

# A Bayesian Approach to Modelling Graphical Vector Autoregressions

Marco Zanotti

University Milano-Bicocca



# Contents

1. VARs
2. Graphical VARs
3. Bayesian Model Assessment for Graphical VARs
4. Conclusions

# 1. VARs

# VAR Model

Vector autoregressive (VAR) models are popular choice for studying the joint dynamics of multiple time series.

Consider  $x_t$  as a  $p$ -dimensional vector of time-series at time  $t$ , a VAR model is just a multivariate normal regression of  $x_t$  on its own-lags

$$x_t = c + \sum_{i=1}^k \Pi_i x_{t-i} + \epsilon_t \quad (1)$$

where  $\Pi_i$  are  $p \times p$  coefficient matrices determining the dynamics of the system,  $c$  is a deterministic vector of  $p$  components and  $\epsilon_t \sim N_p(0, \Sigma)$ .

## Example of VAR(1)

A first order VAR model ( $p = 1$ ) in two variables would be given by

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

If, for example  $\pi_{12} \neq 0$ , this means that the history of  $x_2$  helps explaining  $x_1$ .

## Pros & Cons

- ▶ Main advantage: simplicity. VAR models require no special structure since the outcome variables are regressed on their own lagged variables.
- ▶ Main disadvantage: large number of regression coefficients,  $K(Kp + m)$ . The significant number of regression parameters is proportional to the number of lags, hence interpretation may be difficult and may not be good when fitted to small data.

Several tools have been proposed to aid in the interpretation of VAR models, most notably Impulse Response Functions (IRF) and Granger Causality tests (GC).

# Granger Causality

The general idea behind Granger causality is that a variable  $X$  Granger causes  $Y$  if past values of  $X$  can help explain  $Y$ .

Of course, if Granger causality holds this does not guarantee that  $X$  causes  $Y$ . Nevertheless, if past values of  $X$  have explanatory power for current values of  $Y$ , it at least suggests that  $X$  improves  $Y$ 's predictability.

In the context of a VAR model, testing for Granger causality simply implies to test the significance of the cross-coefficients ( $\pi_{12}$  and  $\pi_{21}$  in the VAR(1) example).

## 2. Graphical VARs



# Graphical Model

Graphical models can be informally defined as statistical models represented in the form of a graph, where the nodes (vertices) represents the variables and the presence of an edge between a pair of vertices means that the variables are in some sense related.

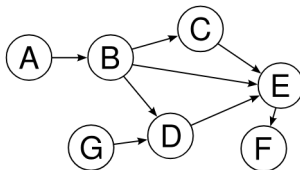
Some advantages of a graphical model are:

- ▶ represents graphically the logical implication of the relationships
- ▶ suitable representation of the causal relationships using directed edges
- ▶ clarity of interpretation when analyzing complex interactions.

## DAG Model

In cross-sectional data, the graph usually represents the conditional independence structure of the system and (by symmetry) the edges are undirected. In time-series data, the time dimension of a process makes it more feasible to consider directed flow.

A Directed Acyclic Graph (DAG) is a collection  $G = \{V, E\}$ , where  $V$  is the set of vertices and  $E$  is the set of edges.



## Connecting DAG and VAR

Given the VAR model in (1), there is a one-to-one relationship between the  $\Pi_i$  matrices and DAGs, that is

$$x_{a,t-i} \rightarrow x_{b,t} \iff \Pi_i(a,b) \neq 0$$

Where  $x_{a,t-i} \rightarrow x_{b,t}$  means that  $x_{a,t-i}$  “causes” somehow  $x_{b,t}$

Hence, directed edges represent the Granger causality relations.

## Characterization of Graphical VARs

For a Gaussian VAR process  $x_t$  with a Granger causality graph  $G(V, E)$ , the following two conditions hold:

1.  $(a, b) \notin E_1 \iff \Pi_i(a, b) = 0, \quad \forall i = 1, \dots, k$
2.  $(a, b) \notin E_2 \iff \Sigma^{-1}(a, b) = 0$

where  $E_1$  and  $E_2$  are the sets of directed and undirected edges, and  $\Pi_i(a, b)$  is the  $(a, b)$ th element of  $\Pi_i$ .

A Gaussian VAR process satisfying these conditions is said to belong to a graphical Granger causal VAR model,  $VAR(G, k)$ .

## 3. Bayesian Graphical VARs

# The $VAR(G, k)$ Problem

The unknown quantities of a  $VAR(G, k)$  process are:

- ▶ the underlying Granger causality graph  $G$
- ▶ the number of lags  $k$
- ▶ the modelling parameters  $\Pi_1, \dots, \Pi_k$  and  $\Sigma$  (from now  $\theta$ )

Inference on  $G$  and  $k$  is essentially a model determination problem and the authors proposed a solution within a Bayesian framework.

## The Posterior of $VAR(G, k)$

Following a Bayesian approach, the authors proposed a joint modeling of Granger causality and lag length. The joint posterior distribution of  $(G, k)$  conditional on the observed time series  $X$  can be written as

$$\pi(G, k|X) = \frac{m(G, k|X) p(G, k)}{\sum_{G \in \mathcal{G}} \sum_{k=0}^K m(G, k|X) p(G, k)}$$

where  $p(G, k)$  is the joint prior of  $G$  and  $k$ , and  $m(G, k|X)$  is the marginal likelihood of the observed time series  $X$ .

## The Prior of $G$ and $k$

The joint prior of  $G$  and  $k$  is over a discrete set and can be chosen in many ways. For instance, if there is no reason for favoring any particular graph in  $\mathcal{G}$  a priori, the following prior can be used

$$p(G, k) = \frac{p(k)}{|\mathcal{G}|}$$

where  $p(k)$  is some discrete distribution over the integers  $k = 0, 1, \dots, K$ .



# The Marginal Likelihood

The marginal likelihood of the observed time series  $X$  and is given by

$$m(G, k|X) = \int L(X|\theta, G, k) p(\theta|G, k) d\theta$$

where  $L(X|\theta, G, k)$  is the usual likelihood function under model  $(G, k)$  with parameters  $\theta$ , and  $p(\theta|G, k)$  is the prior distribution of  $\theta$ .

# The Fractional Bayes Approach

For the  $VAR(G, k)$  family of models, a common choice for  $p(\theta|G, k)$  is the inverse Wishart distribution. However, this is an improper prior and is not directly usable for deriving the joint posterior of  $G$  and  $k$ .

A solution to this problem is to use a Fractional Bayes approach: first, a small part of the sample is sacrificed in updating the improper prior to a proper posterior, then use this posterior as a new prior for the remaining observations.

The authors derived an approximated version of the exact Fractional Marginal Likelihood and this approach produced favorable results in model inference.

# Maximum Likelihood Estimation

- ▶ Discussion of maximum likelihood estimation for VAR( $G, k$ ) models
- ▶ Obtaining the maximum likelihood estimate  $\hat{R}$  of the covariance matrix  $R$

## 4. Conclusions

## Conclusions

- ▶ DAG approach can be used to study the Granger causality relationships in VAR models
- ▶ Since the  $VAR(G, k)$  usually has a very large number of parameters, a fractional Bayes approach can be used to approximate the joint posterior of the process
- ▶ Empirically valid results

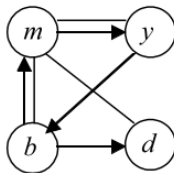


Figure 1: GC graph with the highest posterior probability for macroeconomic data.

# Bibliografy

*Corander, J. & Villani, M. (2006), 'A Bayesian Approach to Modelling Graphical Vector Autoregressions', Journal of Time Series Analysis 27(1), 141–156.*

Thank you!