A note on objective Bayes analysis for graphical vector autoregressive models

Lucia Paci and Guido Consonni

Abstract Vector Autoregressive (VAR) models are widely used to estimate and forecast multivariate time series. However, the large number of parameters of VAR models can lead to unstable inference and inaccurate forecasts, particularly with many variables. For this reason, restrictions supported by the data are usually required. We propose an objective Bayes approach based on graphical VAR models for learning contemporaneous dependencies as well as dynamic interactions among variables. We show that, if the covariance matrix at each time is Markov with respect to the same decomposable graph, then the likelihood of a graphical VAR can be factorized as an ordinary decomposable graphical model. Additionally, 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.

Key words: Bayesian model selection, decomposable graphical model, fractional Bayes factor, multivariate time series

1 Introduction

Vectore Autoregressive (VAR) models represent the workhorse models for estimating and forecasting multiple time series and widely applied in many fields such as macroeconomics, environmental sciences, neuroscience and genomic.VAR models are very flexible and allow to account for both contemporaneous dependencies among variables as well as their evolution over time. However, the large number of parameters of the VAR model usually leads to unstable inference and inaccurate

Lucia Paci
Università Cattolica del Sacro Cuore, Milano, e-mail: lucia.paci@unicatt.it

Guido Consonni
Università Cattolica del Sacro Cuore, Milano, e-mail: guido.consonni@unicatt.it

forecasts, particularly when dealing with many variables. This suggests to introduce parsimonious models.

Several solutions have been proposed in the literature. For instance, the Bayesian stochastic search variable selection approach, introduced by [9], has been extensively applied to select restrictions in VAR models. As an alternative, graphical modeling can be employed for the identification of the VAR model [6, 1, 2].

Following the latter track, we propose an objective Bayes approach for learning contemporaneous dependencies and dynamic interactions among variables under a graphical VAR model. We argue that, if the covariance matrix at each time is Markov with respect to the same decomposable graph, then the likelihood of a graphical VAR can be factorized as an ordinary decomposable graphical model. Additionally, using a fractional Bayes factor methodology, we are able to obtain the marginal likelihood in closed form and perform Bayes graphical model selection with limited computational burden.

2 Vector Autoregressive 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

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

where $\mathbf{B}_i$ are $(q \times q)$ matrices of coefficients or lag matrices, determining the dynamics of the system and ε_t is a $(q \times 1)$ dimensional white noise process, that is $\varepsilon_t \mid \Sigma \sim N_q(\mathbf{0}, \Sigma)$, independently over time. Clearly, the observations depend linearly on the previous k observation vectors, where k is assumed to be known. Exogenous variables can be added to the model, leading to straightforward modifications of the results shown here. For simplicity, the intercept is also omitted in the following.

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 $\mathbf{B} = \left(\mathbf{B}'_1, \ldots, \mathbf{B}'_k\right)'$ be the $(kq \times q)$ obtained by matrix stacking the coefficients. Hence, equation (1) can be written as

$$\mathbf{y}_t = \mathbf{B}' \mathbf{z}_t + \varepsilon_t. \tag{2}$$

For given initial values $\mathbf{Y}_0 = (\mathbf{y}'_0, \mathbf{y}'_{-1}, \ldots, \mathbf{y}'_{-k+1})'$, which we assume throughout to be available, the (conditional) likelihood of VAR(k) in (1) is written in the form

$$f(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \mathbf{B}, \Sigma) = \prod_{t=1}^{T} f(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{B}, \Sigma), \tag{3}$$

where the conditional distribution $f(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{B}, \Sigma)$ in (3) is the multivariate normal distribution $\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{B}, \Sigma \sim \mathscr{N}_q(\mathbf{B}' \mathbf{z}_t, \Sigma)$. 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., $\mathbf{Z} = (\mathbf{z}_t, \ldots, \mathbf{z}_T)'$. Equation (1) can be rewritten in matrix form as

$$\mathbf{Y} = \mathbf{Z}\mathbf{B} + \mathbf{E}, \tag{4}$$

where $\mathbf{E} = (\varepsilon_1, \ldots, \varepsilon_T)'$ is the $(T \times q)$ matrix of errors following a Matrix Normal distribution with zero mean, row identity matrix $\mathbf{I}_T$ and column (or cross) covariance Σ, that is, $\mathbf{E} \mid \Sigma \sim N_{T,q}(\mathbf{0}, \mathbf{I}_T, \Sigma)$. Therefore, we can write the likelihood of VAR(k) as

$$f(\mathbf{Y} \mid \mathbf{B}, \Sigma) = (2\pi)^{-\frac{Tq}{2}} |\Sigma|^{-\frac{T}{2}} \exp\left\{ -\frac{1}{2} \operatorname{tr}\left[\Sigma^{-1} \left(\left(\mathbf{B} - \hat{\mathbf{B}}\right)' \mathbf{Z}'\mathbf{Z} \left(\mathbf{B} - \hat{\mathbf{B}}\right) + \hat{\mathbf{E}}'\hat{\mathbf{E}} \right) \right] \right\} \tag{5}$$

where $\operatorname{tr}(\cdot)$ is the trace operator, $\hat{\mathbf{E}} = \mathbf{Y} - \mathbf{Z}\hat{\mathbf{B}}$ and $\hat{\mathbf{B}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ is the OLS estimator of the coefficient matrix, requiring $T \geq kq$. In other words, the likelihood of the VAR model can be expressed as the likelihood of a multivariate regression model where the predictors are the lagged variables.

3 Fractional Bayes inference

To complete the Bayesian specification of the VAR(k) model in (1), a prior distribution for model parameters $\mathbf{B}$ and Σ is required. In this work, we employ an objective approach for model selection based on the Fractional Bayes Factor (FBF), see [11]. The idea of the FBF is to train a noninformative prior using a small fractional power b of the likelihood function, converting the noninformative prior into a proper prior that is then used to compute the marginal likelihood using the complementary fraction power $1 - b$ of the likelihood.

Starting with a improper prior $p^D(\mathbf{B}, \Sigma) \propto |\Sigma|^{-(a_D+q+1)/2}$ and setting $b = T_0/T$, we can show that the fractional prior for VAR(k) is the Matrix Normal-Inverse Wishart $\mathscr{MNIW}(\underline{\mathbf{B}}, \underline{\mathbf{C}}, \nu, \underline{\mathbf{R}})$, where $\underline{\mathbf{B}} = \hat{\mathbf{B}}$, $\underline{\mathbf{C}} = T/T_0\, (\mathbf{Z}'\mathbf{Z})^{-1}$, $\nu = a_D - kq + T_0$ and $\underline{\mathbf{R}} = T_0/T\, \hat{\mathbf{E}}'\hat{\mathbf{E}}$. Hence, the fractional prior density is given by

$$p^F(\mathbf{B}, \Sigma) = K(\underline{\mathbf{C}}, \underline{\mathbf{R}}, \nu)\, |\Sigma|^{-\left(\frac{a_D+T_0+q+1}{2}\right)} \exp\left\{ -\frac{T_0}{2T} tr \left[\Sigma^{-1} \left(\left(\mathbf{B} - \hat{\mathbf{B}}\right)' (\mathbf{Z}'\mathbf{Z}) \left(\mathbf{B} - \hat{\mathbf{B}}\right) + \hat{\mathbf{E}}'\hat{\mathbf{E}} \right) \right] \right\}, \tag{6}$$

where

$$K(\underline{\mathbf{C}}, \underline{\mathbf{R}}, \nu) = (2\pi)^{-kq^2/2} |\underline{\mathbf{C}}|^{-q/2} |\underline{\mathbf{R}}/2|^{\nu/2} \Gamma_q(\nu/2)^{-1} \tag{7}$$

is the normalizing constant with $\Gamma_q(\nu/2)$ the q-dimensional gamma function evaluated at $\nu/2$. Prior (6) is proper under two conditions: i) $a_D + T_0 - kq + 1 > q$ so that $\nu > q - 1$; ii) $T - kq > q - 1$ so that $\hat{\mathbf{E}}'\hat{\mathbf{E}}$ is (almost surely) positive define. The first condition becomes $T_0 > q + kq - 1$ if $a_D = 0$, or $T_0 > kq$ if $a_D = q - 1$, i.e., a larger T_0 is needed in the case $a_D = 0$. Since b has to be minimal, [5] recommend to set $a_D = q-1$ and $T_0 = kq+1$ such that $\nu = q$. The second condition simplifies to $T > q+kq-1$ that looks realistic when dealing with long time series.

Combining prior (6) with likelihood (5) we obtain a posterior distribution that is a Matrix Normal-Inverse Wishart with updated parameters, i.e., $\mathscr{MNIW}(\overline{\mathbf{B}}, \overline{\mathbf{C}}, \overline{\nu}, \overline{\mathbf{R}})$, where $\overline{\mathbf{B}} = \hat{\mathbf{B}}$, $\overline{\mathbf{C}} = (\mathbf{Z}'\mathbf{Z})^{-1}$, $\overline{\nu} = a_D - kq + T$ and $\overline{\mathbf{R}} = \hat{\mathbf{E}}'\hat{\mathbf{E}}$.

Because of conjugacy, the fractional marginal likelihood of VAR(k) is available in closed form and can be obtained, up to a multiplicative factor, as the ratio of the prior and posterior normalizing constants, leading to

$$m^F(\mathbf{Y}) = \pi^{-\frac{(T-T_0)q}{2}} \left(\frac{T_0}{T}\right)^{\frac{(a_D+T_0)q}{2}} \left|\hat{\mathbf{E}}'\hat{\mathbf{E}}\right|^{-\frac{T-T_0}{2}} \frac{\Gamma_q\left(\left(a_D-kq+T\right)/2\right)}{\Gamma_q\left(\left(a_D-kq+T_0\right)/2\right)}. \tag{8}$$

Let $\mathbf{Y}_J$ be the $T \times |J|$ submatrix which contains selected columns of data matrix $\mathbf{Y}$ belonging to a subset J of cardinality $|J|$ of the full set of q variables. Using the result presented in [5], we can obtain the fractional marginal likelihood $m^F(\mathbf{Y}_J)$ based on the submatrix $\mathbf{Y}_J$ by making the following substitutions in (8):

$$q \to |J|, \qquad a_D \to a_D - |\bar{J}|, \qquad \hat{\mathbf{E}} \to \hat{\mathbf{E}}_J = \mathbf{Y}_J - \mathbf{Z}\hat{\mathbf{B}}_J, \tag{9}$$

where $\bar{J}$ denotes the complementary set of J and $\hat{\mathbf{B}}_J$ is the $kq \times |J|$ submatrix of $\hat{\mathbf{B}}$ whose column contain the OLS estimates of the regression coefficients for the selected responses. To ensure positive definiteness of $\hat{\mathbf{E}}_J'\hat{\mathbf{E}}_J$, the condition $|J| < T - kq + 1$ must be satisfied, when setting $a_D = q-1$ and $T_0 = kq+1$.

4 Graphical VAR

[7] introduced the class of time series chain graphs (TSCG). More specifically, let $Y = \{Y_t(a), t \in \mathbb{Z}, a = 1, \ldots, q\}$ be a q-variate stationary stochastic process and $V = \{1, 2, \ldots, q\}$ be the set of indexes. Let $G = (V_{TS}, E)$, be a graph with $V_{TS} = V \times \mathbb{Z}$ and edge set E, whose edges have at most lag k and which is invariant under translation. If $\mathbf{B}_i(b,a)$ is the (b,a)-element of matrix $\mathbf{B}_i$ in (1) and $\Omega(a,b)$ is the (a,b)-entry of precision matrix $\Omega = \Sigma^{-1}$, then the VAR model with the following constraints on the parameters

$$\begin{aligned} (a, t-i) \;\to\; (b,t) \;&\in E \;\Leftrightarrow\; \mathbf{B}_i(b,a) \neq 0 \qquad i = 1, \ldots, k \\ (a,t) \;—\; (b,t) \;&\in E \;\Leftrightarrow\; \Omega(a,b) \neq 0 \qquad t = 1, \ldots, T \end{aligned}$$

represents a VAR(k, G) model. Thus, a nonzero element in $\mathbf{B}$ corresponds to a directed edge in the graph reflecting the recursive structure of the time series while undirected edges specify contemporaneous interactions among variables, that is a covariance selection model. Hence, our goal is making inference on graph G.

4.1 Covariance selection model

Let $G^u = (V, E^u)$ be the undirected graph representing the contemporaneous dependencies at any time t, and assume that Σ is Markov with respect to G^u. For the graph-theory terminology in this section we refer the reader to [10]. Also, we confine our analysis to the class of decomposable graphs for all time points. Recall that G^u is decomposable when all cycles in G^u admit a chord, that is an edge joining two non-consecutive vertices of the cycle. Let $\mathscr{C}$ and $\mathscr{S}$ denote the set of cliques and separators of the graph G^u, respectively. Then, we can show that the likelihood of a graphical VAR(k,G) factorizes as following:

$$f(\mathbf{Y} \mid \mathbf{B}, \Sigma, G^u) = \frac{\prod_{C\in\mathscr{C}} f\left(\mathbf{Y}_C \mid \mathbf{B}_C, \Sigma_{CC}\right)}{\prod_{s\in\mathscr{S}} f\left(\mathbf{Y}_S \mid \mathbf{B}_S, \Sigma_{SS}\right)}, \tag{10}$$

where $\mathbf{B}_C$ and $\mathbf{B}_S$ are the $kq \times |C|$ and $kq \times |S|$ matrices whose columns contain the coefficients of the selected responses $\mathbf{Y}_C$ and $\mathbf{Y}_S$, respectively.

If $\mathbf{B}$ is unconstrained and Σ is in $M^+(G^u)$, the set of all symmetric positive-definite matrices having elements in Σ^{-1} set to zero for all $(a,b) \notin E^u$, then a natural noninformative prior on $(\mathbf{B}, \Sigma \mid G^u)$ is

$$p^D(\mathbf{B}, \Sigma \mid G^u) \propto \frac{\prod_{c\in\mathscr{C}} |\Sigma|^{-|C|}}{\prod_{s\in\mathscr{S}} |\Sigma|^{-|S|}}, \tag{11}$$

which is a limiting form of an Hyper-Inverse Wishart distribution [8]. Training prior (11) with a fraction $b = T_0/T$ of the likelihood, the fractional prior for a VAR(k,G) becomes a Matrix Normal Hyper-inverse Wishart distribution [3], $\mathscr{M}\mathscr{N}\mathscr{H}\mathscr{I}\mathscr{W}(\underline{\mathbf{B}}, \underline{\mathbf{C}}, \delta, \underline{\mathbf{R}})$, where $\underline{\mathbf{B}}$,$\underline{\mathbf{C}}$, $\underline{\mathbf{R}}$ are defined as above and $\delta = T_0 - kq$. Therefore, the prior density is

$$\begin{aligned} p^F(\mathbf{B}, \Sigma \mid G^u) &= H(\underline{\mathbf{C}}, \underline{\mathbf{R}}, \delta) \\ &\times \frac{\prod_{c\in\mathscr{C}} |\Sigma_{CC}|^{-(|C|+T_0/2)} \exp\left\{-\frac{T_0}{2T} \operatorname{tr}\left[\Sigma_{CC}\left(\left(\mathbf{B}_C - \hat{\mathbf{B}}_C\right)' \mathbf{Z}'\mathbf{Z}\left(\mathbf{B}_C - \hat{\mathbf{B}}_C\right) + \hat{\mathbf{E}}_C'\hat{\mathbf{E}}_C\right)\right]\right\}}{\prod_{s\in\mathscr{S}} |\Sigma_{SS}|^{-(|S|+T_0/2)} \exp\left\{-\frac{T_0}{2T} \operatorname{tr}\left[\Sigma_{SS}\left(\left(\mathbf{B}_S - \hat{\mathbf{B}}_S\right)' \mathbf{Z}'\mathbf{Z}\left(\mathbf{B}_S - \hat{\mathbf{B}}_S\right) + \hat{\mathbf{E}}_S'\hat{\mathbf{E}}_S\right)\right]\right\}}, \end{aligned} \tag{12}$$

where the normalizing constant is

$$H(\underline{\mathbf{C}}, \underline{\mathbf{R}}, \delta) = \frac{\prod_{c\in\mathscr{C}} (2\pi)^{-|C|kq/2} |\underline{\mathbf{C}}|^{-|C|/2} |\underline{\mathbf{R}}_{CC}/2|^{(\delta+|C|-1)/2} \Gamma_{|C|}\left((\delta+|C|-1)/2\right)^{-1}}{\prod_{s\in\mathscr{S}} (2\pi)^{-|S|kq/2} |\underline{\mathbf{C}}|^{-|S|/2} |\underline{\mathbf{R}}_{SS}/2|^{(\delta+|S|-1)/2} \Gamma_{|S|}\left((\delta+|S|-1)/2\right)^{-1}}.$$

In other words, a Markovian structure is assumed a priori for the lag coefficients that follows the structure of the likelihood. Thus, using prior (12), the fractional marginal likelihood has a closed form obtained, again, as the ratio of the prior and posterior normalizing constants. Equivalently, we can write

$$m^F(\mathbf{Y} \mid G^u) = \frac{\prod_{c\in\mathscr{C}} m^F(\mathbf{Y}_C)}{\prod_{s\in\mathscr{S}} m^F(\mathbf{Y}_S)}, \tag{13}$$

where, following [4], $m^F(\mathbf{Y}_C)$ and $m^F(\mathbf{Y}_S)$ are computed by means of (9) with $J = c$ and $J = s$, respectively, when setting $a_D = q-1$. Again, formula (9) provides a valid marginal likelihood if $T > |C| + kq - 1$, for each $c \in \mathscr{C}$.

As a final remark here, we stress that the joint likelihood of a graphical VAR(k,G) factorizes as an ordinary decomposable graph model, even though the decomposable structure is assumed conditionally at each time step. As a result, the closed form of the marginal likelihood allows to perform Bayes graphical model selection of VAR(k,G) models with easy computation.

A simulation study (not presented here for brevity) shows the capability of the approach to recover the underlying graph according to different scenarios (sample size, number of variables, number of lags, lag matrix). Future work will explore the possibility to build a joint prior model that simultaneously accounts for restrictions both on lag coefficients and covariance matrix.

References

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

[2] D. F. Ahelegbey, M. Billio, and R. Casarin. Bayesian graphical models for structural vector autoregressive processes. *J Appl Econom*, 31:357–386, 2016.

[3] C. M. Carvalho and M. West. Dynamic matrix-variate graphical models. *Bayesian Anal*, 2:69–97, 2007.

[4] G. Consonni and L. La Rocca. Objective Bayes factors for Gaussian directed acyclic graphical models. *Scand J Stat*, 39:743–756, 2012.

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

[6] J. Corander and M. Villani. A Bayesian approach to modelling graphical vector autoregressions. *J Time Ser Anal*, 27:141–156, 2006.

[7] R. Dahlhaus and M. Eichler. Causality and graphical models in time series analysis. In P. Green, N. Hjort, and S. Richardson, editors, *Highly structured stochastic systems*, pages 115–137. University Press, Oxford, 2003.

[8] A. P. Dawid and S. L. Lauritzen. Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann Stat*, 21:1272–1317, 1993.

[9] E. I. George, D. Sun, and S. Ni. Bayesian stochastic search for VAR model restrictions. *J Econometrics*, 142:553–580, 2008.

[10] S. L. Lauritzen. *Graphical models*. Oxford University Press, Oxford, 1996.

[11] A. O'Hagan. Fractional Bayes factors for model comparison. *J R Stat Soc B*, 57:99–138, 1995.