Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Estimation and identification of periodic autoregressive models with one exogenous variable

Eugen Ursu [a,b,*], Jean-Christophe Pereau [a]

[a] GREThA UMR-CNRS 5113, Université de Bordeaux, France

[b] ECREB, West University of Timisoara, Romania

ARTICLE INFO

Article history:
Received 13 February 2017
Accepted 11 July 2017
Available online xxxx

MSC 2010:
primary 62M10
secondary 90C59

Keywords:
Periodic time series
Autoregressive models
Genetic algorithms
Identification
Fisheries
Sea surface temperature
Catch-per-unit of effort

ABSTRACT

This paper analyzes the identification and estimation procedures for periodic autoregressive models with one exogenous variable (PARX). The identification of the optimal PARX model is based on the use of a genetic algorithm combined with the Bayes information criterion. The estimation of the parameters relies on the least squares method and their asymptotic properties are studied. Two simulation experiments are performed and indicate the success of the suggested method. A PARX model is used to study the relationship between the catch-per-unit-effort and the sea surface temperature as exogenous variable for the shrimp French Guiana fishery from January 1989 to December 2012.

1. Introduction

The use of periodic models appears to be well-suited to deal with many real life phenomena characterized by a seasonal behavior (Dudek, Hurd, & Wojtowicz, 2014; Lund, Shao, & Basawa, 2006; Tesfaye, Meerschaert, & Anderson, 2006). These models are increasingly used in the climatology or hydrology literature (Hipel & McLeod, 1994; Jones & Brelsford, 1967; Li & Lund, 2012; Lu, Lund, & Lee, 2010; Lund et al., 2007; Ursu & Pereau, 2016; Vecchia, 1985) but also in other disciplines like macroeconomics (Franses & Paap, 2004), engineering (Schlick, Duckwitz, & Schneider, 2013) and marine fisheries (Stoffer, 1986). In many applications of time series analysis, the variable of interest may be affected by other variables, called exogenous or unmodeled variables, which are determined outside the system of interest. In the case study of the paper, climate change and global warming in particular through its effect on the sea surface temperature appears as an exogenous major driver on inter-annual fluctuations in fish abundance (Brander, 2007; Cheung et al., 2009).

Autoregressive models with exogenous variables (ARX) have been extensively used in the econometric literature. The dependent variable is assumed to depend on its past values and the present and lagged values of exogenous variables. These models are also called conditional or partial models (Lütkepohl, 2005), distributed lag models (Reinsel, 1997) or transfer function models (Wei, 2006). The introduction of ARX in a var model (VARX) with systematically varying coefficients in a state-space model form has been studied by Lütkepohl (2005). Paroli and Spezia (2008) introduced a periodic component in an ARX model. Extending ARX models to periodic autoregressive (PAR) models with exogenous variables (PARX) in a Bayesian

* Corresponding author.
E-mail address: eugen.ursu@u-bordeaux.fr (E. Ursu).

http://dx.doi.org/10.1016/j.jkss.2017.07.001

framework has been proposed by Andel (1987, 1989). Maçaira, Oliveira, Ferreira, de Almeida, and Souza (2017) proposed a PAR with exogenous variables to generate scenarios for hydrological inflows for some Brazilian reservoirs and showed that the inclusion of exogenous variables decreases the error measure by 3%. This notation PARX should not be confused with other acronyms like Poisson AutoRegression with eXogenous covariates (Agosto, Cavaliere, Kristensen, & Rahbek, 2016; Angelini & Angelis, 2016) or Polynomial Autoregressive Regression with eXogenous variables (Wojciechowski, 2001).

The main challenge with PARX models relies on the large number of parameters to estimate. Sometimes there exist several sets of model parameters that give reasonable results. A way to reduce the parameter space in PAR models consists in introducing restrictions on the parameters (Ursu & Duchesne, 2009; Ursu & Turkman, 2012). A similar problem occurs in time series analysis with a large number of models which needs to be compared and estimated. Genetic Algorithms (GA) appear to be an useful tool to investigate the space of solutions and to select the combination of parameters that corresponds to the best model (Baragona & Battaglia, 2009). Baragona, Battaglia, and Cucina (2004), Wu and Chang (2002) developed a method which exploits GA with Akaike's Information Criterion (AIC). GA coupled with Minimum Description Length (MDL) have recently been used in Song and Bondon (2013) and Yau, Tang, and Lee (2015). GA in Bayesian context was also used by Jeong and Kim (2013) to locate change points for an autoregressive model. To deal with PARX models, this paper uses an automatic procedure based on GA combined with Bayesian Information Criterion (BIC).

Our contributions to the PARX literature are twofold. A first contribution is to provide theoretical results on the least squares estimators and derives their asymptotic properties in PARX models. Contrary to Maçaira et al. (2017), the estimation and identification of the model are analyzed taking into account constraints on the parameters. Our paper also considers a global optimization strategy relying on genetic algorithms and not a sequential procedure which consists in estimating first a PAR model and then adding exogenous variables. A second contribution refers to the applied fisheries literature concerning the stock assessment of fish species populations which is often performed on the basis of the well-known "Virtual Population Analysis" method (Lassen & Medley, 2001; Sparre & Venema, 1998). This age-structured method aims at providing to the fishery manager information on the recruit abundance, the spawning stock biomass as well as the fishing mortality. However this method appears to be stringent in terms of data collection (Haddon, 2011). PARX models can be more easier to implement and provide short-term forecasts to managers concerning the catch-per-unit-effort for fish species population.

The rest of the article is organized as follows. In Section 2, the PARX model is introduced and a brief review of the estimation techniques is presented. Section 3 develops a genetic algorithm to conduct the identification process and Section 4 reports some Monte Carlo simulation results. In Section 5, an application to the shrimp French Guiana fishery is presented. Section 6 concludes.

2. Periodic models with one exogenous variable

The dependent variable is denoted by Y_t, the exogenous variable by X_t and the error process by ϵ_t. Y_t and X_t are assumed to be periodic stationary processes according the following definition.

A stochastic process W_t is periodic stationary if

$$E(W_{n+s}) = E(W_n) \quad \text{and} \quad cov(W_{n+s}, W_{m+s}) = cov(W_n, W_m),$$

for all integers n and m, where s stands for the period. Periodic series are also called periodically correlated (Gladyshev, 1961) or cyclostationary (Lund & Basawa, 2000).

Based on Andel (1989), we consider the following periodic autoregressive model with one exogenous variable:

$$Y_{ns+\nu} = \sum_{k=1}^{p(\nu)} \phi_k(\nu) Y_{ns+\nu-k} + \sum_{j=0}^{m(\nu)} \theta_j(\nu) X_{ns+\nu-j} + \epsilon_{ns+\nu}, \tag{1}$$

for $n = 0, 1, \ldots, N-1$ and $\nu = 1, 2, \ldots, s$. The autoregressive model order at season ν for Y_t is given by $p(\nu)$, while the terms $\phi_k(\nu)$, $k = 1, \ldots, p(\nu)$, represent the autoregressive model coefficients during season ν. Concerning the exogenous variable X_t, the autoregressive order at season ν is given by $m(\nu)$ while the terms $\theta_j(\nu)$, $j = 0, \ldots, m(\nu)$ are the autoregressive model coefficients during season ν. The mean of Y_t is equal to zero in each of the s seasons, that is $E(Y_{ns+\nu}) = 0$, $\nu = 1, 2, \ldots, s$. Without loss of generality, the mean of process X_t is assumed to be zero. The error process $\epsilon = \{\epsilon_t, t \in \mathbb{Z}\}$ corresponds to a periodic white noise, with $E(\epsilon_t) = 0$ and $\text{var}(\epsilon_{ns+\nu}) = \sigma^2(\nu) > 0$, $\nu = 1, \ldots, s$. In the following we assumed that X_t and ϵ_t are independent processes, although most results can be obtained under less restrictive conditions. Note that if $s = 1$, then Eq. (1) reduces to a classical autoregressive model with exogenous variables (ARX). As in Andel (1989), the model can be extended to several exogenous variables.

2.1. Unconstrained least squares estimators

To estimate the parameters of the model, we consider the time series data $Y_{ns+\nu}$, $n = 0, 1, \ldots, N-1$, $\nu = 1, \ldots, s$ with sample size Ns. Let $\mathbf{z}(\nu) = \left(Y_\nu, Y_{s+\nu}, \ldots, Y_{(N-1)s+\nu}\right)^\top$ and $\mathbf{e}(\nu) = \left(\epsilon_\nu, \epsilon_{s+\nu}, \ldots, \epsilon_{(N-1)s+\nu}\right)^\top$ be $(N \times 1)$ vectors with $\top$ the