameters, so its specification is a complex combinatorial problem. We shall adopt a strategy based on Genetic Algorithms (GAs) (Holland 1992) for optimizing an information criterion, a choice that has been often used for analysing both independent data (Kapetanios 2007) and time series (e. g. Baragona et al. 2004). Use of a GA for identifying a PAR model was first introduced by Ursu and Turkman (2012). An alternative, if extensive parallel computing facilities are available, is complete enumeration (e. g. Kontoghiorghes 2005).

The main contribution of this paper is to develop a flexible procedure that allows to deal with evolving seasonal variations in the most concise way. In fact, the model outlined above is composed by many parameters, referred to both the regimes and the seasonal patterns, but may be not all of them are necessary. The proposed procedure allows to build more parsimonious representations of the model, in order to facilitate the interpretation of results. We employ subset selection to decrease the number of AR parameters, but a more decisive gain in parsimony can be achieved by grouping the seasonal parameters. For example, if we observe a monthly hydrological time series we may find out that the hottest and the coldest months share, respectively, a similar behaviour as the mean level is considered: in that case we could consider only two seasonal means in the model instead of twelve, and the same may not hold true for the autocorrelation structure. Another example would be the so-called Monday effect in daily stock exchange data: that specific day, as the stock market re-opens, may require a set of parameters completely different with respect to the other days. These kind of problems have been generally accounted for in literature by means of hypothesis testing: in Franses and Paap (2004) a test of equality among all the autoregressive parameters is conducted to compare the PAR with a non periodic AR process; a test proposed in Thompson et al. (1985) allows to group together the seasonal means and AR parameters of two adjacent months basing on the equality of residual variances. Instead, our proposal allows to compare all possible arrangements of the seasonal positions (e. g. months) into groups and to select the best according to a specific criterion.

The paper is structured as follows: Section 2 introduces the models; Section 3 details the implementation of the computational method; Section 4 is concerned with a simulation study; Section 5 presents some applications and Section 6 concludes.

## 2 Model

We outline the model employed in the paper, which generalizes some procedures introduced in Cucina et al. (2018) and Battaglia et al. (2018). It concerns a seasonal time series  $X_t$ , ( $t = 1, \dots, N$ ) observed  $S$  times a year, whose behaviour is not homogeneous during the observation period. Thus  $m$  change times  $\tau_1, \dots, \tau_m$  specify the existence of  $m + 1$  regimes, in which the model may change its structure.

In each regime  $j$  ( $j = 1, \dots, m + 1$ ) the model also allows pooling the seasonal parameters, that regulate both the first and the second order properties of the underlying process. According to this parsimonious strategy, the same parameters may be adopted for describing a number of adjacent seasonal positions, avoiding a possible excess of detail. Let us consider the case of monthly data, for which  $S = 12$  and we label January as month 1, February as month 2, and so on. We could provide a description of the mean function using only two parameters, for example the months from 11 to 3 and from 4 to 10: in that case the number of seasons is decreased to  $s_M^{(j)} = 2$ . The labels of the months, to be grouped, are taken relative to the seasonal position of the first observation. But since we allow arbitrary season grouping, an additional parameter is needed to determine the starting months of each season. We indicate with  $f_M^{(j)}$  the first month of the first completely observed season for regime  $j$ . These parameters have an influence on how the seasons will be grouped. The resulting arrangement is unconstrained in terms of number (between 1 and  $S$ ) and size of the groups, and will be independently derived for the seasonal means and the AR models, these latter including the autoregressive parameters and the residual variances (for the season arrangement of the AR models we use the notation  $s_{AR}^{(j)}$  for the number of seasons and  $f_{AR}^{(j)}$  for the first observation of the first completely observed season of regime  $j$ ).

For each regime  $j$  ( $j = 1, \dots, m + 1$ ) we denote by  $\tau_{j-1}$  the starting time and by  $\tau_j - 1$  the last time, with:

$$1 \equiv \tau_0 < \tau_1 < \dots < \tau_m < \tau_{m+1} \equiv N + 1.$$

In order to ensure reliable AR estimates, we require at least  $L$  observations in each regime. The model for regime  $j$  is defined by:

$$X_t = a^{(j)} + b^{(j)}t + \mu^{(j)}(k^*) + W_t, \quad \tau_{j-1} \leq t < \tau_j,$$

$$W_t = \sum_{i=1}^p \phi_k^{(j)}(i)W_{t-i} + \varepsilon_t, \quad \tau_{j-1} \leq t < \tau_j,$$

where  $a^{(j)}$  and  $b^{(j)}$  are trend parameters,  $\mu^{(j)}(k^*)$  are the seasonal means,  $W_t$  is a PAR model, for which

$\phi_k^{(j)}(i)$  ( $i = 1, \dots, p$ ) are the AR parameters and  $\varepsilon_t$  is a periodic white noise with variance  $\sigma_{j,k}^2$ . Indices  $k^*$  and  $k$  denote, respectively, the season for the mean and the AR model at time  $t$ , and the number of seasonal groups  $s_M^{(j)}$  and  $s_{AR}^{(j)}$  are possibly different for each regime  $j$ .

Our model is an extension of that introduced in Franses and Paap (2004) because considers regime changes, and an extension of the model proposed in Lu et al. (2010) because allows regime-varying trend and PAR structures. A similar model, but without regime changes, was considered by So and Chung (2014): they assume a non periodic AR model but allow each seasonal mean to be stochastically varying around a long-term stationary level, rather than deterministic as in our case.

The proposed model is composed by three kind of parameters: the external parameters, which are chosen by the user or are implicitly derived from the available data; the structural parameters, which include the change times that specify the regime structure, the seasonal arrangements features and the indicators of subset models; the regression and autoregression parameters.

The seasonal arrangement is specified by a vector  $v_M^{(j)}(k^*)$  ( $v_{AR}^{(j)}(k)$  for the AR models) that indicates the starting seasonal position of each group. The subset models specification is indicated by binary vectors  $\delta_k^{(j)}$  of size  $p$  that denote presence (zero) or absence (one) of the parameter at each lag. In summary, considering regime  $j = 1, \dots, m+1$ , season of the mean  $k^* = 1 \dots s_M^{(j)}$ , season of the AR model  $k = 1 \dots s_{AR}^{(j)}$ , lag  $i = 1, \dots, p$ , the parameters may be listed as follows:

– **External parameters**

- $N$  Number of observations
- $S$  Number of observations per year
- $p$  Maximum AR order
- $C$  Maximum number of regimes
- $L$  Minimum number of observations per regime

– **Structural parameters**

- $m$  Number of change times
- $\tau_j$  Change times
- $\delta_k^{(j)}$  Subset PAR indicators
- $s_M^{(j)}, s_{AR}^{(j)}$  Number of groups for the means and the AR models
- $v_M^{(j)}(k^*), v_{AR}^{(j)}(k)$  Seasonal arrang. structure for means and AR models
- $f_M^{(j)}, f_{AR}^{(j)}$  Starting observation for means and AR models

– **Regression parameters**

- $a^{(j)}, b^{(j)}$  Trend parameters
- $\mu^{(j)}(k^*)$  Seasonal means
- $\phi_k^{(j)}(i)$  AR parameters

The main difficulty in building the model is the specification of structural parameters, because it is a complex combinatorial problem. We base our procedure on information theoretic ideas for which, informally, a model dominates another if it has a better fit while not being much more complex. It is a popular approach to model selection and requires the adoption of the model that optimizes a form of penalized likelihood, often called information criterion (e. g. Sin and White 1996). Such criteria are widely employed in linear and time series model building: most frequent is Akaike's AIC, but also others (e. g. Schwarz, Hannan and Quinn) are adopted. A generic form of information criterion ( $IC$ ) for dealing with a multi-regime model with seasonality like the present one may be written:

$$IC = \sum_{j=1}^{m+1} \sum_{k=1}^{s_{AR}^{(j)}} [n_{jk} \log \hat{\sigma}_{jk}^2 + \pi_{jk} \cdot P_{jk}], \quad (1)$$

where  $n_{jk}$  is the length of the subseries corresponding to regime  $j$  and season  $k$ ,  $\hat{\sigma}_{jk}^2$  is the average of the squared estimated residuals of that subseries,  $P_{jk}$  is the number of parameters of regime  $j$  and season  $k$ , and  $\pi_{jk}$  is the related penalization.

Following the famous Box's dictum (Box 1976) *All models are wrong but some are useful* we do not assume an underlying data generating process, but rather assign a score to each possible model, and the best one according to such score shall be selected. A wide discussion on this topic is provided in Kapetanios (2007).

### 3 Implementation

A model is specified by external, structural and regression parameters and is evaluated by an information criterion of the form (1). Given the structural parameters, the best values of the regression parameters  $a^{(j)}$ ,  $b^{(j)}$ ,  $\mu^{(j)}(k^*)$  and  $\phi_k^{(j)}(i)$  may be estimated by least squares and the residual variance estimate  $\hat{\sigma}_{jk}^2$  is obtained by the average of the estimated squared residuals  $\{\hat{\varepsilon}_t^2\}$  with  $t$  belonging to regime  $j$  and season  $k$ . One referee suggested to model the mean functions  $\mu^{(j)}(\cdot)$  (that are periodic functions with period  $s_M^{(j)}$ ) by trigonometric polynomials. This may be done without difficulty by a simple modification in the design matrix of the least squares estimation, and may provide additional parsimony, since not all the harmonic cycles may be necessary, when the number of seasons for the mean  $s_M^{(j)}$  is large.

The specification of structural parameters, on the other hand, will be tackled by means of a Genetic Algorithm (GA). The GA was introduced originally by

Holland (1992) as a formal imitation of the evolution of biological systems towards adaptation to the environment. It revealed soon its potential in solving statistical problems (Chatterjee et al. 1996) and has been successfully employed for problems similar to the present one (e. g. Baragona et al. 2004; Davis et al. 2006; Lu et al. 2010). A GA is a population-based evolutionary algorithm and proceeds iteratively in rounds (called generations) through stochastic rules (called genetic operators). Each possible solution is codified into a numeric (generally binary) vector called chromosome, and the goodness of that solution is expressed by a numerical function of the chromosome, called fitness function. The population includes several chromosomes and evolves through generations by reproduction of selected chromosomes (a stochastic mechanism called selection), better chromosomes in terms of fitness having a larger probability to survive in the next generation. Moreover, new chromosomes appear in the new population, originated by genetic operators: mutation (random changes of any binary digit) and recombination of genetic material (crossover operator: exchange of binary code portions between two chromosomes).

The GA, like many metaheuristic algorithms, is a stochastic search device for exploring a large discrete space of solutions, instead of an exhaustive enumeration. Under some assumptions the sequence of the best chromosomes at each generation converges in probability to the optimum (assumed unique), but the GA does not guarantee to reach the best solution in every instance. However in some cases, like the present one, the choice of some structural parameters, whose possible different values are not too many, may be done by complete enumeration, reaching the best result surely. This is possible for the subset indicators  $\delta_k^{(j)}$ : they will not be coded in the chromosome, but all the possible  $2^p$  subset models, for any given arrangement of the other structural parameters, will be estimated inside the fitness function, and the most fit retained. It is suitable because we consider relatively small orders  $p$  and the least squares estimation of subset AR models is fast, and it reduces the solution space by a factor  $2^p$ .

A further considerable simplification of the search process may be achieved by observing that our structural parameters are related to two different structures: the trend and change times on one side, and the seasons arrangement and the PAR models on the other. These two sets of parameters may be evolved by different GAs, thus we shall adopt a procedure in two stages.

In the first stage a GA will be run to select the number of change times  $m$ , the change times  $\tau_j$  and the trend parameters  $a^{(j)}$  and  $b^{(j)}$ . To avoid interactions, in this stage we consider full PAR models, different for

each seasonal position (e. g. months), i. e. with  $s_M^{(j)} = s_{AR}^{(j)} = S$  for any  $j$ . This procedure is more effective than the simpler sequential addition of change points like forward or stepwise methods (as discussed e. g. by Davis et al. 2006; Li and Lund 2012) because it approaches the global optimum examining a much bigger portion of the search space.

In the second stage, given the number of regimes and the change times, a second GA runs for selecting the optimal values of the remaining structural parameters: the number of different seasons in each regime, the length of each season, the subset indicators, the first observations  $f_M^{(j)}$  and  $f_{AR}^{(j)}$ . For these two last kinds of parameters we shall adopt complete enumeration as it will be explained in the next subsection. Thus our proposed procedure uses partly a GA and partly complete enumeration, so it may be considered a hybrid algorithm.

An implementation of GAs requires three kinds of critical choices: the coding procedure, the fitness function and the genetic operators. Moreover, the complexity of the present model selection problem suggests to derive some alternative, faster though sub-optimal, computational strategies.

### 3.1 Coding

The task of encoding the values of the structural parameters into a binary vector (chromosome) is often the most critical step in the implementation of a GA. Ideally, there should be a one-to-one correspondence between chromosomes and admissible solutions, but this is rarely possible when the solution space is not a cartesian product, or the values of the structural parameters interact with each other (this is especially true in grouping problems, see Falkenauer 1998). In particular, an efficient coding should avoid:

- Illegality: existence of chromosomes that do not correspond to feasible sets of structural parameters
- Redundancy: existence of different chromosomes that correspond to the same model.

In the GA of the first stage we encode the number of regime changes  $m$  and the change times  $\tau_j$ . For  $m$  we can simply use the base 2 representation of its decimal value, and if the maximum number of regimes  $C$  is a multiple of 2 this solution is efficient and legal. On the contrary, coding the change times  $\tau_j$  is not so simple, because of the ordering constraints and the lower limitation imposed by the minimum number of observations per regime  $L$ , that implies, given  $m$ , the following

inequalities:

$$\tau_1 \geq L + 1 ; \tau_2 \geq \tau_1 + L ; \dots$$

$$\tau_m \geq \tau_{m-1} + L ; N + 1 \geq \tau_m + L.$$

We adopt the encoding method introduced by Battaglia and Protopapas (2012) based on  $m$  real values between zero and one that determine the length of each regime as a portion of its maximum possible length. Through this procedure any  $m$ -tuple of numbers between zero and one  $(h_1, \dots, h_m)$  may be converted in a valid arrangement of change times  $(\tau_1, \dots, \tau_m)$ . Therefore the chromosome encodes the thresholds  $h_1, \dots, h_m$ , using for each of them a  $n$ -bit sequence. The size  $n$  should be large enough to allow that all integer values in the admissible range are possible for each  $\tau_j$ , but not too large for avoiding excessive redundancy. However redundancy cannot be completely avoided here, because in general it cannot be avoided that two contiguously coded values of  $h_j$  result in an identical integer value for  $\tau_j$ , but no illegality problem arises. The total number of binary digits of the chromosome in the first stage is therefore  $\log_2 C + n(C - 1)$ .

For the second stage we have a known number of regimes and the subseries belonging to each regime are given, so we must evolve the arrangement of the observations into seasons, both for the mean and for the AR structures. In each given regime  $j$  the  $S$  seasonal positions should be grouped into  $s_M^{(j)}$  different seasons for the mean and  $s_{AR}^{(j)}$  different seasons for the PAR structure. We shall explain the coding procedure of the possible season groupings only for the mean, for the PAR structure the method is similar; for ease of exposition we consider monthly data ( $S = 12$ ) starting in January and label the months from the first observation: 1 = January, 2 = February, ..., 12 = December. For any seasonal arrangement, we label with 1 the first season that is completely observed. Suppose for the moment that the first season starts at January (and is therefore labeled as season 1). In this case the arrangements may be described by the labels of the month of the first observation of each season, through the vector  $v_M = [v_M(1), v_M(2), \dots, v_M(s_M^{(j)})]$  with

$$1 = v_M(1) < v_M(2) < \dots < v_M(s_M^{(j)}) \leq S.$$

These vectors are in a one-to-one correspondence with the combinations of the objects  $\{2, 3, \dots, S\}$  taken  $s_M^{(j)}$  at a time. It follows that the total number of different season arrangements is  $2^{S-1}$ , and they may be coded directly and efficiently by a binary vector of  $S - 1$  bits. In such a vector the  $k$ -th bit equal to one denotes that a season starts at the  $(k + 1)$ -th month. Decoding the binary vector is immediate because  $v_M(1) = 1$  and  $v_M(2)$ ,

...,  $v_M(s_M^{(j)})$  are taken equal to one plus the positions of the binary vector where the bit equals one.

As an example, let us consider the following seasonal arrangement: first season includes months from January to March (1-3), second season months from April to July (4-7), third season months from August to October (8-10) and the fourth season includes November and December (11-12). Here  $s_M = 4$ ,  $v_M = [1, 4, 8, 11]$  and the corresponding binary vector is  $(0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0)$ .

This procedure should be repeated for the seasonal arrangements related to the means and to the PAR structures, and for each regime. Therefore the overall length of the chromosome in the second stage is  $2(S - 1)(m + 1)$ . There is no need to code the number of seasons  $s_M^{(j)}$  and  $s_{AR}^{(j)}$  because their values are implicit, and no illegal chromosomes nor redundancy arise.

We now release the assumption that there is a season starting in January. Then any grouping associated to a vector  $v_M$  may be associated to more than one actual season arrangement, because the first position of the first season, indicated by  $v_M(1) = 1$ , could now be associated to a month that is not the first of the observed data (i. e. January). Our convention is to label with 1 the first season which is completely observed, and to denote by  $f_M$  the starting month of that season. It follows that the month of the first observation for the first season, in the arrangement described by the vector  $v_M$ , equals  $f_M$  and not 1. Consequently, each season labeled  $k$  starts at month  $v_M(k) + f_M - 1$  (always assuming that the first observation of the series relates to month 1). The possible values of  $f_M$  are determined by the arrangement  $v_M$ . First of all, if the number of seasons is  $S$  or 1, the only possible value for  $f_M$  is 1. On the contrary, when  $1 < s_M < S$ , more cases are possible. However, since the first completely observed season starts at  $f_M$  rather than at 1, it follows that the preceding season (whose label is  $s_M$ ) was not completely observed, thus its starting month was earlier than month 1, or  $v_M(s_M) + f_M - 1 \leq S$  that implies  $f_M \leq S - v_M(s_M) + 1$ . We conclude that for a given season arrangement  $v_M$ , the possible values for the first observation  $f_M$  are contained in the range  $[1, S - v_M(s_M) + 1]$ . Note further that there is a first observation  $f_M$  for the means and one,  $f_{AR}$ , for the PAR structure, and they may be different in each regime. In order to avoid excessive complexity in the GA, we decided to completely enumerate all possible configurations of  $f_M$  and  $f_{AR}$  into the fitness computation step, and to retain the fittest solution.

In the second stage the trend parameters  $a^{(j)}$  and  $b^{(j)}$  are estimated again by least squares together with the seasonally grouped means (see next subsection).

### 3.2 Fitness function

The fitness function measures adaptation to the environment of each chromosome and has to be maximized, therefore it is natural to take a monotonically decreasing transformation of the information criterion  $IC$ . Our choice is simply

$$\text{fitness} = \exp\{-IC/\beta\}. \quad (2)$$

The convergence speed of a GA usually depends on the behaviour of the fitness function near the optimum, and the constant  $\beta$  may be used for avoiding an excessive flatness (for further details on this issue, also known as fitness scaling, see e. g. Baragona et al. 2011, p. 53). However, the most serious differences in the GA results are determined by the choice of the information criterion itself.

First of all, the value of  $IC$ , and consequently of the fitness, is extremely sensible to the choice of the penalizing constants. The values corresponding to the original Akaike criterion (constant equal to 2) may result in a relatively large number of structural parameters; an alternative may be linking the penalizing constant to the sample size, in a similar way to the Schwarz criterion ( $\log N$ ) or the Hannan-Quinn criterion ( $2c \log \log N$ ). Several implementations of model selection related to structural breaks adopted the minimum description length (MDL) criterion (e. g. Davis et al. 2006, 2008; Lu et al. 2010; Li and Lund 2012). A further option is selecting the penalizing constant in such a way that the information criterion behaves asymptotically like a hypothesis test where the null model is nested into the alternative model (for example testing the absence of seasonality, or testing that seasonality is constant across regimes), see Appendix.

For evaluating a model with  $m$  regime changes,  $s_M^{(j)}$  seasons for the mean and  $s_{AR}^{(j)}$  for the PAR models in the  $j$ -th regime, and the subset indicators  $\delta_k^{(j)}$ , we use the negative exponential transformation (2) of the following criterion:

$$IC = \sum_{j=1}^{m+1} \sum_{k=1}^{s_{AR}^{(j)}} [n_{jk} \log(\hat{\sigma}_{jk}^2) + \pi_{jk}(p - |\delta_k^{(j)}|)] + \pi \left( \sum_{j=1}^{m+1} s_M^{(j)} + m + 1 \right), \quad (3)$$

where the last term of the right hand side accounts for the seasonal means and the slopes.

In the first stage the chromosome includes only the number of regimes and the change times, the number of seasons is assumed equal to  $S$  (no season grouping

considered) and the PAR models are estimated without subset constraints, thus we apply (3) with  $s_M^{(j)} = s_{AR}^{(j)} = S$  and  $|\delta_k^{(j)}| = 0, \forall j, k$ . In order to simplify computations, we regress the data on a linear homogeneous trend with coefficient  $b^{(j)}$  and  $S$  monthly averages  $c_k^{(j)}$ , then we give to the seasonal means  $\mu^{(j)}(k)$  the meaning of deviations from the general regime trend letting  $a^{(j)} = (c_1^{(j)} + c_2^{(j)} + \dots + c_S^{(j)})/S$  and  $\mu^{(j)}(k) = c_k^{(j)} - a^{(j)}$ .

In the second stage we assume the number of regimes and the change times as given; the chromosome includes only coding for the season arrangement, while the optimal values of the subset indicators are obtained by complete enumeration. Complete enumeration is also used for computing the optimal values of the first observations  $f_M^{(j)}$  and  $f_{AR}^{(j)}$ , as illustrated previously. For the second stage also only a homogeneous trend and  $s_M^{(j)}$  seasonal averages  $c_k^{(j)}$  are estimated for each regime, the intercept  $a^{(j)}$  and the seasonal means  $\mu^{(j)}(k)$  are obtained from the constraint  $\mu^{(j)}(1) + \mu^{(j)}(2) + \dots + \mu^{(j)}(s_M^{(j)}) = 0$ . Lastly, simple linear constraints may be added in the least squares estimation of  $a^{(j)}, b^{(j)}, \mu^{(j)}(k)$  to ensure that the piecewise linear trend is continuous, though allowing discontinuity may help to identify possible level changes in the series.

Among possible choices of the information criterion, we shall apply that corresponding to the original Akaike's proposal, setting the penalizing constants  $\pi = \pi_{jk} = 2 \forall j, k$  (labeled  $AIC$ ) and a version corresponding to the Schwarz criterion obtained letting  $\pi_{jk} = \log(n_{jk})$  and  $\pi = \log N$  (labeled  $BIC$ ).

In any case, it is important to stress that different choices of the penalizing constants (or the information criterion itself) yield different optimal models (corresponding to the largest fitness).

Finally we observe that the information criterion in (3) is the simple average of the values for each sub-series related to a regime and a season. In some cases (for example when forecasting future observations is of special interest) it may be appropriate to assign different weights to the different regimes and modify the criterion accordingly.

### 3.3 Genetic operators

Both for the first and the second stage our GA implementation uses only the three classical operators of selection, crossover and mutation. Other more elaborated techniques introduced in the literature (e. g. inversion) were not found necessary to obtain satisfactory results. Like in all optimization applications of GA we adopted the so-called elitism: the best chromosome in each population survives to the next generation surely.

Selection is operated using the so-called roulette wheel rule: each chromosome is assigned a survival probability equal to the ratio of its fitness to the sum of the fitness on the entire current population. A simple alternative is using, for computing survival probabilities, not the fitness itself but its rank in the population (rank selection).

For crossover we use two alternative types: random one-cut-point and uniform. Both are applied to pairs of chromosomes. The first one consists in choosing at random one specific position (bit) in the two chromosomes, and exchanging between them the bits following that position (cut-point). The uniform crossover, on the contrary, acts on each single bit position, choosing the bit value of one of the two chromosomes at random (with probability  $1/2$ ). There is a predefined probability ( $P_{cross}$ ) that each unit of the population undergoes a crossover.

The mutation operator is also applied in a standard way: with a predefined probability ( $P_{mut}$ ) each bit of each chromosome in the population may be subject to flip from 0 to 1 or from 1 to 0 (bit-flip mutation).

In our experience we found that the results are not strongly influenced by the type of crossover operator, while the probabilities  $P_{mut}$  and  $P_{cross}$  have a larger effect on the performance. By means of a pilot experiment we chose the values  $P_{cross} = 0.7$  and  $P_{mut} = 0.2$ .

Two other important decisions concern the number of chromosomes in the population and the number of generations before stopping the algorithm. As to the population size, we follow Reeves (1993) who suggested to consider the probability that, in a random population, both values 0 and 1 are represented at each fixed bit position of the chromosome. If we want that this probability be at least  $Q$ , and the chromosome length is  $G$ , the population size should be at least  $1 + \log(-G/\log Q)/\log 2$ . For a dimension similar to our problems, and probability  $Q = 0.99$ , this rule gives sizes around 20.

Lastly, the number of generations is also very important. Few theoretical guidelines are available, except that the number of generations required for convergence depends on the dimension of the solution space. Several stopping rules were proposed, but is important basing on pilot experiments or preceding experience on problems of a similar nature.

### 3.4 Computational strategies

The complexity of a GA and the computational time required to converge depend on the size of the solution space and generally increase exponentially, therefore it

is always advisable, whenever possible, to decompose the optimization problem into subproblems with a solution space of lower dimension.

In the algorithm of the first stage we use  $n = 10$  bits for encoding the real thresholds  $h_j \in (0, 1)$ , as they ensure a sufficient precision for series up to a few thousands observations; with a maximum number of regimes  $C$ , the chromosome length is  $10(C - 1) + \log_2 C$  bits and the solution space may be reasonably explored by a standard GA implementation. Besides, there is no obvious meaningful decomposition of the solution space.

On the contrary, in the second stage where the chromosome length may be up to  $2C(S - 1)$ , especially with monthly data ( $S = 12$ ), the solution space is large and high-dimensional; moreover the fitness function is computationally heavier, since it includes complete enumeration of subset models and first observations  $f_M$  and  $f_{AR}$ . Thus, the derivation of alternative sub-optimal strategies, that may ensure a faster computation, appears advisable.

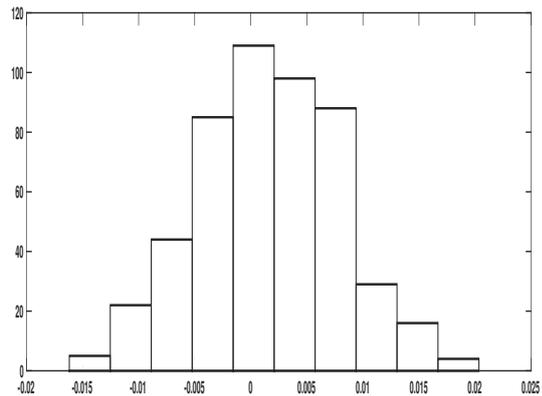
A first useful observation is that the information criterion is additive in the regimes, and the seasonal arrangement for one regime is independent from those of the other regimes. Thus the second stage GA may be applied independently and sequentially with respect to the regimes. This splits a search in a  $2^{(m+1)\ell}$  space into  $(m + 1)$  independent searches in  $2^\ell$  spaces, and reduces considerably the required overall number of generations. To be more precise, there is a certain amount of dependence between the model of one regime and that of the preceding regime, because the estimation of the AR parameters involves the detrended observations  $W_t$  of  $p$  preceding data: thus in the parameter estimation of the model for regime  $j$  we use the last  $p$  detrended observations  $W_t$  belonging to regime  $j - 1$  (and depending on the model selected for regime  $j - 1$ ). However we believe that the influence of such a feature on the overall results is negligible, therefore the proposed strategy may be considered essentially optimal rather than sub-optimal.

Another possibility of modifying the search is to evolve separately the seasonal arrangement for the means and that for the AR structures. More precisely, we may consider in a first step only chromosomes that encode different seasonal arrangements for the mean and a full seasonal splitting for the PAR models, i. e.  $S$  different models for the  $S$  seasonal positions. Once obtained the best seasonal grouping for the means, in a second step a GA is run with chromosomes that encode the optimal arrangement found for the means, and evolve the seasonal arrangement for the AR structure. Such a strategy, that we name conditional, requires the use of two GAs in sequence, each operating on a solution space

whose dimension is half the original. This idea may be applied in two different ways: evolving simultaneously the season grouping for the means of all regimes, and then, conditional on the results, finding the season groupings for the PAR models of all regimes (we call it a full conditional strategy), or applying the conditional procedure separately and sequentially for each regime (we call it a sequential conditional strategy). On a theoretical ground such strategies are sub-optimal because in each step the search for new chromosomes is limited to a linear subspace of the solution space. However in our experience the sequential conditional strategy seldom missed to find the optimal solution, and required a considerably smaller computational time.

#### 4 Simulation studies

We shall conduct five simulation experiments to evaluate the performance of the proposed procedure. The first three simulations are designed to evaluate the ability of detecting change times, while the last two experiments will be devoted to grouping seasons with similar autocorrelation structure or seasonal means. A summary of the parameters used to generate the models can be found in Table 1. In all the scenarios the  $\{\epsilon_t\}$  is generated from normal random variables with zero mean and  $\sigma_{jk}^2 = 1, \forall j, k$ . In the first experiment (Model I) the data are simulated from a PAR(1) model with two regimes: the seasonal means, indicated with  $\mu^A(k)$  in Table 1, are the same in the two regimes, whereas both the trend and the AR parameters change at time  $\tau_1 = 481$ . In the first regime there is no trend and the PAR structure is given by the parameters  $\phi_1^A(k)$ , while in the second the trend slope and intercept are  $a^{(2)} = -5$  and  $b^{(2)} = 0.007$  and the PAR parameters are all zero. In the second simulation (Model II) the true model contains two change times located at  $\tau_1 = 481$  and  $\tau_2 = 841$ . The seasonal means and the autocorrelation structures do not change in the three regimes and are specified by  $\mu^A(k)$  and  $\phi_1^A(k)$  respectively. The trend parameters in the three regimes are:  $a = (0, -5, 1)$  and  $b = (0, 0.0138, 0.0033)$ . In the third simulation (Model III) the true process contains one change time located at  $\tau_1 = 1801$ . This model has no trend and a constant PAR(1) structure with respect to the regimes specified by  $\phi_1^A(k)$ , while the seasonal means are equal to  $\mu^A(k)$  in the first regime and are reduced by a factor of 0.25 in the second. In the fourth simulation (Model IV) the data is simulated from a PAR(1) with two regimes without trend: in the first one there are 11 different seasonal means and 4 different PAR models denoted by  $\mu^A(k)$  and  $\phi_1^B(k)$  (monthly means of July and August are equal), whereas in the



**Fig. 1** Differences between the fitness of the selected model and that of the actual data generating process for Model IV and BIC criterion

second there are 4 different seasonal means specified by  $\mu^B(k)$  and a white noise autocorrelation structure. In the last scenario (Model V) the true process is a single regime PAR without trend with 3 different seasonal means and 4 different PAR models specified by  $\mu^C(k)$  and  $\phi_1^C(k)$ .

From each model we simulate 500 monthly ( $S = 12$ ) time series of length  $N = 1200$ , except for the Model III in which we use  $N = 3600$ , and apply our method to each series with two different fitness functions based, respectively, on the AIC and BIC criteria (see subsection 3.2).

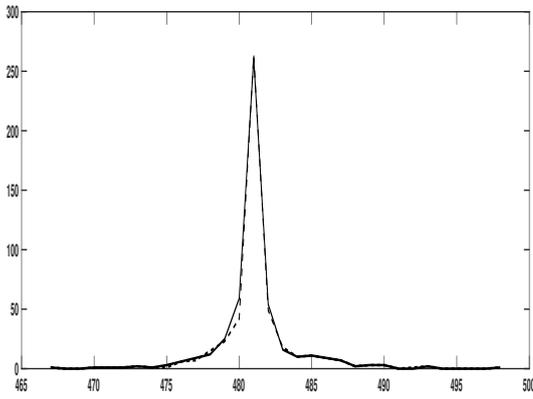
As far as the GA settings are concerned, the population size is fixed at 50 individuals, the roulette wheel is used as selection operator, the uniform crossover as recombination operator with rate 0.7 and the standard bit-flip mutation with probability set to 0.2. The elitist strategy is also employed, and the reaching of 500 generations is adopted as stopping criterion for the algorithm.

We focus on an important aspect before describing the simulation results. For some replications the best model provided by the GA can achieve a larger fitness than the one of the actual data generating process. Such finding will not be considered an error, because the aim of our proposal is not to discover the generating process but rather to select the best parsimonious model both in terms of fit and parsimony according to a specific criterion. For example Figure 1 displays the histogram of the observed differences between the best fitness reached in each replication and fitness of the actual data generating process for Model IV and BIC criterion.

For the data simulated from Model I, the percentage of simulations for which the correct number of regimes is detected is 96.2% for AIC and 100.0% for the BIC. The estimated change time locations are reported in Figure 2. This plot reports the distribution of the de-

Month	$\mu^A(k)$	$\phi_1^A(k)$	$\mu^B(k)$	$\phi_1^B(k)$	$\mu^C(k)$	$\phi_1^C(k)$
1	-0.61	0.30	0.0	0.5	0.0	0.7
2	0.99	0.30	0.0	0.5	0.0	0.7
3	2.35	0.50	0.0	0.5	0.0	0.7
4	4.91	0.30	8.0	0.3	0.0	0.3
5	8.74	0.35	8.0	0.3	6.0	0.3
6	12.15	0.30	8.0	0.3	6.0	0.3
7	15.55	0.25	15.0	0.3	6.0	0.3
8	15.47	0.10	15.0	-0.1	6.0	-0.2
9	12.79	0.10	15.0	-0.1	2.0	-0.2
10	7.82	0.10	3.0	-0.1	2.0	-0.2
11	2.32	0.20	3.0	0.2	2.0	0.0
12	-0.25	0.20	3.0	0.2	2.0	0.0

**Table 1** Summary of the values of monthly means and AR parameters used in simulations

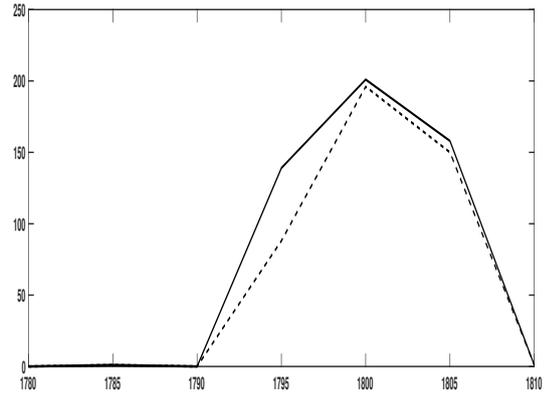


**Fig. 2** Change times detected for Model I (continuous line: BIC; dashed line: AIC)

tected change times in the 500 replications. We can observe that both the curve of AIC and the one related to BIC have a mode around the true change time location  $\tau = 481$  without large differences.

In the second simulation experiment, Model II has three regimes different only for the trend, while the seasonal means and the PAR structure do not change. With the fitness based on AIC the number of regimes is correctly specified nearly always (92.2%). The change times  $\tau_1 = 481$  and  $\tau_2 = 841$  are correctly specified (with an error of  $\pm 1$ ) in 56.8% and 89.2% of the replications, respectively; it seems that the first change time was more difficult to detect. With the fitness based on the more parsimonious BIC criterion the number of regimes is underestimated: a single regime is found in 67.2% of the replications, two regimes in 19%, and three regimes in 13.8%.

For the third scenario (Model III) the number of regimes is always correctly specified for the BIC criterion, and in 87.2% of the replications for AIC. Figure 3 displays a count plot of the detected change times for the 500 replications with intervals size equal to 5. We can see that, for example, approximately in 201 repli-



**Fig. 3** Change times detected for Model III (continuous line: BIC; dashed line: AIC)

Number of seasons	Model IV				Model V	
	$s_M^{(1)}$	$s_M^{(2)}$	$s_{AR}^{(1)}$	$s_{AR}^{(2)}$	$s_M^{(1)}$	$s_{AR}^{(1)}$
1	-	-	-	75	-	-
2	-	-	14	46	-	-
3	-	-	53	79	-	-
4	-	21	104	117	101	38
5	-	88	135	83	158	145
6	-	152	107	60	139	178
7	-	139	63	28	74	102
8	-	60	18	9	21	33
9	-	26	1	3	6	2
10	64	5	-	-	1	1
11	173	2	-	-	-	1
12	263	7	-	-	-	-

**Table 2** Number of estimated season groups, Models IV and V, fitness based on AIC

Number of seasons	Model IV				Model V	
	$s_M^{(1)}$	$s_M^{(2)}$	$s_{AR}^{(1)}$	$s_{AR}^{(2)}$	$s_M^{(1)}$	$s_{AR}^{(1)}$
1	-	-	24	470	-	-
2	-	-	222	25	-	-
3	-	-	182	4	-	12
4	-	36	66	1	269	240
5	-	194	6	-	216	224
6	-	219	-	-	14	22
7	-	50	-	-	1	2
8	-	1	-	-	-	-
9	-	-	-	-	-	-
10	155	-	-	-	-	-
11	171	-	-	-	-	-
12	174	-	-	-	-	-

**Table 3** Number of estimated season groups, Models IV and V, fitness based on BIC

cations a time in the interval 1799 – 1803 is detected as change time.

For Model IV the two regimes were nearly always correctly selected (94% of the replications for AIC, 100% for BIC) and similar findings for the single regime Model V (97% of the replications for AIC, 100% for BIC). The detailed results on the number of specified seasons for means ( $s_M$ ) and the PAR structure ( $s_{AR}$ ) are reported

in Table 2 for AIC and Table 3 for BIC. It may be observed that the number of specified seasons tends to be larger than expected, slightly with the BIC and more widely with the AIC criterion. Such over-selection is more evident for the AR models than for the seasonal means.

In order to evaluate the results on season grouping more synthetically, a measure of agreement between the true arrangement and the estimated one is needed. We adopt the Rand index ( $R$ ) (Rand 1971), which is a standard measure in the literature of cluster analysis. It ranges from 0 (no pairs classified in the same way under both arrangements) to 1 (identical groups). The value of  $R$  depends on both the number of clusters and the number of elements. A drawback of the Rand index is that its expected value does not take a constant value under an appropriate null model. Therefore the adjusted Rand index ( $R^*$ ) (Hubert and Arabie 1985) will be also computed. The expected value of  $R^*$  under a completely random classification is 0. Such measure will be computed separately for the means and the PAR structure.

Concerning Model IV, combining the GA with AIC criterion the average (on 500 replications) of the Rand index is 0.974 for the seasonal means and 0.778 for the PAR structure. The Adjusted Rand index is 0.843 for the seasonal means and is 0.491 for the PAR structure. Using BIC criterion we obtain an average  $R$  equal to 0.982 for the seasonal means and equal to 0.939 for the PAR structure. The corresponding average of  $R^*$  is 0.899 for the seasonal means and is 0.875 for the PAR structure.

For Model V, our methods combined with AIC criterion provide an average  $R$  equal to 0.898 for the seasonal means and equal to 0.932 for the PAR structure. The average of  $R^*$  is 0.749 for the seasonal means and is 0.81 for the PAR structure. With BIC we obtain an average  $R$  of 0.937 for the seasonal means and equal to 0.939 for the PAR structure, while the corresponding values for  $R^*$  are 0.852 and 0.849.

As an additional evaluation, we shall consider also the differences between the estimated partition of the observations set into seasons and that of the true models. If the former is a refinement of the latter, no pair of observations is erroneously split into two true seasons. We observe that the selected partition is always a refinement of the true one in the 500 replications for the means in every simulation experiment, while the season arrangement of the AR structures is less precise. To be more specific, we computed (only for the seasonal arrangement of the AR structure) for each replication the percentage of observation pairs that, belonging to the same true season, are correctly classified into the same

Criterion	AIC		BIC	
	Model IV	Model V	Model IV	Model V
0 – 60%	1	0	0	13
60 – 70%	6	2	2	18
70 – 80%	47	9	24	14
80 – 90%	105	116	266	283
90 – 100%	341	373	208	172

**Table 4** Percentages for each replication of the observations pairs belonging to the same true season for the AR structures and classified into the same estimated season

estimated season. We report the results in Table 4 for both Model IV and V. The frequencies are large, better for the AIC than the BIC, and slightly better for Model IV than Model V.

We considered also, for comparison, two standard changepoint tests: the original Cusum test based on recursive residual of Brown et al. (1975) and the Cusum test based on ordinary least squares (ols) residuals of Ploberger and Krämer (1992), including in the regression a constant and the index of time, with size 0.05. We referred to the R package *strucchange* (Kleiber et al. 2002) for the computations. Such methods are not designed for dealing with evolving seasonality, and the results were not satisfactory. On the series simulated according to Models I and IV all tests suggest no break in more than 99% of the replications. For Model II (3 regimes) the Cusum test based on recursive residuals detects, in about 58% of the replications, a break around the second actual change time ( $\tau_2 = 841$ ) and never the first one, while the ols-based Cusum test detects a unique break around the first change time ( $\tau_1 = 481$ ) but only for 24 of the 500 replications. For the simulations according to Model III, the ols-based Cusum test detects a break around the actual date  $\tau_1 = 1801$  in 43 series, while the Cusum based on recursive residuals suggests in 137 replications a break towards the last observations (located between 2500 and 2900). Similar results were found for Model V (that has no regime change): the Cusum test based on recursive residuals detected one break at the last observations in 210 series, while the ols-based test never suggested any break.

## 5 Applications

We used our proposed model to analyse time series in various fields such as climatology, hydrology and economics, and found that results are comparable with more specific models presented in literature. The proposed model privileges parsimony but allows for reproducing both long term and seasonal inhomogeneity, and the only critical decision left to the user is the choice of the information criterion.

In the present applications we adopted the hybrid algorithm strategy: the regime changes are searched on complete models, then the season groupings and subset choices are evaluated by the second stage GA and the best one retained. The GA population size was 50 and the number of generations 200. We used roulette wheel selection with elitist strategy and uniform crossover.

Various specifications of the proposed models were fitted: a *complete* model with no restrictions on the AR parameters and no season grouping; a *subset* model with no season grouping but choosing the best subset AR model for each month according to fitness; a *grouped* model where the seasonal parameters for the mean and for the AR parameters are grouped together in the best arrangement with respect to fitness; a *non periodic* AR model, where the AR parameters are assumed equal for each month and regime; and finally (if more than one regime is detected) a *constant season* grouped model where we assume that the seasonal components (number and values of the seasonal means, number, structure and parameter values of the AR models) are the same in each regime. The fitness comparison of the different models helps to highlight the importance of the seasonality and its evolving behaviour.

### 5.1 Italian Industrial Production Index

We focus on the time series of monthly Industrial Production Index (IPI) in Italy. It is among the most important macroeconomic indicators, so understanding and forecasting its behaviour is a fundamental need for many decision makers. Most of the recent research on this topic (e. g. Bodo et al. 1991; Bruno and Lupi 2004; Bulligan et al. 2010; Girardi et al. 2016) is devoted to forecasting using various proposed leading indicators, and the forecasts are compared with those of a univariate linear model, generally a seasonal multiplicative  $SARIMA(3, 0, 0) \times (0, 1, 0)_{12}$  model (see Bulligan et al. 2010). We will show that our proposed procedure is superior with respect to such univariate models which do not use other indicators.

We considered the monthly time series of the IPI (Ateco class 020, base 2015) from January 1990 to December 2017 (336 observations)<sup>1</sup>. The series is displayed in Figure 4 (left panel). Our proposed models (with order  $p = 3$ ) were fitted under various specifications and fitness based on BIC; the results are summarized in Table 5.

In all cases two regimes are detected, with a change point in November, 2008 (the Cusum test with ols residuals detects a break around the same date and an ad-

ditional break during 1996, while the test based on recursive residuals does not suggest breaks). Grouping means does not provide a substantial advantage in terms of fitness, since the only suggested merging is May with June in the first regime and October with November in the second regime (models with 11 different monthly means). The estimated monthly means for the two regimes are reported in Figure 4 (right panel). The variance of the series is 313, the variance of the detrended demeaned series  $W_t$  is 23.5, and the global residual variance ranges from 8 to 9.8 (according to the model), therefore the variability is explained for more than 90% by trend and seasonal means, and the periodic autoregression accounts for about 60% of the remaining variability.

The comparison of the non periodic AR and constant season models with the other more flexible alternatives, both in terms of residual variance and of fitness, suggests that as far as the second moments are concerned, the seasonal pattern is also relevant and changes according to the regime.

The best model in terms of fitness is the grouped season model. For the first regime only four different AR structures are specified, associated to the months of April, May to October, November, December to March. In the second regime, a more articulated PAR model is suggested, with parameters equal only for December-January, April-May, August-September. The usual portmanteau tests applied to global residuals and residuals of each season are not significant at level 0.05.

The last column in Table 5 shows the results obtained for the SARIMA model  $(1 - 0.167B - 0.269B^2 - 0.426B^3)(1 - B^{12})X_t = -0.162 + \varepsilon_t$  that are uniformly far less satisfying.

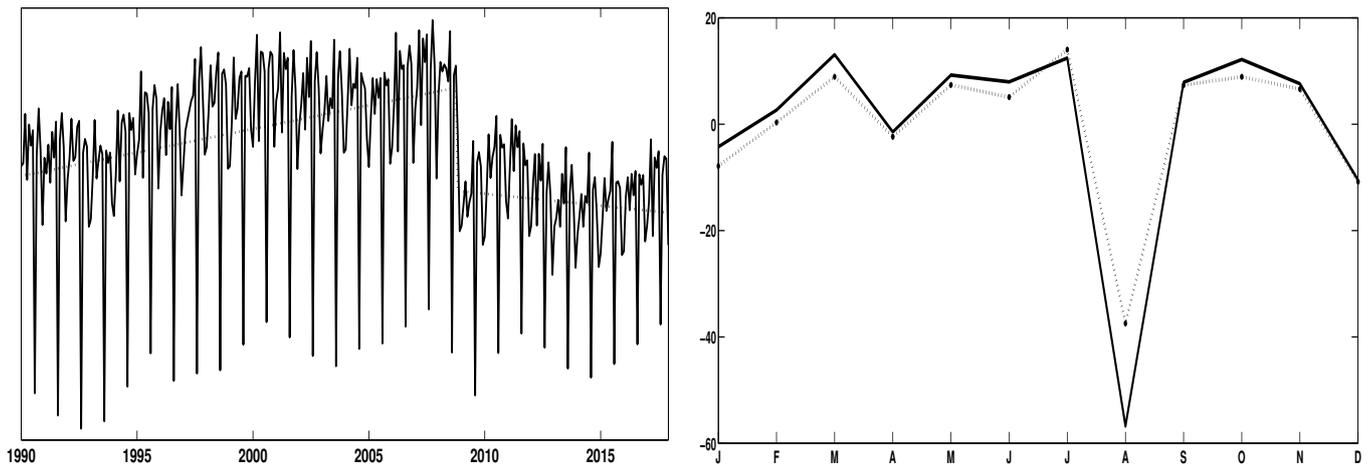
The results of a small forecasting experiment appear in the last two rows of Table 5. Employing also fresh data from January to March 2018, we computed one to six-step-ahead forecasts for origins from January to September 2017, and the average square forecast errors for the 9 forecasts at each fixed horizon are reported. There is a large variability due to the reduced number of instances, but in general we may conclude that the forecast ability of the SARIMA and that of the constant season models are smaller than the other models.

### 5.2 Central England Temperature monthly data

The CET series of monthly mean surface temperature for a location in the Midlands region<sup>2</sup> is one of the longest time series available for monthly temperature.

<sup>1</sup> Downloaded from <http://dati-congiuntura.istat.it/>

<sup>2</sup> Downloaded from <http://www.metoffice.gov.uk/hadobs/hatcet/data/download.html>



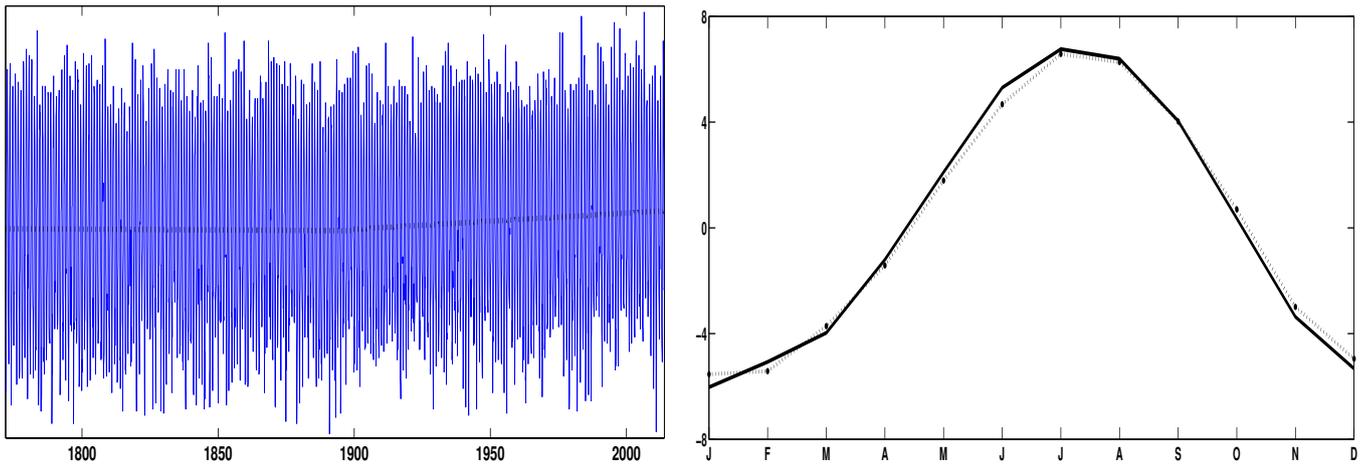
**Fig. 4** Left panel: Italian industrial production index 1990-2017 (dotted line: estimated trend). Right panel: monthly means, continuous line: first regime; dotted line: second regime

Model	Complete	Subset	Grouped	Const. season	AR	SARIMA
Regimes	2, 11/2008	2, 11/2008	2, 11/2008	2, 11/2008	2, 11/2008	1
Fitness ( $\times 100$ )	3.70	4.91	5.70	4.24	3.13	3.99
Means	(12,12)	(12,12)	(11,11)	11	11	
AR models	(12,12)	(12,12)	(4,9)	8	1	
Res. Var.	8.07	8.63	9.86	20.39	23.84	23.35
MSFE(1:3)	34.4, 27.4, 21.2	38.1, 26.1, 28.5	32.8, 27.9, 24.2	45.2, 42.5, 63.6	26.5, 20.3, 24.8	34.4, 36.5, 31.9
MSFE(4:6)	33.2, 42.3, 41.5	29.2, 42.8, 42.6	23.2, 34.2, 40.1	15.2, 35.5, 50.6	24.9, 29.0, 36.7	31.4, 33.7, 28.9

**Table 5** Estimated models for the industrial production index series

It starts at year 1659, but since a quality issue arises for data before 1772, we analysed data from January 1772 to December 2013 (2904 observations). This series was studied in a recent paper (Proietti and Hillebrand 2017), where a detailed analysis and review of the results in literature may also be found. The main features of the series are a trend change around the beginning of 20-th century, due to global warming, and an evolving seasonal structure appearing as a shift of the seasonal cycle toward an earlier inception of the Spring, that has been identified in the literature as precession of Earth's axis of rotation. Proietti and Hillebrand (2017) consider a state-space model with a separate stochastic trend for each month and deviations from trend following a PAR(1) model with white noise variance different for each month. To achieve parsimony, for each parameter type, the 12 monthly values are assumed to be a linear combination of only five representative months (January, March, July, October, December). As a basic measure of goodness of fit, Proietti and Hillebrand (2017) report the estimated one-step-ahead prediction error variance (averaged over the 12 months) equal to 1.8486, and the results of a forecasting exercise for multi-step-ahead forecasts with origin moving from January 1979 to December 2014; the observed average squared forecast errors appear in the last column

of Table 6. We fitted our proposed models (with order  $p = 1$  as used in Proietti and Hillebrand) under various specifications. For the fitness the AIC information criterion ( $\pi = 2$ ) was used since the very large number of observations makes the BIC exceedingly parsimonious. The results are summarized in Table 6. Two regimes were selected, with change time at October 1897 and a larger positive slope in the second regime, agreeing with the assessed increase in temperature starting at the beginning of the 20-th century (the Cusum tests did not detect any break). Most of the variability (around 90%) is accounted for by the trend and the monthly means, and grouping the monthly means has not a positive effect on fitness. However, the fitness of the constant season model and of the non periodic model are smaller than those of the other models, suggesting that the autocorrelation exhibits a seasonal pattern. Figure 5 (right panel) reports the graph of the monthly means estimated in the two regimes, and partly supports the hypothesis of a shift in the seasonal cycle. The effect of the AR structure is limited, therefore the various models do not yield much different results, but the best model in terms of fitness is the grouped season model, with only 5 different AR structures for the first regime (months: March-April, May, June-September, October-November, December-February) and 8 for the second



**Fig. 5** Left panel: CET series and trend (dotted line). Right panel: monthly means, continuous line: first regime; dotted line: second regime

Model	Complete	Subset	Grouped	Const. season	AR	Proietti Hillebrand
Regimes	2, 10/1897	2, 10/1897	2, 10/1897	2, 10/1897	2, 10/1897	1
Fitness	.566	.599	.603	.595	.557	.590
Means	(12,12)	(12,12)	(12,11)	12	12	5
AR models	(12,12)	(12,12)	(5,8)	8	1	5
Res. Var.	1.717	1.718	1.724	1.759	1.776	1.848
MSFE(1:3)	1.71, 1.81, 1.83	1.72, 1.81, 1.83	1.70, 1.80, 1.82	1.72, 1.81, 1.83	1.72, 1.81, 1.83	1.67, 1.79, 1.81
MSFE(4:6)	1.84, 1.85, 1.87	1.84, 1.85, 1.87	1.83, 1.84, 1.86	1.84, 1.85, 1.87	1.84, 1.85, 1.87	1.81, 1.81, 1.81

**Table 6** Estimated models for the central England temperature series

regime (months: January-February, March, April-July, August, September, October, November, December). None of the usual portmanteau tests applied to global residuals and residuals of each season suggests to reject this model. The last rows for each model in Table 6 display the estimated residual variance and the 1 to 6-step-ahead mean square forecast errors for the forecast exercise proposed in Proietti and Hillebrand (2017), labeled MFSE(1-6). Results are similar, with our models having a slightly better fit and a slightly worse forecast ability.

### 5.3 Saugeen river flow

We shall analyse the Saugeen river flow data, measured at Walkerton, Ontario. Observations are average monthly flows in  $m^3/sec$  from January 1915 to December 1976<sup>3</sup>.

River flows of Saugeen have already been analysed in Noakes et al. (1985); Thompstone et al. (1985); Wong et al. (2007). The third paper adapts a functional coefficient autoregression for the analysis, while the first two employ PAR models. Thompstone et al. (1985) also

propose a parsimonious version of the PAR, introduced to reduce the number of parameters, which allows to group together pairs of adjacent months by means of tests on equality of variances of fitted residuals. If two months are joined, this is assumed both for the mean and AR parameters; moreover, no trend is allowed. We consider 708 observations from January 1915 to December 1973, order  $p = 3$  and fitness based on BIC. Data are plotted in Figure 6 (left panel).

No regime change was detected (also by the Cusum tests) but a slowly increasing trend. Here also various model specifications were fitted and the results are summarized in Table 7. For this series, the total variance is explained for nearly 60% by trend and seasonal means, and the periodic autoregression accounts for about 35% of the remaining variability.

First of all, the considerably smaller fitness (and larger residual variance) of the non periodic model suggests that the choice of periodic AR models for this series is appropriate. It may also be seen that the subset model provides a substantial advantage, in terms of fitness, on the complete model. The best model is the grouped seasonal, with 9 seasonal means (only means of August-September, and December to February are grouped) and only 5 different AR structures (months of March, April, May to September, October-November,

<sup>3</sup> Downloaded from <http://www.stats.uwo.ca/faculty/mcleod/epubs/mhsets/MHSETS.zip>

Model	Complete	Subset	Grouped	AR
Fitness	3.92	4.36	4.82	4.11
Means	12	12	9	12
AR models	12	12	5	1
Res. Var.	0.175	0.177	0.182	0.210
MSFE ( $\times 10^3$ )	175	158	163	212

**Table 7** Estimated models for the Saugeen river flow series

December to February). None of the usual portman-teau tests applied to global residuals and residuals of each season suggests to reject this model. The monthly means are reported in Figure 6, right panel. We have also computed out-of-sample one-step-ahead forecasts for the same range considered by Noakes et al. (1985): 36 months from January 1974 to December 1976, the observed average square forecast errors are shown in the last row of Table 7. Noakes et al. (1985) report corresponding figures of 186 for the  $SARIMA(1, 0, 0) \times (0, 1, 1)_{12}$  model and 177 for their best fitted PAR by BIC criterion (with no trend and no season grouping). For their model, based on 8 seasons, Thompstone et al. (1985) report a corresponding error equal to 182.

## 6 Conclusions

The idea of allowing different regimes in time where both the trend, the seasonal means and the autocorrelations may change seems adequate for modelling several phenomena with evolving features inside a large time span.

More parsimonious but satisfactorily fit models were obtained by arbitrarily grouping the seasonal patterns. The balance between goodness of fit and parsimony is controlled by the information criterion on which the fitness function is based. The choice of the information criterion is very important and critical: we have considered the original Akaike's and the Schwarz criterion, but several other forms may be used.

For selecting and estimating a model inside the very large space of solutions we proposed a procedure based on genetic algorithms with an efficient coding, that proved to be effective and successful. Alternative sub-optimal strategies were also outlined in case the solution space is exceedingly large and high-dimensional (as may be the case with weekly or decadal observations).

Finally we observe that many extensions are possible. A natural generalization is considering smooth transitions between regimes, as in Teräsvirta (1994). A further interesting extension would be allowing regime changes driven not by time, but by the values assumed by a preceding observation (like in Tong 1990), or both (like in Battaglia and Protopapas 2012).

## Appendix Asymptotic equivalence of a test of hypothesis and an information criterion

Let us consider two nested models  $M_0$  and  $M_1$ , where  $M_0$  has  $p$  parameters and  $M_1$  has  $q$  additional parameters that under the null hypothesis  $H_0$  are assumed equal to zero. A standard  $F$  test under gaussianity for  $H_0$  is obtained by the test statistic:

$$F = \frac{(SS_0 - SS_1)/q}{SS_1/(N - p - q)} \quad (4)$$

where  $SS_0$  and  $SS_1$  are the residual sum of squares of the model  $M_0$  and  $M_1$  fitted to  $N$  observations.

Under the null hypothesis the statistic (4) follows a  $F$  distribution with  $q$  and  $N - p - q$  degrees of freedom. Let  $\bar{F}_\alpha$  denote the  $(1 - \alpha)$ -quantile of that distribution, then hypothesis  $H_0$  is rejected at level  $1 - \alpha$  whenever  $F > \bar{F}_\alpha$ , or

$$\log(SS_0) - \log(SS_1) > \log\left(1 + \bar{F}_\alpha \frac{q}{N - p - q}\right).$$

For  $N$  large the right hand side may be approximated by  $\bar{F}_\alpha q/(N - p - q)$  and rejection is equivalent to

$$\log(SS_0) - \log(SS_1) > \bar{F}_\alpha \frac{q}{N - p - q}$$

that may be also written  $IC_0^* - IC_1^* > 0$  where  $IC_0^* = \log(SS_0) + \pi^* p$ ,  $IC_1^* = \log(SS_1) + \pi^* (p + q)$  and

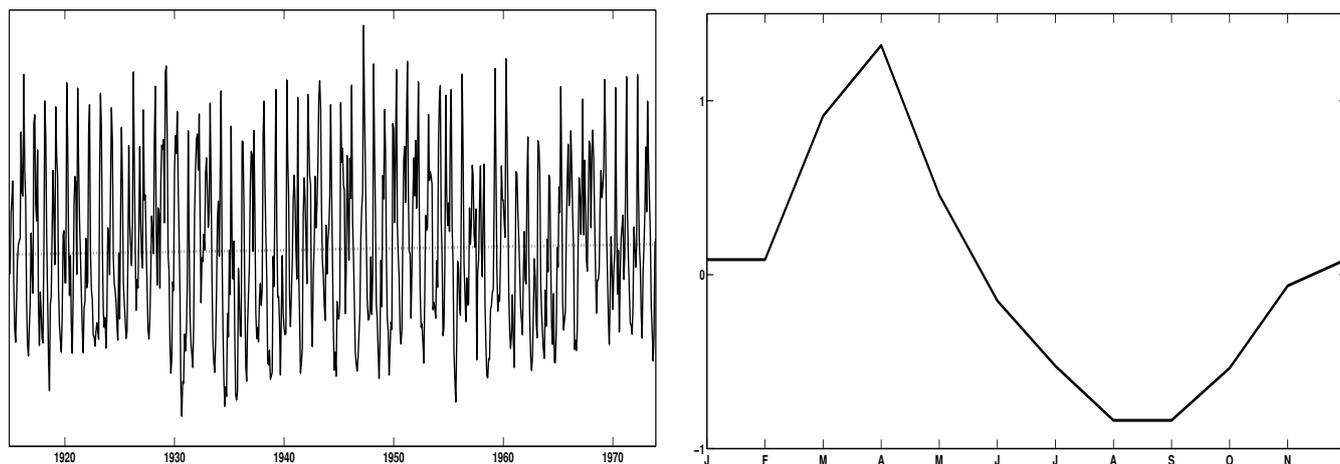
$$\pi^* = \frac{N}{N - p - q} \bar{F}_\alpha. \quad (5)$$

It follows that choosing a penalizing constant equal to  $\pi^*$  in (5), model  $M_1$  will be preferred to model  $M_0$  by the information criterion whenever the test rejects the null hypothesis  $H_0$ .

**Acknowledgements** We thank the Associate Editor and the Referees for their useful comments and suggestions. We are grateful to Eugen Ursu for involving us in research on periodic autoregressive models. This research was supported by University La Sapienza, Rome under grant C26A15NK4H.

## References

- Bai J, Perron P (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66(1):47–78
- Baragona R, Battaglia F, Cucina D (2004) Fitting piecewise linear threshold autoregressive models by means of genetic algorithms. *Computational Statistics & Data Analysis* 47(2):277–295
- Baragona R, Battaglia F, Poli I (2011) *Evolutionary Statistical Procedures*. Springer, Berlin
- Battaglia F, Protopapas M (2012) Multi-regime models for nonlinear nonstationary time series. *Computational Statistics* 27(2):319–341
- Battaglia F, Cucina D, Rizzo M (2018) A generalization of periodic autoregressive models for seasonal time series. Tech. Rep. 2, Department of Statistical Sciences, University La Sapienza, Rome, Italy, ISSN 2279-798X
- Bodo G, Cividini A, Signorini LF (1991) Forecasting the italian industrial production index in real time. *Journal of Forecasting* 10(3):285–299



**Fig. 6** Left panel: Saugeen series (logs) and trend (dotted line). Right panel: monthly means

- Box GEP (1976) Science and Statistics. *Journal of the American Statistical Association* 71(356):791–799
- Brown RL, Durbin J, Evans JM (1975) Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Ser B* 37(2):149–192
- Bruno G, Lupi C (2004) Forecasting industrial production and the early detection of turning points. *Empirical Economics* 29(3):647–671
- Bulligan G, Golinelli R, Parigi G (2010) Forecasting monthly industrial production in real-time: from single equations to factor-based models. *Empirical Economics* 39(2):303–336
- Caporale G, Cunado J, Gil-Alana LA (2012) Deterministic versus stochastic seasonal fractional integration and structural breaks. *Statistics and Computing* 22(2):349–358
- Chatterjee S, Laudato M, Lynch LA (1996) Genetic algorithms and their statistical applications: an introduction. *Computational Statistics & Data Analysis* 22(6):633–651
- Cucina D, Rizzo M, Ursu E (2018) Multiple changepoint detection for periodic autoregressive models with an application to river flow analysis. arXiv, 1801.01697
- Davis R, Lee T, Rodriguez-Yam G (2008) Break detection for a class of nonlinear time series models. *Journal of Time Series Analysis* 29(5):834–867
- Davis RA, Lee TCM, Rodriguez-Yam GA (2006) Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association* 101(473):223–239
- Falkenauer E (1998) *Genetic Algorithms and Grouping Problems*. Wiley, Chichester
- Franses PH (1995) The effects of seasonally adjusting a periodic autoregressive process. *Computational Statistics & Data Analysis* 19(6):683–704
- Franses PH, Paap R (2004) *Periodic Time Series Models*. Oxford University Press, New York
- Girardi A, Guardabascio B, Ventura M (2016) Factor-augmented bridge models (FABM) and soft indicators to forecast Italian industrial production. *Journal of Forecasting* 35(6):542–552
- Gladyshev E (1961) Periodically correlated random sequence. *Soviet Math* 2:385–388
- Hipel KW, McLeod AI (1994) *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, Amsterdam
- Holland JH (1992) *Adaptation in Natural and Artificial Systems*, 2nd edn. MIT Press, Cambridge
- Hubert L, Arabie P (1985) Comparing partitions. *Journal of Classification* 2(1):193–218
- Kapetanios G (2007) Variable selection in regression models using nonstandard optimisation of information criteria. *Computational Statistics & Data Analysis* 52(1):4–15
- Kleiber C, Hornik K, Leisch F, Zeileis A (2002) strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software* 7(2):1–38
- Kontoghiorghes EJ (2005) *Handbook of Parallel Computing and Statistics*. Chapman & Hall/CRC, Boca Raton
- Li S, Lund R (2012) Multiple changepoint detection via genetic algorithm. *Journal of Climate* 25:674–686
- Lu Q, Lund R, Lee T (2010) An MDL approach to the climate segmentation problem. *The Annals of Applied Statistics* 4(1):299–319
- Noakes DJ, McLeod AI, Hipel KW (1985) Forecasting monthly riverflow time series. *International Journal of Forecasting* 1(2):179–190
- Ploberger W, Krämer W (1992) The CUSUM test with OLS residuals. *Econometrica* 60(2):271–285
- Proietti T, Hillebrand E (2017) Seasonal changes in Central England temperatures. *Journal of the Royal Statistical Society Ser A* 180(3):769–791
- Rand WM (1971) Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* 66(336):846–850
- Reeves CR (1993) *Modern Heuristic Techniques for Combinatorial Problems*. Wiley, New York
- Sin CY, White H (1996) Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics* 71(1-2):207–225
- So MK, Chung RS (2014) Dynamic seasonality in time series. *Computational Statistics & Data Analysis* 70:212–226
- Teräsvirta T (1994) Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89(425):208–218
- Thompstone R, Hipel K, McLeod I (1985) Grouping of periodic autoregressive models. In: Anderson O, Ord J, Robinson E (eds) *Time Series Analysis: Theory and Practice* 6. Proc. Conference Toronto 1983, North-Holland, Amsterdam, pp 35–49

- 
- Tong H (1990) *Non Linear Time Series: a Dynamical System Approach*. Oxford University Press, Oxford
- Ursu E, Turkman KF (2012) Periodic autoregressive model identification using genetic algorithms. *Journal of Time Series Analysis* 33(2):398–405
- Wong H, Ip WC, Zhang R, Xia J (2007) Non-parametric time series models for hydrological forecasting. *Journal of Hydrology* 332(3-4):337–347
- Zeileis A, Kleiber C, Krämer W, Hornik K (2003) Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis* 44(1-2):109–123