Structural learning of contemporaneous dependencies in graphical VAR models

Lucia Paci

joint work with *Guido Consonni*

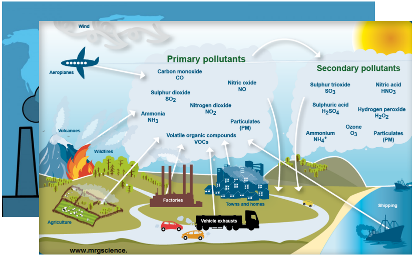
Department of Statistical Sciences
Università Cattolica del Sacro Cuore, Milan

O'BAYES 2019
University of Warwick, 28 June - 2 July 2019

Air quality time series

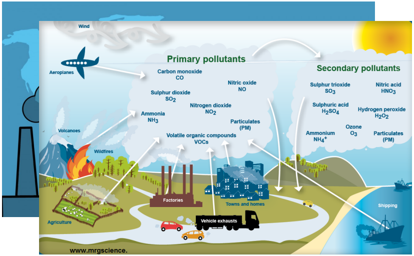
- Air pollution is a major global environmental risk to human health (WHO, 2018)

Air quality time series



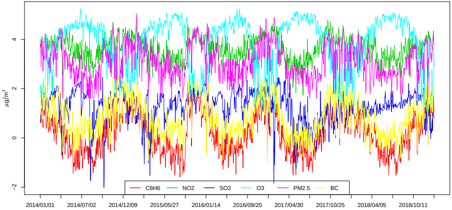
- Air pollution is a major global environmental risk to human health (WHO, 2018)
- We are simultaneously exposed to a complex mixture of air pollutants
- Moving toward a multi-pollutant approach to air quality (Dominici et al., 2010)

Air quality time series



- Air pollution is a major global environmental risk to human health (WHO, 2018)
- We are simultaneously exposed to a complex mixture of air pollutants
- Moving toward a multi-pollutant approach to air quality (Dominici et al., 2010)

- A better understanding of the interactions between air pollutants is critical
- *Learning dependencies among multiple time series*



Graphical modeling

Graph theory

$$G = (V, E)$$

- finite set of **vertices** $V = \{1, \ldots, q\}$
- subset of **edges** $E \subseteq V \times V$

Nodes $\Leftrightarrow$ Random variables

Edges $\Leftrightarrow$ Probabilistic relationships

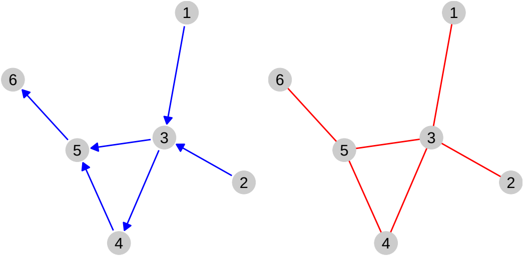


Figure: Directed (left) and undirected (right) graphs.

Graphical modeling

Graph theory

$$G = (V, E)$$

- finite set of **vertices** $V = \{1, \ldots, q\}$
- subset of **edges** $E \subseteq V \times V$

Nodes $\Leftrightarrow$ Random variables

Edges $\Leftrightarrow$ Probabilistic relationships

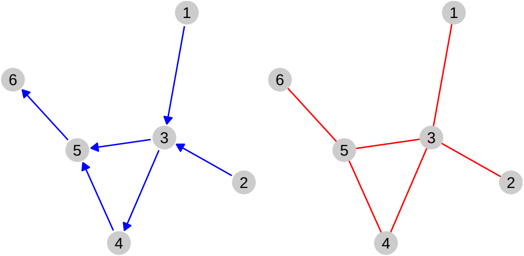


Figure: Directed (left) and undirected (right) graphs.

Graphical model

Family of probability distributions for the q random variables which factorizes according to a given graph. *Conditional independencies* are read from the graph.

Vector autoregressive models

- VAR models are widely used to analyze multivariate time series, with application in many research fields (economics, finance, genomics, neuroscience, environmental sciences, ...)
- VARs account for contemporaneous dependencies among variables as well as their evolution over time
- VARs are flexible and easy to interpret

Vector autoregressive models

- VAR models are widely used to analyze multivariate time series, with application in many research fields (economics, finance, genomics, neuroscience, environmental sciences, ...)
- VARs account for contemporaneous dependencies among variables as well as their evolution over time
- VARs are flexible and easy to interpret
- VARs are over-parametrized models: investigating parameters restrictions that are supported by the data. Some solutions proposed in the literature:
    - Bayesian SSVS (**George et al. 2008**, Wang 2010; Korobilis 2013)
    - BNP approach (Billio et al., 2019)
    - Graphical modeling (Corander and Villani, 2006; Abegaz and Wit, 2013; Ahelegbey et al., 2016)

Vector autoregressive models

VAR(k)

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

- $\boldsymbol{y}_t$: $(q \times 1)$ vector of observations at time t
- $\boldsymbol{B}_i$: $(q \times q)$ lag-i matrix of coefficients
- $\boldsymbol{\epsilon}_t$: $(q \times 1)$ vector of error, where $\boldsymbol{\epsilon}_t \mid \boldsymbol{\Sigma} \sim \mathrm{N}(0, \boldsymbol{\Sigma})$
- k : known number of lags

To simplify notation, no intercept is considered here.

In general, exogenous variables can be included in the model.

VAR (k)

Matrix form

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{B} + \boldsymbol{E}$$

- $\boldsymbol{Y}$: $(T \times q)$ matrix of observations
- $z_t = (\boldsymbol{y}'_{t-1}, \dots, \boldsymbol{y}'_{t-k})'$: $(kq \times 1)$ vector of lagged observations at time t
- $\boldsymbol{Z}$: $(T \times kq)$ matrix of all lagged observations
- $\boldsymbol{B}$: $(kq \times q)$ matrix of coefficients
- $\boldsymbol{E}$: $(T \times q)$ matrix of errors, $\boldsymbol{E} \mid \boldsymbol{\Sigma} \sim \text{MN}(\boldsymbol{0}, I_T, \boldsymbol{\Sigma})$

VAR (k)

Matrix form

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{B} + \boldsymbol{E}$$

- $\boldsymbol{Y} : (T \times q)$ matrix of observations
- $z_t = (\boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-k})' : (kq \times 1)$ vector of lagged observations at time t
- $\boldsymbol{Z} : (T \times kq)$ matrix of all lagged observations
- $\boldsymbol{B} : (kq \times q)$ matrix of coefficients
- $\boldsymbol{E} : (T \times q)$ matrix of errors, $\boldsymbol{E} \mid \boldsymbol{\Sigma} \sim \mathrm{MN}(\boldsymbol{0}, I_T, \boldsymbol{\Sigma})$

Likelihood

$$f(\boldsymbol{Y} \mid \boldsymbol{B}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{Tq}{2}} |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\left(\boldsymbol{B} - \hat{\boldsymbol{B}}\right)' \boldsymbol{Z}'\boldsymbol{Z}\left(\boldsymbol{B} - \hat{\boldsymbol{B}}\right) + \hat{\boldsymbol{E}}'\hat{\boldsymbol{E}}\right)\right]\right\}$$

where $\hat{\boldsymbol{E}} = \boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{B}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$.

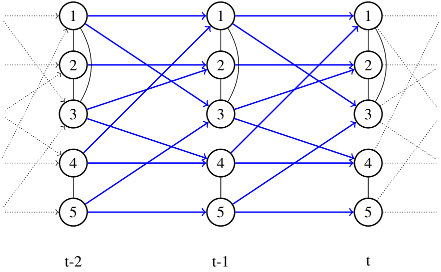
VAR(k, G)

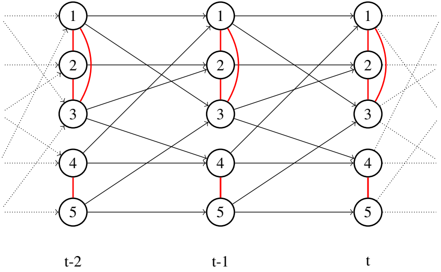
- $V = \{1, 2, \ldots, q\}$ and $V_{TS} = V \times \{1, \ldots, T\}$
- $G = (V_{TS}, E)$ (edges have at most k lags)

VAR(k, G)

- $V = \{1, 2, \ldots, q\}$ and $V_{TS} = V \times \{1, \ldots, T\}$
- $G = (V_{TS}, E)$ (edges have at most k lags)

$$
\begin{array}{lllllll}
(v, t-i) & \rightarrow (w, t) & \in E & \Leftrightarrow & (\boldsymbol{B}_i)_{vw} & \neq 0 & i = 1, \ldots, k \\
(v, t) & — (w, t) & \in E & \Leftrightarrow & \left(\boldsymbol{\Sigma}^{-1}\right)_{vw} & \neq 0 & t = 1, \ldots, T
\end{array}
$$





VAR(k, G)

- $V = \{1, 2, \ldots, q\}$ and $V_{TS} = V \times \{1, \ldots, T\}$
- $G = (V_{TS}, E)$ (edges have at most k lags)

$$\begin{aligned} (v, t-i) \; &\rightarrow (w,t) \in E \Leftrightarrow (\boldsymbol{B}_i)_{vw} \neq 0 \qquad i = 1, \ldots, k \\ (v, t) \; &— (w,t) \in E \Leftrightarrow \left(\boldsymbol{\Sigma}^{-1}\right)_{vw} \neq 0 \qquad t = 1, \ldots, T \end{aligned}$$

- Nonzero elements in $\boldsymbol{B}$ correspond to directed edges in the graph
  dynamic interactions - *inter* time dependencies
- Nonzero elements in the precision matrix $\boldsymbol{\Sigma}^{-1}$ correspond to undirected edges in the graph
  contemporaneous interactions - *intra* time dependencies

VAR(k, G)

- $V = \{1, 2, \ldots, q\}$ and $V_{TS} = V \times \{1, \ldots, T\}$
- $G = (V_{TS}, E)$ (edges have at most k lags)

$$
\begin{array}{lllllll}
(v, t-i) & \rightarrow (w, t) & \in E & \Leftrightarrow & (\boldsymbol{B}_i)_{vw} & \neq 0 & i = 1, \ldots, k \\
(v, t) & — (w, t) & \in E & \Leftrightarrow & \left(\boldsymbol{\Sigma}^{-1}\right)_{vw} & \neq 0 & t = 1, \ldots, T
\end{array}
$$

- Nonzero elements in $\boldsymbol{B}$ correspond to directed edges in the graph
  dynamic interactions - *inter* time dependencies

  $\Rightarrow$ **VARIABLE SELECTION PROBLEM**

- Nonzero elements in the precision matrix $\boldsymbol{\Sigma}^{-1}$ correspond to undirected edges in the graph
  contemporaneous interactions - *intra* time dependencies

VAR(k, G)

- $V = \{1, 2, \ldots, q\}$ and $V_{TS} = V \times \{1, \ldots, T\}$
- $G = (V_{TS}, E)$ (edges have at most k lags)

$$\begin{array}{lcl} (v, t-i) \;\rightarrow\; (w,t) \in E \Leftrightarrow (\boldsymbol{B}_i)_{vw} \neq 0 & & i = 1, \ldots, k \\ (v, t) \;-\; (w,t) \in E \Leftrightarrow \left(\boldsymbol{\Sigma}^{-1}\right)_{vw} \neq 0 & & t = 1, \ldots, T \end{array}$$

- Nonzero elements in $\boldsymbol{B}$ correspond to directed edges in the graph
  dynamic interactions - *inter* time dependencies

  $\Rightarrow$ **VARIABLE SELECTION PROBLEM**

- Nonzero elements in the precision matrix $\boldsymbol{\Sigma}^{-1}$ correspond to undirected edges in the graph
  contemporaneous interactions - *intra* time dependencies

  $\Rightarrow$ **COVARIANCE SELECTION PROBLEM**

VAR(k, G)

- $\boldsymbol{\Gamma}$: binary connectivity matrix such that $(\boldsymbol{\Gamma})_{vw} = 1 \Leftrightarrow (\mathbf{B})_{vw} \neq 0$
- $\mathbf{B}^{(\boldsymbol{\Gamma})}$: the associated matrix of nonzeros coefficients

VAR(k, G)

- $\mathbf{\Gamma}$: binary connectivity matrix such that $(\mathbf{\Gamma})_{vw} = 1 \Leftrightarrow (\mathbf{B})_{vw} \neq 0$
- $\mathbf{B}^{(\mathbf{\Gamma})}$: the associated matrix of nonzeros coefficients
- $G^u = (V, E^u)$: undirected ***decomposable*** graph representing contemporaneous dependencies among variables at any time t
- $\mathbf{\Sigma}^{(G^u)}$: the associated covariance matrix which is Markov with respect to G^u

VAR(k, G)

- $\boldsymbol{\Gamma}$: binary connectivity matrix such that $(\boldsymbol{\Gamma})_{vw} = 1 \Leftrightarrow (\mathbf{B})_{vw} \neq 0$
- $\mathbf{B}^{(\boldsymbol{\Gamma})}$: the associated matrix of nonzeros coefficients
- $G^u = (V, E^u)$: undirected ***decomposable*** graph representing contemporaneous dependencies among variables at any time t
- $\boldsymbol{\Sigma}^{(G^u)}$: the associated covariance matrix which is Markov with respect to G^u

Likelihood

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

$\mathbf{Y}_C$ and $\mathbf{B}_C^{(\boldsymbol{\Gamma})}$ are submatrices corresponding to clique $C \in \mathcal{C}$ of G^u; similarly for $S \in \mathcal{S}$ of G^u.

NOTE: the likelihood of a graphical VAR factorizes as the likelihood of an ordinary decomposable graphical model!

Fractional Bayes Factor (FBF)

- $\{M_l\}$: collection of models for $\boldsymbol{Y}$, each consisting of a family of sampling densities $f_l(\boldsymbol{Y} \mid \boldsymbol{\theta}_l)$, indexed by $\boldsymbol{\theta}_l$
- $p_l^D(\boldsymbol{\theta}_l)$: default noninformative prior density on $\boldsymbol{\theta}_l$
- $b \in (0, 1)$

Fractional Bayes Factor (FBF)

- $\{M_l\}$: collection of models for $\boldsymbol{Y}$, each consisting of a family of sampling densities $f_l(\boldsymbol{Y} \mid \boldsymbol{\theta}_l)$, indexed by $\boldsymbol{\theta}_l$
- $p_l^D(\boldsymbol{\theta}_l)$: default noninformative prior density on $\boldsymbol{\theta}_l$
- $b \in (0, 1)$

"To show you how well I understand fractions, I only did half of my homework."

The FBF (O'Hagan, 1995) of model M_l against $M_{l'}$ is

$$\mathrm{FBF}_{ll'} = m_l^F(\boldsymbol{Y})/m_{l'}^F(\boldsymbol{Y}),$$

where $m_l^F(\boldsymbol{Y})$ is the **fractional marginal likelihood** of M_l given by

$$m_l^F(\boldsymbol{Y}) = \frac{\int f_l(\boldsymbol{Y} \mid \boldsymbol{\theta}_l) p_l^D(\boldsymbol{\theta}_l) d\boldsymbol{\theta}_l}{\int f_l^b(\boldsymbol{Y} \mid \boldsymbol{\theta}_l) p_l^D(\boldsymbol{\theta}_l) d\boldsymbol{\theta}_l} = \int f_l^{1-b}(\boldsymbol{Y} \mid \boldsymbol{\theta}_l) p_l^F(\boldsymbol{\theta}_l) d\boldsymbol{\theta}_l,$$

where $p_l^F(\boldsymbol{\theta}_k) \propto f_l^b(\boldsymbol{Y} \mid \boldsymbol{\theta}_l) p_l^D(\boldsymbol{\theta}_l)$ is the induced **fractional prior**.

Fractional prior for VAR(k, G)

We start with the noninformative prior

$$p^D\left(\boldsymbol{B}^{(\boldsymbol{\Gamma})}, \boldsymbol{\Sigma}^{(G^u)}\right) \propto \frac{\prod_{C \in \mathcal{C}} \left|\boldsymbol{\Sigma}^{(G^u)}\right|^{-|C|}}{\prod_{S \in \mathcal{S}} \left|\boldsymbol{\Sigma}^{(G^u)}\right|^{-|S|}}$$

the marginal distribution of $\boldsymbol{\Sigma}^{(G^u)}$ is a limiting form of an Hyper-inverse Wishart distribution (Dawid and Lauritzen, 1993).

Fractional prior for VAR(k, G)

We start with the noninformative prior

$$p^D\left(\boldsymbol{B}^{(\boldsymbol{\Gamma})}, \boldsymbol{\Sigma}^{(G^u)}\right) \propto \frac{\prod_{C \in \mathcal{C}} \left|\boldsymbol{\Sigma}^{(G^u)}\right|^{-|C|}}{\prod_{S \in \mathcal{S}} |\boldsymbol{\Sigma}^{(G^u)}|^{-|S|}}$$

the marginal distribution of $\boldsymbol{\Sigma}^{(G^u)}$ is a limiting form of an Hyper-inverse Wishart distribution (Dawid and Lauritzen, 1993).

Training the prior with a fraction $b = T_0/T$ of the likelihood, the fractional prior of a VAR(k, G) becomes a *Matrix Normal Hyper-inverse Wishart* distribution:

$$\left(\mathbf{B}^{(\boldsymbol{\Gamma})}, \boldsymbol{\Sigma}^{(G^u)}\right) \sim \text{MNHIW}\left(\hat{\boldsymbol{B}}^{(\boldsymbol{\Gamma})}, \boldsymbol{D}, d, \boldsymbol{R}\right)$$

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

$\hat{\mathbf{B}}^{(\boldsymbol{\Gamma})}$ is the OLS estimate of nonzero coefficients, $d = T_0 - kq$, $\boldsymbol{D} = (T/T_0)\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}$ and $\underline{\mathbf{R}} = T_0/T\, \hat{\mathbf{E}}'\hat{\mathbf{E}}$ with $\hat{\mathbf{E}} = \mathbf{Y} - \mathbf{Z}\hat{\mathbf{B}}^{(\boldsymbol{\Gamma})}$.

Fractional marginal likelihood

Because of conjugacy, the fractional marginal likelihood has a closed form obtained as the ratio of the prior and posterior normalizing constants, that is

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

where

$$m^F(\boldsymbol{Y}_J \mid \boldsymbol{\Gamma}) = \pi^{-\frac{(T-T_0)|J|}{2}} \left(\frac{T_0}{T}\right)^{\frac{(a-q-|J|+T_0)|J|}{2}} \left|\hat{\boldsymbol{E}}_J' \hat{\boldsymbol{E}}_J\right|^{-\frac{T-T_0}{2}} \frac{\Gamma_{|J|}\left((a-kq+T)/2\right)}{\Gamma_{|J|}\left((a-kq+T_0)/2\right)}$$

with $J = \{C, S\}$, $a = q - 1$ and $T_0 = kq + 1$ (Consonni et al., 2017).

Valid marginal likelihood if $T > |C| + kq - 1$, for each $C \in \mathcal{C}$.

Prior distribution on graph space

- Dynamic component

$$p(\mathbf{\Gamma}) \propto \binom{kq^2}{p^*} \Gamma\left(1+p^*\right) \Gamma\left(1+kq^2-p^*\right)$$

where $p^* = \sum_{h=1}^{kq^2} \text{vec}(\mathbf{\Gamma})_h$.

Prior distribution on graph space

- Dynamic component

$$p(\boldsymbol{\Gamma}) \propto \binom{kq^2}{p^*} \Gamma\left(1 + p^*\right) \Gamma\left(1 + kq^2 - p^*\right)$$

where $p^* = \sum_{h=1}^{kq^2} \text{vec}(\boldsymbol{\Gamma})_h$.

- Contemporaneous component

$$p(G^u) \propto \binom{q(q-1)/2}{|G^u|} \Gamma\left(1 + |G^u|\right) \Gamma\left(b + q(q-1)/2 - |G^u|\right),$$

where $|G^u|$ denotes the number of edges in G^u and $b = (2q-2)/3 - 1$ to favor sparsity.

Computational details

Posterior inference on the space of decomposable graphs is carried out via a collapsed Gibbs sampling. At each step, we *locally* modify $\mathbf{\Gamma}$ and G^u and then update through the following Metropolis-Hasting steps:

Computational details

Posterior inference on the space of decomposable graphs is carried out via a collapsed Gibbs sampling. At each step, 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

$$r(\mathbf{\Gamma}, \mathbf{\Gamma}_*) = min\left\{1, \frac{m^F(\boldsymbol{Y} \mid \mathbf{\Gamma}_*, G^u)p(\mathbf{\Gamma}_*)q(\mathbf{\Gamma} \mid \mathbf{\Gamma}_*)}{m^F(\boldsymbol{Y} \mid \mathbf{\Gamma}, G^u)p(\mathbf{\Gamma})q(\mathbf{\Gamma}_* \mid \mathbf{\Gamma})}\right\}$$

  where $q(\mathbf{\Gamma}_* \mid \mathbf{\Gamma}) = \alpha_\Gamma$ when adding a directed edge, and $q(\mathbf{\Gamma}_* \mid \mathbf{\Gamma}) = 1 - \alpha_\Gamma$ when deleting a directed edge;

Computational details

Posterior inference on the space of decomposable graphs is carried out via a collapsed Gibbs sampling. At each step, 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

$$r(\mathbf{\Gamma}, \mathbf{\Gamma}_*) = min\left\{1, \frac{m^F(\boldsymbol{Y} \mid \mathbf{\Gamma}_*, G^u)p(\mathbf{\Gamma}_*)q(\mathbf{\Gamma} \mid \mathbf{\Gamma}_*)}{m^F(\boldsymbol{Y} \mid \mathbf{\Gamma}, G^u)p(\mathbf{\Gamma})q(\mathbf{\Gamma}_* \mid \mathbf{\Gamma})}\right\}$$

  where $q(\mathbf{\Gamma}_* \mid \mathbf{\Gamma}) = \alpha_\Gamma$ when adding a directed edge, and $q(\mathbf{\Gamma}_* \mid \mathbf{\Gamma}) = 1 - \alpha_\Gamma$ when deleting a directed edge;

- we move from G^u to G^u_* with acceptance probability

$$r(G^u, G^u_*) = min\left\{1, \frac{m^F(\boldsymbol{Y} \mid \mathbf{\Gamma}, G^u_*)p(G^u_*)q(G^u \mid G^u_*)}{m^F(\boldsymbol{Y} \mid \mathbf{\Gamma}, G^u)p(G^u)q(G^u_* \mid G^u)}\right\},$$

where $q(G^u_* \mid G^u) = \alpha_G$ when adding an undirected edge, and $q(G^u_* \mid G^u) = 1 - \alpha_G$ when deleting an undirected edge (Bhadra and Mallick, 2013).

Posterior summaries

Given the MCMC output:

- we can approximate the posterior probability of each distinct graph G_l^u $(l = 1, \ldots, L)$

$$p(G_l^u \mid \boldsymbol{Y}) = \left[1 + \sum_{l' \neq h} \frac{p(G_{l'}^u)}{p(G_l^u)} \mathrm{FBF}_{l'l}(\boldsymbol{Y}) \right]^{-1}$$

Posterior summaries

Given the MCMC output:

- we can approximate the posterior probability of each distinct graph G_l^u ($l = 1, \dots, L$)

$$p(G_l^u \mid \boldsymbol{Y}) = \left[1 + \sum_{l' \neq h} \frac{p(G_{l'}^u)}{p(G_l^u)} \mathrm{FBF}_{l'l}(\boldsymbol{Y}) \right]^{-1}$$

- we can approximate the posterior inclusion probability of any directed and undirected edges through the proportion of MCMC iterations wherein the edges appear;

Posterior summaries

Given the MCMC output:

- we can approximate the posterior probability of each distinct graph G_l^u ($l = 1, \ldots, L$)

$$p(G_l^u \mid \boldsymbol{Y}) = \left[1 + \sum_{l' \neq h} \frac{p(G_{l'}^u)}{p(G_l^u)} \mathrm{FBF}_{l'l}(\boldsymbol{Y})\right]^{-1}$$

- we can approximate the posterior inclusion probability of any directed and undirected edges through the proportion of MCMC iterations wherein the edges appear;
- provide an estimate of the data generating graph. Here, we use the (approximate) expected *false discovery rate* (FDR; Peterson et al. 2015), 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%.

Simulation results

Contemporaneous graph ($k = 1$)

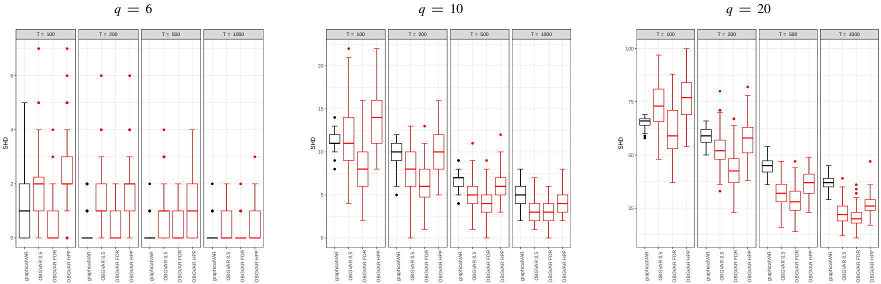
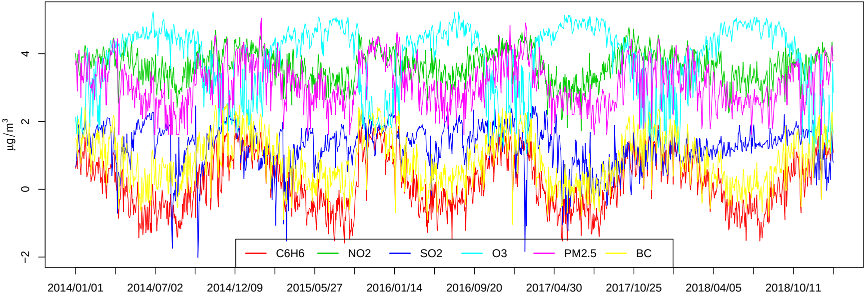


Figure: Structural Hamming distances (SHDs) between the estimated undirected graph and the true one, over 100 replicates.

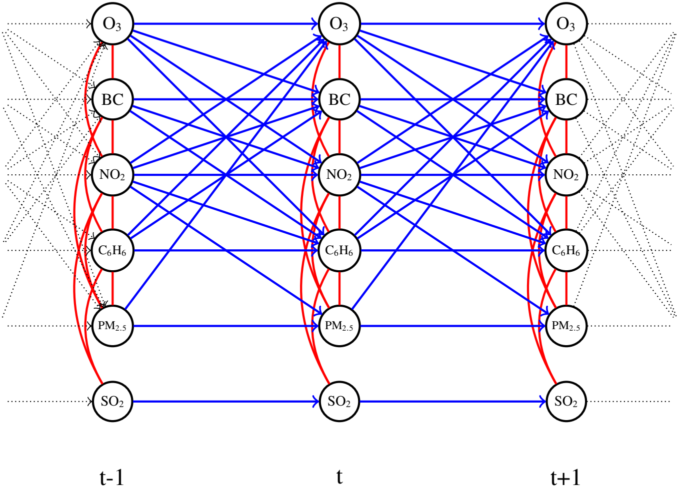
Analysis of air quality data

- Daily average measurements of 6 pollutants from Jan 2014 to Dec 2018, collected by a monitoring station located in Milan (Italy)
- Meteorological covariates (daily temperature and humidity average)



Results

VAR$(1, G)$



Summary

- Structure learning of dynamic and contemporaneous dependencies of multiple time series
- Fractional Bayes factor approach tailored to graphical VARs
- The likelihood of a graphical VAR factorizes as an ordinary decomposable graphical model
- The fractional marginal likelihood is available in closed form, expediting the MCMC computation
- Posterior summaries of the graph beyond decomposable models
- Full posterior inference can be also accomplished by sampling from the joint full conditional distribution (Carvalho et al., 2007)

Summary

- Structure learning of dynamic and contemporaneous dependencies of multiple time series
- Fractional Bayes factor approach tailored to graphical VARs
- The likelihood of a graphical VAR factorizes as an ordinary decomposable graphical model
- The fractional marginal likelihood is available in closed form, expediting the MCMC computation
- Posterior summaries of the graph beyond decomposable models
- Full posterior inference can be also accomplished by sampling from the joint full conditional distribution (Carvalho et al., 2007)
- What's next ?
  - Clustering of VAR coefficients
  - Spatio-temporal analysis

Thank you for your attention

References

Abegaz, F. and Wit, E. (2013). Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics*, 14:586.

Ahelegbey, D. F., Billio, M., and Casarin, R. (2016). Bayesian graphical models for structural vector autoregressive processes. *Journal of Applied Econometrics*, 31:357–386.

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

Billio, M., Casarin, R., and Rossini, L. (2019). Bayesian nonparametric sparse var models. *Journal of Econometrics*.

Carvalho, C. M., Massam, H., and West, M. (2007). Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*, 94:647–659.

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

Corander, J. and Villani, M. (2006). A Bayesian approach to modelling graphical vector autoregressions. *Journal of Time Series Analysis*, 27:141–156.

Dawid, A. P. and Lauritzen, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21:1272–1317.

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

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142:553–580.

Korobilis, D. (2013). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, 28(2):204–230.

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

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

Wang, H. (2010). Sparse seemingly unrelated regression modelling: Applications in finance and econometrics. *Computational Statistics & Data Analysis*, 54:2866 – 2877.