Periodic Autoregressive Models with Multiple Structural Changes by Genetic Algorithms

Francesco Battaglia, Domenico Cucina, and Manuel Rizzo

Abstract We present a model and a computational procedure for dealing with seasonality and regime changes in time series. In this work we are interested in time series which in addition to trend display seasonality in mean, in autocorrelation and in variance. These type of series appears in many areas, including hydrology, meteorology, economics and finance. The seasonality is accounted for by subset PAR modelling, for which each season follows a possibly different Autoregressive model. Levels, trend, autoregressive parameters and residual variances are allowed to change their values at fixed unknown times. The identification of number and location of structural changes, as well as PAR lags indicators, is based on Genetic Algorithms, which are suitable because of high dimensionality of the discrete search space. An application to Italian industrial production index time series is also proposed.

Keywords Time series · Seasonality · Nonstationarity

1 Model Description and Estimation

In this work we are interested in economic time series showing a trend and seasonal fluctuations that are not very stable over time. This may be caused by economic agents who have preferences, technologies, constraints, and expectations which are not constant over the seasons [1]. In these kind of series it may happen that linear seasonal adjustment filters, as in $SARIMA$ models, are not likely to remove the intrinsic seasonality. In that case the residuals from these models show patterns that can be explained by the presence of dynamic periodicity [1]. Recently, economic

F. Battaglia · M. Rizzo (✉)
Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy
e-mail: francesco.battaglia@uniroma1.it; manuel.rizzo@uniroma1.it

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

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_19

models in which seasonally varying parameters are allowed attracted some attention [2]. The present paper considers PAR models with linear piecewise trends, allowing nonstationarity and seasonality simultaneously. There is empirical evidence of the existence of discontinuities due to structural changes possibly originated by policy changes [3]. Consequently, detecting the existence of structural changes and estimating their number and locations in periodic time series is an important stage. In this paper we propose an automatic procedure based on Genetic Algorithms (GAs) for estimating number and locations of change points.

We refer to a seasonal time series of period s, observed s times a year for N complete years ($s = 12$ for monthly data). This series is possibly divided in a number of regimes up to M, specified by $m = M - 1$ change points $\tau_1, \ldots, \tau_m$ occurring at the end of the year $\tau_j - 1$, and set $\tau_0 = 1$ and $\tau_M = N + 1$. In order to ensure reasonable estimates, it is required that each regime contains at least a minimum number mrl of years, therefore $\tau_j \geq \tau_{j-1} + mrl$ for any regime j. We let $R^j = \{\tau_{j-1}, \tau_{j-1} + 1, \ldots, \tau_j - 1\}$, $j = 1, \ldots, M$, so that if year n belongs to set R^j then observation at time $(n-1)s + k$ is in regime j, where $k = 1, \ldots, s$. For the observation in season k of year n the model is:

$$X_{(n-1)s+k} = a^j + b^j[(n-1)s + k] + \mu_k^j + Y_{(n-1)s+k}, \; n \in R^j, \; 1 \leq k \leq s, 1 \leq j \leq M \tag{1}$$

where process $\{Y_{(n-1)s+k}\}$ is a PAR given by:

$$Y_{(n-1)s+k} = \sum_{i=1}^{p} \phi_i^j(k) Y_{(n-1)s+k-i} + \epsilon_{(n-1)s+k}, \quad n \in R^j, \; 1 \leq k \leq s, 1 \leq j \leq M. \tag{2}$$

Model (1) is a pure structural change model where the trend parameters a^j and b^j depend only on the regime, whereas means μ_k^j are allowed to change also with seasons. The parameters $\phi_i^j(k)$, $i = 1, \ldots, p$, represent the PAR coefficients during season k of the j-th regime, and some of them may be allowed to be constrained to zero, in order to get more parsimonious subset models. The innovations process $\{\epsilon_t\}$ in Eq. (2) corresponds to a periodic white noise, with $E(\epsilon_{(n-1)s+k}) = 0$ and $Var(\epsilon_{(n-1)s+k}) = \sigma_j^2(k) > 0, n \in R^j$.

Our model is characterized by both structural parameters: the changepoints number m, their locations $\tau_1, \tau_2, \ldots, \tau_m$ and PAR lags indicator, which specifies presence or absence of PAR parameters; and by regression parameters: the trend intercepts a^j, the slopes b^j, the seasonal means μ_k^j, the AR parameters $\phi_i^j(k)$ and the innovation variances $\sigma_j^2(k)$. These regression parameters, assuming that structural parameters are given, can be estimated by Ordinary Least Squares.

Identification of this complex model requires only selection of structural parameters. We adopt a procedure based on a GA [4] for optimization. The GA is a nature-inspired metaheuristic algorithm, for which a population of solutions evolves through iterations by use of so-called genetic operators, until a stopping rule is reached. Out implementation optimizes structural parameters simultaneously by considering an objective function based on NAIC criterion, introduced in [5] for threshold models selection, given by:

$$g = [\sum_{j=1}^{M}\sum_{k=1}^{s} n_{j,k} \log(\hat{\sigma}_j^2(k)) + 2\sum_{j=1}^{M}\sum_{k=1}^{s} P_{j,k}]/(Ns),$$

where $\hat{\sigma}_j^2(k)$ is the model residual variance of series in regime j and season k, $n_{j,k}$ is related sample size, $P_{j,k}$ is related number of parameters. Other options of penalization could be adopted basing on identification criteria literature. The GA objective function to be maximized, named *fitness*, is a scaled exponential transformation of g given by $f = \exp\{-g/\beta\}$, where β is a positive scaling constant.

2 Application and Conclusions

The industrial production index is an important macroeconomic variable since it can reflect business cycle behaviour and changing directions of an underlying trend. Forecasts of this variable are used by many decision makers [1]. Given that decisions sometimes concern time intervals shorter than 1 year, forecasts for the monthly observed industrial production index can be useful. The time series of industrial production index in Italy (Ateco category C) for months 1990.1–2016.12 is displayed in Fig. 1. From this figure it can be observed that the time series shows trend and a non-stable seasonal patterns.

The GA procedure, when applied to the original data, segmented the series in two regimes with one structural change at $\tau_1 = 2009$ and an opposite trend behaviour (see Fig. 1). To examine the effectiveness of subset piecewise PAR models, we compared their performance, in terms of fitness, with that of some other models used for seasonal data. Four types of models were considered in this study and

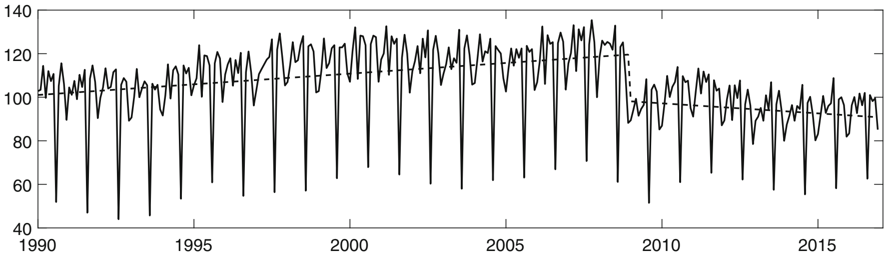

Fig. 1 Italy Industrial Production Index series with estimated trend and changepoint at end 2008

Table 1 Performance of different models

Model	Fitness	Prediction MSE
Subset piecewise $PAR(3)$	0.799	20.22
Complete piecewise $PAR(3)$	0.789	26.64
Piecewise $AR(3)$	0.769	20.37
$SARIMA(2, 0, 1) \times (0, 1, 0)_{12}$	0.738	27.11

the corresponding fitness values are reported in the first column of Table 1. The first two models are a complete and subset piecewise PAR, as described above. The third model is a non-periodic $AR(3)$ applied on data after removing the break, the trend and the seasonal means. The fourth model is a more conventional $SARIMA(2, 0, 1) \times (0, 1, 0)_{12}$ but estimated separately for each regime. The best model, in terms of fitness, is the subset piecewise PAR, while the complete model has a slightly smaller fitness. The non-periodic piecewise AR reaches an even smaller fitness, while the $SARIMA$ model is markedly less fit.

As far as diagnostic checking is concerned, the Box-Pierce statistics with 12 or 18 lags computed on residuals were not significant at level 0.01 for the piecewise $PAR(3)$ models, and largely significant for the other two models.

We computed also out-of-sample forecasts for the first 9 months of 2017. The mean square prediction errors for the four models are reported in the second column of Table 1. We may conclude that for the industrial production index dataset the proposed procedure outperforms the other models both in terms of fitting and forecasting.

References

1. Franses, P.H.: Periodically integrated subset autoregressions for dutch industrial production and money stock. J. Forecast. **12**, 601–613 (1993)
2. Franses, P.H., Paap, R.: Periodic Time Series Models. Oxford University Press, Oxford (2004)
3. Giordani, P., Kohn, R., van Dijk, D.: A unified approach to nonlinearity, structural change and outliers. J. Economet. **137**, 112–133 (2007)
4. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, New York (1989)
5. Tong, H.: Non-linear Time Series. A Dynamical System Approach. Oxford University Press, Oxford (1990)