Journal of Statistical Planning and Inference 155 (2014) 93–106

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Robust modelling of periodic vector autoregressive time series

Eugen Ursu *, Jean-Christophe Pereau

GREThA UMR-CNRS 5113, Université de Bordeaux, Avenue Léon Duguit, 33608 Pessac CEDEX, France

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form 15 May 2014
Accepted 2 July 2014
Available online 15 July 2014

Keywords:
Periodic time series
Parameter constraints
Robust estimation
Identification
Genetic algorithms

ABSTRACT

This paper develops a robust estimation and identification method for periodic vector autoregressive models (hereafter PVAR) with linear constraints set on parameters for a given season. Since the least squares estimators are extremely sensitive to additive outliers, this paper suggests a robust estimation based on residual autocovariances (RA) and analyses the asymptotic properties of these RA estimates. To identify the optimal order of the PVAR, this paper also uses a genetic algorithm with Bayes information criterion (BIC). The proposed procedure is applied to a small simulation study for PVAR models in the case of four seasons. Empirical results show that the robust estimators perform better than the least squares estimators when the contamination rate of the additive outliers is at random or at fixed positions.

1. Introduction

Time series with periodically varying parameters are frequently used in hydrology (Vecchia, 1985a,b), climatology (Jones and Brelsford, 1967) and many other disciplines. Holan et al. (2010) provides a recent review of periodic models.

There are two convenient and efficient estimation techniques available for PVAR models, namely the least squares (LS) method in the univariate case (Franses and Paap, 2004) or the multivariate case (Lütkepohl, 2005) and the method of moments based on Yule–Walker equations (McLeod, 1994). It is worth noting that the identification, the estimation and the diagnostic tests can be modelled independently for each season (Pagano, 1978; Ursu and Duchesne, 2009).

A general problem in PVAR modelling lies with the large number of parameters to estimate. To obtain parsimonious models, it is of interest to study situations in which there are linear constraints on the parameters of a given season (Ursu and Duchesne, 2009). As the LS estimation method is sensitive to outliers, the main goal of this paper is to propose a robust methodology for PVAR models. Several authors have studied the effects of the outliers on the parameter estimates of time series models. Denby and Martin (1979) discussed the robust estimation of the parameters for an autoregressive model of first order. Ben et al. (1999) proposed robust estimators in vector autoregressive moving average time series models (VARMA), which are an extension of the residual autocovariances (RA) estimators of Bustos and Yohai (1986). Recent work on robustness in periodic time series include the estimator of periodic autoregressive (PAR) models proposed by Sarnaglia et al. (2010), which is a generalization of the robust scale estimator developed by Ma and Genton (2000). Furthermore, they studied the effects of additive outliers on the correlation structure of a PAR process. Another robust estimation for PAR models was discussed in Shao (2007). However, most of the effort on robustness in periodic time series concentrates on the univariate case.

From a model-building point of view, it is well recognized that model identification is typically the most difficult aspect. McLeod (1994) used an automatic selection criterion, such as the BIC (Schwarz, 1978) to identify PAR models. A

* Corresponding author. Tel.: +33 556842587.
E-mail addresses: eugen.ursu@u-bordeaux.fr (E. Ursu), jean-christophe.pereau@u-bordeaux.fr (J.-C. Pereau).

http://dx.doi.org/10.1016/j.jspi.2014.07.005
0378-3758/

possible difficulty with this procedure is the large number of models to be investigated. Ursu and Turkman (2012) proposed an automatic procedure to model selection by using the genetic algorithms (GA). For a recent review of the usage of GA for time series analysis, see Baragona et al. (2011). Therefore, we propose an identification methodology of PVAR models, based on BIC and genetic algorithms. In the case of multivariate models, the advantage of using genetic algorithms over traditional methods may be emphasized, due to the increasing number of involved parameters.

Our results improve the literature in two directions. Firstly, they allow a generalization of the robust estimation of VAR processes (Li and Hui, 1989). Secondly, we propose an automatic method for the identification of relevant variables of PVAR models based on BIC, by means of genetic algorithms.

This article is organized as follows. In Section 2, the PVAR model is introduced and least squares estimators are given. In Section 3, we extend the methodology of robust estimation used by Ben et al. (1999) for vector autoregressive moving average (VARMA) models to PVAR models. Section 4 deals with model selection tools. Some simulation results are reported in Section 5. Finally, Section 6 offers some concluding remarks.

2. Periodic vector autoregressive time series models

This section presents principal results on the least squares estimators in the unconstrained and constrained case of PVAR models. For a detailed discussion, see Ursu and Duchesne (2009).

Let $\mathbf{Y} = \{\mathbf{Y}_t, t \in \mathbb{Z}\}$ be a stochastic process, where

$$\mathbf{Y}_t = (Y_t(1), \ldots, Y_t(d))^\top$$

represents a random vector of dimension d. The process $\mathbf{Y}$ is a PVAR process of order $p(\nu)$, $\nu \in \{1, \ldots, s\}$ (s is a predetermined value), if there exist $d \times d$ matrices $\mathbf{\Phi}_k(\nu) = \left(\Phi_{k,ij}(\nu)\right)_{i,j=1,\ldots,d}$, $k = 1, \ldots, p(\nu)$ such that

$$\mathbf{Y}_{ns+\nu} = \sum_{k=1}^{p(\nu)} \mathbf{\Phi}_k(\nu)\mathbf{Y}_{ns+\nu-k} + \boldsymbol{\epsilon}_{ns+\nu}. \tag{1}$$

In Eq. (1), the random vector $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}_t, t \in \mathbb{Z}\}$ stands for a zero mean periodic white noise, that is $\boldsymbol{\epsilon}$ is composed of independent $d \times 1$ random vectors $\boldsymbol{\epsilon}_t = (\epsilon_t(1), \ldots, \epsilon_t(d))^\top$, such that $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$ and $E(\boldsymbol{\epsilon}_{ns+\nu}\boldsymbol{\epsilon}_{ns+\nu}^\top) = \mathbf{\Sigma}_{\boldsymbol{\epsilon}}(\nu)$, where the error covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}(\nu) = \left(\sigma_{\boldsymbol{\epsilon},ij}(\nu)\right)_{i,j=1,\ldots,d}$ is assumed to be non singular, $\nu = 1, \ldots, s$. The PVAR process (1) is supposed to have a zero mean, that is $E(\mathbf{Y}_t) = \mathbf{0}$. In practical applications, seasonal means are first removed from the series, meaning that a model is formulated by examining $\mathbf{Y}_{ns+\nu} - \boldsymbol{\mu}_\nu$, say, where in general the mathematical expectation $E(\mathbf{Y}_{ns+\nu}) = \boldsymbol{\mu}_\nu$ may be a function of season ν. Unless otherwise stated we assume that PVAR models are stationary in the periodic sense. Periodic stationarity is discussed in Gladyshev (1961). Typically, the periodic models used in water resources and environmental systems are stationary, in the sense that they do not need to be differenced to achieve stationarity (or, otherwise put, data do not have unit roots). However, treating unit roots in periodic models exceeds the scope of this paper. For an extensive discussion of this topic see Franses and Paap (2004) (and various other studies by the same authors).

Let $\boldsymbol{\beta}(\nu) = (\text{vec}^\top\{\mathbf{\Phi}_1(\nu)\}, \ldots, \text{vec}^\top\{\mathbf{\Phi}_{p(\nu)}(\nu)\})^\top$ be a $\{d^2p(\nu)\} \times 1$ vector of parameters, where vec($\mathbf{A}$) corresponds to the vector obtained by stacking the columns of $\mathbf{A}$. The PVAR model in (1) has $\sum_{\nu=1}^{s} p(\nu)$ autoregressive parameters $\mathbf{\Phi}_k(\nu)$, $k = 1, \ldots, p(\nu)$, $\nu = 1, \ldots, s$, and s additional $d \times d$ covariance matrices $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}(\nu)$, $\nu = 1, \ldots, s$. The number of parameters can be quite large; for example, in the case of bivariate monthly data where $d = 2$, $s = 12$, and, in the simplest case $p(\nu) \equiv 1$, this means that 48 independent autoregressive parameters must be estimated. In view of these considerations, we consider estimation in the unrestricted case but also in the situation where the parameters of the same season ν satisfy the following relation:

$$\boldsymbol{\beta}(\nu) = \mathbf{R}(\nu)\boldsymbol{\xi}(\nu) + \mathbf{b}(\nu), \tag{2}$$

where $\mathbf{R}(\nu)$ is a known $\{d^2p(\nu)\} \times K(\nu)$ matrix of rank $K(\nu)$, $\mathbf{b}(\nu)$ is a known $\{d^2p(\nu)\} \times 1$ vector and $\boldsymbol{\xi}(\nu)$ represents a $K(\nu) \times 1$ vector of unknown parameters. Letting $\mathbf{R}(\nu) = \mathbf{I}_{d^2p(\nu)}$, where $\mathbf{I}_d$ denotes the $d \times d$ identity matrix, $\mathbf{b}(\nu) = \mathbf{0}$, $\nu = 1, \ldots, s$ gives what we call the full unconstrained case. This linear constraint includes the special case of parameters set to zero on certain components of $\mathbf{\Phi}_k(\nu)$, $\nu = 1, \ldots, s$.

Consider the time series data $\mathbf{Y}_{ns+\nu}$, $n = 0, 1, \ldots, N-1$, $\nu = 1, \ldots, s$, giving a sample size equal to $n = Ns$. For a given value of $\boldsymbol{\beta}(\nu)$ we define the residuals:

$$\hat{\boldsymbol{\epsilon}}_{ns+\nu} = \begin{cases} \mathbf{Y}_{ns+\nu} - \displaystyle\sum_{k=1}^{p(\nu)} \mathbf{\Phi}_k(\nu)\mathbf{Y}_{ns+\nu-k}, & ns+\nu > p(\nu), \\ \mathbf{0}, & ns+\nu \le p(\nu), \end{cases}$$

which are well-defined for $n = 0, 1, \ldots, N-1$. Using the relation

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A})\text{vec}(\mathbf{B}),$$