Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

On constrained estimation of graphical time series models

T.P. Yuen, H. Wong, K.F.C. Yiu *

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

A R T I C L E I N F O

Article history:
Received 22 December 2016
Received in revised form 23 January 2018
Accepted 30 January 2018
Available online xxxx

Keywords:
Graphical models
Time series
Estimation
Optimization
Air pollution

A B S T R A C T

Graphical time series models encode the conditional independence among the variables of a multivariate time series. An iterative method is proposed to estimate a graphical time series model based on a sparse vector autoregressive process. The method estimates both the autoregressive coefficients and the inverse of noise covariance matrix under sparsity constraints on both the coefficients and the inverse covariance matrix. This iterative method estimates a sparse vector autoregressive model by considering maximum likelihood estimation with the sparsity constraints as a biconcave problem, where the optimization problem becomes concave when either the autoregressive coefficients or the inverse noise covariance matrix is fixed. The method also imposes fewer restrictions in the estimation comparing to the use of a structural vector autoregressive model to study the dynamic interdependencies between time series variables.

1. Introduction

Graphical models represent the conditional independencies among random variables in multivariate data. These independence relationships can be visualized by an undirected graph where vertices represent the variables and edges between vertices illustrate that the corresponding variables of the connected vertices are conditionally dependent. Since the introduction of log-linear models on discrete data, researchers have attempted to link up graphical models with log-linear models (Darroch et al., 1980). By analogy with the log-linear models for contingency tables, models based on the multivariate normal distribution have been introduced. Edwards (1995) and Lauritzen (1996) give good introduction to graphical modelling.

Consider a K-dimensional random variable $X \sim N(0, \Sigma)$; a Gaussian graphical model can be established by calculating the precision matrix, $\Theta = \Sigma^{-1}$. With the precision matrix, the conditional independence between variables is determined. For example, two components of X are independent conditioning on the remaining components if and only if the corresponding entry in the precision matrix is zero, i.e., X_i and X_j are conditionally independent if and only if $\Theta_{ij} = 0$. With prior information on the conditional independence between variables, the estimation of a Gaussian graphical model can be formulated as the covariance selection problem (Dempster, 1972), as formulated in (1).

$$\begin{aligned} &\text{Maximize} \quad \log\det(\Theta) - \operatorname{tr}(\mathbf{S}\Theta) \\ &\text{subject to} \quad \Theta_{ij} = 0, \quad (i,j) \in \Omega, \end{aligned} \tag{1}$$

where $\mathbf{S}$ is the sample covariance matrix, Ω is a set consisting of pairs of known conditionally independent nodes.

The increasing interest in data science has heightened the need for the development of Gaussian graphical models with sparse coefficients on high dimension data (see Banerjee et al. (2008); Dahl et al. (2008); Friedman et al. (2008)). To achieve

* Corresponding author.
E-mail address: cedric.yiu@polyu.edu.hk (K.F.C. Yiu).

https://doi.org/10.1016/j.csda.2018.01.019
0167-9473/

sparsity, researchers have considered the penalized likelihood methods shown below.

$$\begin{aligned} &\text{Maximize} \quad \log\det(\Theta) - \operatorname{tr}(\mathbf{S}\Theta) \\ &\text{subject to} \quad \rho(\Theta) \le k, \end{aligned} \tag{2}$$

where $\mathbf{S}$ is the sample covariance matrix, $\rho(\cdot)$ is a regularization term, and k is a tuning parameter.

Brillinger (1996) and Dahlhaus (2000) extended the use of graphical models to multivariate time series to explore the interrelationship between variables of a multivariate time series process. A recent summary of graphical time series models can be found in Tunnicliffe Wilson et al. (2015). The partial correlation structure of the components of the process given the remaining components can be identified by the partial spectral coherence or the inverse of the spectral density matrix. These frequency domain statistics measure the linear association of two components of a process given the linear effects of the remaining components. Similar to Gaussian graphical models, two components of a multiple time series is conditionally uncorrelated given the other components if and only if the corresponding partial spectral coherence is zero at all frequencies (Brillinger, 1981; Dahlhaus, 2000). The interrelationships are visualized by an undirected partial correlation graph, where each vertex represents a component of the process, and the edges are characterized by the partial spectral coherences. In particular, the partial correlation graph exploits the conditional dependence structure of the components if the time series is Gaussian.

With the partial correlation graph, the complexity of fitting a time series model with large dimension can be reduced by imposing sparsity constraints on the VAR model based on the partial correlation graph. Songsiri et al. (2009) discussed the VAR model estimation problem, subject to conditional independence constraints based on the inverse of the spectral density matrix, using convex optimization methods. Davis et al. (2016) proposed a two-stage approach for fitting sparse VAR models in which non-zero autoregressive coefficients are selected according to the partial spectral coherence together with the Bayesian information criterion (BIC) (Schwarz, 1978). The model is then refined in the second stage to reduce the number of parameters further by using the t-ratios of the estimated autoregressive coefficients. These two articles (Davis et al., 2016; Songsiri et al., 2009) also investigated the penalized likelihood methods in the maximum likelihood estimation by imposing regularization term, like L_1 regularization, to achieve sparsity. Other related research (Hsu et al., 2008; Ren et al., 2013; Songsiri, 2013; Jung et al., 2015) also discussed penalized regression methods for VAR modelling. The penalized regression approaches for VAR modelling ignore the contemporaneous dependence in the time series (Song and Bickel, 2011) since the noise covariance matrix is not taken into account when a loss function of the sum of squared residuals is used.

To our knowledge, very few research have been done on fitting a VAR model with sparsity constraints on both the autoregressive coefficients and the inverse of the noise covariance matrix. These constraints become important when a graphical VAR model is constructed. Similar to the partial correlation graph, each vertex of a graphical VAR model represents a component of the multivariate time series. The autoregressive coefficients characterize the directed edges and the undirected edges are determined by the non-diagonal entries of the inverse of the noise covariance matrix, see Eichler (2012). An alternative to study the dynamic interdependencies among the components of a multivariate time series is to consider a structural vector autoregressive (SVAR) model and represent them by a directed acyclic graph (DAG) (Oxley et al., 2004). The estimation of an SVAR model, however, requires restrictions, so that it is identifiable and a DAG can be built to represent the model.

To avoid such restrictions, and impose sparsity on both the autoregressive coefficients and the inverse of innovation covariance matrix, we propose an iterative algorithm to estimate a sparse VAR model. The algorithm considers the maximum likelihood estimation with the sparsity constraints as a "biconcave" problem in the sense that the optimization problem becomes concave when either the autoregressive coefficients or the inverse of noise covariance matrix is fixed (Gorski et al., 2007). We solve the alternating maximization problem, assuming the sparsity structure is known, using the alternating convex search (ACS) method and compare with the interior point method (fmincon in Matlab) and the direct search method (patternsearch in Matlab). To identify the structure, we present two methods, namely the time domain and frequency domain methods, and compare these two methods by simulation experiments.

Section 2 presents the sparse vector autoregressive model. Section 3 provides the proposed algorithm and visualization method on the fitted VAR model. Section 4 gives simulation studies. Section 5 exemplifies the proposed method by real data applications. Section 6 concludes.

2. Sparse vector autoregressive model

2.1. Vector autoregressive model

Consider a K-dimensional VAR(p) process:

$$y_t = \nu + \mathbf{A}_1 y_{t-1} + \cdots + \mathbf{A}_p y_{t-p} + u_t, \tag{3}$$

where $y_t = \left(y_{1,t}, \ldots, y_{K,t}\right)'$ is a $(K \times 1)$ vector, $\mathbf{A}_l, l = 1, \ldots, p$, are $(K \times K)$ autoregressive coefficient matrices, ν is a $(K \times 1)$ vector of intercepts, $u_t = \left(u_{1,t}, \ldots, u_{K,t}\right)'$ is a K-dimensional Gaussian noise vector, with mean $\mathbf{0}$ and a $(K \times K)$ nonsingular

Please cite this article in press as: Yuen T.P., et al., On constrained estimation of graphical time series models. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.01.019.