Computational Statistics and Data Analysis 106 (2017) 103–126

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Sparse seasonal and periodic vector autoregressive modeling[☆]

Changryong Baek[a,*], Richard A. Davis[b], Vladas Pipiras[c]

[a] Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul, 110-745, Republic of Korea
[b] Department of Statistics, Columbia University, 1255 Amsterdam Avenue, MC 4690, New York, NY 10027, USA
[c] Department of Statistics and Operations Research, UNC at Chapel Hill, CB#3260, Hanes Hall, Chapel Hill, NC 27599, USA

ARTICLE INFO

Article history:
Received 16 October 2015
Received in revised form 3 September 2016
Accepted 13 September 2016
Available online 16 September 2016

Keywords:
Seasonal vector autoregressive (SVAR) model
Periodic vector autoregressive (PVAR) model
Sparsity
Partial spectral coherence (PSC)
Adaptive lasso
Variable selection

ABSTRACT

Seasonal and periodic vector autoregressions are two common approaches to modeling vector time series exhibiting cyclical variations. The total number of parameters in these models increases rapidly with the dimension and order of the model, making it difficult to interpret the model and questioning the stability of the parameter estimates. To address these and other issues, two methodologies for sparse modeling are presented in this work: first, based on regularization involving adaptive lasso and, second, extending the approach of Davis et al. (2015) for vector autoregressions based on partial spectral coherences. The methods are shown to work well on simulated data, and to perform well on several examples of real vector time series exhibiting cyclical variations.

1. Introduction

In this work, we introduce methodologies for sparse modeling of stationary vector (q-dimensional) time series data exhibiting cyclical variations. Sparse models are gaining traction in the time series literature for similar reasons sparse (generalized) linear models are used in the traditional setting of i.i.d. errors. Such models are particularly suitable in a high-dimensional context, for which the number of parameters often grows as q^2 (as for example with vector autoregressive models considered below) and becomes prohibitively large compared to the sample size even for moderate q. Sparse models also ensure better interpretability of the fitted models and numerical stability of the estimates, and tend to improve prediction. See, for example, Lange (2010, Section 16.5, p. 312) for a discussion related to numerical stability, and Hastie et al. (2013, Section 3.4), related to predictions (in the general context of shrinkage estimators).

In the vector time series context, sparse modeling has been considered for the class of vector autoregressive (VAR) models:

$$X_n - \mu = A_1(X_{n-1} - \mu) + \cdots + A_p(X_{n-p} - \mu) + \epsilon_n, \quad n \in \mathbb{Z}, \tag{1.1}$$

where $X_n = (X_{1,n}, \ldots, X_{q,n})'$ is a q-vector time series, $A_1, \ldots, A_p$ are $q \times q$ matrices, μ is the overall constant mean vector and ϵ_n are white noise (WN) error terms. Regularization approaches based on lasso and its variants were taken in Hsu et al. (2008),

☆ The work of the first author was supported in part by the Basic Science Research Program from the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1006025). The second author was supported in part by ARO MURI grant W911NF-12-1-0385. The third author was supported in part by NSA grant H98230-13-1-0220.
* Corresponding author.
E-mail addresses: crbaek@skku.edu (C. Baek), rdavis@stat.columbia.edu (R.A. Davis), pipiras@email.unc.edu (V. Pipiras).

http://dx.doi.org/10.1016/j.csda.2016.09.005
0167-9473/

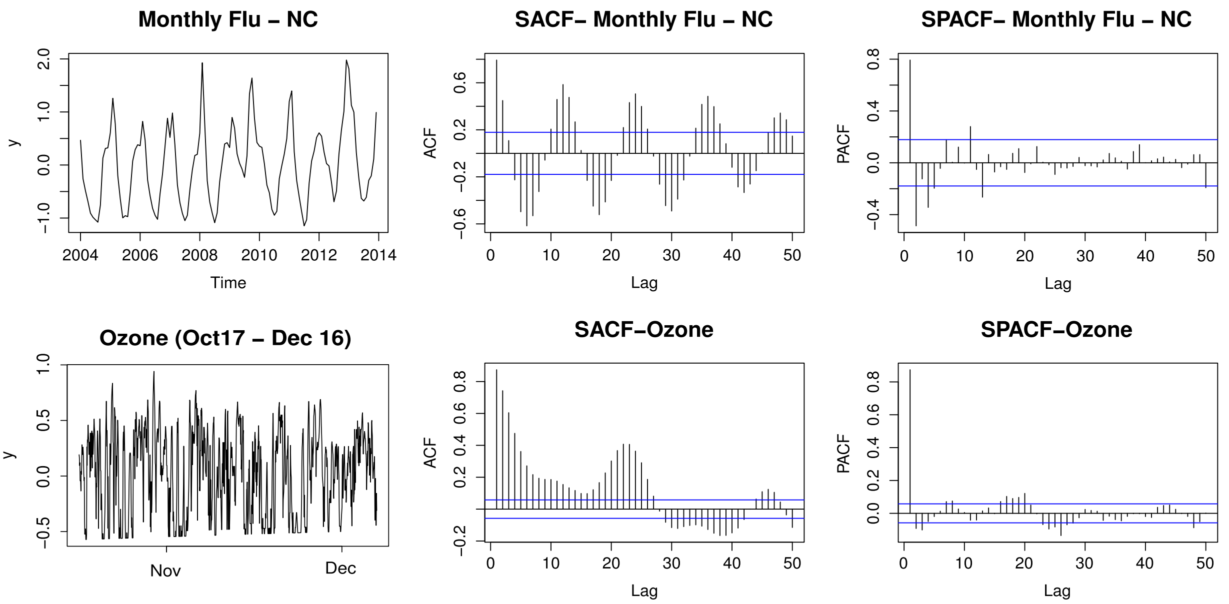

Fig. 1. Top: Monthly flu trend in NC. Bottom: 23 h ozone levels at a CA location. Respective sample ACFs and PACFs are given.

Shojaie and Michailidis (2010), Song and Bickel (2011), Medeiros and Mendes (2012), Basu and Michailidis (2015), Nicholson et al. (2015) and Kock and Callot (2015), with applications to economics, neuroscience (e.g. functional connectivity among brain regions), biology (e.g. reconstructing gene regulatory network from time course data), and environmental science (e.g. pollutants levels over time). As usual, the model (1.1) will be abbreviated as

$$\Phi(B)(X_n - \mu) = \epsilon_n, \quad n \in \mathbb{Z}, \tag{1.2}$$

where $\Phi(B) = 1 - A_1 B - \cdots - A_p B^p$ and B is the backshift operator.

In a different approach, Davis et al. (2015) introduced an alternative 2-stage procedure for sparse VAR modeling. In the first stage, all pairs of component series are ranked based on the estimated values of their partial spectral coherences (PSCs), defined as

$$\sup_{\lambda} |\mathrm{PSC}^X_{jk}(\lambda)|^2 := \sup_{\lambda} \frac{|g^X_{jk}(\lambda)|^2}{g^X_{jj}(\lambda) g^X_{kk}(\lambda)}, \quad j, k = 1, \ldots, q,\ j \neq k, \tag{1.3}$$

where $g^X(\lambda) = f^X(\lambda)^{-1}$ with f^X being the spectral density matrix of X. Then, the order $\widetilde{p}$ and the top $\widetilde{M}$ pairs are found which minimize the BIC(p, M) value, and the coefficients of matrices A_r are set to 0 for all pairs of indices j, k not included in $\widetilde{M}$. In the second stage, the estimates of the remaining non-zero coefficients are ranked according to their t-statistic values. Again, the top m^* of the coefficients are selected that minimize a suitable BIC, and then the rest of the coefficients are set to 0. As shown in Davis et al. (2015), this 2-stage procedure outperforms regular lasso. The basic idea of this approach is that small PSCs do not increase the likelihood sufficiently to warrant the inclusion of the respective coefficients of matrices A_r in the model. Partial spectral coherences have been used extensively in the time series literature, especially in connection to graphical modeling (see e.g. Dahlhaus, 2000; Bach and Jordan, 2004; Fried and Didelez, 2005; Eichler, 2012).

We shall extend here the regularization approach based on lasso and the approach of Davis et al. (2015) based on PSCs to sparse modeling of vector time series data exhibiting cyclical variations. The motivation here is straightforward. Consider, for example, the benchmark flu trends and pollutants series studied through sparse VAR models by Davis et al. (2015), and others. Fig. 1 depicts the plots of (the logs of) their two component series with the respective sample ACFs and PACFs. The cyclical nature of the series can clearly be seen from the figure. The same holds for other component series (not illustrated here).

Cyclical features of component series are commonly built into a larger vector model by using one of the following two approaches. A *seasonal* VAR model (SVAR(P, p) model, for short; not to be confused with the so-called structural VAR) is one possibility, defined as

$$\Phi(B)\Phi_s(B^s)(X_n - \mu) = \epsilon_n, \quad n \in \mathbb{Z}, \tag{1.4}$$

where $\Phi(B)$ and ϵ_n are as in (1.2), $\Phi_s(B^s) = 1 - A_{s,1}B^s - \cdots - A_{s,P}B^{Ps}$ with $q \times q$ matrices $A_{s,1}, \ldots, A_{s,P}$, μ denotes the overall mean and s denotes the period. This is the vector version of the multiplicative seasonal AR model proposed by Box and Jenkins (1976). Note that the number of parameters of the SVAR(P, p) model (including the covariance matrix of innovations terms)