Sparse BGVAR models for Systemic Risk Analysis

Modelli BGVAR sparsi per l'analisi del rischio sistemico

Daniel Felix Ahelegbey and Monica Billio and Roberto Casarin

Abstract Measuring systemic risk requires the joint analysis of large sets of time series which calls for the use of high-dimensional models. In this context, inference and forecasting may suffer from lack of efficiency. In this paper we provide a solution to these problems based on a Bayesian graphical approach and on recently proposed prior distributions which induces sparsity in the graph structure. The application to the European stock market shows the effectiveness of the proposed methods in extracting the most central sectors during periods of high systemic risk level.

Abstract *La misurazione del rischio sistemico comporta l'analisi congiunta di un numero elevato di serie storiche e all'utilizzo di modelli di grandi dimensioni. In questo contesto l'inferenza e la previsione possono essere inefficienti. In questo lavoro viene proposta una soluzione a questi problemi fondata sull'utilizzo di modelli grafici bayesiani e sull'utilizzo di una distribuzione a priori che induce sparisità nella struttura del grafo. L'applicazione al mercato azionario europeo mostra l'efficacia del metodo proposto nell'estrarre i settori più rilevanti durante i periodi di elevato rischio sistemico.*

Key words: High-dimensional Models, Large Vector Autoregression, Systemic Risk, Sparse Graphical Models

Daniel Felix Ahelegbey
Department of Economics, University Ca' Foscari of Venice, Cannaregio 873, Fondamenta San Giobbe, 30121 Venezia, e-mail: dfkahey@unive.it

Monica Billio
Department of Economics, University Ca' Foscari of Venice, Cannaregio 873, Fondamenta San Giobbe, 30121 Venezia, e-mail: billio@unive.it

Roberto Casarin
Department of Economics, University Ca' Foscari of Venice, Cannaregio 873, Fondamenta San Giobbe, 30121 Venezia, e-mail: r.casarin@unive.it

1 Bayesian Graphical VAR (BGVAR) Models

Graphical modeling is a class of multivariate analysis that uses graphs to represent statistical models. The graph consists of nodes and edges, where nodes denote variables and edges show interactions ([7]). They can be represented by the pairs $(G,\theta) \in (\mathscr{G} \times \Theta)$, where G is a graph of relationships among variables, θ is the model parameters, $\mathscr{G}$ is the space of the graphs and Θ is the parameter space.

Let $\mathbf{x}_t$ be $n \times 1$ vector of observations at time t and assume $\mathbf{x}_t = (\mathbf{y}_t', \mathbf{z}_t')$, where $\mathbf{y}_t$, the $n_y \times 1$ vector of endogenous variables, and $\mathbf{z}_t$, a $n_z \times 1$, $n_z = (n - n_y)$ vector of exogenous predictors. In a VAR model with exogenous variables, the variables of interest $\mathbf{y}_t$, is determined by the equation

$$\mathbf{y}_t = \sum_{i=1}^{p} B_i \mathbf{x}_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim N_{n_y}(\mathbf{0}, \Sigma_\varepsilon) \tag{1}$$

$t = 1, \ldots, T$, independent and identically normal; p is the maximum lag order; B_i, $1 \le i \le p$ is $n_y \times n$ matrix of coefficients.

The temporal dependence structure in (1) can be expressed in a graphical framework with the relation $B_s = (G_s \circ \Phi_s)$, where G_s is a $n_y \times n$ binary adjacency matrix, Φ_s is a $n_y \times n$ coefficients matrix, and the operator $(\circ)$ is the element-by-element Hadamard's product. Based on this definition, we identify a one-to-one correspondence between B_s and Φ_s conditional on G_s, such that $B_{s,ij} = \Phi_{s,ij}$, if $G_{s,ij} = 1$; and $B_{s,ij} = 0$, if $G_{s,ij} = 0$ (see [1]). The above relationship can be presented in a stacked matrix form. Let $B = (G \circ \Phi)$, where $B = (B_1, \ldots, B_p)$, $G = (G_1, \ldots, G_p)$, $\Phi = (\Phi_1, \ldots, \Phi_p)$, and $\mathbf{w}_t = (\mathbf{x}_{t-1}', \ldots, \mathbf{x}_{t-p}')'$, $\mathbf{v}_t = (\mathbf{y}_t', \mathbf{w}_t')'$. Suppose the joint, $\mathbf{v}_t$, follows the distribution, $\mathbf{v}_t \sim N_{n_y+np}(\mathbf{0}, \Omega)$, then the joint distribution of the variables in $\mathbf{v}_t$ can be summarized with a graphical model, (G, θ), where $G \in \mathscr{G}$ consists of directed edges. See [2] for further details on the relationship between, Ω, Σ_ε and B. In this paper, we focus on modeling directed edges from elements in $\mathbf{w}_t$ to elements in $\mathbf{y}_t$, capturing the temporal dependence among the observed variables. More specifically, $G_{ij} = 0$, means the i-th element of $\mathbf{y}_t$ and j-th element of $\mathbf{w}_t$ are conditionally independent given the remaining variables in $\mathbf{v}_t$, and $G_{ij} = 1$ corresponds to a conditional dependence between the i-th and j-th elements of $\mathbf{y}_t$ and $\mathbf{w}_t$ respectively given the remaining variables in $\mathbf{v}_t$.

2 A Sparse BGVAR Model

The description of our graphical VAR for high dimensional multivariate time series is completed with the elicitation of the prior distributions for the lag order p, a sparsity prior on the graph, and the prior on G and Ω.

We allow for different lag orders for the different equations of the VAR model. We denote with p_i the lag order of the i-th equation. We assume for each p_i, $i = 1, \ldots, n_y$, a discrete uniform prior on the set $\{\underline{p}, \ldots, \bar{p}\}$.

We follow [2] to model the sparsity on the graph by introducing a prior on the maximal number of explanatory variables in a DAG model. We denote with $\bar{\eta}$, $0 \leq \bar{\eta} \leq 1$, the measure of the maximal density, i.e. the fraction of the predictors that explains the dependent variables. Thus the level of sparsity is given by $(1-\bar{\eta})$. We set the upper bound on the number of predictors for each equation (fan-in) to $f = \lfloor \bar{\eta} m_p \rfloor$, where $m_p = \min\{np, T-p\}$ and $\lfloor x \rfloor$ is the largest integer less than x. To allow for different levels of sparsity for the equations in the VAR model, we assume independent prior distributions for the maximal density in the different equations. We denote $\bar{\eta}_i$ the maximal density of the i-th equation and assume the prior on $\bar{\eta}_i$, given lag order p_i is beta distributed with parameters $a, b > 0$, $\bar{\eta}_i \sim \mathscr{B}e(a,b)$, on the interval $[0,1]$

$$P(\bar{\eta}_i) = \frac{1}{B(a,b)} \bar{\eta}_i^{a-1} (1-\bar{\eta}_i)^{b-1} \tag{2}$$

We define the graph prior for each equation in the VAR model conditional on the sparsity prior. We refer to the prior on the graph of each equation as the local graph prior, denoted by $P(\pi_i | p_i, \gamma, \bar{\eta}_i)$. Following [8], we consider the inclusion of predictors in each equation as exchangeable Bernoulli trials with prior probability

$$P(\pi_i | p_i, \gamma, \bar{\eta}_i) = \gamma^{|\pi_i|} (1-\gamma)^{np_i - |\pi_i|} \chi_{\{0,\ldots,f_i\}}(|\pi_i|) \tag{3}$$

where $\gamma \in (0,1)$ is the Bernoulli parameter, $|\pi_i|$ is the number of selected predictors out of np_i and $f_i = \lfloor \bar{\eta}_i m_p \rfloor$ is the fan-in restriction for the i-th equation and $\chi_A(x)$ is the indicator function which takes value 1 if $x \in A$ and zero otherwise. We assign to each variable inclusion a prior probability, $\gamma = 1/2$, which is equivalent to assigning the same prior probability to all models with predictors less than the fan-in f_i, i.e,

$$P(\pi_i | p_i, \bar{\eta}_i) = \frac{1}{2^{np_i}} \chi_{\{0,\ldots,f_i\}}(|\pi_i|) \tag{4}$$

Following [5], we assume a prior distribution on the unconstrained precision matrix, Ω, conditional on any complete DAG, G, for a given lag order p, is Wishart distributed. Based on the assumption that the conditional distribution of the dependent variables given the set of predictors, is described by equation (1), with parameters $\{B, \Sigma_\varepsilon\}$, we assume the prior distribution on $(B, \Sigma_\varepsilon | p, G)$ is an independent normal-Wishart. This is one of the prior distributions extensively applied in the Bayesian VAR literature. Specifically, we assumed the coefficients matrix B is independent and normally distributed, $B | p, G \sim N_{n_y np}(\underline{B}, \underline{V})$, and Σ_ε^{-1} is Wishart distributed, $\Sigma_\varepsilon^{-1} \sim \mathscr{W}(\underline{\nu}, \underline{S}/\underline{\nu})$. The prior expectation, $\underline{B} = \underline{0}_{n_y \times np}$, is a zero matrix, and the prior variance of the coefficient matrix, $\underline{V} = I_{n_y np}$, where I_k is a k-dimensional identity matrix. Also, the prior expectation of Σ_ε is $\frac{1}{\underline{\nu}} \underline{S}$ where $\underline{S} = \underline{\nu} I_{n_y}$ and $\underline{\nu} = n_y + 1$ is the degrees of freedom parameter.

3 Computational Details

In order to approximate the posterior distributions of the equations of interest, we consider the collapsed Gibbs sampler proposed in [2]:

1. Sample jointly, $p^{(j)}$, $\bar{\eta}^{(j)}$ and $\mathbf{G}_p^{(j)}$ from $P(p,\bar{\eta},\mathbf{G}_p|\mathscr{X})$.
2. Estimate $B^{(j)}$ and $\Sigma_\varepsilon^{(j)}$ directly from $P(B,\Sigma_\varepsilon|p^{(j)},\mathbf{G}_p^{(j)},\mathscr{X})$.

where $\mathscr{X}=(\mathbf{v}_1,\dots\mathbf{v}_T)$ is the set of observations. At the j-th iteration of the Gibbs, we consider for each equation $i=1,\dots,n_y$ and each lag order $p_i=\underline{p},\dots,\bar{p}$, a sample of $\bar{\eta}_i^{(j)}$ and $\mathbf{G}_{p,i}^{(j)}$ from $P(\bar{\eta}_i,\mathbf{G}_{p,i}|p_i,\mathscr{X})\propto P(\bar{\eta}_i|p_i)P(\pi_i|p_i,\bar{\eta}_i)P(\mathscr{X}|p_i,\mathbf{G}_{p,i})$. By conditioning on each possible lag order, we are able to apply standard MCMC algorithm. As regards the first step we use the following pseudo-marginal likelihood

$$P(\mathscr{X}|p,\mathbf{G}_p)\approx\prod_{i=1}^{n_y}P(\mathscr{X}|p_i,\mathbf{G}_{p,i}(y_i,\pi_i))=\prod_{i=1}^{n_y}\frac{P(\mathscr{X}^{(y_i,\pi_i)}|p_i,\mathbf{G}_{p,i})}{P(\mathscr{X}^{(\pi_i)}|p_i,\mathbf{G}_{p,i})}\tag{5}$$

where $\mathbf{G}_p(y_i,\pi_i)$ is the local graph of the i-th equation with y_i as dependent variable and π_i as the set of predictors; $\mathscr{X}^{(y_i,\pi_i)}$ is the sub-matrix of $\mathscr{X}$ consisting of y_i and π_i; and $\mathscr{X}^{(\pi_i)}$ is the sub-matrix of π_i. This approximation allows us to develop search algorithms to focus on local graph estimation. More specifically we apply the Markov chain Monte Carlo (MCMC) algorithm proposed in [2] and which use the approximated marginal pseudo-likelihood.

$$P(\mathscr{X}^{d_i}|p_i,\mathbf{G}_{p,i})=\pi^{\frac{-T_i|d_i|}{2}}\,\frac{|\bar{\Sigma}_{d_i}|^{-\frac{(\nu+T_i)}{2}}}{|\underline{\Sigma}_{d_i}|^{-\frac{\nu}{2}}}\prod_{i=1}^{|d_i|}\frac{\Gamma\left(\frac{\nu+T_i+1-i}{2}\right)}{\Gamma\left(\frac{\nu+1-i}{2}\right)}\tag{6}$$

where $d_i\in\{(y_i,\pi_i),\ \pi_i\}$, and $\mathscr{X}^{d_i}$ is a sub-matrix of $\mathscr{X}$ consisting of $|d_i|\times T_i$ realizations, where $|d_i|$ is the dimension of d_i, $T_i=T-p_i$, $|\underline{\Sigma}_{d_i}|$ and $|\bar{\Sigma}_{d_i}|$ are the determinants of the prior and posterior covariance matrices associated with d_i.

4 Systemic Risk Measures

Volatility connectedness also referred to as "fear connectedness" by [3] has received a lot of attention due to the evidence that volatilities track the fear of investors and reflect the extent to which markets evaluate arrival of information. They have become important for analyzing contagion and risk propagation in the financial system.

The dataset in this application is intra-day high-low price indexes of 119 institutions of the financial sector of Euro Stoxx 600 from November 1, 2005 to December 13, 2012 from Datastream. These are the largest financial institutions which consists of 41 Banks, 24 Financial Service institutions, 33 Insurance companies and 21 Real

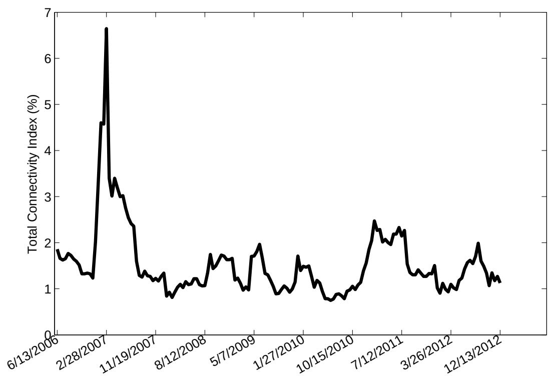

Fig. 1 Dynamics of total connectedness index over the period 2006-2012 obtained from a rolling estimation with windows size of 150-days.

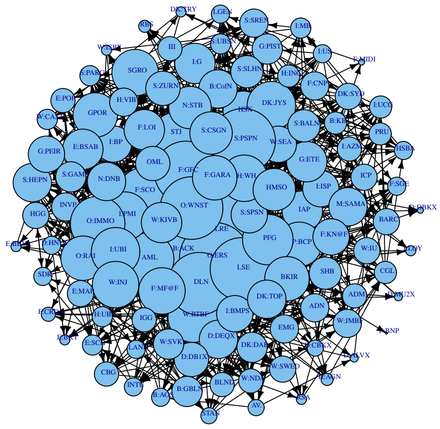

Fig. 2 Volatility network among the financial institutions for the period ending February 28, 2007. Size of the variable shows the degree of connectedness in the network.

Estates in the Euro-area covering countries like Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain.

We present the connectedness as the dependence pattern from a VAR(1) model with lag order based on testing the appropriate lag length using the BIC criteria and the available dataset (see [2]). We characterize the dynamics of the volatility connectedness (see Figure 1) using a rolling estimation with window size of 150-days over the sample period. From the figure, the highest total connectedness started early in the first quarter of 2007. We present in Figure 2, the graphical representation of the volatility network for the period ending February 28, 2007 which characterized the highest connectedness over the sample period. Table 1 report the top 10 institutions by eigenvector centrality from Figure 2. From the table, we observed that the top 10 central institutions as at the time was dominated by Banks, Financial Services and Real Estates. More specifically, we notice that prior to the global financial crisis between 2007-2009, the first quarter of 2007 shows evidence of some Banks, notably Credit Suisse Group in Switzerland, Raiffeisen Bank in Austria and Banco de Sabadell in Spain, acting as systemically important institutions in the "fear connectedness" expressed by market participants in the financial sector of the Euro-area.

Table 1 Volatility Network Centrality Ranking

Rank	Name	Tick	Type[a]	Eigen[b]	In-Deg[c]	Out-Deg[d]	T-Deg[e]
1	Credit Suisse Group	S:CSGN	BK	0.1885	19	2	21
2	Raiffeisen Bank Intl.	O:RAI	BK	0.1804	16	6	22
3	Banco de Sabadell	E:BSAB	BK	0.1775	13	7	20
4	London Stock Ex. Group	LSE	FS	0.1716	16	15	31
5	Immofinanz Group	O:IMMO	RE	0.1619	13	14	27
6	Zurich Insurance Group	S:ZURN	IN	0.1602	16	1	17
7	Kinnevik ‘B’	W:KIVB	FS	0.1596	13	7	20
8	Derwent London	DLN	RE	0.1595	13	18	31
9	Gecina	F:GFC	RE	0.1579	14	17	31
10	PSP Swiss Property AG	S:PSPN	RE	0.1519	14	18	32

Note: The table report the top 10 institutions by eigenvector centrality for the period ending February 28, 2007; [a] The financial super-sectors, BK (Banks), FS (Financial Services), RE (Real Estates), and IN (Insurance); [b] Eigenvector Centrality; [c] In-Degree; [d] Out-Degree; [e] Total Degree

5 Conclusion

We applied sparse Bayesian graphical VAR model to the analysis of systemic risk on the European stock market. We found evidence of increased number of linkages between institutions during the 2007-2009 financial crisis. Our sparse method allows us to extract a reduced number of systemically relevant institutions with respect non-sparse approaches.

References

1. Ahelegbey D. F., Billio M, Casarin R.: Bayesian Graphical Models for Structural Vector Autoregressive Processes. Journal of Applied Econometrics (forthcoming) (2015)
2. Alehegbey D. F., Billio, M. and Casarin, R.: Sparse Graphical Vector Autoregression: A Bayesian Approach, Working Paper N. 24/WP/2014, Dept. of Economics, University Ca’ Foscari of Venice (2014)
3. Diebold F., Yilmaz K.: On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. Journal of Econometrics 182(1): 119–134 (2014)
4. Donoho D.L.: High–Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension. Discrete Computational Geometry 35: 617–652 (2006)
5. Geiger D., Heckerman D.: Parameter Priors for Directed Acyclic Graphical Models and the Characterization of Several Probability Distributions. Annals of Statistics 30: 2002 (1999).
6. Giudici P., Castelo R.: Improving Markov chain Monte Carlo Model Search For Data Mining. Machine Learning 50: 127–158 (2003)
7. Lauritzen S. L., Wermuth N.: Graphical Models for Associations Between Variables, Some of Which Are Qualitative and Some Quantitative. Annals of Statistics 17: 31–57 (1989)
8. Scott J. G., Berger J. O.: Bayes and Empirical-Bayes Multiplicity Adjustment in the Variable-Selection Problem. The Annals of Statistics 38: 2587–2619 (2010)