Graphical model selection for air quality time series

L. Paci[1,*] and G. Consonni[1]

[1] *Department of Statistical Sciences, Università Cattolica del Sacro Cuore; lucia.paci@unicatt.it, guido.consonni@unicatt.it;*
[*] *Corresponding author*

Abstract. *We propose an objective Bayes approach based on graphical models for learning dependencies among multiple air quality time series within the framework of Vector Autoregressive (VAR) models. Using a fractional Bayes factor approach, we obtain the marginal likelihood in closed form and construct an MCMC algorithm for Bayesian graphical model determination with limited computational burden. We apply our method to study the interactions between four air pollutants over the municipality of Milan (Italy).*

Keywords. *Decomposable graphical model; Fractional Bayes factor; Multiple pollutants.*

1 Introduction

Air pollution is a major global environmental risk to human health. Because humans are simultaneously exposed to a complex mixture of air pollutants, many organizations are moving toward a multi-pollutant approach to air quality [4]. Key aspects of such approach are the estimation of the health risk of multiple pollutants, the setting of regulatory standards and the design of compliance strategies for multiple pollutants. For example, a strategy to reduce levels of one pollutant, say particulate matter, may also affect the levels of other pollutants, say ozone. To take on these challenges, a better understanding of the interactions between air pollutants is required.

Pollutant measurements or numerical model estimates usually arise a multivariate time series collected at fixed locations or aggregated over a given spatial domain. Vector Autoregressive (VAR) models offer a suitable framework for analyzing multiple time series, such as air quality data. VAR models can be naturally represented by graphs, with directed edges reflecting the autoregressive structure over time while undirected edges describe the contemporaneous interactions among variables.

In this paper we describe an objective Bayes methodology to learn dynamic and contemporaneous dependencies among multiple pollutants modeled through a graphical VAR. Using a fractional Bayes factor approach, we are able to obtain the marginal likelihood in closed form and perform Bayes graphical model selection with limited computational burden because we focus on marginal likelihood, and disregard inference on model parameters. We apply our method to analyze the time series of four pollutants over the municipality of Milan (Italy). Results offer helpful insights about the relationship between these pollutants.

1.1 VAR model

Let $\mathbf{y}_t$ be a $(q \times 1)$ vector of observations collected at time t, $t = 1, \ldots, T$. The reduced form of a stable VAR of order k, VAR(k), is given by

$$\mathbf{y}_t = \sum_{i=1}^{k} \mathbf{B}_i \mathbf{y}_{t-i} + \boldsymbol{\epsilon}_t, \qquad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{B}_i$ are $(q \times q)$ matrices of coefficients or lag matrices, determining the dynamics of the system and $\boldsymbol{\epsilon}_t$ is a $(q \times 1)$ dimensional white noise process, that is $\boldsymbol{\epsilon}_t \mid \boldsymbol{\Sigma} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma})$, independently over time. The observation vector at time t depends linearly on the previous k observations, where k is assumed to be known. The intercept and exogenous variables can be added to the model, leading to straightforward modifications of the results shown here; for simplicity we omit details. Let $\mathbf{z}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-k})'$ denote the $(kq \times 1)$ vector of lagged observations at time t and $\mathbf{B}' = (\mathbf{B}_1, \ldots, \mathbf{B}_k)$ be the $(q \times kq)$ matrix obtained by collecting together the corresponding coefficient matrices. Let $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)'$ be the $(T \times q)$ matrix collecting all observations and $\mathbf{Z}$ be the $(T \times kq)$ matrix containing all the lagged variables, i.e., $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T)'$. Equation (1) can be rewritten in matrix form as

$$\mathbf{Y} = \mathbf{Z}\mathbf{B} + \mathbf{E}, \tag{2}$$

where $\mathbf{E} = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T)'$ is the $(T \times q)$ matrix of errors following a Matrix Normal distribution with zero mean, cross-covariance matrix between vector column j and vector column j' of $\mathbf{Y}$ equal to $\sigma_{jj'}\mathbf{I}_T$ and covariance matrix of vector row i equal to $\boldsymbol{\Sigma}$, which we write $\mathbf{E} \mid \boldsymbol{\Sigma} \sim \mathcal{N}_{T,q}(\mathbf{0}, \mathbf{I}_T, \boldsymbol{\Sigma})$.

2 Model selection for graphical VAR

Let $G = (V_{TS}, E)$, be a graph with node set $V_{TS} = V \times \mathbb{Z}$, $V = \{1, 2, \ldots, q\}$, and edge set E, whose edges have at most k lags and which is invariant under translation. If $(\mathbf{B}_i)_{vw}$ is the (v,w)-element of matrix $\mathbf{B}_i$ in (1) and $(\boldsymbol{\Omega})_{vw}$ is the (v,w)-entry of precision matrix $\boldsymbol{\Sigma}^{-1}$, then the VAR model with the following constraints on the parameters

$$\begin{aligned} &i) \quad (v, t-i) \to (w,t) \in E \Leftrightarrow (\mathbf{B}_i)_{vw} \neq 0 \qquad i = 1, \ldots, k \\ &ii) \quad (v,t) — (w,t) \in E \Leftrightarrow (\boldsymbol{\Sigma}^{-1})_{vw} \neq 0 \qquad t = 1, \ldots, T \end{aligned} \tag{3}$$

represents a VAR(k, G) model [6]. It follows from (3) that nonzero elements in $\mathbf{B}$ correspond to directed edges in the graph reflecting the recursive structure of the time series, while nonzero elements in $\boldsymbol{\Sigma}^{-1}$ correspond to undirected edges that specify conditional independencies at any given time t. In other words, learning the dynamic structure of a graphical VAR translates into a variable selection problem while learning the interactions among variables translates into a covariance selection problem.

Let $\boldsymbol{\Gamma}$ the binary connectivity matrix such that $(\boldsymbol{\Gamma})_{vw} = 1 \Leftrightarrow (\mathbf{B})_{vw} \neq 0$. Let $G^u = (V_{TS}, E^u)$ denote the undirected graph corresponding to the contemporaneous dependencies. We assume that $\boldsymbol{\Sigma}$ is Markov with respect to G^u, i.e., condition ii) is satisfied. We confine our analysis to the class of decomposable graphs for all time points, although we provide posterior graph summaries that go beyond this assumption. Given $\boldsymbol{\Gamma}$ and G^u, we denote $\mathbf{B}^{(\boldsymbol{\Gamma})}$ the associated coefficient matrix and $\boldsymbol{\Sigma}^{(G^u)}$ the associated covariance matrix. Then the likelihood of a graphical VAR(k, G) factorizes as

$$f\left(\mathbf{Y} \mid \mathbf{B}^{(\boldsymbol{\Gamma})}, \boldsymbol{\Sigma}^{(G^u)}\right) = \frac{\prod_{C \in \mathcal{C}} f\left(\mathbf{Y}_C \mid \mathbf{B}_C^{(\boldsymbol{\Gamma})}, \boldsymbol{\Sigma}_{CC}^{(G^u)}\right)}{\prod_{S \in \mathcal{S}} f\left(\mathbf{Y}_S \mid \mathbf{B}_S^{(\boldsymbol{\Gamma})}, \boldsymbol{\Sigma}_{SS}^{(G^u)}\right)}, \tag{4}$$

where $\mathcal{C}$ and $\mathcal{S}$ denote the set of cliques and separators of the undirected decomposable graph, while $\mathbf{B}_C^{(\mathbf{\Gamma})}$ and $\mathbf{B}_S^{(\mathbf{\Gamma})}$ are the matrices whose columns contain the nonzeros coefficients of the selected responses $\mathbf{Y}_C$ and $\mathbf{Y}_S$, respectively. Notice that the set of cliques C and separators S depend on G^u, which is omitted for simplicity.

In this work, we employ an objective approach for model selection based on the Fractional Bayes Factor (FBF), originally presented in [2]. The idea of the FBF is to train a noninformative, typically improper, prior using a small fractional power b of the likelihood, thus converting the noninformative prior into a proper prior. The latter is then used to compute the marginal likelihood based on the complementary fractional power $(1-b)$ of the likelihood.

We start with a prior for $\left(\mathbf{B}^{(\mathbf{\Gamma})}, \mathbf{\Sigma}^{(G^u)}\right)$ that is a limiting form of a Matrix Normal Hyper-Inverse Wishart distribution. Combining such prior with a fraction $b = T_0/T$ of likelihood (4), we obtain the fractional prior of a VAR(k,G) that is a Matrix Normal Hyper-Inverse Wishart distribution, $\mathcal{MNHIW}\left(\hat{\mathbf{B}}^{(\mathbf{\Gamma})}, \underline{\mathbf{C}}, d, \underline{\mathbf{R}}\right)$, where $\hat{\mathbf{B}}^{(\mathbf{\Gamma})}$ is the ordinary least square estimate of nonzero coefficients, $\underline{\mathbf{C}} = T/T_0\,(\mathbf{Z}'\mathbf{Z})^{-1}$, $d = T_0 - kq$ and $\underline{\mathbf{R}} = T_0/T\,\hat{\mathbf{E}}'\hat{\mathbf{E}}$ with $\hat{\mathbf{E}} = \mathbf{Y} - \mathbf{Z}\hat{\mathbf{B}}^{(\mathbf{\Gamma})}$. Hence, the fractional prior factorizes as

$$p^F\left(\mathbf{B}^{(\mathbf{\Gamma})}, \mathbf{\Sigma}^{(G^u)}\right) = \frac{\prod_{C\in\mathcal{C}} \mathcal{N}_{kq,|C|}\left(\mathbf{B}_C^{(\mathbf{\Gamma})}, \underline{\mathbf{C}}, \mathbf{\Sigma}_{CC}^{(G^u)}\right) \mathcal{IW}_{|C|}\left(d+|C|-1, \mathbf{R}_{CC}\right)}{\prod_{S\in\mathcal{S}} \mathcal{N}_{kq,|S|}\left(\mathbf{B}_S^{(\mathbf{\Gamma})}, \underline{\mathbf{C}}, \mathbf{\Sigma}_{SS}^{(G^u)}\right) \mathcal{IW}_{|S|}\left(d+|S|-1, \mathbf{R}_{SS}\right)}. \tag{5}$$

Because of conjugacy of prior (5), we can write

$$m^F(\boldsymbol{Y} \mid \mathbf{\Gamma}, G^u) = \frac{\prod_{C\in\mathcal{C}} m^F\left(\mathbf{Y}_C \mid \mathbf{\Gamma}\right)}{\prod_{S\in\mathcal{S}} m^F\left(\mathbf{Y}_S \mid \mathbf{\Gamma}\right)}, \tag{6}$$

where, $m^F(\mathbf{Y}_C \mid \mathbf{\Gamma})$ and $m^F(\mathbf{Y}_S \mid \mathbf{\Gamma})$ can be obtained in closed form using the results in [5].

2.1 Computational details

Posterior inference on the space of decomposable graphs is carried out through Markov Chain Monte Carlo (MCMC) methods. In particular, at each step of our collapsed Gibbs sampling, we locally modify $\mathbf{\Gamma}$ and G^u and then update through the following Metropolis-Hasting steps:

- we move from $\mathbf{\Gamma}$ to $\mathbf{\Gamma}_*$ with acceptance probability $r(\mathbf{\Gamma}, \mathbf{\Gamma}_*) = min\left\{1, \frac{m^F(\boldsymbol{Y}|\mathbf{\Gamma}_*, G^u) p(\mathbf{\Gamma}_*) q(\mathbf{\Gamma}|\mathbf{\Gamma}_*)}{m^F(\boldsymbol{Y}|\mathbf{\Gamma}, G^u) p(\mathbf{\Gamma}) q(\mathbf{\Gamma}_*|\mathbf{\Gamma})}\right\}$;
- we move from G^u to G^u_* with acceptance probability $r(G^u, G^u_*) = min\left\{1, \frac{m^F(\boldsymbol{Y}|\mathbf{\Gamma}, G^u_*) p(G^u_*) q(G^u|G^u_*)}{m^F(\boldsymbol{Y}|\mathbf{\Gamma}, G^u) p(G^u) q(G^u_*|G^u)}\right\}$.

We compute $m^F(\boldsymbol{Y} \mid \mathbf{\Gamma}_*, G^u)$ using (6) while a multiplicity-correction prior for both the directed dynamic graph and the undirected contemporaneous graph is assumed [5]. Finally, $q(\mathbf{\Gamma}_* \mid \mathbf{\Gamma}) = \alpha$ when adding an edge, and $q(\mathbf{\Gamma}_* \mid \mathbf{\Gamma}) = 1-\alpha$ when deleting an edge; same proposal is employed for G^u, see [1].

Given the MCMC output we can approximate the posterior inclusion probability of edge (v,w) as the proportion of MCMC iterations, after the burn-in, wherein the edge (v,w) appears. A variety of summaries of the MCMC output can be adopted to estimate the data generating graph. Here, we employ a Bayesian version of the (approximate) expected false discovery rate (FDR; [3]), i.e., we estimate the graph considering those edges whose posterior probability of inclusion is greater than $1-r$, where r is determined so that the FDR is at most 5%.

3 Analysis of air quality data

We analyze a 4-dimensional time series of air pollutants from January 2016 to December 2018 over the municipality of Milan (Italy). Data are provided and aggregated by the air quality system of ARPA (the regional environmental protection agency of Lombardia). In particular, we study the time series of daily average of nitrogen dioxide (NO_2), daily 8-hour maximum ozone (O_3), daily particulate matter (PM_{10}) and daily fine particulate matter ($PM_{2.5}$). Left panel of Figure 1 displays such time series, highlighting the cyclical pattern of the pollutants.

A VAR$(1, G)$ is fitted using the approach described in Section 2. Daily average of temperature and precipitation over the city are employed as covariates. Preliminary results of the variable selection algorithm applied to these variables show that temperature is a relevant predictor for both NO_2 and O_3 while precipitation is relevant for particulate matter and nitrogen dioxide. Right panel of Figure 1 shows

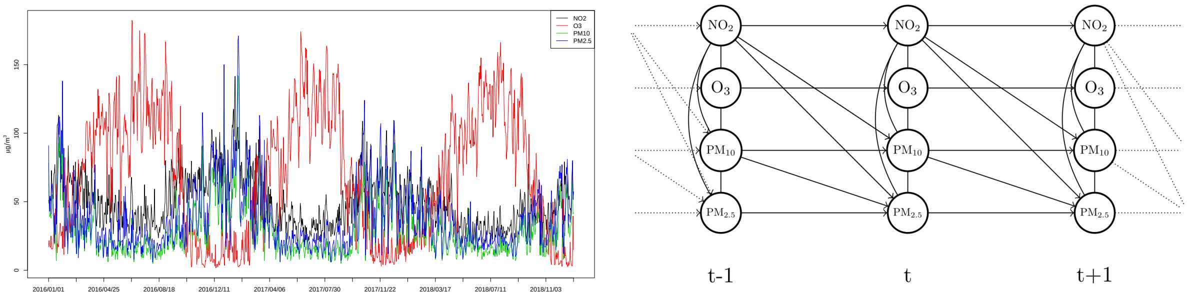

Figure 1: Left panel: daily averages of the pollutants. Right panel: estimated VAR graph.

the estimated VAR graph obtained using the FDR criterion described in Section 2. Some of the links of the graph can be explained by the chemical and physical transformation of the pollutants, e.g., NO_2 is a precursor to ozone while particulate matter and NO_2 are both indicators of urban pollution. However, interactions between air pollutants are very complex and require further investigations.

Acknowledgments. The research of the authors has been partially supported by a grant from UCSC (track D1) and by the EU COSTNET project (CA15109).

References

[1] Bhadra A. and Mallick B. K. Joint high-dimensional Bayesian variable and covariance selection with an application to eQTL analysis. *Biometrics*, 69:447–457, 2013.

[2] O'Hagan A. Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society. Series B*, 57:99–138, 1995.

[3] Peterson C., Stingo F. C., and Vannucci M. Bayesian inference of multiple Gaussian graphical models. *Journal of the American Statistical Association*, 110:159–174, 2015.

[4] Dominici F., Peng R., Barr C. D., and Bell M. Protecting human health from air pollution: Shifting from a single-pollutant to a multi-pollutant approach. *Epidemiology*, 21:187–194, 2010.

[5] Consonni G., La Rocca L., and Peluso S. Objective Bayes covariate-adjusted sparse graphical model selection. *Scandinavian Journal of Statistics*, 44:741–764, 2017.

[6] Eichler M. Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153:233–268, 2012.