Sparse Bayesian Graphical Vector Autoregression For Risk Analysis

Daniel Felix Ahelegbey* Monica Billio† Roberto Casarin‡

Abstract

This paper considers a sparsity approach for inference in large vector autoregressive (VAR) models. The approach is based on a Bayesian procedure and a graphical representation of VAR models. We discuss a Markov chain Monte Carlo algorithm for sparse graph selection, parameter estimation, and equation-specific lag selection. We show the efficiency of our algorithm on simulated data and illustrate the effectiveness of our approach in measuring contagion risk among financial institutions.

Key Words: High-dimensional models, Large VAR, Model Selection, Prior Distribution, Sparse Graphical Models, Risk Analysis

1. Introduction

High dimensional modeling has received considerable attention in recent years. It is well known that useful information is often scattered among many variables. Developing models that extract such information enhances a better understanding of the modern economic and financial systems. Many studies have shown that combining financial and macroeconomic variables in large VAR models produces better forecasts than standard approaches (see Banbura et al., 2010; Stock and Watson, 2012; Koop, 2013). Many applications involving large datasets of financial time series also report evidence of a highly interconnected system, where linkages play a fundamental role in the spread of risk (see Billio et al., 2012; Huang et al., 2012; Diebold and Yilmaz, 2014; Hautsch et al., 2015).

In this paper, we propose a model selection approach for large VAR model that is based on a Bayesian procedure and network representation. We consider graphical models that represent the causal relationships among the variables via directed edges (Pearl, 2000). Such models have been applied in time series analysis for estimating causal relationships in VAR models (see Corander and Villani, 2006; Demiralp and Hoover, 2003; Swanson and Granger, 1997) and for identifying restrictions in structural VAR (Ahelegbey et al., 2016). They have shown to be a promising tool for the analysis of financial interconnectedness and contagion (see Billio et al., 2012; Ahelegbey and Giudici, 2014; Diebold and Yilmaz, 2014). See also Ahelegbey (2016) for a review with focus on financial time series analysis.

As described in the following, we contribute to the literature in many ways. In a typical large VAR model, there are often too many parameters to estimate, compared to the available observations. The standard techniques discussed in the literature to overcome such difficulties are: parameter shrinkage methods (Doan et al., 1984; De Mol et al., 2008; Banbura et al., 2010); factor models (Forni et al., 2000; Bai and Ng, 2002; Stock and Watson, 2002); factor augmented VARs (Bernanke et al., 2005); sparse factor models (Bhattacharya and Dunson, 2011; Carvalho et al., 2012; Kaufmann and Schumacher, 2013); Bayesian model averaging (Koop and Potter, 2004; Jacobson and Karlsson, 2004); and sparse VAR model estimation (De Mol et al., 2008; Gefang, 2014; Basu and Michailidis, 2015; Davis et al., 2015; Kock and Callot, 2015; Medeiros and Mendes, 2016). In the literature, a key empirical finding is that many explanatory variables in large VAR models tend to be highly

*Department of Mathematics and Statistics, University of Boston, USA
†Department of Economics, University of Venice, Italy
‡Department of Economics, University of Venice, Italy

correlated and, hence, parsimony can be obtained by assuming that the dependent variables are driven by a small number of "common factors" or independent variables. We contribute to this stream of literature by considering a Bayesian graphical approach to estimating sparse VAR models.

We model sparsity in the temporal dependence among variables and uncertainty on the lags of a graphical VAR model. In standard applications, the common practice is to estimate an unrestricted model for a given range of lags and to apply a criterion to select a universal lag order. Such an approach may be inefficient for high-dimensional models because using an unrestricted VAR means including many irrelevant variables. Moreover, lag selection in high dimensional time-varying multivariate stochastic volatility (TV-MSV) models is still an open issue. In many VAR and TV-MSV models, the interaction between variable selection and equation-specific lag selection is not considered.

This paper extends the graphical VAR model, the inference approach and the posterior approximation algorithm in Ahelegbey et al. (2016). The approach adopted in this paper is based on a new hierarchical prior for the equation-specific lags, the graph sparsity parameter and the parameters of the graphical model. The idea is to impose a restriction on the number of predictors (fan-in restriction) to effectively estimate the model (see Friedman et al., 1998). Setting an a-priori hard fan-in might be too restrictive for large VAR applications. We therefore allow for different prior information levels on the maximal number of explanatory variables per equation in the model. We consider a random equation specific fan-in via a prior on the number of potential covariates for each equation. Since there is a duality between the prior distribution and the penalty term in the information criterion, we show that our prior leads to a modified Bayesian information criterion (BIC) which can be used to select the graph structure and the equation-specific lag order. See Friedman and Goldszmidt (1998) for learning local networks with standard BIC. See also Bogdan et al. (2004); Chen and Chen (2008); Foygel and Drton (2010) for extended BIC. We also propose a new Markov Chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution, the sparse graph structure, the lag order and the parameters of the VAR model.

We demonstrate the efficiency of our approach on simulated experiments and with an application to analyze the risk connectedness in the European financial market based on daily realized volatilities. We find evidence of higher level of systemic vulnerability during the global financial crisis of 2007-2008 than the European sovereign debt crisis period.

The paper proceeds as follows. We present graphical VAR models in Section 2. Section 3 presents the hierarchical prior distribution and posterior approximations. Section 4 outlines the Bayesian inference procedure. Section 5 presents the simulation experiment. Sections 6 discuss the empirical applications to financial data.

2. Graphical VAR Models

Graphical models is a class of multivariate analysis that uses graphs to represent statistical models (Lauritzen, 1996). They are formally represented by $(G, \theta) \in (\mathcal{G} \times \Theta)$, where G is a graph of relationships between variables, θ is the model parameter, $\mathcal{G}$ is the space of graphs and Θ is the parameter space. The graph, G, is defined in terms of a set of nodes, denoting variables, joined by a set of edges, depicting interactions.

In a VAR(p) model, an $n \times 1$ vector of time series, Y_t, is modeled by

$$Y_t = \sum_{s=1}^{p} B_s Y_{t-s} + \varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \Sigma_\varepsilon) \tag{1}$$

$t = 1, \ldots, T$, where ε_t is $n \times 1$ vector of errors; p is the maximum lag order; B_s, $1 \leq s \leq p$

is $n \times n$ matrix of coefficients. Equation (1) can be expressed as a graphical VAR model (see Ahelegbey et al., 2016), with a network that is directly related to the coefficient matrices.

Let $X_t = (Y'_{t-1}, \ldots, Y'_{t-p})'$ be a $np \times 1$ vector, $B = (B_1, \ldots, B_p)$, $G = (G_1, \ldots, G_p)$ and $\Phi = (\Phi_1, \ldots, \Phi_p)$ denote stacked representations of B_s, G_s and Φ_s, $1 \leq s \leq p$, respectively where each matrix is of dimension $n \times np$. In general, B_{ij} measures the impact of changes in the j-th element of X_t (i.e., $X_{t,j}$) on the i-th element of Y_t (i.e., $Y_{t,i}$). Thus, $B_{ij} = 0$ if there is no direct effect of $X_{t,j}$ on $Y_{t,i}$. We define $B = (G \circ \Phi)$, where G is a binary connectivity matrix, Φ is a coefficients matrix, and $(\circ)$ is the element-by-element Hadamard's product (i.e., $B_{ij} = G_{ij}\Phi_{ij}$). There is a one-to-one correspondence between B and Φ conditional on G, such that $B_{ij} = \Phi_{ij}$, if $G_{ij} = 1$; and $B_{ij} = 0$, if $G_{ij} = 0$.

Let $D_t = (Y'_t, X'_t)'$ be a $(n + np) \times 1$ vector, and $D_t \sim \mathcal{N}(0, \Omega^{-1})$, where $\Sigma = \Omega^{-1}$ is a $d \times d$ is the covariance matrix, where $d = n + np$. Then, the joint distribution of D_t can be summarized with a graphical model, (G, θ), where G is the network among the variables and Ω consists of the VAR parameters, $\{B, \Sigma_\varepsilon\}$, of model (1). The relationship between Ω and $\{B, \Sigma_\varepsilon\}$ is as follows. Assume $D_t = (Y'_t, X'_t)' \sim \mathcal{N}(0, \Omega^{-1})$, $X_t \sim \mathcal{N}(0, \Sigma_{xx})$ and $Y_t | X_t \sim \mathcal{N}(BX_t, \Sigma_\varepsilon)$, $\{B, \Sigma_\varepsilon\}$ can be obtained from $\Sigma = \Omega^{-1}$ by $B = \Sigma_{yx}\Sigma_{xx}^{-1}$, and $\Sigma_\varepsilon = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$, where Σ_{xy} is $np \times n$ covariances between X_t and Y_t, and Σ_{yy} is $n \times n$ covariances among Y_t. Given $\{B, \Sigma_\varepsilon\}$ and Σ_{xx}, $\Omega = \Sigma^{-1}$ can be obtained using the well-known Sherman-Morrison-Woodbury formula, (see Woodbury, 1950)

$$\Omega = \begin{pmatrix} \Sigma_\varepsilon^{-1} & -\Sigma_\varepsilon^{-1}B \\ -B'\Sigma_\varepsilon^{-1} & \Sigma_{xx}^{-1} + B'\Sigma_\varepsilon^{-1}B \end{pmatrix} \tag{2}$$

Based on the relationship between B and G defined above, equation (2) shows how Ω is related G through B. In this paper, we model the direct effects from elements in X_t on elements in Y_t. Thus, $G_{ij} = 0$, if $Y_{t,i}$ and $X_{t,j}$ are conditionally independent given the rest of the elements in D_t, and $G_{ij} = 1$ otherwise. Formulating (1) in a graphical form allows us to identify the causal dependencies in the model. Moreover, for large VAR models, the number of parameters to be determined is often too large compared to the number of observations, and this possibly leads to overfitting and loss of inference accuracy. The graphical form can be used to deal with these issues and to obtain more efficient estimators. Thus, we propose a sparse graphical approach to identify the set of covariates that explains the dependent variables. The quantities to estimate consist of the lag order, the sparse graph structure and the parameters. Estimating all these jointly is very challenging. We therefore adopt a Bayesian procedure that allows us to incorporate prior information where necessary, to include simulation based approximation techniques, and to apply model averaging.

3. Sparse Bayesian Graphical VAR Models

The specification of our sparse Bayesian graphical VAR model is completed with the choice of a hierarchical prior on the lag order p, the sparse graph structure G and the parameter Ω.

3.1 Lag Order Prior Distribution

Let p_i be the i-th equation lag order, and $\mathbb{I}_A(x)$, the indicator function, i.e., unity if $x \in A$ and zero otherwise. We assume for p_i, a discrete uniform prior on the set $\{\underline{p}, \ldots, \bar{p}\}$, i.e.,

$$Pr(p_i) = \frac{1}{(\bar{p} - \underline{p} + 1)} \mathbb{I}_{\{\underline{p}, \ldots, \bar{p}\}}(p_i), \qquad \text{for some} \quad 0 < \underline{p} < \bar{p} < \infty \tag{3}$$

3.2 Sparse Graph Prior Distribution

We model sparsity in the graph structure via fan-in, i.e., imposing a restriction on the number of regressors that can predict a dependent variable (see Friedman and Koller, 2003). For a large VAR(p) model with n time series each of length T, the number of coefficients is np_i for each equation, which is typically larger than T. We assume a maximum fan-in m where $m < np_i$. Specifically, we set $m = T - p_i$ for each equation.

Let f_i be the equation-specific fan-in, $0 \leq f_i \leq m$, and $\pi_i = \{j = 1, \ldots, np_i : G_{ij} = 1\}$ the set of indexes of explanatory variables in the i-th equation, with $|\pi_i|$ as the cardinality of π_i. Setting a-priori the value of f_i might be too restrictive in many applications. We introduce a prior distribution on f_i that allows for different prior degrees of information on the equation-specific fan-in. We denote with $\eta_i = |\pi_i|/np_i$, $0 \leq \eta_i \leq 1$, the fraction of covariates that explains the i-th dependent variable. We set $f_i = \min(\lfloor \eta_i np_i \rfloor, m)$, where $\lfloor A \rfloor$ is the largest integer less than A. Given $|\pi_i|$ and n, η_i decreases with p_i. Thus, the conditional prior on η_i given p_i, is assumed to be the beta distribution, $\eta_i | p_i \sim \mathcal{B}e(a_i, b_i)$, on $[0, 1]$, where a_i and b_i are functions of p_i. The parameters a_i and b_i can be chosen following the researcher prior belief on η_i. We consider an uninformative $\mathcal{B}e(1, 1)$ prior.

Following Scott and Berger (2010), we consider the inclusion of explanatory variables in each equation as exchangeable Bernoulli trials with prior probability

$$Pr(\pi_i | p_i, \eta_i, \gamma) \propto \gamma^{|\pi_i|}(1-\gamma)^{np_i - |\pi_i|} \, \mathbb{I}_{\{0,\ldots,f_i\}}(|\pi_i|) \tag{4}$$

where $\gamma \in (0, 1)$ is the Bernoulli parameter. 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 a number of explanatory variables less than the fan-in f_i.

3.3 Parameter Prior Distribution

Following the standard practice in graphical modeling (see Geiger and Heckerman, 2002), we assume the prior distribution on the unconstrained Ω, given p and a complete graph with no missing edges, G, is Wishart distributed with density function

$$Pr(\Omega | p, G) = \frac{1}{C_d(\nu, S_0)} |\Omega|^{\frac{(\nu-d-1)}{2}} \exp\Big\{ -\frac{1}{2} \langle \Omega, S_0 \rangle \Big\} \tag{5}$$

where $\langle A, B \rangle = tr(A'B)$ denotes the trace, $\nu > d+1$ is the degrees of freedom parameter, S_0 is a $d \times d$ prior scale matrix, and $C_d(\nu, S_0)$ is the normalizing constant:

$$C_d(\nu, S_0) = 2^{\frac{\nu d}{2}} |S_0|^{-\frac{\nu}{2}} \Gamma_d\Big(\frac{\nu}{2}\Big), \qquad \text{with} \qquad \Gamma_q(r) = \pi^{\frac{q(q-1)}{4}} \prod_{i=1}^{q} \Gamma\Big(r + \frac{1-i}{2}\Big) \tag{6}$$

where $\Gamma_q(\cdot)$ is the q-variate generalization of the gamma function, $\Gamma(\cdot)$. We further assume $(B, \Sigma_\varepsilon | p, G)$ is an independent normal-Wishart, which is a standard prior in the VAR literature. Specifically, we assume B is independent and normally distributed, $B | p, G \sim \mathcal{N}(\underline{B}, \underline{V})$, and Σ_ε^{-1} is Wishart distributed. The prior expectation, $\underline{B}$, is a null matrix, and the prior variance, $\underline{V}$, is the identity matrix. The prior expectation of Σ_ε is $\frac{1}{\underline{\nu}}\underline{S}$ where $\underline{S}$ is the $n \times n$ scale matrix and $\underline{\nu} > n + 1$ the degrees of freedom parameter.

4. Bayesian Inference

Let G_s be a $n \times n$ matrix of temporal relationship between variables at time $t - s$ and variables at time t. In order to emphasize the dependence of the graph size on the lag order,

we denote with $\vec{G}_p = G = (G_1, \ldots, G_p)$, the collection of all graphs such that $\vec{G}_p$ is of dimension $n \times np$. We define $\vec{G}_{p,i}$ as the local graph of the i-th equation which is the i-th row of $\vec{G}_p$. Let $\mathcal{D} = (D_1, \ldots, D_{T_l})$ be the collection of all observations, where $T_l = T - p$. The likelihood function $Pr(\mathcal{D}|p, \vec{G}_p, \Omega)$ is multivariate Gaussian with density

$$Pr(\mathcal{D}|p, \vec{G}_p, \Omega) = (2\pi)^{-\frac{dT_l}{2}} |\Omega|^{\frac{T_l}{2}} \exp\Big\{ -\frac{1}{2}\langle \Omega, \hat{S} \rangle \Big\} \tag{7}$$

where $\hat{S} = \sum_{t=1}^{T_l} D_t D_t'$ is the $d \times d$ sample sum of squares matrix. Ω can be integrated out analytically with respect to its prior distribution to obtain the marginal likelihood function

$$Pr(\mathcal{D}|p, \vec{G}_p) = \frac{C_d(\nu + T_l, S_0 + \hat{S})}{(2\pi)^{\frac{dT_l}{2}} C_d(\nu, S_0)} = \frac{C_d(\nu + T_l,\ (\nu + T_l)\bar{\Sigma})}{(2\pi)^{\frac{dT_l}{2}} C_d(\nu,\ \nu\underline{\Sigma})} \tag{8}$$

where $\underline{\Sigma} = \frac{1}{\nu} S_0$ and $\bar{\Sigma} = \frac{1}{\nu + T_l}(S_0 + \hat{S})$, with S_0 and $S_0 + \hat{S}$ as the prior and posterior sum of square matrices respectively. In VAR models, the errors are correlated across equations which makes the factorization of (8) into local marginal likelihoods quite problematic. However, following the large-sample approximation of Kass et al. (1988) and Chickering and Heckerman (1997), we treat the errors as unobserved variables that can be considered as independent when dealing with a large set of observed variables. We therefore approximate (8) with a pseudo-marginal likelihood given by the product of local densities

$$Pr(\mathcal{D}|p, \vec{G}_p) \approx \prod_{i=1}^{n} Pr(\mathcal{D}|p_i, \vec{G}_{p,i}(i, \pi_i)) = \prod_{i=1}^{n} \frac{Pr(\mathcal{D}^{(i,\pi_i)}|p_i, \vec{G}_{p,i})}{Pr(\mathcal{D}^{(\pi_i)}|p_i, \vec{G}_{p,i})} \tag{9}$$

where $\vec{G}_p(i, \pi_i)$ is the sub-graph of $\vec{G}_p$ with links from W_{t,π_i} to $Y_{t,i}$, i.e., $W_{t,\pi_i} \in W_t$ denotes the vector of predictors of $Y_{t,i}$, $\mathcal{D}^{(i,\pi_i)}$ and $\mathcal{D}^{(\pi_i)}$ are sub-matrices of $\mathcal{D}$ consisting of $(Y_{t,i}, W_{t,\pi_i})$ and W_{t,π_i} respectively. This allows us to focus on the local graph estimation. According to standard practice, since (8) factorizes into local scores, a search algorithm can be applied to estimate local graphs (Friedman and Goldszmidt, 1998; Bach and Jordan, 2004). For any $n \times n$ matrix A, and any scalar ρ, we have $|\rho A| = \rho^n |A|$, thus, the product elements on the left-hand side of (9) become

$$Pr(\mathcal{D}^{w_i}|p_i, \vec{G}_{p,i}) = \frac{\pi^{-\frac{1}{2}T_i|w_i|} \nu^{\frac{1}{2}\nu|w_i|}}{(\nu + T)^{\frac{1}{2}(\nu+T_i)|w_i|}} \frac{|\underline{\Sigma}_{w_i}|^{\frac{1}{2}\nu}}{|\bar{\Sigma}_{w_i}|^{\frac{1}{2}(\nu+T_i)}} \prod_{i=1}^{|w_i|} \frac{\Gamma\big(\frac{\nu+T_i+1-i}{2}\big)}{\Gamma\big(\frac{\nu+1-i}{2}\big)} \tag{10}$$

where $w_i \in (\{i\} \cup \pi_i)$, and $\mathcal{D}^{w_i}$ is a sub-matrix of $\mathcal{D}$ consisting of $|w_i| \times T_i$ observations, where $|w_i|$ is the dimension of w_i, $T_i = T - p_i$, $|\underline{\Sigma}_{w_i}|$ and $|\bar{\Sigma}_{w_i}|$ are the determinants of the prior and posterior covariance matrices associated with w_i.

4.1 Posterior Approximation

The standard approach to approximate the graph and parameters joint posterior distribution is to consider a collapsed Gibbs sampler. At the j-th iteration, the sampler consists of the following steps:

1. Sample jointly, $p^{(j)}$, $\eta^{(j)}$ and $\vec{G}_p^{(j)}$ from $Pr(p, \eta, \vec{G}_p|\mathcal{D})$.

2. Sample $B^{(j)}$ and $\Sigma_\varepsilon^{(j)}$ directly from $Pr(B, \Sigma_\varepsilon|p^{(j)}, \vec{G}_p^{(j)}, \mathcal{D})$.

Regarding the first step, since the lag order is unknown the dimension of the sample space changes with the number of lags at each Gibbs iteration. Thus, standard MCMC algorithms (e.g., Madigan and York, 1995) cannot be applied given they are designed for fixed dimension posterior distributions. The commonly discussed approach in the literature for this problem is the reversible jump (RJ) MCMC (Green, 1995). However, in a typical graph estimation problem, the graph space dimension increases super-exponentially with the number of variables (Chickering et al., 2004). Therefore, using the RJ algorithm will require a higher number of iterations to thoroughly visit the space of all candidates. In our model, the inferential difficulty increases due to the random fan-in restriction.

We propose to adopt the standard approach to lag selection and to sample the equation-specific sparse graph structure for each lag order. Thus, for the i-th equation and for each lag $p_i = \underline{p}, \ldots, \bar{p}$, we sample at the j-th iteration, $\eta_i^{(j)}$ from $Pr(\eta_i|p_i, \vec{G}_{p,i}, \mathcal{D})$ and $\vec{G}_{p,i}^{(j)}$ from $Pr(\vec{G}_{p,i}|p_i, \eta_i, \mathcal{D})$. By conditioning on each possible lag, the dimension of the model remains fixed, which allows us to apply standard MCMC algorithms to sample the graph. This enables us to avoid MCMC moves between parameter spaces of different dimensions. After J iterations, we apply the metric in Ahelegbey et al. (2016) to monitor convergence of the MCMC chain and to estimate $\hat{\vec{G}}_{p,i}$. Next, we select $(\hat{p}_i, \hat{G}_{\hat{p},i})$ which minimizes the criterion in (22). From $\hat{p}_i$ and $\hat{G}_{\hat{p},i}$, we select the relevant explanatory variables per equation and estimate B and Σ_ε.

4.2 Sparse Graph Selection

Our MCMC algorithm differs from the one described in Grzegorczyk and Husmeier (2011) in two aspects: the initialization and the inclusion of the random fan-in restriction. In MCMC search algorithms, the space exploration crucially depends on the choice of the starting point of the MCMC chain. A set of burn-in iterations is often used to obtain a good starting point. Brooks et al. (2011) showed that any sample representative of the equilibrium distribution is considerable. Following this view, we initialize the search by extracting explanatory variables with reliable information to improve predictions of the dependent variables. Let $\vec{G}_{p,i}$ denote the local graph of the i-th equation, $\mathbf{V}_{p,x}^i$, the vector of all possible covariates with lags up to p_i, $p_i \in \{\underline{p}, \ldots, \bar{p}\}$, and $\mathbf{V}_y$, the vector of dependent variables. We run the following steps:

1. Initialize the graph $\vec{G}_p$ as $n \times np$ null matrix, i.e., $\vec{G}_{p,i}$ is $1 \times np$ null vector.

2. For each equation $i = 1, \ldots, n$ and each covariate index $k = 1, \ldots, np_i$: test whether predictions of $y_i \in \mathbf{V}_y$ is improved by incorporating information from $x_k \in \mathbf{V}_{p,x}^i$, i.e., $Pr(y_i|x_k) > Pr(y_i)$. Following a Minnesota type of prior (see Doan et al., 1984), we assume recent lags of y_i are more reliable to influence current realizations. Based on this concept, we set $\vec{G}_p(i,k) = 1$, if $x_k = Y_{t-1,i}$. For $x_k \neq Y_{t-1,i}$, we compare the null hypothesis, $H_0 : Pr(\mathcal{D}|p_i, \vec{G}_p(i,\emptyset)) \geq Pr(\mathcal{D}|p_i, \vec{G}_p(i,k))$ against $H_1 : Pr(\mathcal{D}|p_i, \vec{G}_p(i,k)) > Pr(\mathcal{D}|p_i, \vec{G}_p(i,\emptyset))$, where $\emptyset$ is the empty set. If H_0 is rejected, we set $\vec{G}_p(i,k) = 1$ and retain x_k in $\mathbf{V}_{p,x}^i$. Otherwise, set $\vec{G}_p(i,k) = 0$ and remove x_k from $\mathbf{V}_{p,x}^i$.

3. We denote with $N_p(\pi_i)$ the set of indexes of the candidate covariates in the i-th equation which consists of elements retained in $\mathbf{V}_{p,x}^i$. We investigate the combination of candidates in $N_p(\pi_i)$ that produces the highest score networks.

In our experience, the above initialization provides a good starting point for the MCMC algorithm. At the j-th iteration, let $\vec{G}_{p,i}^{(j-1)}$ be the current local graph and $\pi_i^{(j-1)}$, the current

set of indexes of covariates in $\vec{G}_{p,i}^{(j-1)}$, then for each equation, the Gibbs iterates as follows:

1. Draw the sparsity parameter for the forward proposal probability, $\eta_i^{(*)}$ from a $\mathcal{B}e(a_i + |\pi_i^{(j-1)}|, b_i + np_i - |\pi_i^{(j-1)}|)$ and set the fan-in $f_i^{(*)} = \min(\lfloor \eta_i^{(*)} np \rfloor, m)$.

2. If $|\pi_i^{(j-1)}| < f_i^{(*)}$, randomly draw an index k from the set of candidate covariates, $N_p(\pi_i)$, and add/remove the link between y_i and x_k, i.e., set $\vec{G}_p^{(*)}(i,k) = 1 - \vec{G}_p^{(j-1)}(i,k)$. The forward proposal probability $Q(\vec{G}_{p,i}^{(*)} | \vec{G}_{p,i}^{(j-1)}, \eta_i^{(*)}) = 1/|N_p(\pi_i)|$. If $|\pi_i^{(j-1)}| \geq f_i^{(*)}$, then randomly draw an index k from $\pi_i^{(j-1)}$, and remove the link between y_i and x_k, i.e., $\vec{G}_p^{(*)}(i,k) = 0$. The forward proposal probability is $Q(\vec{G}_{p,i}^{(*)} | \vec{G}_{p,i}^{(j-1)}, \eta_i^{(*)}) = 1/|\pi_i^{(j-1)}|$.

3. We denote with $\pi_i^{(*)}$, the set of indexes of covariates in $\vec{G}_{p,i}^{(*)}$ which accounts for the changes made in step 2. Next, we draw the sparsity parameter for the reverse proposal probability, $\eta_i^{(**)}$ from a $\mathcal{B}e(a_i + |\pi_i^{(*)}|, b_i + np_i - |\pi_i^{(*)}|)$ and set $f_i^{(**)} = \min(\lfloor \eta_i^{(**)} np \rfloor, m)$.

4. If $|\pi_i^{(*)}| < f_i^{(**)}$, the reverse move will draw of an index from $N_p(\pi_i)$ to add or delete from $\vec{G}_{p,i}^{(*)}$. The reverse proposal probability is $Q(\vec{G}_{p,i}^{(j-1)} | \vec{G}_{p,i}^{(*)}, \eta_i^{(**)}) = 1/|N_p(\pi_i)|$. If $|\pi_i^{(*)}| \geq f_i^{(**)}$, the reverse will randomly draw an index from $\pi_i^{(*)}$ to delete from $\vec{G}_{p,i}^{(*)}$. Here the reverse proposal probability is $Q(\vec{G}_{p,i}^{(j-1)} | \vec{G}_{p,i}^{(*)}, \eta_i^{(**)}) = 1/|\pi_i^{(*)}|$.

5. From (4), the ratio of the local graph priors simplifies to 1 and the acceptance probability becomes $A(\vec{G}_{p,i}^{(*)}, \eta_i^{(*)} | \vec{G}_{p,i}^{(j-1)}, \eta_i^{(**)}) = \min\{1,\ R_A\}$ where

$$R_A = \frac{Pr(\mathcal{D}|p_i, \vec{G}_{p,i}^{(*)})}{Pr(\mathcal{D}|p_i, \vec{G}_{p,i}^{(j-1)})} \ \frac{Q(\vec{G}_{p,i}^{(j-1)} | \vec{G}_{p,i}^{(*)}, \eta_i^{(**)})}{Q(\vec{G}_{p,i}^{(*)} | \vec{G}_{p,i}^{(j-1)}, \eta_i^{(*)})} \tag{11}$$

   and $Pr(\mathcal{D}|p_i, \vec{G}_{p,i}) = Pr(\mathcal{D}|p_i, \vec{G}_p(i, \pi_i))$ can be computed from (9) and (10).

6. Sample $u \sim \mathcal{U}_{[0,1]}$ and if $u < \min\{1, R_A\}$, then accept changes made in the local graph and set $\vec{G}_{p,i}^{(j)} = \vec{G}_{p,i}^{(*)}$, otherwise set $\vec{G}_{p,i}^{(j)} = \vec{G}_{p,i}^{(j-1)}$.

4.3 Duality between Priors and Penalties

In graphical models, the most plausible graph is the one that minimizes the following criterion: (see Heckerman et al., 1995; Friedman and Goldszmidt, 1998)

$$-2 \log Pr(G|\mathcal{D}) \ \approx \ BIC(G) = -2 \log Pr(\mathcal{D}|\hat{\Omega}_G, G) + |E_G| \log T \tag{12}$$

where $\approx$ indicates the approximation obtained by replacing the likelihood with the pseudo-likelihood (see equation (9)), $\hat{\Omega}_G$ is the estimate of Ω under G, $Pr(\mathcal{D}|\hat{\Omega}_G, G)$ is the likelihood evaluated at $\hat{\Omega}_G$, and $|E_G|$ is the total number of edges in G. The BIC score decomposes as $BIC(G) = \sum_{i=1}^{n} BIC(i, \pi_i)$ (see Friedman and Goldszmidt, 1998), where

$$BIC(i, \pi_i) = -2 \log Pr(\mathcal{D}|\hat{\Omega}_{G,i}, G(i, \pi_i)) + |\hat{\pi}_i| \log T \tag{13}$$

is the local BIC score, $Pr(\mathcal{D}|\hat{\Omega}_{G,i}, G(i, \pi_i))$ is the local likelihood evaluated at $\hat{\Omega}_{G,i}$, and $|\hat{\pi}_i|$ is the number of selected variables. See Bach and Jordan (2004) for the AIC version.

Due to the duality between prior distributions and the penalization of likelihood functions, we define an information criterion for selecting $\hat{p}_i$ and $\hat{\pi}_i$ by solving the following

$$(\hat{p}_i, \hat{\pi}_i) = \arg\max_{p_i, \pi_i} \; Pr(p_i) Pr(\mathcal{D}|p_i, \vec{G}_{p,i}) \int Pr(\pi_i|p_i, \eta_i, \gamma) Pr(\eta_i|p_i) d\eta_i \tag{14}$$

Proposition 1. *Let $Pr(\eta_i|p_i)$ be the probability density function of the beta distribution $\mathcal{B}e(a_i, b_i)$ and let $Pr(\pi_i|p_i, \eta_i, \gamma)$ be as in (4), then setting $\gamma = 1/2$, η_i can be integrated out to obtain*

$$\begin{aligned} Pr(\pi_i|p_i) &= Pr(\pi_i|p_i, \gamma = 1/2) \\ &\propto \frac{1}{2^{np_i}} \sum_{j=0}^{m} \mathbb{I}_{\{0,\ldots,j\}}(|\pi_i|) \Big(I_{\frac{np_i - m + j}{np_i}}(a_i, b_i) - I_{\frac{j}{np_i}}(a_i, b_i) \Big) \end{aligned} \tag{15}$$

where $I_z(a_i, b_i) = \int_0^z (B(a_i, b_i))^{-1} \, \eta_i^{a_i - 1} (1 - \eta_i)^{b_i - 1} \, d\eta_i$, is the incomplete beta function (see Abramowitz and Stegun, 1964, p. 263).

Proof. From the prior distributions in (4), setting $\gamma = 1/2$, η_i can be marginalized as

$$Pr(\pi_i|p_i, \gamma = 1/2) \propto \frac{1}{2^{np_i}} \int_0^1 \mathbb{I}_{\{0,\ldots,f_i\}}(|\pi_i|) \frac{1}{B(a_i, b_i)} \frac{(\eta_i)^{a_i - 1}}{(1 - \eta_i)^{b_i - 1}} \, d\eta_i \tag{16}$$

where $f_i = \min(\lfloor \eta_i np \rfloor, m)$, $m = \min\{np_i, T - p_i\}$, $\mathbb{I}_{\{0,\ldots,f_i\}}(|\pi_i|)$ is such that

$$\mathbb{I}_{\{0,\ldots,f_i\}}(|\pi_i|) = \begin{cases} \mathbb{I}_{\{0\}}(|\pi_i|), & 0 \leq \eta_i < \frac{1}{np_i} \\ \quad \vdots & \vdots \\ \mathbb{I}_{\{0,\ldots,m-1\}}(|\pi_i|), & \frac{m-1}{np_i} \leq \eta_i < \frac{m}{np_i} \\ \mathbb{I}_{\{0,\ldots,m\}}(|\pi_i|), & \frac{m}{np_i} \leq \eta_i \leq 1 \end{cases} \tag{17}$$

$$\begin{aligned} Pr(\pi_i|p_i) \propto & \frac{1}{2^{np_i}} \left[\mathbb{I}_{\{0\}}(|\pi_i|) \int_0^{\frac{1}{np_i}} H(\eta_i) d\eta_i + \ldots + \mathbb{I}_{\{0,\ldots,m\}}(|\pi_i|) \int_{\frac{m}{np_i}}^1 H(\eta_i) \, d\eta_i \right] \\ = & \frac{1}{2^{np_i}} \sum_{j=0}^{m} \mathbb{I}_{\{0,\ldots,j\}}(|\pi_i|) \Big(I_{\frac{np_i - m + j}{np_i}}(a_i, b_i) - I_{\frac{j}{np_i}}(a_i, b_i) \Big) \end{aligned} \tag{18}$$

where $H(\eta_i) = (B(a_i, b_i))^{-1} \, \eta_i^{a_i - 1} (1 - \eta_i)^{b_i - 1}$, and $I_z(a_i, b_i) = \int_0^z H(\eta_i) d\eta_i$ is the incomplete beta function (Abramowitz and Stegun, 1964, p. 263). □

Corollary 4.1. *For a uniform prior on η_i, i.e., $a_i = b_i = 1$, equation (15) becomes*

$$Pr(\pi_i|p_i) \propto \frac{(np_i - m)}{np_i \, 2^{np_i}} \sum_{j=0}^{m} \mathbb{I}_{\{0,\ldots,j\}}(|\pi_i|) = \frac{1}{2^{np_i}} \Big(1 - \frac{m}{np_i} \Big) \Big(m + 1 - |\pi_i| \Big) \tag{19}$$

Proof. By assuming a uniform prior on η_i, i.e., $H(\eta_i) = 1$, the difference between the incomplete beta functions in (18) is $I_{\frac{np_i - m + j}{np_i}}(a_i, b_i) - I_{\frac{j}{np_i}}(a_i, b_i) = 1 - \frac{m}{np_i}$. Thus

$$Pr(\pi_i|p_i) \propto \frac{(np_i - m)}{np_i \, 2^{np_i}} \sum_{j=0}^{m} \mathbb{I}_{\{0,\ldots,j\}}(|\pi_i|) = \frac{1}{2^{np_i}} \Big(1 - \frac{m}{np_i} \Big) \Big(m + 1 - |\pi_i| \Big) \tag{20}$$

□

Proposition 2. *Let $Pr(\pi_i|p_i)$ be as in (19). If $\varphi(|\pi_i|)$ is an approximation asymptotically equivalent to $-\log Pr(\pi_i|p_i)$, where $0 \leq |\pi_i| \leq np_i$, then $\varphi(|\pi_i|)$ is a convex function given $p_i > 0$ and $n > 0$.*

Proof. The function $\varphi(|\pi_i|)$ is convex if and only if $\varphi''(|\pi_i|) > 0$, $\forall|\pi_i|$. By defining $\varphi(|\pi_i|)$ as an approximation asymptotically equivalent to $-\log Pr(\pi_i|p_i)$, then

$$\varphi''(|\pi_i|) = \frac{1}{(m+1-|\pi_i|)^2} > 0 \tag{21}$$

□

From (19), it follows that $Pr(\pi_i|p_i) < \frac{1}{2^{np_i}}, \forall|\pi_i| \leq m$. Given that $Pr(p_i)$ is constant according to (3) and (13), we define a modified BIC for local graph and lag selection as

$$BIC(p_i, i, \pi_i) = -2\log Pr(\mathcal{D}|p_i, \hat{\Omega}_{G,i}, \vec{G}_p(i, \pi_i)) + |\hat{\pi}_i|\log T + 2np_i \log 2 \tag{22}$$

Following Chib and Greenberg (1995), we use the estimated local graph to evaluate the score and to select the lag order. Selecting the local graph and the lag order for each equation may produce asymmetric lags for the different equations. Closely related to our criterion is the extended BIC by Bogdan et al. (2004); Chen and Chen (2008); Foygel and Drton (2010). A significant difference between our BIC and the ones discussed in the above papers is that our additional penalty term depends on the number of potential covariates (np_i) per equation and not on the estimated number of explanatory variables ($|\hat{\pi}_i|$). Thus, in comparing sparse graphs with the same lags, but different configuration, our criterion differs from the standard BIC (13) by a constant term.

4.4 Model Estimation

We estimate B and Σ_ε by exploiting their relationship with Ω (see equation (2)). We assume an independent normal-Wishart and by conditioning on $\hat{G}_p$, we estimate the coefficients associated with the non-zero elements of the i-th equation graph $\hat{G}_{p,i}$. We define the selection matrix $E_i = (e_{j_1}, \ldots, e_{j_{|\pi_i|}})$, where E_i is of dimension $np_i \times |\pi_i|$, $j_k \in \pi_i$ is an element of the set of predictor indexes for the i-th equation, and e_k is the k-th element of the standard orthonormal basis of the space of real np_i-dimensional vectors. The posterior mean and variance of B are

$$\bar{B}_{G,i} = \bar{V}_{G,i}(\underline{V}^{-1}_{G,i}\underline{B}_{G,i} + \bar{\sigma}_i^{-2}W'_{G,i}Y_i), \qquad \bar{V}_{G,i} = (\underline{V}^{-1}_{G,i} + \bar{\sigma}_i^{-2}W'_{G,i}W_{G,i})^{-1} \tag{23}$$

with $W_{G,i} = WE_i$, $\underline{B}_{G,i} = \underline{B}_i E_i$, $V_{G,i} = E_i'\underline{V}_i E_i$, where $W_{G,i} \in W'$, is the set of selected predictors of the i-th equation; W' is stacked $W_1', \ldots, W_{T_l}'$, with dimension $T_l \times np_i$; Y is stacked $Y_1', \ldots, Y_{T_l}'$, with dimension $T_l \times n$; Y_i is the i-th column of Y; $\underline{B}_{G,i}$ and $\underline{V}_{G,i}$, are the prior mean and variance of B respectively; $\bar{\sigma}_i^2$, is the i-th diagonal of Σ_ε, and the posterior of Σ_ε^{-1} is Wishart distributed with scale matrix, $\bar{S} = \underline{S} + (Y' - \bar{B}W')'(Y - W\bar{B}')$, and degrees of freedom $\bar{\nu} = \underline{\nu} + T_l$. Here $\bar{B} = (\bar{B}_{G,1}, \ldots, \bar{B}_{G,n})$, is the stacked posterior mean of the coefficients, such that $\bar{B}_{ij} \neq 0$, if $\hat{G}_{p,ij} = 1$; and $\bar{B}_{ij} = 0$, otherwise.

5. Simulation Study

We investigate the effectiveness of our approach through simulated experiments. The data generating process (DGP) of the experiment is a VAR with exogenous variables given by

$$Y_t = BX_{t-1} + \varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma_\varepsilon) \tag{24}$$

$t = 1, \ldots, T$, where Σ_ε, is a full matrix drawn from an inverse-Wishart distribution with $n_1 + 1$ degrees of freedom and an identity scale matrix, I_{n_1}, B is $n_1 \times n$ coefficient matrix, Y_t and X_t are is a $n_1 \times 1$ and $n \times 1$ respectively. To analyze different sparsity levels, we generate the coefficients matrix such that the number of non-zero coefficients for each equation is drawn from a uniform on $\{0, \ldots, 40\}$. We set $n_1 = 10$ and $n = 100$. We replicate the simulation and estimation exercise 100 times. For each replication, we generate a sample size $T = 60$ and use $T_0 = 50$ samples to estimate the model and 10 samples for out-sample forecast analysis.

We compare our sparse Bayesian graphical VAR (SBGVAR), against the standard graphical model (referred to as Bayesian graphical VAR - BGVAR), and the standard Lasso-type methods, i.e., LASSO (Tibshirani, 1996) and Elastic-net (ENET; Zou and Hastie, 2005). We evaluate the efficiency of the methods in terms of graph estimation accuracy and predictive performances.

We set $\underline{p} = 1$ and $\bar{p} = 4$ and implement a parallel estimation for the LASSO and ENET. We apply a five-fold cross validation to select the regularization parameter. We run 20,000 Gibbs iterations for the graph estimation and 2000 iterations for parameter estimations. The replications are conducted on a cluster multiprocessor system which consists of 4 nodes; each comprises four Xeon E5-4610 v2 2.3GHz CPUs, with 8 cores, 256GB ECC PC3-12800R RAM, Ethernet 10Gbit, 20TB hard disk system with Linux.

DGP Links = 201.5	LASSO	ENET	BGVAR	SBGVAR
Links	149.8400	181.1200	238.2000	50.0400
ACC	96.2675	95.8375	94.5135	**96.5935**
BIC_G	4233.6897	4315.3526	4447.7941	**4053.1426**
MSFE	0.3055	**0.3010**	0.5038	0.4027
LPS	-5.9148	**-5.7482**	-14.0147	-9.1776
AIC_M	311.5096	373.7364	504.4293	**118.4352**

Table 1: Graph and model performance. Note: $Links$ is Predicted number of edges; ACC (graph accuracy); BIC_G (graph BIC); LPS (log predictive score); and AIC_M (predictive AIC). Results are averaged over 100 replications. Boldface values indicate the best choice for each metric.

We proceed by comparing the effectiveness of the methods in estimating the graph of the DGP. Over all the simulation exercises, the DGP reported an average of 201.5 links. The table shows that, on average, the BGVAR predicted more links than in the DGP and the other competing methods. In terms of graph estimation accuracy, the SBGVAR outperforms the other methods. The network BIC score favors the SBGVAR which confirms the outcome of the graph accuracy metric. In conclusion, the SBGVAR network provides a better representation of the temporal dependence in the simulated dataset than the other methods. In terms of estimated model performance, the MSFE and the log predictive score favor the ENET, while the predictive AIC favors the SBGVAR. The results show that the sparsity restriction on the model selection enables us to better identify the significant set of the most influential explanatory variables and that the SBGVAR produces a more parsimonious model with competitive point and density forecasts.

6. Risk Connectedness in European Financial Markets

Risk connectedness among financial institutions and markets is currently recognized as an important source of vulnerability for the financial system, since it can magnify the impact of the triggers, as evidenced by the recent global financial crisis (Diebold and Yilmaz, 2014; Hautsch et al., 2015). Policy makers and regulators are increasingly focusing on interconnectedness in order to identify the vulnerability in the system and areas of risk concentrations in the build-up and after-math of the crises. See, e.g., Moghadam and Viñals

(2010); Viñals et al. (2012) for the role of interconnectedness in policy making.

We analyze the stability of the volatility connectedness in the financial sector of the European stock market. We construct daily realized volatilities using intraday high-low-close price indexes of 118 institutions of Euro Stoxx 600, obtained from Datastream, covering from January 3, 2005 to September 19, 2014. These are the largest euro area financial institutions consisting of 42 Banks, 31 Financial Services, 31 Insurance companies and 22 Real Estates. The dataset covers the following countries: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain.

Let H_t, L_t and C_t denote the high, low and closing price of a given stock on day t respectively. Following Garman and Klass (1980), we construct the realized volatility as

$$RV_t = 0.5\big(\log H_t - \log L_t\big)^2 - (2\log 2 - 1)\big(\log C_t - \log C_{t-1}\big)^2 \tag{25}$$

In stochastic volatility models, higher-order autoregressive terms are usually considered (e.g., see Asai, 2008). Lag selection in high dimensional time-varying multivariate stochastic volatility (TV-MSV) models is still a challenging issue. Our MCMC procedure is designed to deal with this problem. A closely related paper to our approach is Loddo et al. (2011), where a Bayesian stochastic search is applied to select regressors and non-zero covariances. However, in the above mentioned paper and many TV-MSV papers (e.g., see Asai et al., 2006, for a review), the interaction between variable selection and equation-specific lag selection is not considered. Our approach accounts for this interaction.

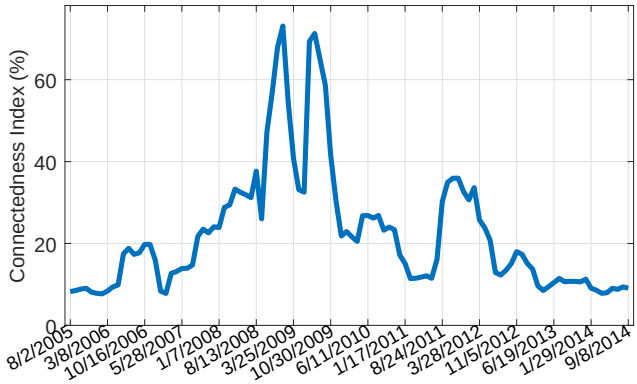
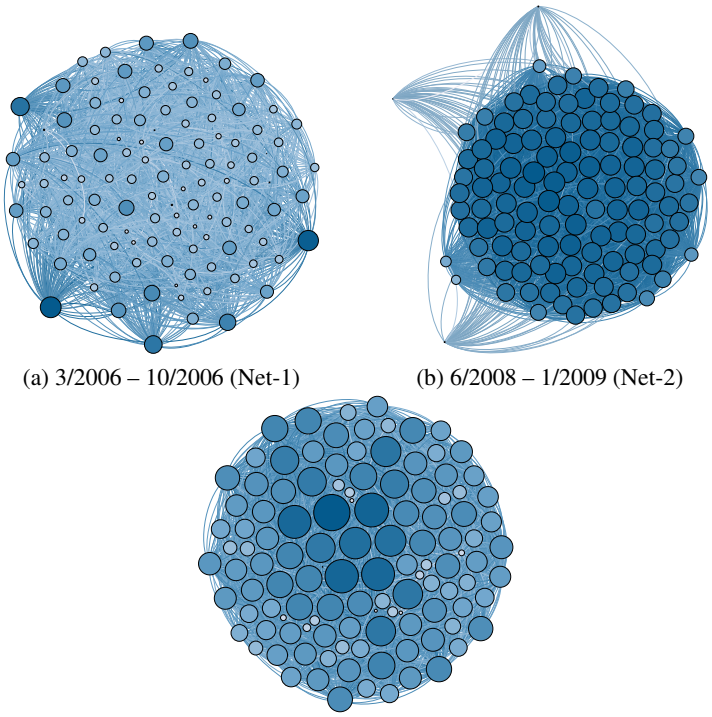


Figure 1: Volatility connectedness index obtained from a rolling window estimation of the SBG-VAR with a window size of 150 days, covering the period January, 2005 – September, 2014.

We define connectedness through the lagged linkages among institutions captured by the autoregressive coefficients matrix of a VAR model. We set the minimum and maximum lag order to $\underline{p} = 1$ and $\bar{p} = 5$ respectively. We characterize the total connectedness, i.e., the sum of the inter-linkages (see Figure 1), using a rolling estimation with window size of 150-days. The total connectedness index is formally used to summarize the degree of inter-connection for each window. This index attains a minimum value of zero when there are no linkages and a maximum value of 100 when all the institutions are fully connected.

From Figure 1, we notice a steady rise in the connectedness index with a small peak in 2006, rising through 2007 with a higher peak in 2008–2009. The index then declined steadily after 2009 and rose again with another peak in 2011–2012. The closeness between the peak in 2006 and the steady rise in 2007, coupled with the steady decline after 2009 only to record another rise in 2010–2011, somehow shows that volatility connectedness seems to occur in clusters. Our results support the findings of previous studies (see Tang et al., 2010) about clustering effects in financial crisis occurrence. Interestingly, these three peaks correspond to periods during the Iceland and Turkey crisis in 2006, the global financial

crisis in 2007–2009 and the European sovereign debt crisis in 2010–2013, respectively. The higher level of interconnectedness has been shown to indicate vulnerability of the system, where linkages play a significant role in financial risk propagation (see, Tang et al., 2010; Billio et al., 2012; Diebold and Yilmaz, 2014; Hautsch et al., 2015).



(a) 3/2006 – 10/2006 (Net-1) (b) 6/2008 – 1/2009 (Net-2)

(c) 4/2011 – 11/2011 (Net-3)

Figure 2: Networks of volatility connectedness in the European financial market during (2a) the Iceland and Turkey crisis in 2006, (2b) the global financial crisis in 2007–2009, and (2c) the European sovereign debt crisis in 2010–2013. The size of the nodes is proportional to their degree.

	Links	Density	Average Degree	Number of Communities	Average Path Length
Net-1 (2006)	2731	0.351	40.752	51	1.652
Net-2 (2008–2009)	10093	0.909	106.39	58	1.092
Net-3 (2011–2012)	4965	0.583	68.153	55	1.417

Table 2: The network statistics for the three graphs. The average path length represents the average graph-distance between all pairs of nodes. Connected nodes have graph distance 1.

Figure 2 shows the networks at the volatility connectedness peaks: (2a) the Iceland and Turkey crisis in 2006 (Net-1), (2b) the global financial crisis in 2007–2009 (Net-2), and (2c) the European sovereign debt crisis in 2010–20013 (Net-3). The topology of the three networks shows some differences with a few highly connected nodes in Net-1 and a large number of highly connected nodes in Net-2 and Net-3. These features indicate that different spreading mechanisms occurred in the three crises. A more detailed analysis of the network statistics confirms these differences.

Table 2 shows the network statistics for the three structures. As regards the number of communities or cohese groups of nodes (see Girvan and Newman, 2002), we used a resolution coefficients of 0.2 (see Blondel et al., 2008). We observe that Net-2 (2008–2009) recorded about four times the estimated links in Net-1 (2006) and more than twice that of Net-3 (2011–2012). Net-1, on the other hand, recorded the lowest density (0.351), number of communities (51) and average degree (40.752) followed by Net-3 and Net-2 has

the highest density (0.909), number of communities (58) and average degree (106.39).

We detect an increase in the number of communities during and after the financial crisis, thus, providing evidence in favor of the use of stochastic block models in contagion analysis (see Hałaj and Kok, 2013). The average path length represents the average graph-distance between all pair of nodes, where connected nodes have graph distance equals to 1. The higher the graph distance the longer time it takes for a default cascade to cause a systemic breakdown. Net-1 recorded the highest distance between nodes (average path length of 1.652), followed by Net-3 with 1.417 and Net-2 with 1.09.

In conclusion, the result shows that the 2008–2009 sub-period recorded the highest volatility (fear) connectedness over the entire sample period. Thus, the vulnerability of the system during the 2007–2009 crisis was much higher than that in 2006 and in 2010–2013. Hence, the severity in the impact of the global crisis (recession) through the large number of closely connected communities affected a much broader aspect of the European financial market, in a very short propagation time and involving several important financial institutions, than the Iceland and Turkey crises and the European sovereign debt crisis.

7. Discussion

We consider a sparsity approach to inference for large vector autoregressive (VAR) models. The approach is based on a Bayesian procedure and graphical representation of VAR models. We present a new prior which allows for the equation-specific lags, and graph sparsity. We also discuss an efficient Markov chain Monte Carlo algorithm for sparse graph selection, parameter estimation, and lag selection. We demonstrate the effectiveness of our approach through simulation experiments and empirical applications to finance. An empirical analysis of systemic risk based on realized volatilities in the European financial market shows a higher level of systemic vulnerability and a shorter propagation time during the global finance crisis than in the sovereign debt crisis period.

Acknowledgments

Authors' research is supported by funding from the European Union, Seventh Framework Programme FP7/2007-2013, "Systemic Risk Tomography", under grant agreement SYRTOSSH-2012-320270; the Global Risk Institute in Financial Services and the Institut Europlace de Finance, "Systemic Risk grant"; the Italian Ministry of Education; University and Research (MIUR) PRIN 2010-11 grant MISURA; and the US Air Force Office for Scientific Research (AFOSR), whose grant number is FA9550-12-1-0102. This research used the SCSCF multiprocessor cluster system at University Ca' Foscari of Venice.

REFERENCES

Abramowitz M, Stegun IA. 1964. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Courier Dover Publications.

Ahelegbey DF. 2016. The Econometrics of Bayesian Graphical Models: A Review With Financial Application. *Journal of Network Theory in Finance* **2**: 1–33.

Ahelegbey DF, Billio M, Casarin R. 2016. Bayesian Graphical Models for Structural Vector Autoregressive Processes. *Journal of Applied Econometrics* **31**: 357–386.

Ahelegbey DF, Giudici P. 2014. Bayesian Selection of Systemic Risk Networks. *Advances in Econometrics: Bayesian Model Comparison* **34**: 117–153.

Asai M. 2008. Autoregressive Stochastic Volatility Models with Heavy-Tailed Distributions: A Comparison with Multifactor Volatility Models . *Journal of Empirical Finance* **15**: 332 – 341.

Asai M, McAleer M, Yu J. 2006. Multivariate Stochastic Volatility: A Review. *Econometric Reviews* **25**: 145–175.

Bach FR, Jordan MI. 2004. Learning Graphical Models for Stationary Time Series. *IEEE Transactions on Signal Processing* **52**: 2189–2199.

Bai J, Ng S. 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica* **70**: 191–221.

Banbura M, Giannone D, Reichlin L. 2010. Large Bayesian Vector Autoregressions. *Journal of Applied Econometrics* **25**: 71 – 92.

Bastian M, Heymann S, Jacomy M. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Third International AAAI Conference on Weblogs and Social Media*, volume 8. 361–362.

Basu S, Michailidis G. 2015. Regularized Estimation in Sparse High-dimensional Time Series Models. *The Annals of Statistics* **43**: 1535–1567.

Bernanke B, Boivin J, Eliasz PS. 2005. Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics* **120**: 387–422.

Bhattacharya A, Dunson DB. 2011. Sparse Bayesian Infinite Factor Models. *Biometrika* **98**: 291–306.

Billio M, Getmansky M, Lo AW, Pelizzon L. 2012. Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors. *Journal of Financial Economics* **104**: 535 – 559.

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. 2008. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**: 1–12.

Bogdan M, Ghosh JK, Doerge R. 2004. Modifying the Schwarz Bayesian Information Criterion to Locate Multiple Interacting Quantitative Trait Loci. *Genetics* **167**: 989–999.

Brooks S, Gelman A, Jones G, Meng XL. 2011. *Handbook of Markov Chain Monte Carlo*. CRC press.

Carvalho CM, Chang J, Lucas JE, Nevins JR, Wang Q, West M. 2012. High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics. *Journal of the American Statistical Association* **103**: 1438–1456.

Chen J, Chen Z. 2008. Extended Bayesian Information Criteria for Model Selection with Large Model Spaces. *Biometrika* **95**: 759–771.

Chib S, Greenberg E. 1995. Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-varying Parameter Models. *Journal of Econometrics* **68**: 339–360.

Chickering DM, Heckerman D. 1997. Efficient Approximations for the Marginal Likelihood of Bayesian Networks with Hidden Variables. *Machine Learning* **29**: 181–212.

Chickering DM, Heckerman D, Meek C. 2004. Large-Sample Learning of Bayesian Networks is NP-Hard. *Journal of Machine Learning Research* **5**: 1287–1330.

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

Davis RA, Zang P, Zheng T. 2015. Sparse Vector Autoregressive Modeling. *Journal of Computational and Graphical Statistics (Forthcoming)* .

De Mol C, Giannone D, Reichlin L. 2008. Forecasting Using a Large Number of Predictors: Is Bayesian Shrinkage a Valid Alternative to Principal Components? *Journal of Econometrics* **146**: 318–328.

Demiralp S, Hoover KD. 2003. Searching for the Causal Structure of a Vector Autoregression. *Oxford Bulletin of Economics and Statistics* **65**: 745–767.

Diebold F, Yilmaz K. 2014. On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. *Journal of Econometrics* **182**: 119–134.

Doan T, Litterman R, Sims C. 1984. Forecasting and Conditional Projection Using Realistic Prior Distributions. *Econometric Reviews* **3**: 1–100.

Forni M, Hallin M, Lippi M, Reichlin L. 2000. The Generalized Dynamic-Factor Model: Identification and Estimation. *Review of Economics and Statistics* **82**: 540–554.

Foygel R, Drton M. 2010. Extended Bayesian Information Criteria for Gaussian Graphical Models. In *Advances in Neural Information Processing Systems*. 604–612.

Friedman N, Goldszmidt M. 1998. Learning Bayesian Networks with Local Structure. In *Learning in Graphical Models*. Springer, 421–459.

Friedman N, Koller D. 2003. Being Bayesian About Network Structure. *Journal of Machine Learning* **50**: 95–125.

Friedman N, Murphy K, Russell S. 1998. Learning the Structure of Dynamic Probabilistic Networks. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 139–147.

Garman MB, Klass MJ. 1980. On the Estimation of Security Price Volatilities from Historical Data. *Journal of Business* **53**: 67–78.

Gefang D. 2014. Bayesian Doubly Adaptive Elastic-net Lasso for VAR Shrinkage. *International Journal of Forecasting* **30**: 1–11.

Geiger D, Heckerman D. 2002. Parameter Priors for Directed Acyclic Graphical Models and the Characterization of Several Probability Distributions. *Annals of Statistics* **30**: 1412–1440.

Girvan M, Newman ME. 2002. Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences* **99**: 7821–7826.

Green PJ. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* **82**: 711–732.

Grzegorczyk M, Husmeier D. 2011. Non-homogeneous Dynamic Bayesian Networks for Continuous Data. *Machine Learning* **83**: 355–419.

Hałaj G, Kok C. 2013. Assessing Interbank contagion Using Simulated Networks. *Computational Management Science* **10**: 157–186.

Hautsch N, Schaumburg J, Schienle M. 2015. Financial Network Systemic Risk Contributions. *Review of Finance* **19**: 685–738.

Heckerman D, Geiger D, Chickering DM. 1995. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning* **20**: 197–243.

Huang X, Zhou H, Zhu H. 2012. Systemic Risk Contributions. *Journal of Financial Services Research* **42**: 55–83.

Jacobson T, Karlsson S. 2004. Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach. *Journal of Forecasting* **23**: 479–496.

Kass R, Tierney L, Kadane J. 1988. Asymptotics in Bayesian Computation. *Bayesian Statistics* **3**: 261–278.

Kaufmann S, Schumacher C. 2013. Bayesian Estimation of Sparse Dynamic Factor Models with Order-Independent Identification. Working paper, Studienzentrum Gerzensee.

Kock AB, Callot L. 2015. Oracle Inequalities for High Dimensional Vector Autoregressions. *Journal of Econometrics* **186**: 325–344.

Koop G. 2013. Forecasting with Medium and Large Bayesian VARs. *Journal of Applied Econometrics* **28**: 177 – 203.

Koop G, Potter S. 2004. Forecasting in Dynamic Factor Models Using Bayesian Model Averaging. *The Econometrics Journal* **7**: 550–565.

Lauritzen SL. 1996. *Graphical Models*. Oxford University Press, Oxford.

Loddo A, Ni S, Sun D. 2011. Selection of Multivariate Stochastic Volatility Models via Bayesian Stochastic Search. *Journal of Business & Economic Statistics* **29**: 342–355.

Madigan D, York J. 1995. Bayesian Graphical Models for Discrete Data. *International Statistical Review* **63**: 215–232.

Medeiros MC, Mendes EF. 2016. L1-regularization of High-dimensional Time-series Models with Non-Gaussian and Heteroskedastic Errors. *Journal of Econometrics* **191**: 255–271.

Moghadam R, Viñals J. 2010. Understanding Financial Interconnectedness. Mimeo, International Monetary Fund.

Pearl J. 2000. *Causality: Models, Reasoning and Inference*. Cambridge University Press, London, UK.

Scott JG, Berger JO. 2010. Bayes and Empirical-Bayes Multiplicity Adjustment in the Variable-Selection Problem. *The Annals of Statistics* **38**: 2587–2619.

Stock JH, Watson MW. 2002. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* **97**: 1167–1179.

Stock JH, Watson MW. 2012. Generalized Shrinkage Methods for Forecasting using Many Predictors. *Journal of Business & Economic Statistics* **30**: 481–493.

Swanson NR, Granger CWJ. 1997. Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions. *Journal of the American Statistical Association* **92**: 357–367.

Tang C, Dungey MM, Martin MV, González-Hermosillo MB, Fry MR. 2010. Are Financial Crises Alike? Working Paper 10–14, International Monetary Fund.

Tibshirani R. 1996. Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society. Series B* **58**: 267–288.

Viñals J, Tiwari S, Blanchard O. 2012. *The IMF'S Financial Surveillance Strategy*. International Monetary Fund.

Woodbury MA. 1950. Inverting Modified Matrices. *Memorandum Report* **42**: 106.

Zou H, Hastie T. 2005. Regularization and Variable Selection via the Elastic-Net. *Journal of the Royal Statistical Society. Series B* **67**: 301–320.