BAYESIAN GRAPHICAL MODELS FOR STRUCTURAL VECTOR AUTOREGRESSIVE PROCESSES

DANIEL FELIX AHELEGBEY[a]*, MONICA BILLIO[a] ROBERTO CASARIN[a]
[a] *Department of Economics, University of Venice, Italy*

SUMMARY

This paper proposes a Bayesian, graph-based approach to identification in vector autoregressive (VAR) models. In our Bayesian graphical VAR (BGVAR) model, the contemporaneous and temporal causal structures of the structural VAR model are represented by two different graphs. We also provide an efficient Markov chain Monte Carlo algorithm to estimate jointly the two causal structures and the parameters of the reduced-form VAR model. The BGVAR approach is shown to be quite effective in dealing with model identification and selection in multivariate time series of moderate dimension, as those considered in the economic literature. In the macroeconomic application the BGVAR identifies the relevant structural relationships among 20 US economic variables, thus providing a useful tool for policy analysis. The financial application contributes to the recent econometric literature on financial interconnectedness. The BGVAR approach provides evidence of a strong unidirectional linkage from financial to non-financial super-sectors during the 2007-2009 financial crisis and a strong bidirectional linkage between the two sectors during the 2010-2013 European sovereign debt crisis.

Keywords: Bayesian Graphical Models, Granger Causality, Markov Chain Monte Carlo, Structural VAR, Vector Autoregression.

1. INTRODUCTION

Since the seminal paper of Sims (1980), vector autoregressive (VAR) models have been widely used to estimate and forecast multivariate time series in macroeconomics. Despite the success of the VAR model, two of the challenges of econometricians are the problems of over-parametrization and identification of the VAR models. Various solutions to these problems have been discussed and criticized in many papers (e.g., Cooley and Leroy, 1985, Bernanke, 1986, King et al., 1991, Rubio-Ramirez et al., 2010, Doan et al., 1984). In particular, for structural VAR (SVAR) model identification, the standard approach relies on shocks for the dynamic analysis of the model through impulse response functions. To achieve this,

*Address: Department of Economics, University Ca' Foscari of Venice, Fondamenta San Giobbe 873, 30121, Venice, Italy. Corresponding author: Daniel Felix Ahelegbey, dfkahey@unive.it. Other contacts: billio@unive.it (Monica Billio); r.casarin@unive.it (Roberto Casarin).

This is a preprint of an article accepted for publication in *Journal of Applied Econometrics*.

some researchers impose structures provided by a specific economic model in which case the empirical results will be only as credible as the underlying theory (Kilian, 2013). Moreover, in many cases, there are not enough credible exclusion restrictions to achieve identification.

In this paper, we propose an identification approach for SVAR models based on a graph representation of the conditional independence among variables (see Pearl, 1988, Lauritzen and Wermuth, 1989, Whittaker, 1990, Wermuth and Lauritzen, 1990). The graph inference discussed in this paper is in the spirit of Corander and Villani (2006), but differs substantially from it. A first relevant difference is that our approach considers acyclic graphs, whereas in Corander and Villani (2006), the dependence structures are not acyclic and cannot be used to achieve identification. A second major difference is that we propose a joint inference of the graphs and the parameters, whereas they focus only on the causal structures. Furthermore, Corander and Villani (2006) apply a fractional Bayes approach, which is a questionable methodology for a variety of models (e.g., see Berger and Pericchi, 1998, de Santis and Spezzaferri, 1999). Finally, it is not at all clear in general, how one should define the factorization of the likelihood, or which fractions should be used for each component. Thus, in this paper we follow Madigan and York (1995) and Grzegorczyk and Husmeier (2008) and apply an efficient MCMC algorithm.

The SSVS of George et al. (2008) is, perhaps, the closest approach to the model inference discussed in this paper. The authors use two separate sets of restrictions for the contemporaneous and lagged interactions, as in our BGVAR model. However, the SSVS procedure and the BGVAR model differ substantially in the way the restrictions are introduced. The BGVAR handles the restrictions directly on the structural model, whereas the SSVS deals with the reduced-form model. This represents one of the most important contributions of our paper, since we are able to solve the identification problem of the SVAR using the natural interpretation of the graph structures and the acyclic constraints on the contemporaneous relationships. Consequently, the BGVAR model provides a convenient framework for policy analysis, as the contemporaneous graph reveals the presence and direction of the effects of policy actions. Moreover, the BGVAR model allows the researcher to learn about relationships among variables in the absence of indications from economic theory.

Another major difference between the BGVAR model and the SSVS regards the algo-

rithm used for the posterior approximation. The algorithm proposed in George et al. (2008) for the SSVS inference is a single-move Gibbs sampler, whereas our MCMC sampler for BGVAR is a collapsed and multi-move Gibbs sampler that has been proven to be more efficient both in the MCMC literature (e.g., see Liu, 1994, Roberts and Sahu, 1997) and in our simulation comparisons.

We provide some applications of our approach to well-known data sets studied in macroeconomics and finance. The macroeconomic application focuses on modeling and forecasting US-macroeconomic time series following the moderate-dimension VAR approach in Stock and Watson (2008) and Koop (2013). Our BGVAR approach provides a data-driven identification of the structural relationships among economic variables, thus offering a useful tool for policy analysis. The financial application focuses on the empirical investigation of the linkages among economic sectors in the Euro-zone. The use of graphical models in financial time series analysis have been investigated in Carvalho and West (2007) and Carvalho et al. (2007), and receives a lot of attention in the recent years (e.g., see Billio et al., 2012, Diebold and Yilmaz, 2014, 2015). In our application, the BGVAR produces a better representation of the linkages between the financial and non-financial super-sectors than the Granger-causal (GC) inference approach previously used in the literature.

The paper is organized as follows: Section 2 presents BGVAR models. Section 3 discusses the model inference scheme. Section 4 provides an illustration of BGVAR on synthetic datasets, and a comparison with alternative approaches. Section 5 and 6 present macroeconomic and financial applications, respectively.

2. BAYESIAN GRAPHICAL VECTOR AUTOREGRESSION

In a SVAR model, the dynamics of the variable of interest Y_t is

$$Y_t = B_0 Y_t + \sum_{i=1}^{p} B_i Y_{t-i} + \sum_{i=1}^{p} C_i Z_{t-i} + \varepsilon_t, \tag{1}$$

$t = 1, \ldots, T$, where Y_t is n_y vector of response variables, Z_t is n_z vector of predictor variables, ε_t is n_y vector of structural error terms, independent and identically normal, i.e., $\varepsilon_t \overset{iid}{\sim}$

$\mathcal{N}(0, \Sigma_\varepsilon)$, p is the maximum lag order; B_0 is $n_y \times n_y$ matrix of structural contemporaneous coefficients, with zero diagonals; B_i and C_i, with $1 \leq i \leq p$, are $n_y \times n_y$ and $n_y \times n_z$ matrices of structural coefficients, respectively.

The identification problem of SVAR is that, (1) is not directly estimable from which to derive the 'true' model parameters. This is because, a set of values exists for the coefficient matrices such that the likelihood function takes the same value at all points of this set. Thus, the true values for the coefficient matrices cannot be directly estimated from the data.

2.1 Over-Parametrized VAR Models

A general approach to model and forecast dynamics in multivariate time series is the reduced-form VAR model. Let $X_t = (Y_t, Z_t)'$ be a $(n = n_y + n_z)$-dimensional vector of observed variables at time t, and $B_i^* = (B_i, C_i), 1 \leq i \leq p$, the $n_y \times n$ matrices of unknown coefficients for the response and predictor variables. The reduced form of (1) can be expressed as

$$Y_t = \sum_{i=1}^{p} A_i X_{t-i} + A_0^{-1} \varepsilon_t \tag{2}$$

for $t = 1, \ldots, T$, where $A_0 = (\mathbf{I}_{n_y} - B_0)$ is a $n_y \times n_y$ matrix, $\mathbf{I}_{n_y}$ is the n_y-dimensional identity matrix, $A_i = A_0^{-1} B_i^*$, $1 \leq i \leq p$ are the reduced-form lag coefficient matrices such that A_i, $1 \leq i \leq p$, is of dimension $n_y \times n$ and $u_t = A_0^{-1} \varepsilon_t$ is a n_y-dimensional vector of reduced-form errors independent and identically normal, $u_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma_u)$. Alternatively, equation (2) can be rewritten in a matrix form as:

$$Y = X' A_+' + U \tag{3}$$

where Y is stacked $Y_{p+1}', \ldots, Y_{T-p}'$, such that Y is of dimension $(T-p) \times n_y$, X' is stacked $X_{p+1-s}', \ldots, X_{T-s}'$, $1 \leq s \leq p$, such that X' is of dimension $(T-p) \times np$, $A_+ = (A_1, A_2, \ldots, A_p)$ is of dimension $n_y \times np$ and U is stacked $u_{p+1}', \ldots, u_{T-p}'$, such that U is of dimension $(T-p) \times n_y$.

Estimating (3) with a full, lagged dependence structure across equations and a high lag order may result in a very large number of parameters relative to the number of data points

at hand. This phenomenon is referred to as over-parametrization and could lead to a loss of degrees of freedom that affects inference and the reliability of predictions.

To deal with an over-parametrized VAR model, several approaches have been discussed in the econometric literature. Early works by Doan et al. (1984) proposed a prior distribution (e.g., the Minnesota prior) to shrink the coefficients toward a random walk model. George et al. (2008) proposed the SSVS prior distribution to identify the relevant variables for predicting the response variables. The SSVS incorporates a latent variable, γ, which is an indicator matrix with elements interpreted as and indicator of whether to include or exclude a variable from the model.

Although the Minnesota prior distribution has proved efficient in handling over-parametrized VAR models, its effectiveness is limited (see e.g., McNees, 1986, Kadiyala and Karlsson, 1993, George et al., 2008). The SSVS has also proved efficient in selecting relevant variables in over-parameterized VAR models. However, the estimated SSVS coefficient matrix often consists of elements with values significantly different from zero, whereas the rest concentrate around zero but are not ignored. Parsimony is, therefore, not guaranteed.

2.2 Identification of Structural Dynamics

The reduced-form does not offer much information regarding the structural dynamics of the VAR model. The challenging problem of econometricians relates to learning about the structural dynamics from the reduced-form estimates. A standard approach to this problem relies on the role of shocks for the dynamics of the model. This approach is done through impulse response functions. Consider the covariance matrix of the errors related to (3)

$$\Sigma_u \;=\; E(A_0^{-1}\varepsilon_t\varepsilon_t'(A_0^{-1})') \;=\; A_0^{-1}\Sigma_\varepsilon(A_0^{-1})' \tag{4}$$

then, the identification problem relies on estimating A_0 and Σ_ε. The standard SVAR, however, assumes the covariance matrix of the structural errors is diagonal (normalized), $\Sigma_\varepsilon = \mathbf{I}_{n_y}$. This means that the main challenge lies in estimating A_0 or B_0. Thus, in the SVAR framework, B_0 is interpreted as a contemporaneous relationship among shocks rather than the observed variables. To have an identifiable model, some researchers impose

the structure provided by a specific economic model, although in that case, the empirical results will be only as credible as the underlying theory (Kilian (2013)). Moreover, in many cases, there are not enough credible exclusion restrictions to achieve identification. Various approaches to this problem have been discussed and criticized (e.g., Cooley and Leroy, 1985, Bernanke, 1986, King et al., 1991, Rubio-Ramirez et al., 2010, Kilian, 2013).

The inability to provide convincing and enough credible exclusion restrictions and to achieve identification has stimulated interest in alternative identification methods (Kilian, 2013). Swanson and Granger (1997) argued that contemporaneous correlation among errors is not appropriate for the impulse response studies. Stock and Watson (2001) also pointed out that the identification problem relates to differentiating between correlation and causation. As an alternative to imposing restrictions on B_0, Demiralp and Hoover (2003) showed that the application of graph-theoretic methods and stochastic search algorithms can reduce or even eliminate the need for prior information or to appeal to restrictions from an economic theory when identifying the causal order of structural models. In their SSVS prior, George et al. (2008) also incorporate a latent variable, ω, as an indicator matrix to learn about the contemporaneous correlations among shocks. Identification of the structural dynamics using shocks is subject to the specification of the reduced-form residuals or the contemporaneous covariance matrix of the residuals alone. Although reliance on shocks for the structural dynamics proves useful, it is limited in some ways. A possible limitation is based on the assumption that the VAR is correctly estimated. Even the variable selection of the SSVS does not totally ignore irrelevant variables. Therefore, misspecification of the model can affect the estimation of the reduced-form error covariance matrix that may affect the relationship identification. Moreover, policy actions are not necessarily shocks and, therefore, the idea of structural analysis from the assessment of reduced-form residuals may affect conclusions on dependence among the response variables.

2.3 Graphical Models and Structural VAR

Graphical models are statistical models that summarize the marginal and conditional independences among random variables by means of graphs (Brillinger, 1996). Specifically,

a graph is characterized by nodes and edges, where the nodes represent variables and the edges depict the nature of the interaction among variables. For instance, the relationship $P \to Q$ means the variable P causes the variable Q. The node P from which a directed edge originates is the parent (explanatory variable), and its end Q is the child (response variable).

One of the appealing features of the graphical approach to multivariate time series analysis is the possibility of giving a graphical representation of the logical implications of models as well as the conditional independence relationships. As an example, assume $P \to Q \to R$, then P and R would be probabilistically dependent in the absence of Q; but conditional on Q, they would be independent. Such kind of dependence is common in structural models, and a graphical approach provides a simple framework to represent and estimate these relationships. In this paper, we employ directed graphs, which present an unambiguous direction of causation among the variables. Using such class of models provides information on the structural dynamics among the variables by means of the directed edges.

Let $X_t = (X_t^1, X_t^2, \ldots, X_t^n)$, where X_t^i is a realization of the i-th variable at time t. Equation (1) can be represented in the form of a graphical model with a one-to-one correspondence between the coefficient matrices and a directed acyclic graph (DAG):

$$X_{t-s}^j \to X_t^i \iff B_{s,ij}^* \neq 0 \qquad 0 \leq s \leq p \tag{5}$$

where $B_0^* = B_0$, for $s = 0$, and $B_s^* = (B_s, C_s)$, for $1 \leq s \leq p$. By considering the structural dynamics as a causal dependence among variables, the relationship in (5) for $1 \leq s \leq p$ can be referred to as lagged (temporal) dependence, and as contemporaneous dependence for $s = 0$. Temporal dependence is based on the time flow and relies on the assumption that causes precede effects in time. Contemporaneous causal relationships are based on distinguishing between instantaneous causation from correlations.

Let $\mathcal{X} = (X_1, \ldots, X_T)$ be a time series of n variables and length T. The joint distribution of the variables in $\mathcal{X}$ can be described by a graphical model $(G, \theta) \in \{\mathcal{G}, \Theta\}$, where G is a graph representing the structural relationships, θ is a vector of structural parameters, $\mathcal{G}$ is the space of the graphs, and Θ is the parameter space. We represent $G \in \mathcal{G}$ as a DAG

composed of directed edges defining the contemporaneous and temporal dependence among the variables, $\theta \in \Theta$ is the structural parameters, $\theta \equiv \{\mu, \Sigma_x\} \equiv \{\mu, B^*, \Sigma_\varepsilon\}$, where μ is n vector of means of $X_t = (X_t^1, \ldots, X_t^n)$, $\forall t$; Σ_x is the covariance matrix of the observed time series that decomposes into $\{B^*, \Sigma_\varepsilon\}$, where $B^* = (B_0, B_1, \ldots, B_p, C_1, \ldots, C_p)$, is the matrix of structural coefficients and Σ_ε is the structural error covariance matrix. Without loss of generality, we assume the data is generated by a stationary process and that $\mu = 0$.

2.4 Bayesian Graphical VAR Models

Following the representation in equation (5), we define:

$$B_s^* = (G_s \circ \Phi_s), \qquad 0 \leq s \leq p \tag{6}$$

where for $s = 0$, $B_0^* = B_0$ is $n_y \times n_y$ structural coefficients of contemporaneous dependence, G_0 is $n_y \times n_y$, binary connectivity matrix and Φ_0 is a $n_y \times n_y$ matrix of coefficients. For $1 \leq s \leq p$, $B_s^* = (B_s, C_s)$ is a $n_y \times (n_y + n_z)$ matrix of structural coefficients of temporal dependence, G_s is a $n_y \times (n_y + n_z)$ binary connectivity matrix and Φ_s is a $n_y \times (n_y + n_z)$ matrix of coefficients. The operator $(\circ)$ is the element-by-element Hadamard's product (i.e., $B_{s,ij}^* = G_{s,ij}\ \Phi_{s,ij}$). We refer to G_0 as the connectivity matrix of contemporaneous dependence and to G_s, $1 \leq s \leq p$, as the matrix of the temporal dependence. Elements in G_s, $0 \leq s \leq p$, are indicators such that $G_{s,ij} = 1 \iff X_{t-s}^j \to X_t^i$ and 0 otherwise. Elements in Φ_s, $0 \leq s \leq p$, are structural regression coefficients, such that $\Phi_{s,ij} \in \mathbb{R}$ represents the value of the effect of X_{t-s}^j on X_t^i. There is a one-to-one correspondence between Φ_s and B_s^* conditionally on G_s:

$$B_{s,ij}^* = \begin{cases} \Phi_{s,ij} & if \qquad G_{s,ij} = 1 \\ 0 & if \qquad G_{s,ij} = 0 \end{cases} \tag{7}$$

Based on our notation in (6), equation (1) can be expressed as:

$$Y_t = (G_0 \circ \Phi_0) Y_t + \sum_{i=1}^{p} (G_i \circ \Phi_i) X_{t-1} + \varepsilon_t \tag{8}$$

where $(G_j \circ \Phi_j)$ are the graphical model structural coefficient matrices whose non-zero elements describe the value associated with the contemporaneous and temporal dependences, respectively. We assume the prior distribution for B^* is normal, i.e., $B^* \sim \mathcal{N}(\underline{B}^*, \underline{V}_B)$. Estimating (8) involves specification of the lag order, p, inference of the causal structure, $G = (G_0, G_1, \ldots, G_p)$, and the set of parameters, $\{B_0^*, B_1^*, \ldots, B_p^*, \Sigma_\varepsilon\}$ that are estimated from Σ_x. In this paper, specification of p is based on testing the appropriate lag order using the sample data and the BIC. Following Madigan and York (1995), we assume that the prior on G is uniform, $P(G) \propto 1$, and that given a complete graph, the prior on $\Omega_x = \Sigma_x^{-1}$ is a conjugate Wishart. See APPENDIX A for details.

The objective of this paper is twofold. First, we provide insight into the structural VAR dynamics by inferring G from the observed time series. This step is necessary to handle the identification issues of SVAR. To achieve this, we follow the conventional Bayesian graphical model approach of integrating the likelihood with respect to the unknown random parameters θ and obtaining the marginal likelihood function over the graphs $P(\mathcal{X}|G)$. See Heckerman and Geiger (1994) for details on the marginal likelihood of Gaussian graphical models. The second objective is to contribute to solving the problem of over-parameterization in VAR models. To achieve this, we incorporate the inferred structural relationships to select the relevant variables to estimate a reduced-form VAR. Following (2), (3), (6), and (8), the reduced-form parameters of the standard VAR model can be mapped to that of the Bayesian graphical model as follows:

$$A_0 = \mathbf{I}_{n_y} - (G_0 \circ \Phi_0), \qquad A_i = (\mathbf{I}_{n_y} - G_0 \circ \Phi_0)^{-1}(G_i \circ \Phi_i), \qquad i = 1, \ldots, p \tag{9}$$

where A_0 is $n_y \times n_y$ coefficient matrix conditional on G_0, u_t is n_y vector of reduced-form error terms, and A_+ is a stacked reduced-form coefficient matrix, B_+^* is a stacked form of $B_1^*, \ldots, B_p^*$. The connectivity matrix associated with B_+^* can be expressed as G_+, a stacked form of $G_1, \ldots, G_p$, and that of A_+ can be represented as $G_+^* = (\mathbf{I}_{n_y} - G_0)^{-1} G_+$, such that G_+ and G_+^* are of dimension $n_y \times np$. We notice that $\Sigma_u = (\mathbf{I}_{n_y} - G_0 \circ \Phi_0)^{-1} \Sigma_\varepsilon (\mathbf{I}_{n_y} - G_0 \circ \Phi_0)^{-1'}$. Following the general concept of the SVAR model, we assume the structural errors are a priori independent, which means Σ_ε is a diagonal matrix. By nor-

malizing Σ_ε to an identity matrix, the identification problem reduces to estimating B_0^* whose structure of dependence is given by G_0. The inference of G_0 from the observed time series offers insight into the contemporaneous dependence of the response variables. Based on the assumption that B_0^* follows a normal density, we can estimate the signs of the contemporaneous relationships from the partial correlations of the observed time series. The inference on the sign, together with the inferred contemporaneous structure, G_0, offers some insight into the presence and causal directions of policy actions on key variables of the system. We shall notice that following the Markov equivalence principle of contemporaneous directed graphs (see Andersson et al. (1997) two or more graphs with similar correlation structures, but different edge directions, may have the same marginal likelihood. As suggested in Andersson et al. (1997), the modeller should focus his attention on the class of essential graphs rather than DAGs. In this case, our graphical approach is not able to provide a unique solution to the SVAR identification problem and the researcher should choose between one of the graphs in the equivalent classes, by using other arguments, from economic theory, or sources of informations. In estimating A_+, non-zero (zero) entries in G_+^* can be interpreted as indicators of relevant (not-relevant) variables to be included (excluded) in (from) the model. These zero entries restrictions allow us to avoid estimating all the coefficients in the VAR model. It presents an automatic way to achieve dimensionality reduction and variable selection. An advantage of this approach to VAR inference is that by combining G_0 and G_+ to obtain G_+^*, we are able to identify relevant variables with a causal interpretation. The prior distributions and the posterior computation for the parameters of our reduced-form VAR are discussed in Section 3.

2.5 Statistical Inference on Graphical VAR Models

Statistical inference on graphical models can be a challenging goal as shown in the following. Let Y_t and Z_t be the n_y and n_z vectors of the response and predictor variables of the SVAR, respectively. The structural dynamics can be decomposed as contemporaneous and temporal dependences. For directed graphs, the number of possible structures of the temporal dependence, $F(p, n_y, n)$, is a function of p, the lag order, n_y, the number of response variables,

and $n = n_y + n_z$, the total number of explanatory variables, while the contemporaneous dependence, $H(n_y)$, is a recursive function of only n_y (Robinson, 1977):

$$F(p, n_y, n) = 2^{pn_y n}, \qquad H(n_y) = \sum_{i=1}^{n_y} (-1)^{i+1} \binom{n_y}{i} 2^{i(n_y - i)} \; H(n_y - i)$$

where $\binom{n_y}{i}$ is the binomial coefficient and $H(0) = 1, H(1) = 1$. Following Friedman et al. (1998), we represent the contemporaneous and temporal dependences separately because the learning processes of the two structures are distinct. Based on this assumption, the number of possible structures for a SVAR of order p, with n_y response variables, and n_z predictor variables, can be expressed as $G(p, n_y, n) = H(n_y)F(p, n_y, n)$. Figure 1 shows that

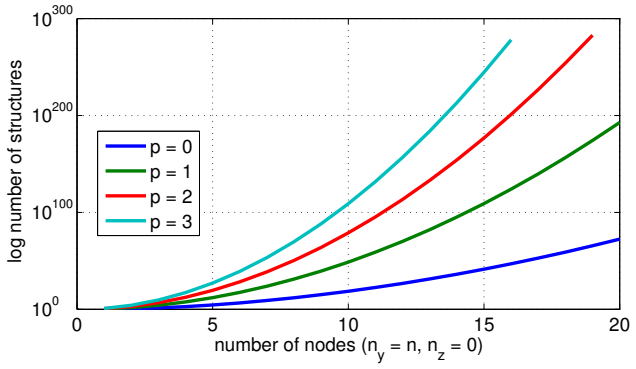


Figure 1: Logarithmic estimates of the number of possible structures for a SVAR model of order p, $0 \leq p \leq 3$, with the number of response variables n_y equal to the number of nodes n, $(0 \leq n \leq 20)$, and no predictors $(n_z = 0)$.

the number of possible structures increases super-exponentially with the number of nodes, $(n = n_y$ and $n_z = 0)$, and the lag order (different lines).

This challenge has been discussed extensively as a model determination problem (see Chickering et al., 2004, Corander and Villani, 2006). However, we follow the Bayesian paradigm of Madigan and York (1995), Giudici and Green (1999), and Dawid and Lauritzen (2001), that allows us to take into account structure and parameter uncertainty.

3. EFFICIENT MODEL INFERENCE SCHEME

Under the Bayesian framework of Geiger and Heckerman (1999), the structural parameters can be integrated out analytically to obtain a marginal likelihood function over graphs. This allows us to apply an efficient Gibbs sampling algorithm (e.g., Casella and Robert, 2004) to sample the graph structure and the model parameters in blocks (e.g., Roberts and Sahu, 1997). At the iteration t, the resulting collapsed Gibbs sampler (Liu, 1994) consists of the following steps:

1. Sample the structural relationships $G_0^{(t)}$ and $G_+^{(t)}$ from the full conditional distribution $P(G_0, G_+|\mathcal{X})$, by using a Metropolis-Hastings (MH) algorithm with random walk proposal distribution $Q(G_0^{(t)}, G_+^{(t)}|\mathcal{X}, G_0^{(t-1)}, G_+^{(t-1)})$

2. Sample the reduced-form parameters $A_+^{(t)}$ and $\Sigma_u^{(t)}$ directly from the full conditional distribution $P(A_+, \Sigma_u|G_0^{(t)}, G_+^{(t)}, \mathcal{X})$

Sampling the structural relationships, $(G_0,\ G_+)$, from the joint distribution is computationally intensive, since the number of possible structures, as shown in Figure 1, increases super-exponentially with the number of nodes and lags. In addition, the acyclic constraint on the contemporaneous structure requires a verification scheme, which is not required in the temporal structure and can negatively affect the mixing of the MCMC.

Following Friedman et al. (1998), we collapse further the Gibbs sampler, drawing the the temporal structure, G_+ and G_0, from their approximate marginal posterior distributions. Thus, the first step of the Gibbs sampler is:

1.1 Sample $G_0^{(t)}$ from $P(G_0|\mathcal{X})$ by a random walk MH

1.2 Sample $G_+^{(t)}$ from $P(G_+|\mathcal{X})$ by a random walk MH

We will refer to G_+ as the multivariate autoregressive (MAR) structure and to G_0 as the multivariate instantaneous (MIN) structure. To sample the graph structures, we use a modified and more efficient version of the MCMC algorithm proposed by (Madigan and York, 1995, Grzegorczyk et al., 2010). See APPENDIX B for further details of our MCMC scheme. In the following we describe the different steps of the Gibbs sampler.

3.1 Sampling the MAR Structure

The likelihood of the MAR structure is given by the probability density function of the normal distribution $\mathcal{N}(0, \Sigma_{x,+})$, where $\Sigma_{x,+}$ is the temporal covariance matrix. Specification of the maximum lag order, p, is based on the BIC. Based on the specification of p, we estimate $G_+ = (G_1, \ldots, G_p)$, a single structure of dimension $n_y \times np$, that comprises all the temporal structures stacked together.

The sampling approach is such that at each iteration, we randomly draw a candidate explanatory variable for each of the response variables and either add or delete an edge between them and account for potential interactions among the explanatory variables. Since edges flow only forward and not backward, edge addition or removal results in an acyclic graph. The probability of selecting a node is strictly positive for all nodes, therefore, this guarantees irreducibility, since it is possible to reach other configurations in finite time regardless of the present state. Furthermore, with a positive probability, the chain can remain in the current state, which satisfies aperiodicity. Hence, this proposal scheme guarantees the ergodicity of the MCMC chain. See APPENDIX B for details.

3.2 Sampling the MIN Structure

To learn the MIN structure from the observed time series, we assume the stationarity of the VAR model and denote with $\mathcal{N}(0, \Sigma_{y,0})$, the distribution of the contemporaneous variables, with $\Sigma_{y,0}$ as the covariance matrix. We allow for acyclic constraints to identify the causal directions in the system and to produce an identifiable model. To sample acyclic graphs, we modify the concept of Giudici and Castelo (2003) by exploiting the following condition. Let X_t^i and X_t^j be two nodes in a MIN structure. $X_t^j \to X_t^i$ is legal if and only if the intersection between the descendants of X_t^i and the ancestors of X_t^j is empty. For example, assume $P \to \ldots \to Q \to \ldots \to R$, then P and Q are ancestors of R, and Q and R are descendants of P. We see that adding an edge $R \to P$ is illegal since it produces a cycle. Our proposal of a directed edge $X_t^j \to X_t^i$ can be implemented in two steps. First, we verify if there is a directed edge from X_t^i to X_t^j. If such relationship exists, we remove the link. Secondly, we add the directed edge $X_t^j \to X_t^i$ only if a directed path from X_t^i to X_t^j,

$(X_t^i \to \ldots \to X_t^j)$ does not exist. This can be verified using the reachability matrix (see e.g., Wasserman, 1994). Alternatively, the second step can be handled by adding $X_t^j \to X_t^i$ and testing whether the resulting structure is a directed acyclic graph.

3.3 Estimating Reduced-Form Graphical VAR

After sampling the structures (G_0, G_+) from the observed data, we proceed to sample the parameter of the associated reduced-form VAR model. Following the expression in (9), we notice that the graph structure associated with the reduced-form coefficients matrix A_+ is given by $G_+^* = (\mathbf{I}_{n_y} - G_0)^{-1} G_+$. Thus, we incorporate the MIN and MAR structures to obtain the graph structure of the reduced-form VAR (G_+^*). Clearly, if G_0 is empty (no contemporaneous dependencies), then the reduced-form coincides with the structural dynamics. We select the non-zero elements of G_+^*, to indicate the relevant variables of the model. To estimate (A_+, Σ_u), we consider two typical prior distributions extensively applied in the Bayesian VAR literature, i.e., the Minnesota (MP) and the normal-Wishart (NW) prior.

Minnesota Prior

The MP prior was proposed by Doan et al. (1984) to shrink the VAR model toward a random walk model. Here, the diagonal elements of A_1 were shrunk toward one and the remaining coefficients in $A_+ = (A_1, \ldots, A_p)$ toward zero. The basic idea is that more recent lags provide more reliable information than distant ones and that one's own lags should explain more of the variation of a given variable than the lags of the other variables. The prior expectation, $\mathbb{E}[(A_k)_{ij}]$, is equal to δ if $j = i$ and $k = 1$, and takes value 0 otherwise. The variance of the coefficient matrices is $\mathbb{V}[(A_k)_{ij}] = \alpha \sigma_i^2 k^{-2} \sigma_j^{-2}$. The coefficients in A_+ are assumed to be a priori independent and normally distributed, $A_+ \sim \mathcal{N}(\underline{A}, \underline{V})$. Conditional on G_+^*, we estimate the posterior mean $(\overline{A}_i)$ and variance $(\overline{V}_i)$ of the coefficient of relevant variables in each equation by $\overline{A}_i = \overline{V}_i(\underline{V}_i^{-1}\underline{A}_i + \sigma_i^{-2} W_i' Y^i)$ and $\overline{V}_i = (\underline{V}_i^{-1} + \sigma_i^{-2} W_i' W_i)^{-1}$, where $\underline{A}_i$ and $\underline{V}_i$, $i = 1, \ldots, n_y$, are the prior mean and variance of the relevant variables in each equation, and $W_i \in X'$ is the set of relevant variables that influence the response

variable Y^i. The covariance matrix of the residuals, Σ_u, is assumed to be diagonal, fixed, and known; $\Sigma_u = \text{diag}(\sigma_1^2, \ldots, \sigma_{n_y}^2)$. Here, $\sigma_i^2,\ i = 1, \ldots, n_y$, is the estimated variance of the residuals from a univariate autoregressive model of order p for variable Y^i (see Banbura et al., 2010, Karlsson, 2013). Following the recent literature, we set $\delta = 0.9$ and $\alpha = 0.5$.

Normal-Wishart Prior

The independent NW is the commonly used prior distribution in the estimation of seemingly unrelated regression (SUR) models. This prior assumes: $A_+ \sim \mathcal{N}(\underline{A}, \underline{V})$ and $\Sigma_u^{-1} \sim \mathcal{W}(\underline{\nu}, \underline{S}^{-1})$, where $\underline{S}$ is the prior sum of squares and $\underline{\nu}$ is the associated degrees of freedom. In this application, we deviate slightly from the standard approach of estimating the posterior mean and variance of A_+ by considering coefficient updates similar to the Minnesota approach. Conditional on G_+^*, we select the relevant variables, $W_i,\ i = 1, \ldots, n_y$, of each equation by $\overline{A}_i = \overline{V}_i(\underline{V}_i^{-1}\underline{A}_i + \overline{\sigma}_i^{-2}W_i'Y^i)$ and $\overline{V}_i = (\underline{V}_i^{-1} + \overline{\sigma}_i^{-2}W_i'W_i)^{-1}$, where $\overline{\sigma}_i^2,\ i = 1, \ldots, n_y$, is the variance of residuals from the posterior of Σ_u. The posterior of Σ_u^{-1} is Wishart distributed with $\overline{S} = \underline{S} + (Y' - X\overline{A}')'(Y - X\overline{A}')$ and $\overline{\nu} = \underline{\nu} + (T - p)$ degrees of freedom. $\overline{A}$ is the posterior of A_+ with dimension $n_y \times np$ such that elements of relevant variables in $\overline{A}$ store the corresponding elements of $\overline{A}_i, i = 1, \ldots, n_y$, and the rest (representing not-relevant variables) are restricted to zero.

4. SIMULATION EXPERIMENTS

We study the efficiency of our inference approach on simulated datasets generated from a n-node graphical model. We consider the following data generating process (DGP):

$$X_t = \sum_{i=0}^{p} (\mathbf{I}_k \otimes B_i) X_{t-i} + \varepsilon_t, \tag{10}$$

$t = 1, \ldots, T$, where X_t is a n-dimensional vector, p is the lag order, and $B_i,\ i = 0, 1, 2, 3$, is the sequence of 5×5 coefficient matrices given in Table I. Note that k controls the size n of the model, since $n = 5k$. For a more general validity of our simulation study, we consider six different settings, which correspond to the lag order and the model dimension one can

commonly find in the empirical applications. We set $p = 1, 2, 3$ in the 5-node (i.e., $k = 1$) and 20-node (i.e., $k = 4$) models. We generate $T = 110$ data points for each k and p, and use 100 observations for the model estimation and 10 for the out-sample forecast analysis. For each setting, we replicate the simulation and estimation exercises 20 times, with random draws from the DGP. All the results reported in the following are averages over the replications.

Table I: Structural coefficients of the data generating process. B_0 is the contemporaneous coefficient matrix, and B_1, B_2, and B_3 are the temporal coefficient matrices at lag 1, 2 and 3, respectively.

$$(B_0)\begin{pmatrix} 0 & 0 & -0.8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & -0.5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (B_1)\begin{pmatrix} -0.8 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.5 & 0 & 0 \\ 0.7 & 0 & -0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.7 & 0 \\ 0 & -0.6 & 0 & 0 & 0.6 \end{pmatrix}$$

$$(B_2)\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.4 \\ -0.6 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.4 & 0 & 0 \end{pmatrix} \quad (B_3)\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 & 0 \end{pmatrix}$$

We compare the MIN structure with the contemporaneous structure of the PC algorithm (see APPENDIX C) and SSVS ω matrix (SSVS(ω)). See George et al. (2008) for details on the implementation of the SSVS approach. The MAR structure is also compared with the temporal dependence structure given by a modified conditional Granger-causal (C-GC) inference and the SSVS γ matrix (SSVS(γ)). For this comparison, we modified the C-GC of (Ding et al., 2006) to select significant Granger-causal variables at different lags (see APPENDIX C). As an alternative to the C-GC the Granger causal priority (see Sims (2010)) can be applied, which is an approach to discriminating between mediated and direct Granger-causal effects.

We evaluate the accuracy of our estimates by comparing the BGVAR model with the BVAR and the SSVS models. The reduced-form model is estimated by considering the BGVAR under the Minnesota (BGV-MP) and the normal-Wishart (BGV-NW) prior distributions, and the BVAR model under the Minnesota (BV-MP) and the normal-Wishart conjugate (BV-NW) prior distributions. The prediction accuracy of the models is evaluated using the log-predictive score and the AIC score. See APPENDIX C for details on the graph accuracy assessment and the model prediction accuracy evaluation.

For the small- (moderate-) dimension models, we run a total of 20,000 (40,000) Gibbs iterations and exclude 50% burn-in samples for our BGVAR model. For the predictive model

estimation, we run a total of 2200 Gibbs iterations with 200 burn-in samples for both BVAR and BGVAR. Following the suggestions in George et al. (2008), we run a total of 20,000 Gibbs iterations for the SSVS and exclude 10,000 burn-in samples.

Table II: Comparing inference on the 5- and 20-node models with lag order $p = 1$, 2, and 3. The comparison is done in terms of accuracy of the contemporaneous structures (first panel), accuracy of the temporal dependence structures (second panel), forecast accuracy (third panel), and computational time (fourth panel).

	Small-size ($n = 5$)			Moderate-size ($n = 20$)		
Scheme	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
Contemporaneous Structure Inference (Accuracy)						
PC	80.00	72.00	72.00	95.75	93.75	96.25
MIN	88.00	88.00	80.00	96.75	96.25	96.25
Temporal Structure Inference (Accuracy)						
C-GC	100.00	100.00	89.33	100.00	99.04	96.67
MAR	100.00	100.00	100.00	100.00	99.47	99.75
Forecast Accuracy (log predictive score)						
BV-MP	-47.12	-50.16	-11.96	-241.13	-256.57	-41.39
BV-NW	-43.48	-45.90	4.09	-308.74	-484.03	-111.82
SSVS	-42.33	-41.66	20.19	-224.77	-226.72	26.97
BGV-MP	-45.55	-47.96	-13.04	-208.21	-206.61	-9.23
BGV-NW	-41.14	-40.14	6.75	-221.44	-223.03	-0.13
Prediction Accuracy Adjusted (AIC)						
BV-MP	144.24	200.32	173.92	1282.27	2113.13	2482.79
BV-NW	136.95	191.81	141.82	1417.48	2568.05	2623.64
SSVS	134.66	183.32	109.62	1249.55	2053.44	2346.06
BGV-MP	117.09	149.92	100.08	640.43	1087.22	1230.47
BGV-NW	108.28	134.29	60.50	666.87	1120.06	1212.26
Computational Time (in seconds)						
SSVS(γ, ω)	42.26	51.38	67.46	735.38	3340.88	9587.92
BGVAR	16.09	17.04	17.47	153.24	187.54	229.01

We report in Table II the results of the comparison between the different inference schemes. The top panel of the table shows that the MIN achieves a higher accuracy than the PC. In both small and moderate-dimension settings, we notice that the MIN inference achieves an accuracy above the 80%. This evidence shows that our inference of the contemporaneous dependence from the observed time series offers some insight into the structural dynamic of the VAR. See also APPENDIX D for further details. The second panel of Table II shows that both MAR and C-GC perform well at inferring the temporal dependence relationships. However, we notice that MAR achieves a slightly higher accuracy than C-GC when $p = 3$, in both small (5-node) and moderate (20-node) dimension models. See also

APPENDIX D for further details. The third panel of Table II shows the forecast accuracy of the models based on log predictive scores - sum of log predictive likelihood over the forecast period. The log predictive score strongly favors BGVAR with normal-Wishart (Minnesota) prior in small (moderate) dimension DGP with lag order $p < 3$, and favors SSVS in small and moderate dimension DGP with lag order $p = 3$. When adjusted for the number of estimated coefficients, (see the fourth panel) BGVAR (with both the Minnesota and normal-Wishart priors) achieves a higher predictive accuracy and fits the simulated data better than BVAR and SSVS. The bottom panel of Table II shows the computational time (average over the different prior settings) for inference of the SSVS and the BGVAR connectivity matrices and parameters. Since only BGVAR and SSVS are concerned with joint inference on parameters, temporal and contemporaneous structures, the computational times for PC, C-GC and BV are not reported. For the small dimension DGP we set 20,000 iterations for the SSVS and the BGVAR inference. Since the number of possible structures increases super-exponentially with the number of nodes and lags, for the moderate dimension DGP, we consider 40,000 iterations for BGVAR, and maintain 20,000 iterations for SVSS. Our results show that, in both small- and moderate-dimension settings, the collapsed Gibbs sampler used for the BGVAR posterior approximation is computationally less expensive than the algorithm used for the SSVS posterior approximation. Although we obtain similar results over different experiments and settings, our limited investigations confirm the good mixing of MCMC chain for our BGVAR model.

5. MODELING AND FORECASTING MACROECONOMIC TIME SERIES

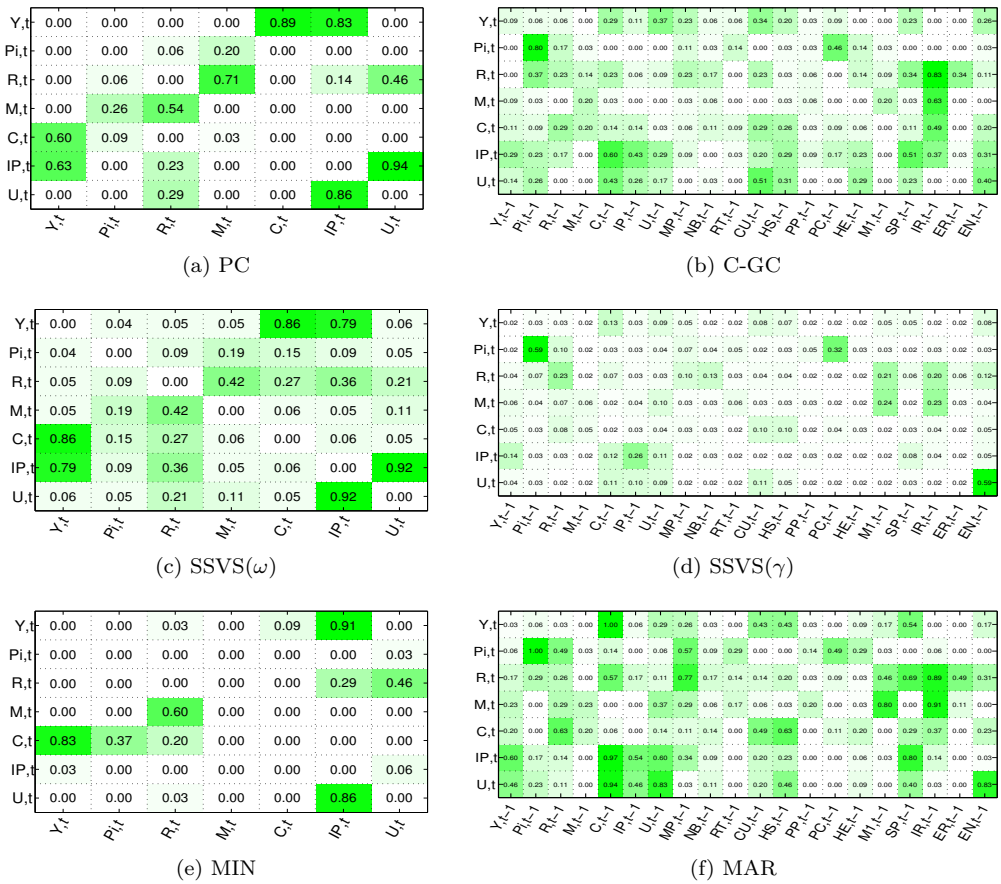
The work by Banbura et al. (2010) has motivated a growing interest in the application of high dimensional BVAR models to forecast macroeconomic time series. In their empirical application, the authors showed that high dimensional BVAR models produce better forecasts than the traditional (factor methods) approach. These findings were recently corroborated by Koop (2013) in his study on forecasting with Bayesian VARs. According to Banbura et al. (2010), most of the gains in forecast performance of high dimensional mod-

els was achieved using medium VARs ($n = 20$). Based on this, we apply our inference to model a moderate dimension VAR of 20 macroeconomic variables. The first objective of this exercise is to offer an interpretations of the structural dynamics by comparing our results with both the PC algorithm and the Granger-causal inference. Secondly, we compare the estimated BGVAR model with both the BVAR model and the SSVS procedure, and evaluate their predictive performance.

The dataset consists of quarterly observations, from 1959Q1 to 2008Q4, of 20 US-macroeconomic variables which were originally used by Koop (2013). We transform the data as in Koop (2013). See APPENDIX E for the list of series and their transformation. The specification of the lag order ($p = 1$) is based on testing the appropriate lag length using BIC. We estimate a model with the following 7 response variables: (Y) - real gross domestic product (GDP), (Pi) - consumer price index, (R) - Federal funds rate, (M) - money stock M2, (C) - real personal consumption, (IP) - industrial production index, and (U) - unemployment. We consider the following 13 additional variables as predictor variables: (MP) - real spot market price index for all commodities, (NB) - non-borrowed reserves, (RT) - total reserves, (CU) - capacity utilization, (HS) - housing, (PP) - producer price index, (PC) - personal consumption expenditure, (HE) - real average hourly earnings, $(M1)$ - money stock M1, (SP) - S&P500 index, (IR) - 10-yr US treasury bill rate, (ER) - US effective exchange rate, (EN) - employment.

To allow for model (structure and coefficient) changes, we apply a moving window to estimate, for each window, the structure and the associated predictive model. The moving window uses the most recent observations to estimate the model. Thus new data are added to the exiting dataset and the oldest observations are deleted. We initialize the first sample from $1960Q1 - 1974Q4$, and then move the window forward by 4-quarters. We produce forecasts for all the horizons up to 4-quarters ahead. Our last sample is set to $1993Q1 - 2007Q4$. Thus the forecast period ranges from $1975Q1$ to $2008Q4$.

We report in the left column of Figure 2 the results of PC, SSVS(ω) and MIN averaged over $1960Q1-2006Q4$. The results of the PC algorithm show strong evidence of the following edges: $C_t - Y_t - IP_t - U_t$, and $M_t - R_t$. SSVS(ω) shows $C_t - Y_t - IP_t - U_t$, and MIN reveals: $Y_t \to C_t$, $IP_t \to (Y_t, U_t)$, and $R_t \to M_t$. Thus all the structures capture contemporaneous



(a) PC (b) C-GC

(c) SSVS(ω) (d) SSVS(γ)

(e) MIN (f) MAR

Figure 2: Contemporaneous structure of PC, SSVS and MIN (left column) and temporal dynamic structure of C-GC, SSVS and MAR (right column), averaged over the period $1960Q1-2006Q4$. In each plot, the light (dark) green color indicates weak (strong) evidence of dependence. The variables are: (Y) - real GDP, (Pi) - consumer price index, (R) - Federal funds rate, (M) - money stock M2, (C) - real personal consumption, (IP) - industrial production index, (U) - unemployment, (MP) - real spot market price, (NB) - non-borrowed reserves, (RT) - total reserves, (CU) - capacity utilization, (HS) - housing, (PP) - producer price index, (PC) - personal consumption expenditure, (HE) - real average hourly earnings, $(M1)$ - money stock M1, (SP) - S&P500 index, (IR) - 10-yr US treasury bill rate, (ER) - US effective exchange rate, (EN) - employment.

relationships between: consumption (C_t) and real GDP (Y_t); industrial production (IP_t) and unemployment (U_t); money supply (M_t) and interest rates (R_t); industrial production (IP_t) and GDP (Y_t).

Though PC shows a stronger evidence of the effects of consumption (C_t) and industrial production (IP_t) on real GDP (Y_t), it does not rule out the possibility of a reverse effect of

real GDP (Y_t) on consumption (C_t) and industrial production (IP_t). A possible explanation for the different findings of MIN regarding this relationship could stem from issues related to Markov equivalence of graphical models (see Section 2) or from issues related to the exclusion of other relevant variables. The results in APPENDIX E show that MIN and PC dependence structures are sensitive to the exclusion of relevant variables, as opposed to the SSVS procedure. On the contrary in the SSVS procedure, the edges between variables may be wrongly inferred, also when all relevant variables are included in the model.

In the right column of Figure 2 we report the temporal dependence structure of C-GC, SSVS(γ) and MAR, averaged over $1960Q1 - 2006Q4$. The figure reveals that C-GC and MAR identify stronger evidence of dependence than SSVS(γ). More specifically, SSVS(γ) finds strong evidence of dependence only for two edges, whose probabilities are close to 60%. MAR, on the other hand, detects edges that are persistent over time. For example, current real GDP (Y_t) strongly depends on previous level of consumption (C_{t-1}); current level of inflation (Pi_t) strongly depends on previous level of inflation (Pi_{t-1}); and current level of money stock M2 (M_t) strongly depends on previous levels of money stock M1 ($M1_{t-1}$) and the 10-year US treasury bill rate (IR_{t-1}). Finally, we shall note that these response variables may also weakly or temporarily depend on other variables, but the edge probabilities in Figure 2 may be close to zero since they are averages over a sequence of rolling estimates. Similar conclusions hold true for the comparison between the estimated coefficient matrices of the BVAR (see APPENDIX E).

The top-left chart in Figure 3 compares the evolution of the BIC scores of PC (in blue) and MIN (in green) over the period $1960Q1 - 2006Q4$. The figure shows that the BIC score favors MIN over PC, giving an indication that MIN provides a better representation of the contemporaneous dependence in the observed time series than PC. The top-right chart in Figure 3 presents the evolution of the BIC scores of C-GC (in blue) and MAR (in green) over the period $1960Q1 - 2006Q4$ for the macroeconomic application. Clearly, the figure shows that the BIC favors the MAR over the C-GC. This seems to indicate that the MAR structure provides a better representation of the temporal dependence in the observed time series than the C-GC.

In the last row of Figure 3, we report the evolution of the log predictive scores and

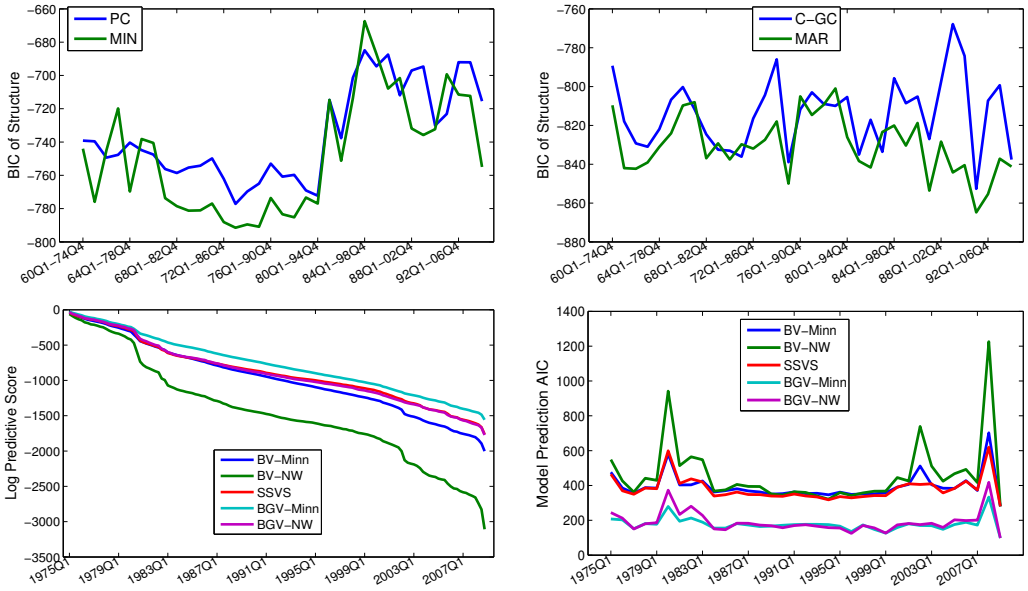


Figure 3: The BIC of the contemporaneous (top left) and temporal (top right) dependence structures of the PC (blue), C-GC (blue) and the MAR (green) over the period $1960Q1 - 2006Q4$. Log predictive score (bottom left) and predictive AIC Score (bottom right) of BVAR, SSVS and BGVAR over $1975Q1 - 2008Q4$. BV-Minn represents BVAR under Minnesota prior (blue), BV-NW - BVAR under normal-Wishart (green), SSVS - Stochastic Search Variable Selection (red), BGV-Minn - BGVAR under Minnesota prior (cyan), and BGV-NW - BGVAR under normal Wishart (pink).

AIC scores of the competing models over the period $1975Q1 - 2008Q4$. The log predictive scores - the sum of the log predictive, measure the forecast performance of the models. The AIC scores on the other hand measure the predictive performance adjusted for number of estimated coefficients. The figure shows that the log predictive score strongly favors BGVAR Minnesota (in cyan), followed by BGVAR normal-Wishart (pink) and SSVS (in red). The two BVAR models record the lowest log predictive scores. When adjusted for the number of estimated coefficients, the AIC scores strongly favor the BGVAR model (under both Minnesota and normal-Wishart prior distributions), followed by SSVS and BVAR. Thus, BGVAR achieves a higher predictive accuracy and fits the data better than BVAR and SSVS.

6. MEASURING FINANCIAL INTERCONNECTEDNESS

The level of interconnectedness of the financial system received a lot of attention by many parts in the aftermath of the 2007-2009 financial crisis. While the greater interconnectedness can increase the systemic risk and the probability of contagion, it can also have a positive impact on the system, provided the authorities take steps to prevent the systemic risk. For this reason, several studies on financial networks have empirically assessed the linkages and the exposures within financial institutions (e.g., see Hautsch et al. (2012), Billio et al. (2012), and Diebold and Yilmaz (2014)). Non-financial institutions, on the other hand, have increasingly gained awareness of the need to adopt financial strategies to avoid being vulnerable to instabilities in financial markets. The objective of the financial application presented in this paper is to investigate empirically, by means of BGVAR, the linkages between financial and non-financial institutions. Ultimately the aim is to assess the interconnectedness of the system and thus its vulnerability.

The dataset consists of the return indexes of the 19 super-sectors of Euro Stoxx 600, sampled at a monthly frequency from January 2001 to August 2013 (source: Datastream). The dataset also represents the largest Eurozone companies in each of the 19 super-sectors, as defined by the Industry Classification Benchmark (ICB). The countries covered are Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain (see APPENDIX E). The specification of the lag order ($p = 1$) is based on testing the appropriate lag length using BIC. Following Billio et al. (2012), we apply a 36-month moving window to analyze the evolution of the linkages among the super-sectors.

In this application we focus only on the temporal dependence among the variables. To this purpose, we compare the structure of our MAR with the modified conditional Granger-causality (C-GC) and with the modified pairwise Granger-causality (P-GC).

We compute the number of linkages by summing all the edges in the graph structure for each window. Top-left panel of Figure 4 compares the evolution of the number of links of P-GC (blue line), C-GC (green line) and MAR (red line) as a percentage of all possible edges among the super-sectors over the sample period. As shown in Figure 4, the percentage of links obtained P-GC is relatively higher than that of C-GC and MAR. C-GC on the other hand, moves in accordance with MAR except in the period 2005-2007, where MAR records a lower number of edges.

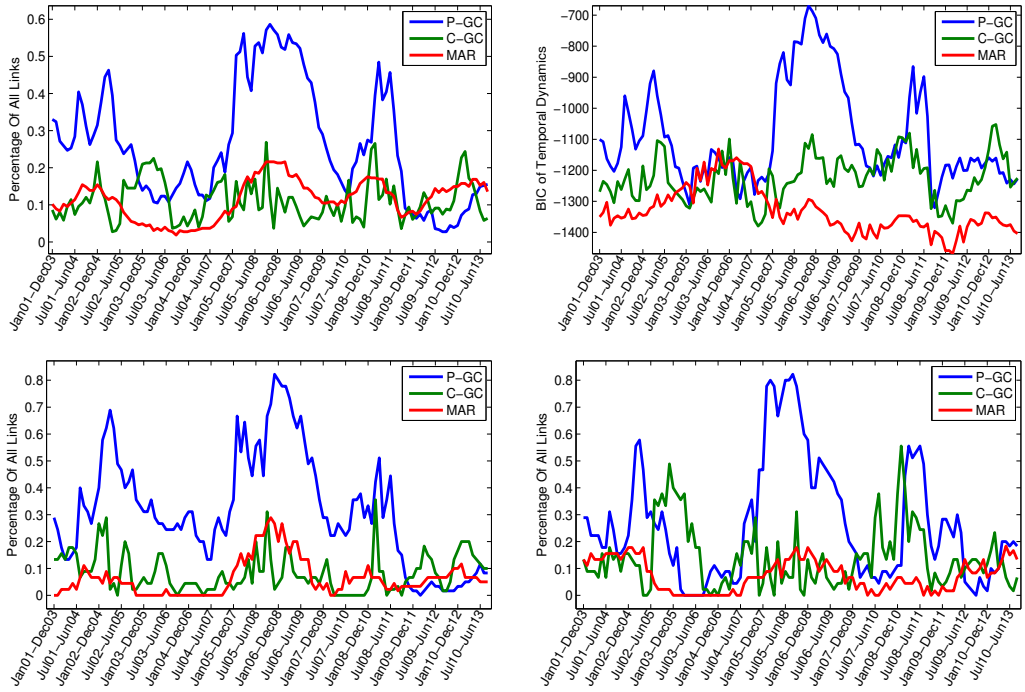


Figure 4: Linkages among the super-sectors of Euro Stoxx 600 from January 2001 to August 2013 based on a 36-month moving window. Evolution of the total links (top left), BIC scores (top right), percentage of links from financial to non-financial (bottom left) and from non-financial to financial (bottom right) super-sectors, of P-GC (blue), C-GC (green) and MAR (red).

Using MAR, we observe four periods of large interconnectedness. The periods identified are pre-2005, 2007-2009, 2010-2011, and 2011-2013. By matching these periods to notable global and European events, we notice that the pre-2005 can arguably be linked to the aftermath of scandals such as Enron and Worldcom; whereas 2007-2009 and 2010-2013 capture the recent financial crisis and the European sovereign crisis, respectively.

Figure 4 compares the evolution of the percentage of links of P-GC (blue line), C-GC (green line) and MAR (red line) from financial to non-financial super-sectors and from non-financial to financial super-sectors. Again, the P-GC overestimates the linkages compared to C-GC and MAR. MAR still identifies four peaks of high connectedness in both figures. Surprisingly, these periods coincide with the results shown in the top-left panel of Figure 4. This similarity suggests that periods of financial market turbulence also experienced a higher number of linkages between financial and non-financial super-sectors. In the pre-2005 period,

we observe a stronger linkage from non-financial to financial super-sectors than the reverse. The 2007-2009 financial crisis sees a stronger linkage from financial to non-financial super-sectors compared to the linkage from non-financial to financial. For the 2010-2013 European sovereign crisis, we observe an equally strong linkage between the two super-sectors.

The BIC scores (Figure 4, top-right) strongly favors the MAR structure in the greater part of the sample period. The P-GC approach is the least favored among the three schemes. This result is, to some degree, expected since the pairwise Granger causality approach deals only with bivariate time series and does not consider the conditioning on relevant covariates. We also observe that the BIC score favors C-GC above P-GC. This is expected since the modified conditional Granger considers the conditioning on relevant covariates. However, with a higher number of variables relative to the number of data point, the C-GC approach encounters a problem of over-parameterization, that leads to a loss of degrees of freedom and to inefficiency in correctly gauging the causal relationships. The BGVAR approach allows the researcher to handle such situations, by conditioning on all relevant variables, carrying out joint inference on all quantities of interest, and achieving model parsimony. The BIC score confirms that BGVAR provides a more accurate representation of the linkages among the institutions than P-GC and C-GC, offering a better approach to studying systemic risk.

7. CONCLUSION

This paper develops a new Bayesian graph-based approach to identification and over-parameterization issues in structural VAR models. Our inference procedure of the BGVAR causal relationships provides also a variable selection procedure. Moreover, we propose an efficient Markov chain Monte Carlo algorithm to infer the BGVAR causal structures and the unknown parameters from observed multiple time series. In both simulation experiments and real data applications, the BIC score indicates that our BGVAR produces a better representation of the structural causal relationships than several competing standard approaches. Our comparison results show that the BGVAR model is more parsimonious and interpretable than both the classical Bayesian VAR (BVAR) and the stochastic search

variable selection (SSVS) models.

In the macroeconomic application, the BGVAR provides a data-driven identification of the structural VAR, thus offering a useful tool for policy analysis. The predictive accuracy, adjusted for number of estimated coefficients, strongly favors BGVAR over BVAR and SSVS. In the financial application the BIC score indicates that our BGVAR inference produces a better representation of the linkages among economic sectors than the Granger inference, thus offering a better approach to studying systemic risk.

ACKNOWLEDGMENTS

We would like to thank the editor, the referees, Fabio Canova, Fulvio Corsi, Alessio Moneta, Lorenzo Frattarolo, Francesco Ravazzolo, Herman K. van Dijk for their comments on an earlier version of this paper. Furthermore, we have benefited much from comments by conference participants of the 68th European meeting of the Econometric Society, Toulouse, 2014, the QED meeting, Bielefeld, 2014, the 5th CSDA International Conference, London, 2013, the Stats meeting, Paris, 2013, the ENSAE ParisTech, 2013, and the Workshop on Modern Tools in Macro-Econometrics, Venice, 2013. Authors' research is supported by funding from the European Union, Seventh Framework Programme FP7/2007-2013 under grant agreement SYRTO-SSH-2012-320270, by the Global Risk Institute in Financial Services and the Institut Europlace de Finance, Systemic Risk grant, and by the Italian Ministry of Education, University and Research (MIUR) PRIN 2010-11 grant MISURA.

REFERENCES

Andersson SA, Madigan D, Perlman MD, et al. 1997. A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics* **25**: 505–541.

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

Berger JO, Pericchi LR. 1998. On criticisms and comparisons of default Bayes factors for model selection and hypothesis testing. In *Proceedings of the Workshopo on Model Selection*. Bologna, Italy: Edizioni Pitagora.

Bernanke BS. 1986. Alternative explanations of the money-income correlation. *Carnegie-Rochester Conference Series on Public Policy,* **25**: 49–100.

Billio M, Getmansky M, Lo A, Pellizzon L. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* **104**: 535 – 559.

Brillinger DR. 1996. Remarks concerning graphical models for time series and point processes. *Revista de Econometria* **16**: 1–23.

Brooks SP, Gelman A. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**: 434–455.

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

Carvalho CM, West M. 2007. Dynamic matrix-variate graphical models. *Bayesian Analysis* **2**: 69–98.

Casella G, Robert CP. 2004. *Monte Carlo Statistical Methods*. New York: Springer Verlag.

Chickering DM, Heckerman D, Meek C. 2004. Large-sample learning of Bayesian networks is np-hard. *Journal of Machine Learning Research* **5**: 1287–1330.

Cooley TF, Leroy SF. 1985. A theoretical macroeconometrics: A critique. *Journal of Monetary Economics, Elsevier,* **16**: 283–308.

Corander J, Villani M. 2006. A Bayesian approach to modelling graphical vector autoregressions. *Journal of Time Series Analysis* **27(1)**: 141–156.

Dawid AP, Lauritzen SL. 2001. Compatible prior distributions. *In Bayesian methods with application to science, policy and official statistics; Selected papers from ISBA 2000* .

de Santis F, Spezzaferri F. 1999. Methods for default and robust Bayesian model comparison: The fractional Bayes factor approach. *International Statistical Review* **67**: 267–286.

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(1)**: 119–134.

Diebold F, Yilmaz K. 2015. *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press.

Ding M, Chen Y, Bressler SL. 2006. Granger causality: basic theory and applications to neuroscience. In *Schelter B., Winterhalder M., Timmer J. (eds) Handbook of time series analysis*. Wiley, Weinheim, 437–460.

Doan T, Litterman R, Sims C. 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* **3**: 1–100.

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.

Geiger D, Heckerman D. 1999. Parameter priors for directed acyclic graphical models and the characterization of several probability distributions. *Annals of Statistics* **30**: 2002.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2013. *Bayesian data analysis*. CRC press.

Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences, (with discussion). In *Statistical Science*, volume 7. 457–511.

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

Giudici P, Castelo R. 2003. Improving Markov chain Monte Carlo model search for data mining. *Machine Learning* **50**: 127–158.

Giudici P, Green PJ. 1999. Decomposable graphical Gaussian model determination. *Biometrika* **86**: 785–801.

Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**: 424–438.

Grzegorczyk M, Husmeier D. 2008. Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. *Journal of Machine Learning* **71**: 265–305.

Grzegorczyk M, Husmeier D, Rahnenführer J. 2010. Modelling nonstationary gene regulatory processes. *Advances in Bioinformatics* : 1–17.

Hautsch N, Schaumburg J, Schienle M. 2012. Financial network systemic risk contributions. Discussion paper 2012-053, crc 649, Humboldt-Universität zu Berlin.

Heckerman D, Geiger D. 1994. Learning Gaussian networks. In *Uncertainty in Artificial Intelligence*. 235–243.

Kadiyala KR, Karlsson S. 1993. Forecasting with generalized bayesian vector auto regressions. *Journal of Forecasting* **12**: 365–378.

Karlsson S. 2013. Forecasting with Bayesian vector autoregressions. In *Handbook of Economic Forecasting, G. Elliott and A. Timmermann, (eds.)*, volume 2. Elsevier, 689–1324.

Kilian L. 2013. Structural vector autoregressions. In *Handbook of Research Methods and Applications in Empirical Macroeconomics, Hashimzade N., Thornton M. (eds.)*. Edward Elgar: Cheltenham, UK, 515 – 554.

King RG, Plosser CI, Stock JH, Watson MW. 1991. Stochastic trends and economic fluctuations. *American Economic Review* **81**: 819–840.

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

Lauritzen SL, Wermuth N. 1989. Graphical models for associations between variables, some of which are quantitative and some qualitative. *Annals of Statistics* **17**: 31–57.

Liu J. 1994. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association* **89**: 958 – 966.

Madigan D, York J. 1995. Bayesian graphical models for discrete data. *International Statistical Review* **63**: 215–232.

McNees S. 1986. Forecasting accuracy of alternative techniques: a comparison of U.S. macroe-conomic forecasts. *Journal of Business and Economic Statistics* **4**: 5–15.

Pearl J. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann publishers, San Mateo, CA, USA,.

Roberts G, Sahu S. 1997. Updating schemes, covariance structure, blocking and parametrisation for the Gibbs sampler. *Journal of the Royal Statistical Society* **59**: 291 – 318.

Robinson RW. 1977. Counting unlabeled acyclic digraphs. In *Combinatorial Mathematics V*. Springer, 28–43.

Rubio-Ramirez JF, Waggoner DF, Zha T. 2010. Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies* **77**: 665–696.

Sims CA. 1980. Macroeconomics and reality. *Econometrica, Econometric Society,* **48**: 1–48.

Sims CA. 2010. Lecture notes: causal orderings and exogeneity.
URL `http://sims.princeton.ftp.edu/yftp/times09/addednotes/gcp09.pdf`

Spirtes P, Glymour C, Scheines R. 2000. *Causation, Prediction, and Search*. MIT Press.

Stock JH, Watson MW. 2001. Vector Autoregressions. *Journal of Economic Perspectives* **15**: 101–115.

Stock JH, Watson MW. 2008. Forecasting in dynamic factor models subject to structural instability. In *The Methodology and Practice of Econometrics: A Festschrift in Honour of Professor David F. Hendry, Castle J., Shephard N. (eds)*. Oxford University Press: Oxford, 173 – 205.

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.

Wasserman S. 1994. *Social network analysis: Methods and applications*, volume 8. Cambridge University Press.

Wermuth N, Lauritzen SL. 1990. On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *Journal of the Royal Statistical Society, Series B* **52**: 21–50.

Whittaker J. 1990. Graphical models in applied multivariate statistics. *John Wiley, Chichester* .

APPENDIX A. BGVAR PRIOR AND POSTERIOR DISTRIBUTIONS

A graphical model is defined by a graph structure G and a collection of parameters, θ. Let $X_t = (X_t^1, X_t^2, \ldots, X_t^n)$, where X_t^i is a realization of the variable X^i at time t. Let $\mathcal{X} = (X_1, \ldots, X_T)$ be a time series of n observed variables. We assume that $\mathcal{X}$ follows a multivariate normal distribution, and define $\theta \equiv \{\mu, \Omega_x\}$, $\Omega_x = \Sigma_x^{-1}$. For simplicity, we assume that the data is generated by a stationary process and, without loss of generality, we set $\mu = 0$. The likelihood of $\mathcal{X}$ is given by

$$P(\mathcal{X}|\Omega_x, G) = (2\pi)^{-\frac{nT}{2}} |\Omega_x|^{\frac{T}{2}} \exp\Big\{ -\frac{1}{2}\langle \Omega_x, \hat{S}_x \rangle \Big\} \tag{A.11}$$

where $\langle A, B\rangle = tr(A'B)$ denotes the trace inner product and $\hat{S}_x = \sum_{t=1}^{T} X_t X_t'$. From a Bayesian perspective, the joint prior distribution over (Ω_x, G) is given as $P(G, \Omega_x) = P(G)P(\Omega_x|G)$. Two sources of uncertainty are associated with the model: the graph structure G and the parameter Ω_x. The graph structure G considered in this paper is characterized by $0-1$ elements, G_{ij}, where $G_{ij} = 1$ means $X^j \to X^i$, and $G_{ij} = 0$ means that no link exists between the two variables. An independent Bernoulli prior with parameter β is assumed on each edge. We assume $\beta = 0.5$, that leads to a uniform prior on the graph

space, i.e. $P(G) \propto 1$.

Following the standard Bayesian paradigm, we assume that Ω_x conditional on a complete graph G is Wishart distributed, $\Omega_x \sim \mathcal{W}(\nu, S_0^{-1})$, with density

$$P(\Omega_x|G) = \frac{1}{K_n(\nu, S_0)} |\Omega_x|^{\frac{(\nu-n-1)}{2}} \exp\Big\{ -\frac{1}{2}\langle \Omega_x, S_0\rangle \Big\} \tag{A.12}$$

where $\nu > n+1$ is the degree of freedom parameter, S_0 is a symmetric positive definite matrix, and $K_n(\nu, S_0)$ is the normalizing constant given by

$$K_n(\nu, S_0) = \int |\Omega_x|^{\frac{(\nu-n-1)}{2}} \exp\Big\{ -\frac{1}{2}\langle \Omega_x, S_0\rangle \Big\}\, d\Omega_x \;=\; 2^{\frac{\nu n}{2}} |S_0|^{-\frac{\nu}{2}} \Gamma_n\Big(\frac{\nu}{2}\Big)$$

with $\Gamma_n(a) = \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\Big(a - \frac{i+1}{2}\Big)$ and $\Gamma(\cdot)$ the gamma function. Following Geiger and Heckerman (1999), we assume that the independence and modularity conditions are satisfied. In our SVAR model, the independence assumption means that the structural coefficients and the error terms, within and across equations, are a priori independent. The modularity assumption states that if a response variable has the same set of explanatory variables in two graphs, then the associated parameters must have the same prior distribution. These assumptions on the prior distributions allow us to factorize the likelihood, to integrate out the parameters analytically, and to obtain the following expression for the marginal likelihood:

$$\begin{aligned} P(\mathcal{X}|G) &= \int P(\mathcal{X}|\Omega_x, G) P(\Omega_x|G) d\Omega_x \\ &= \frac{(2\pi)^{-\frac{nT}{2}}}{K_n(\nu, S_0)} \int |\Omega_x|^{\frac{T+\nu-n-1}{2}} \exp\Big\{ -\frac{1}{2}\langle \Omega_x, S_0 + \hat{S}_x\rangle \Big\} d\Omega_x = \frac{K_n(\nu+T, S_0+\hat{S}_x)}{(2\pi)^{nT/2} K_n(\nu, S_0)} \end{aligned}$$

Based on the uniform prior assumption over structures, maximizing the posterior probability of G is equivalent to maximizing the marginal likelihood metric. For graphical model selection purposes, we sample G in the space of all possible graphs from the marginal posterior distribution, $G \sim P(G|\mathcal{X}) \propto P(\mathcal{X}|G)$. We assume S_0 and $S_0 + \hat{S}_x$ are the prior and posterior covariance matrices and define $\underline{\Sigma}_x = S_0/\nu$, $\overline{\Sigma}_x = (S_0 + \hat{S}_x)/(\nu + T)$. Based on the choice of the prior distribution, the marginal likelihood factorizes into the product of n_y terms, each one involving the response variable (Y^i) and its set of explanatory variables

(π_i) (see Geiger and Heckerman (1999)), i.e.

$$P(\mathcal{X}|G) = \prod_{i=1}^{n_y} P(\mathcal{X}|G(Y^i, \pi_i)) = \prod_{i=1}^{n_y} \frac{P(\mathcal{X}^{(Y^i,\pi_i)}|G)}{P(\mathcal{X}^{(\pi_i)}|G)} \tag{A.13}$$

where n_y is the number of response variables, $G(Y^i, \pi_i)$ is the sub-graph of G with nodes Y^i and the elements of π_i, $\mathcal{X}^{(Y^i,\pi_i)}$ is the sub-matrix of $\mathcal{X}$ consisting of the response variable Y^i and its set of explanatory variables π_i; and $\mathcal{X}^{(\pi_i)}$ is the sub-matrix of the set of explanatory variables of (Y^i). The closed form of the marginal likelihood for a graph G is given as (Heckerman and Geiger, 1994):

$$P(\mathcal{X}^D|G) = (\pi)^{\frac{-n_d T}{2}} \frac{|\overline{\Sigma}_{x,d}|^{-\frac{(\nu+T)}{2}}}{|\underline{\Sigma}_{x,d}|^{-\frac{\nu}{2}}} \prod_{i=1}^{n_d} \frac{\Gamma\left(\frac{\nu+T+1-i}{2}\right)}{\Gamma\left(\frac{\nu+1-i}{2}\right)} \tag{A.14}$$

where $D \in \left\{(Y^i, \pi_i),\ (\pi_i)\right\}$, and $\mathcal{X}^D$ is a sub-matrix of $\mathcal{X}$ consisting of $n_d \times T$ realizations, where n_d is the dimension of D, $|\underline{\Sigma}_{x,d}|$ and $|\overline{\Sigma}_{x,d}|$ are the determinants of the prior and posterior covariance matrices associated with D. The above is referred to as the Bayesian Gaussian equivalent (BGe) metric (Heckerman and Geiger, 1994).

In our application, the parameters of interest are the reduced form coefficients matrix, (A_+), and covariance matrix of the reduced form errors, (Σ_u). These are transformations of the structural parameters, $\{B^*, \Sigma_\varepsilon\}$, which can be estimated from $\Sigma_x = \Omega_x^{-1}$. Following a standard Bayesian approach, we considered two typical BVAR prior settings, i.e. the Minnesota and the normal-Wishart prior. In both cases, we assume $A_+ \sim \mathcal{N}(\underline{A}, \underline{V})$, and compute the posterior mean $(\overline{A}_i)$ and variance $(\overline{V}_i)$ of each VAR equation as follow: $\overline{A}_i = \overline{V}_i(\underline{V}_i^{-1}\underline{A}_i + \overline{\sigma}_i^{-2} W_i' Y^i)$, $\overline{V}_i = (\underline{V}_i^{-1} + \overline{\sigma}_i^{-2} W_i' W_i)^{-1}$, where $\underline{A}_i$ and $\underline{V}_i$, are the prior mean and variance of the relevant variables in each equation, and W_i, is the set of relevant variables that influence the response variable Y^i. Under the assumption of a Minnesota prior distribution, $\overline{\sigma}_i = \sigma_i,,\ i = 1, \ldots, n_y$ are the diagonal elements of $\Sigma_u = \text{diag}(\sigma_1^2, \ldots, \sigma_{n_y}^2)$. Under the normal-Wishart prior distribution, $\overline{\sigma}_i^2$, is the variance of residuals from the posterior of Σ_u, which is assumed to be inverse-Wishart distributed, $\Sigma_u \sim \mathcal{IW}(\underline{\nu}, \underline{S})$.

Algorithm 1 MIN Sampling

Initialize $G^{(1)}$ as $(n_y \times n_y)$ zero matrix
for $t = 1$ to Total iterations **do**
  Pick at random $\rho^{(t)} = (\rho_1^{(t)}, \ldots, \rho_{n_y}^{(t)})$ from the set of the $n_y!$ possible permutations of the integers $\{1, \ldots, n_y\}$
  for $i = 1 : n_y$ **do** set $\tilde{y}_i = y_{\rho_i^{(t)}}$, $\tilde{y}_i \in \mathbf{V}_y$ and $G^* = G^{(t)}$
    Draw a candidate explanatory variable, $y_j \in \mathbf{V}_y \backslash \{y_{\rho_i^{(t)}}\}$
    if $G^*(y_j, \tilde{y}_i) = 1$ **then** set $G^*(y_j, \tilde{y}_i) = 0$
      Add/remove edge; $G^*(\tilde{y}_i, y_j) = 1 - G^*(\tilde{y}_i, y_j)$
      if G^* is acyclic **then** sample $u \sim \mathcal{U}_{[0,1]}$
        Compute the Bayes factor: $BF = P(\mathcal{X}|G^*)/P(\mathcal{X}|G^{(t)})$
        if $u < \min\{1, BF\}$ **then** $G^{(t+1)} = G^*$
        else $G^{(t+1)} = G^{(t)}$
      else $G^{(t+1)} = G^{(t)}$

APPENDIX B. MCMC SAMPLING

The Markov Chain Monte Carlo (MCMC) algorithm described in Madigan and York (1995) is a Metropolis-Hastings (MH). Given a graph G, the algorithm samples a new graph G^* based on a proposal distribution. The new graph is accepted with probability

$$A(G^*|G) = \min\left\{\frac{P(\mathcal{X}|G^*)}{P(\mathcal{X}|G)}\frac{P(G^*)}{P(G)}\frac{Q(G|G^*)}{Q(G^*|G)}, 1\right\} \tag{B.15}$$

where $P(\mathcal{X}|G)$ is the likelihood, $P(G)$ is the prior and $Q(G^*|G)$ is the proposal distribution. For the single edge addition or removal, the proposal distribution assigns a uniform probability to all possible graphs in the neighborhood $nbd(G)$ of G, which is the set of all graphs that can be reached from the current state (G) by adding or deleting a single edge. Following Madigan and York (1995), we considered a symmetric proposal distribution (i.e., $Q(G|G^*) = Q(G^*|G)$). Moreover, by assuming a uniform graph prior the acceptance probability simplifies to the Bayes factor.

We modify the standard MCMC inference scheme to allow for inference of contemporaneous and temporal dependence structures. Let $\mathbf{V}_y$ be n_y vector of response variables, $\mathbf{V}_x$ be np vector of explanatory variables, and $\mathbf{V}_y \backslash \{y_i\}$ the subvector of $\mathbf{V}_y$ obtained by deleting the i-th element of $\mathbf{V}_y$. The pseudo-code for sampling the contemporaneous and temporal structures are given in Algorithms 1 and 2, respectively. To speed up the algorithm, the

Algorithm 2 MAR Sampling

Initialize $G^{(1)}$ as $(n_y \times np)$ zero matrix
for $t = 1$ to Total iterations **do**
  Pick at random $\rho^{(t)} = (\rho_1^{(t)}, \ldots, \rho_{n_y}^{(t)})$ from the set of the $n_y!$ possible permutations of the integers $\{1, \ldots, n_y\}$
  for $i = 1 : n_y$ **do** set $\tilde{y}_i = y_{\rho_i^{(t)}}$, $\tilde{y}_i \in \mathbf{V}_y$ and $G^* = G^{(t)}$
    Draw a candidate explanatory variable, $x_j \in \mathbf{V}_x$
    Add/remove edge; $G^*(\tilde{y}_i, x_j) = 1 - G^*(\tilde{y}_i, x_j)$
    Sample $u \sim \mathcal{U}_{[0,1]}$
    Compute the local Bayes factor: $BF = P(\mathcal{X}|G^*(\tilde{y}_i, \pi_i))/P(\mathcal{X}|G^{(t)}(\tilde{y}_i, \pi_i))$
    if $u < \min\{1, BF\}$ **then** $G^{(t+1)}(\tilde{y}_i, \pi_i) = G^*(\tilde{y}_i, \pi_i)$
    else $G^{(t+1)}(\tilde{y}_i, \pi_i) = G^{(t)}(\tilde{y}_i, \pi_i)$

common approach is to reduce the size of the search space by restricting the maximum number of explanatory variables (fan-in) in the parent set of each response variable. In our application, we do not impose such restriction. However, to reduce the computing time, we pre-compute $\overline{\Sigma}_x$, $\underline{\Sigma}_x$ and the metric in (A.14). Since the proposal involves a single edge addition or removal, we compute the scores of the structures at each iteration by updating only the local scores affected by the move. To sample the MIN structures, we apply the score function in (A.13) and (A.14) by replacing G with the contemporaneous structure G_0 and $\underline{\Sigma}_x$ and $\overline{\Sigma}_x$ with the prior and posterior covariance matrix of the contemporaneous response variables. For the MAR structure inference, we organize our time series into $(1 \times n)$ blocks composed of lagged variables (X) and a dependent variable (Y^i). The matrix X is stacked $X'_{p+1-s}, \ldots, X'_{T-s}$, $1 \leq s \leq p$, such that X is of dimension $np \times (T-p)$, $\mathcal{X}(i) = (X', (Y^i)')'$, where $i = 1, \ldots, n_y$, $\mathcal{X}(i)$ is $(np+1) \times (T-p)$. Let G_+ be a stacked graph of the MAR relationships. Then the marginal likelihood of the MAR structure is

$$P(\mathcal{X}|G_+) = \prod_{i=1}^{n_y} \frac{P\big(\mathcal{X}(i)^{(Y^i, \pi_+^i)}\big|G_+\big)}{P\big(\mathcal{X}(i)^{(\pi_+^i)}\big|G_+\big)} \tag{B.16}$$

where π_+^i is the candidate explanatory variables of Y^i, consisting of the lagged variables in X_+. We score the MAR structure by evaluating (A.13) and (A.14), and by replacing G with G_+, and $\underline{\Sigma_x}$ and $\overline{\Sigma_x}$ with the prior and posterior covariance matrices of $\mathcal{X}(i)$, $i = 1, \ldots, n_y$.

APPENDIX C. GRAPHICAL MODEL EVALUATION

Convergence Diagnostics

We monitor the convergence of the MIN and MAR structures by using the potential scale reduction factor (PSRF) and the multivariate PSRF (MPSRF) of Gelman and Rubin (1992) and Brooks and Gelman (1998), respectively. See also Casella and Robert (2004), ch. 12, for a review on methods for convergence monitoring in MCMC. The PSRF (MPSRF) monitors the within-chain and between-chain covariances of the global (local) log posterior densities of the sampled structures to test whether all the chains converged to the same posterior distribution. The chain is said to have properly converged if $PSRF$ $(MPSRF)$ ≤ 1.2. In all of the simulation and empirical applications, we obtained MPSRF and PSRF for both MAR and MIN less than 1.1, which indicates convergence of the MCMC chain.

Graph Structure Evaluation

We estimate the posterior probability of the edge G_{ij} by $\hat{e}_{ij}$, which is the average of the MCMC samples from the G_{ij} posterior distribution. In order to identify significant explanatory variables for our model, we define a one sided posterior credibility interval for the edge posterior distribution, and find the interval lower bound

$$q_{(1-\alpha)} = \hat{e}_{ij} - z_{(1-\alpha)}\sqrt{\frac{\hat{e}_{ij}(1-\hat{e}_{ij})}{n_{eff}}} \tag{C.17}$$

where n_{eff} is the effective sample size (see Casella and Robert (2004), pp. 499-500) representing the number of independent posterior samples of the graph, and $z_{(1-\alpha)}$ is the z-score of the normal distribution at the $(1-\alpha)$ significance level. Finally we define the estimator $\hat{G}_{ij}$, of the edge from X^j to X^i, as

$$\hat{G}_{ij} = \begin{cases} 1 & \text{if} \quad q_{(1-\alpha)} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{C.18}$$

Clearly as $n_{eff} \to \infty$, $\hat{G}_{ij} = 1$ if $\hat{e}_{ij} > 0.5$.

When the data generating process is known, we refer to the non-zero elements as real positives and the zero elements as real negatives. Based on (C.17), we obtain a connectivity structure, where non-zero elements are referred to as predicted positives and the zero elements as predicted negatives. The metric to evaluate the graph-predictive accuracy is shown in Table III. TP indicates the number of real positives correctly predicted as positives, FP

Table III: Classification table for graph-predictive accuracy evaluation.

	Real Positives	Real Negatives
Predicted Positives	TP	FP
Predicted Negatives	FN	TN

indicates the number of real negatives predicted as positives, TN indicates the number of real negatives correctly predicted as negatives, and FN indicates the number of real positives predicted as negatives. In this paper, we employ the accuracy rate (ACC) to measure the predictive accuracy of the graph structure, that is

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

When the true model is unknown, we compare the graph structure using the BIC:

$$BIC(\hat{G}) = -2\log(P(\mathcal{X}|\hat{G})) + |E|\log(T) \tag{C.19}$$

where $P(\mathcal{X}|\hat{G})$ is the marginal likelihood, $|E|$ is the number of non-zero edges in the estimated graph $\hat{G}$ and T is the number of time series observations.

Model Prediction Accuracy

The evaluation of the accuracy of the estimated model was based on the AIC score (see Gelman et al. (2013)) given by

$$AIC(\hat{M}) = -2\log(P(\mathcal{X}|\hat{M})) + 2|\hat{A}_M| \tag{C.20}$$

where $|\hat{A}_M|$ is the number of estimated coefficients in the model $\hat{M}$, and $\log P(\mathcal{X}|\hat{M})$ is the log predictive score.

PC Algorithm

The PC algorithm, (named after authors Peter Spirtes and Clark Glymour) is a graph-theoretic approach developed by Spirtes et al. (2000) for learning partially directed structures. The algorithm uses conditional independence (e.g the Fisher's z statistic) test to decide whether a particular constraint holds. See Spirtes et al. (2000) for details on implementation of the PC algorithm.

Granger-Causality

Pairwise Granger causality (P-GC) (Granger, 1969) relies on the condition that if the prediction of X_t^i is improved by incorporating lagged observations of X_t^j, then X_t^j has a causal influence on X_t^i. A limitation to this approach in multivariate settings is the inability to discriminate between direct and mediated causal influences. For instance, one variable may influence two other variables with differential time delays, and a pairwise analysis may indicate a causal influence from the variable that receives an early input to the variable that received a late input. Conditional Granger causality (C-GC) (Ding et al., 2006) deals with this limitation by accessing dependence between a pair of time series conditional on other series and their lags. The C-GC relies on block-wise Granger causality test which includes all lags of the pair of variables of interest (see Ding et al. (2006), eqs. 17.30-17.33). In this paper, we modified the C-GC procedure and use pairwise-Granger causality at different lags conditioning on the other series and their lags. In our modified C-GC we estimate the VAR(p) model

$$X_t^i = \sum_{s=1}^{p} \beta_{s,1} X_{t-s}^1 + \ldots + \sum_{s=1}^{p} \beta_{s,i} X_{t-s}^i + \ldots + \sum_{s=1}^{p} \beta_{s,j} X_{t-s}^j + \ldots + \sum_{s=1}^{p} \beta_{s,n} X_{t-s}^n + u_t^i \quad \text{(C.21)}$$

where $i, j \in \{1, \ldots, n\}$, $i \neq j$, and test whether $X^j_{t-s} \to X^i_t$, by checking if the following hypothesis holds true

$$H_0 : \beta_{s,j} = 0, \qquad H_1 : \beta_{s,j} \neq 0$$

with a t-test.

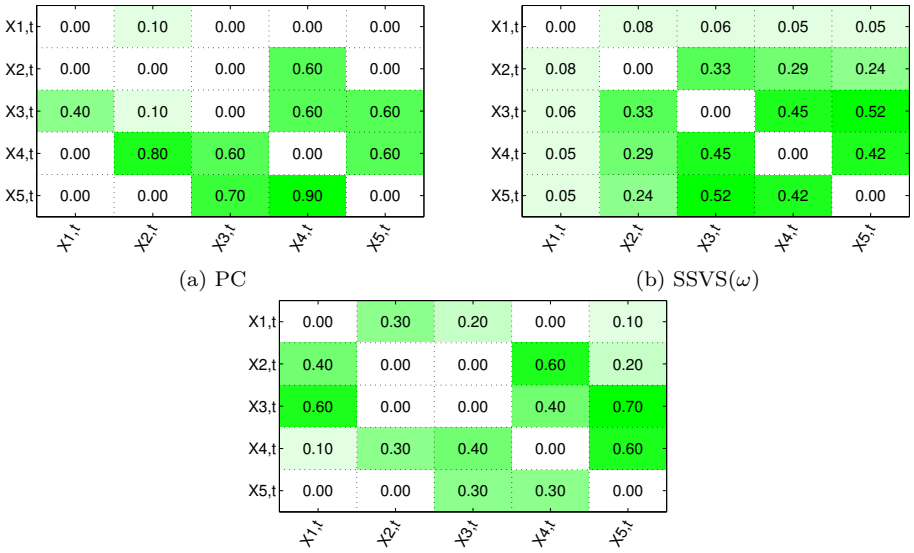
APPENDIX D. APPLICATION TO SIMULATED DATA

In Figure 5, we show the contemporaneous structure of PC, SSVS(ω) and MIN, for the 5-node and $p = 3$ lag model, averaged over the 20 replications. The PC indicates strong evidence of the following relationships: $X^2_t - X^4_t - X^3_t - X^5_t - X^3_t$. The SSVS($\omega$) shows the following: $X^2_t - X^3_t - X^4_t - X^5_t - X^3_t$. The MIN structure indicates the following: $X^1_t \to X^3_t$; $X^4_t \to X^2_t$; $X^5_t \to (X^3_t, X^4_t)$. We conclude that all the three approaches capture similar correlations.

The SSVS(ω) shows only contemporaneous correlations and does not offer the directions of the causal effects. The PC shows partially directed edges with indications that some causal directions are more probable than their reverse, e.g. $P(X^4_t \to X^5_t|\mathcal{X}) = 0.9 > P(X^5_t \to X^4_t|\mathcal{X}) = 0.6$. However the PC indicates a contemporaneous correlations (undirected edge) between the majority of the variables. The MIN structure shows an unambiguous direction of the edges among the variables. From a comparison with the DGP of the contemporaneous structure (B_0), given as: $X^2_t \to X^4_t$; $X^3_t \to X^1_t$; $X^5_t \to X^3_t$, we notice that the MIN shows similar relationships with some edges in the opposite direction of the DGP. A possible explanation relates to issues of Markov equivalence of contemporaneous directed graphical models discussed in Section 2.

Figure 6 shows the temporal structure of C-GC, SSVS(γ) and MAR for the 5-node $p = 3$ lag model averaged over 20 replications. The figure shows that C-GC and MAR detect more strong evidence of dependence than SSVS(γ). MAR presents a better inference of the DGP than C-GC and SSVS(γ).

In Figure 7, we report the estimates of the reduced-form coefficient matrices. For the sake of conciseness, a comparison between BVAR (normal-Wishart), SSVS and BGVAR



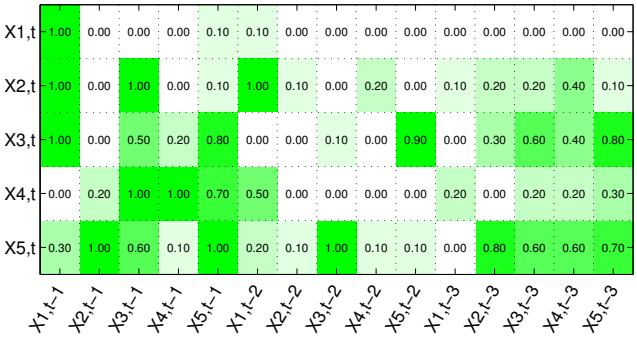
(a) PC (b) SSVS(ω)

(c) MIN

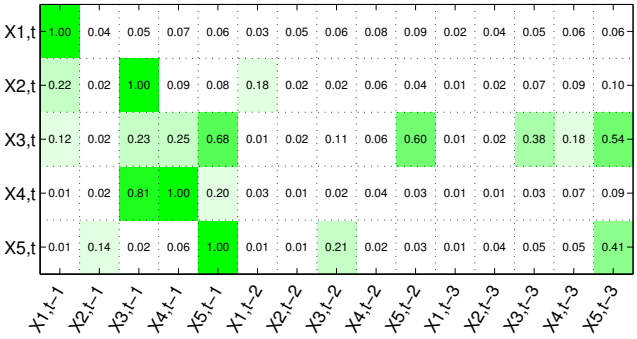
Figure 5: Contemporaneous structure of the 5-node model ($p = 3$) averaged over 20 replications. Response (explanatory) variables are on rows (columns). The light (dark) green indicate weak (strong) dependence.

(normal-Wishart) is provided only for the 5-node DGP with $p = 3$. Negative, positive and null elements are represented in red, green, and white, respectively. The BVAR and the SSVS coefficient matrices look dense whereas that of the BGVAR model are sparse (with a lot of null elements). The SSVS and the BVAR matrices are similar with most of the elements in the BVAR slightly greater (in absolute values) than their SSVS counterparts. This is not surprising since the SSVS draws the coefficient from a mixture of two normal densities. Since a priori one of the two mixture components is peaked at zero, SSVS favors a shrinking-to-zero of the coefficients. The BGVAR model on the other hand is more parsimonious than the BVAR and the SSVS.

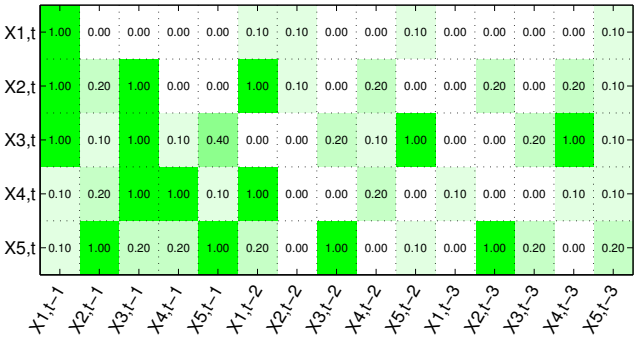
APPENDIX E. ECONOMIC APPLICATIONS



(a) C-GC



(b) SSVS(γ)

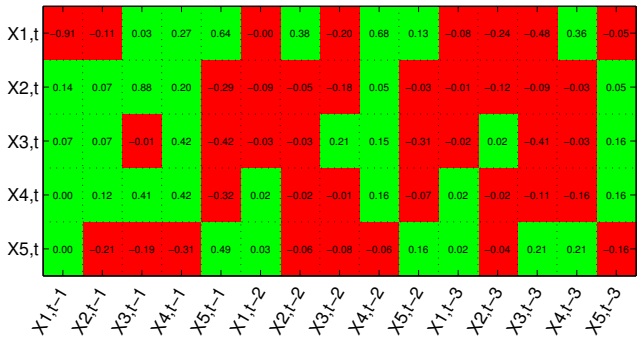


(c) MAR

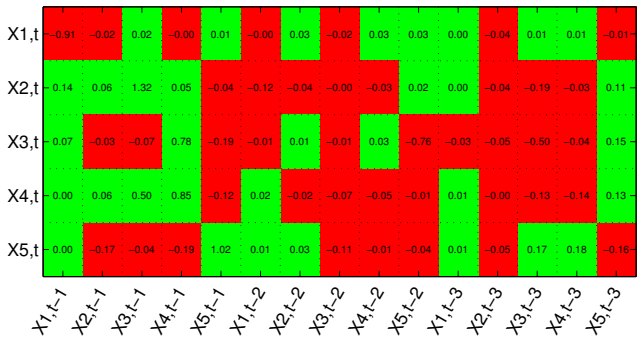
Figure 6: Temporal structure of the 5-node model ($p = 3$) averaged over 20 replications. Response (explanatory) variables are on rows (columns). The light (dark) green indicates weak (strong) dependence.

Macroeconomic and Financial Data Description

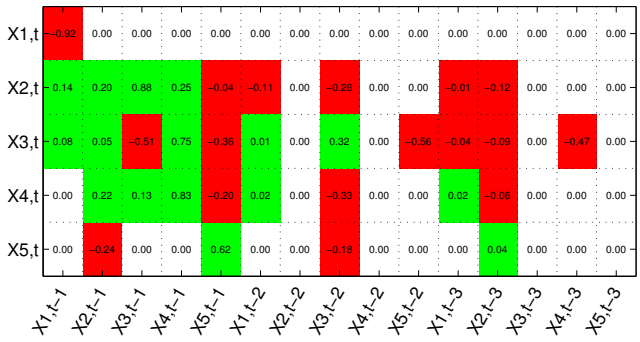
Table IV gives the data description and transformation code from Koop (2013) used for our macroeconomic application. Table V presents the description of the Euro Stoxx 600 super-sectors obtained from Datastream for our financial application.



(a) BVAR-NW Coefficient Matrix



(b) SSVS Coefficient Matrix



(c) BGVAR-NW Coefficient Matrix

Figure 7: Coefficients matrix of BVAR (normal-Wishart), SSVS, and BGVAR (normal-Wishart), for the 5-node model with lag order $p = 3$. Response (explanatory) variables are on the rows (columns). Elements in red (green) represent negative (positive) coefficients, white elements for zeros.

Table IV: Data description and transformation code to achieve stationarity. The transformation code is as follows: 1 = no transformation, 2 = first difference, 3 = second difference, 4 = log, 5 = first difference of the log variable, 6 = second difference of the log variable. (*) Added to augment the response variable vector.

Short ID	Mnemonic	Code	Description
	Response Variables		
Y	GDP251	5	Real GDP, Quantity Index (2000=100)
Π	CPIAUCSL	6	CPI All Items
R	FYFF	2	Interest rate: Federal funds (effective) (% per annum)
M	FM2	6	Money stock: M2 (bil$)
C	GDP252	5	Real Personal Cons. Exp., Quantity Index
IP	IPS10	5	Industrial production index: total
U	LHUR	2	Unemp. rate: All workers, 16 and over (%)
I^*	GDPIC96	5	Real gross domestic private investment
	Predictor Variables		
MP	PSCCOMR	5	Real spot market price index: all commodities
NB	FMRNBA	3	Depository inst reserves: non-borrowed (mil$)
RT	FMRRA	6	Depository inst reserves: total (mil$)
CU	UTL11	1	Capacity utilization: manufacturing (SIC)
HS	HSFR	4	Housing starts: Total (thousands)
PP	PWFSA	6	Producer price index: finished goods
PC	GDP273	6	Personal Consumption Exp.: price index
HE	CES275R	5	Real avg hrly earnings, non-farm prod. workers
$M1$	FM1	6	Money stock: M1 (bil$)
SP	FSPIN	5	S&Ps common stock price index: industrials
IR	FYGT10	2	Interest rate: US treasury const. mat., 10-yr
ER	EXRUS	5	US effective exchange rate: index number
EN	CES002	5	Employees, non-farm: total private

Table V: Description of Euro Stoxx 600 super-sectors. (∗) Included from September 2008.

Short ID	Description	Sector
	Financial	
BK	Banks	Financial
IN	Insurance	Financial
FS	Financial Services	Financial
RE^*	Real Estates	Financial
	Non-Financial	
CM	Construction & Materials	Industrial
IGS	Industrial goods & services	Industrial
AP	Automobiles & Parts	Consumer Goods
FB	Food & Beverage	Consumer Goods
PHG	Personal & Household Goods	Consumer Goods
RT	Retail	Consumer Services
MD	Media	Consumer Services
TL	Travel & Leisure	Consumer Services
CH	Chemicals	Basic Materials
BR	Basic Resources	Basic Materials
OG	Oil & Gas	Oil & Gas
TC	Telecom	Telecom
HC	Health Care	Health Care
TG	Technology	Technology
UT	Utilities	Utilities

Robustness Check for the Macroeconomic Application

In our sensitivity analysis, we augment the set of response variables of the macroeconomic model with the gross domestic private investment. Figure 8 shows the results of PC, SSVS(ω) and MIN for the augmented model.

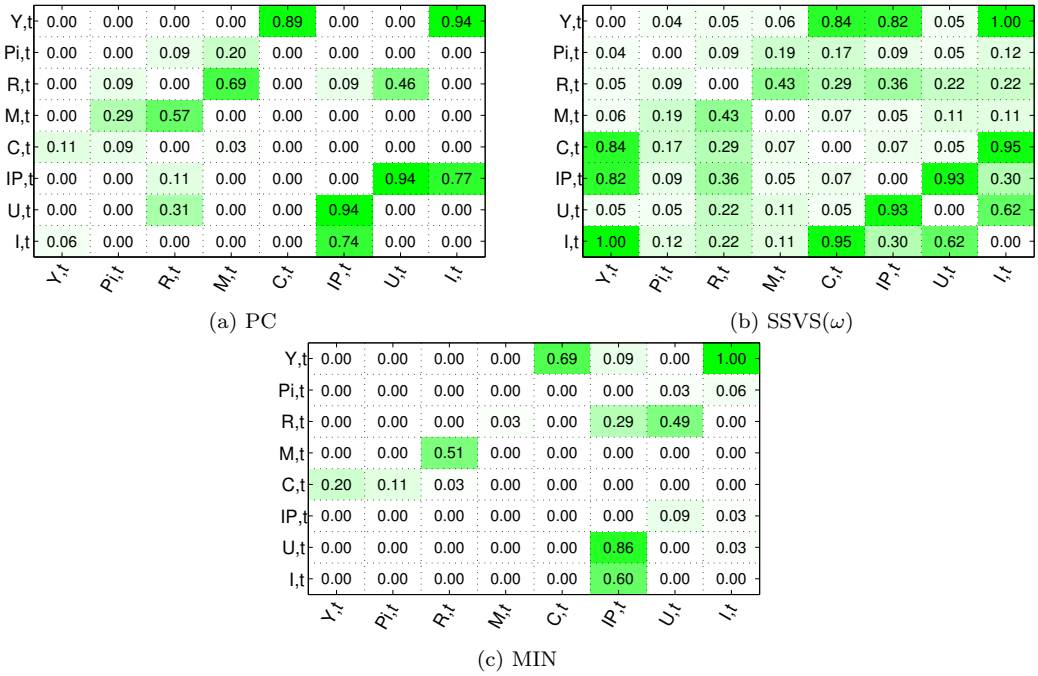


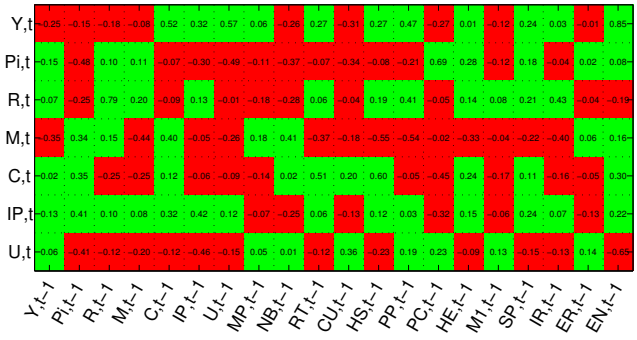
Figure 8: Contemporaneous structure of the PC, SSVS and MIN averaged over $1960Q1 - 2006Q4$. The light (dark) green color indicates weak (strong) evidence of dependence. Response (explanatory) variables are on the rows (columns). The variables are: (Y) - real gross domestic product, (Pi) - consumer price index, (R) - Federal funds rate, (M) - money stock M2, (C) - real personal consumption expenditure, (IP) - industrial production index, (U) - unemployment rate and (I) - gross domestic private investment.

In order to show the effects of the omission of relevant variables on the causal structure estimates, we compare the results of PC, SSVS and MIN, when investment is included as a response variable. The structures of both PC and MIN show no evidence of the effect of real GDP on consumption and no direct effect of industrial production on real GDP (see Figure 8 in APPENDIX E). However, they record strong evidence of the effects of consumption on real GDP. In addition, they capture strong evidence of the effect of industrial production on investment and a direct effect of investment on real GDP. The SSVS(ω) approach on the other hand maintains a strong correlation between industrial production and real gross

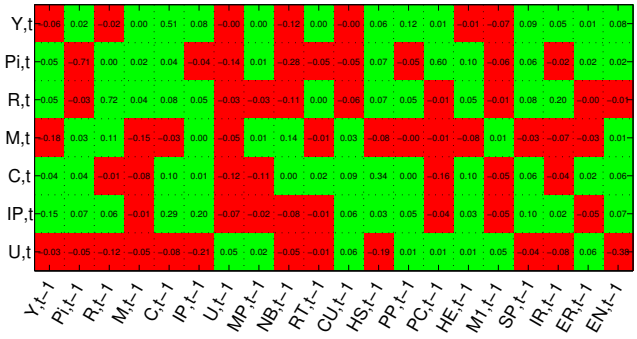
domestic product even when investment is included. Consequently, in our application the MIN and PC dependence structures are sensitive to the exclusion of relevant variables, as opposed to the SSVS procedure. On the contrary in the SSVS procedure, the edges between variables may be wrongly inferred, also when all relevant variables are included in the model. These results are due to the fact that MIN (PC) uses directed (partially directed) edges to represent the contemporaneous dependence structure while SSVS(ω) uses undirected relationships among variables. As a result, SSVS(ω) does not provide any information on the direction of influence among variables and is less sensitive to the choice of the variable to include in the analysis.

Estimated Parameters for the Macroeconomic Application

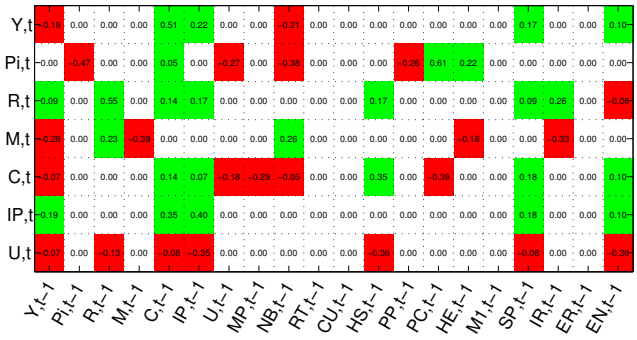
We compare the estimated coefficient matrices of the BVAR (normal-Wishart), SSVS, and BGVAR (normal Wishart prior) for the last window ($1993Q1 - 2007Q4$) of the sample data. This is shown in Figure 9. Elements in red (green) represent negative (positive) coefficients and white elements represent zeros. Most of the BAVR coefficient values are slightly higher (in absolute values) than the corresponding SSVS coefficient values. These results confirm that the SSVS approach acts as a parameter shrinkage and does not ignore unimportant variables. The BGVAR model distinguishes instead the relevant explanatory variables from the unimportant ones. In Figure 9, the coefficients of the unimportant explanatory variables are represented in white color. Having a large number of white elements (zero coefficients) implies that, unlike BVAR and SSVS, BGVAR adopts a framework where response variables can be determined by only a handful of explanatory variables. For instance, forecasting real GDP (Y_{t+1}) for the period $2008Q1 - 2008Q4$ depends only on real GDP (Y_t), consumption (C_t), industrial production (IP_t), non-borrowing reserves (NB_t), S&P500 index (SP_t) and employment (EN_t). A similar observation holds for the other response variables. These results confirm that the BGVAR model is more parsimonious than BVAR and SSVS and reduces the future cost of predictions.



(a) BVAR Coefficients



(b) SSVS Coefficients



(c) BGVAR Coefficients

Figure 9: Coefficients matrix of BVAR (normal-Wishart), SSVS, and BGVAR (normal-Wishart) for the period $1993Q1-2007Q4$. Response (explanatory) variables are on the rows (column). Elements in red (green) represent negative (positive) coefficients, white elements for zeros. The variables are: (Y) - real GDP, (Pi) - consumer price index, (R) - Federal funds rate, (M) - money stock M2, (C) - real personal consumption, (IP) - industrial production index, (U) - unemployment, (MP) - real spot market price, (NB) - non-borrowed reserves, (RT) - total reserves, (CU) - capacity utilization, (HS) - housing, (PP) - producer price index, (PC) - personal consumption expenditure, (HE) - real average hourly earnings, $(M1)$ - money stock M1, (SP) - S&P500 index, (IR) - 10-yr US treasury bill rate, (ER) - US effective exchange rate, (EN) - employment.