Università
Ca'Foscari
Venezia

Scuola Dottorale di Ateneo
Graduate School

Dottorato di ricerca
in Economia
Ciclo XXVII
Anno di discussione 2015

Bayesian Graphical Models with Economic and Financial Applications

Settore scientifico disciplinare di afferenza: SECS-P/05

Tesi di Dottorato di Daniel Felix Ahelegbey, Matricola: 833894

Coordinatore del Dottorato
Prof. Michele Bernasconi

Tutore del Dottorando
Prof. Monica Billio

Co-tutore del Dottorando
Prof. Roberto Casarin

Bayesian Graphical Models with Economic and Financial Applications

by

Daniel Felix Ahelegbey

Matricola 833894

A thesis submitted in partial fulfillment
for the degree of

Doctor of Philosophy in Economics

Department of Economics

Ca' Foscari University of Venice,

Graduate School of Economics and Management, Veneto

Supervisors :

Prof. Monica Billio
Prof. Roberto Casarin

July, 2015

Declaration

I, Daniel Felix Ahelegbey, hereby declare that this Ph.D thesis titled " Bayesian Graphical Models with Economic and Financial Applications" is a presentation of my original research work for the degree of Doctor of Philosophy in Economics under the guidance and supervision of Prof. Monica Billio and Prof. Roberto Casarin. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

I affirm that this work has not been submitted previously in the same or similar form to another examination committee at another department, university or country.

Signed:

_____

Date:

_____

A spider conducts operations that resemble those of a weaver, and a bee puts to shame many an architect in the construction of her cells. But what distinguishes the worst architect from the best of bees is this, that the architect raises his structure in imagination before he erects it in reality.

KARL MARX (1818 – 1883)

Imagination is more important than knowledge. For knowledge is limited to all we now know and understand, while imagination embraces the entire world, and all there ever will be to know and understand.

ALBERT EINSTEIN (1879 – 1955)

Abstract

Recent advances in empirical finance has seen a considerable amount of research in network econometrics for systemic risk analysis. The network approach aims to identify the key determinants of the structure and stability of the financial system, and the mechanism for systemic risk propagation. This thesis contributes to the literature by presenting a Bayesian graphical approach to model cause and effect relationships in observed data. It contributes specifically to model selection in moderate and high dimensional problems and develops Markov chain Monte Carlo procedures for efficient model estimation. It also provides simulation and empirical applications to model dynamics in macroeconomic variables and financial networks.

The contributions are discussed in four self contained chapters. Chapter 2 reviews the literature on network econometrics and the Bayesian approach to graph structural inference with potential applications. Chapter 3 proposes a Bayesian graphical approach to identification in structural vector autoregressive models. Chapter 4 develops a model selection to multivariate time series of large dimension through graphical vector autoregressive models and introducing sparsity on the structure of temporal dependence among the variables. Chapter 5 presents a stochastic framework for financial network models by proposing a hierarchical Bayesian graphical model that can usefully decompose dependencies between financial institutions into linkages between different countries financial systems and linkages between banking institutions, within and/or across countries.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

G	A graph structure or a 0/1 connectivity matrix.
θ	the model parameters.
$\mathcal{G}$	space of the graphs.
Θ	the parameter space.
$G_{ij} = 1$	variable index j influences variable index i ($X_j \rightarrow X_i$).
$G \circ \Phi$	Hadamard's product of two matrices G and Φ
BGVAR	Bayesian graphical vector autoregression.
CAPM	Capital asset pricing model.
DAG	Directed acyclic graph.
ENET	Elastic-net.
LASSO	Least Absolute Shrinkage and Selection Operator
MAR	Multivariate autoregression.
MCMC	Markov chain Monte Carlo.
MIN	Multivariate instantaneous.
PDAG	Partially directed acyclic graph.
SBGVAR	Sparse Bayesian graphical vector autoregression.
SSVS	Stochastic Search Variable Selection.
SVAR	Structural vector autoregression.
UGM	Undirected graphical model.
VAR	Vector autoregression.

Acknowledgements

I ascribe the greatest thanks and give glory to the Almighty God for the gift of life, energy and the wisdom to undertake this study.

It is with great pleasure that I acknowledge my supervisors, Monica Billio and Robert Casarin for their guidance and constant support during the development of this thesis. I am thankful for the opportunity to have me collaborate with professors and researchers during my years as a Ph.D student.

I gratefully acknowledge the PhD scholarship award from the University Ca' Foscari of Venice, funding from the European Union Grant Agreement for SYstemic Risk TOmography (SYRTO) Project, and the Italian Ministry of Education, University and Research (MIUR) PRIN Grant, MISURA.

I acknowledge Fabrizio Lillo (Scuola Normale Superiore, Pisa, Italy), Paolo Giudici (University of Pavia, Italy) and Siem Jan Koopman (Vrije Universiteit Amsterdam, Netherlands) for their support and accepting my visit to their respective institutions. Many thanks goes to editors and referees of Journal of Applied Econometrics and Advances in Econometrics. I am grateful to Domenico Sartore, Herman K. van Dijk, Fabio Canova, Sylvia Frühwirth-Schnatter, Luc Bauwens, Gaetano Carmeci, Marco Del Negro, Francesco Ravazzolo, Fabrizio Leisen, Fulvio Corsi and participants at various conferences, seminars and workshops for stimulating discussions.

Special thanks and blessings to my parents, Gershon K. Ahelegbey and Abla A. Gayina, my siblings, relatives, friends and love ones both in Ghana and abroad who have been part of my dreams and aspirations.

Chapter 1

General Introduction

The functioning of a system is an emergent consequence of the actions of the individuals that constitute it. To analyze systemic action with quantitative data, however, requires a formal theoretical model that relates individual actions to systemic functioning (Coleman, 1986). The global financial crisis of 2007-2009 has stressed the need to understand the world financial system as a network of interconnected institutions, where financial linkages play a fundamental role in the spread of contagion. In trying to measure and address this interconnectedness, researchers have recently proposed network models, that can help model the systemic risk in financial systems which display complex degrees of connectedness.

The thesis contributes to address inferential difficulties in network models by advancing the application of Bayesian graphical models; developing efficient Markov Chain Monte Carlo algorithms for moderate and high-dimensional model selection; and providing simulation and empirical applications to understand modern economic and financial systems.

The contributions are presented in four chapters.

In Chapter 2, we survey the state of the arts for statistical inference and application of networks from a multidisciplinary perspective, and specifically in the context of systemic risk. We contribute to the literature on network econometrics by relating

network models to multivariate analysis with potential applications in econometrics and finance.

In Chapter 3, we propose a Bayesian graphical approach suitable for identification in structural vector autoregression by modeling cross-sectional and temporal dependence in multivariate time series. The novelty of the approach is its ability to handle the restrictions directly on the structural model.

In Chapter 4, we develop a graphical approach to model dependence in high dimensional multivariate time series and to address over-parametrization in large vector autoregressive models. The methodology discussed in this chapter is based on combining graphical model notion of causality with a new sparsity prior distribution on the graph space, able to deal with model selection problems in multivariate time series of large dimension.

Chapter 5 presents a stochastic framework for financial network models by proposing a hierarchical Bayesian graphical model that can usefully decompose dependencies between financial institutions into linkages between different countries financial systems and linkages between banking institutions, within and/or across countries.

Chapter 2 is currently under revision for submission. Chapter 3 has been accepted for publication as: Ahelegbey, D. F., Billio, M. and Casarin, R. (2015). Bayesian Graphical Models for Structural Vector Autoregressive Processes, *Journal of Applied Econometrics*, forthcoming. Chapter 4 is a joint work with Monica Billio and Roberto Casarin and has been submitted to Journal of Econometrics. Chapter 5 has been published as: Ahelegbey, D. F. and Giudici, P. (2014). Bayesian Selection of Systemic Risk Networks, *Advances in Econometrics: Bayesian Model Comparison*, vol. 34, 117–153.

Chapter 2

The Econometrics Aspects Networks: A Review

2.1 Introduction

With an increase in globalization, the concept of networks has become an integral part of every human activity and forms a backbone of our modern society. In recent times, actions of economic agents in some local economies have had great impacts on a global scale due to strong connections among individuals, institutions and markets. These connections play a fundamental role in the spread of information and risk. Many authors have come to the same conclusion that the actions of major financial institutions have significant consequences on the stability or public confidence in the entire system (see, e.g. Billio et al., 2012; Battiston et al., 2012; Acharya et al., 2010; Brunnermeier and Pedersen, 2009).

Network models have become a center of attraction for modeling dependencies in real world phenomena due to simplicity in presentation, and their ability to provide an intuitive way to visualize and interpret complex relationships. For introduction to these models, (see Pearl, 2000; Whittaker, 1990; Edwards, 2000; Lauritzen, 1996;

This chapter is based on: Ahelegbey (2015). *The Econometrics Aspects Networks: A Review.* (Working Paper).

Spirtes et al., 2000). This paper surveys the state of the arts for inference on networks from a multidisciplinary perspective, as motivated by applications in forensic science, genetics, machine learning, etc.

The financial crisis of 2007-2009 has prompted new research interests on the need to understand the structure of the financial system and its risk propagation mechanisms. In providing a framework for strengthening financial stability, policy makers are currently not only refining the regulatory and institutional set-up, but also looking for analytical tools to better identify, monitor and address systemic risk. According to Bernanke (2013), the 2007-2009 crisis was triggered by losses suffered by holders of subprime mortgages and amplified by the vulnerabilities (structural weaknesses) in the financial system. He pointed out that in the absence of these vulnerabilities, the triggers might produce sizable losses to certain firms, investors, or asset classes but would generally not lead to full-blown financial crises. Many authors have found the complex interconnectedness among financial institutions and markets as the potential channels that magnified the initial shocks to the system (see, Bernanke, 2010; Billio et al., 2012; Diebold and Yilmaz, 2014).

To understand the vulnerabilities in the financial system, the idea of network analysis has been shown to be a promising tool with the potential to help in monitoring the interconnectedness of financial institutions and markets. This has led to considerable increase in research on statistical properties of network measures for systemic risk analysis (see Battiston et al., 2012; Billio et al., 2012; Diebold and Yilmaz, 2014; Barigozzi and Brownlees, 2014; Hautsch et al., 2014). While interconnectedness have been widely studied, only few papers review network analysis in finance and they do not focus on the econometrics aspects. This paper contributes to the recent and expanding stream of literature on network econometrics (e.g. Diebold and Yilmaz, 2014), by reviewing the methodological and applied aspects of networks specifically in the context of systemic risk. We present potential applications of networks to

econometrics and financial time series.

Despite the advancement in research tools for network analysis, issues of inferential difficulties has not been adequately addressed. In many real world problems, relationships among variables of interest are more complex than pairwise. In a system characterized by complex interactions, the standard practice of identifying a single model that summarize relationships among variables often ignore the problem of model uncertainty. It is well known that inference of a network is a model determination problem where the number of candidates increases super-exponentially with the number of variables (Chickering et al., 2004). Therefore, selecting a single model from these set of candidates is a challenging problem.

We present the Bayesian approach to network inference that takes into account the model uncertainty problem, which also allows us to incorporate prior information where necessary and to perform model averaging. The approach discussed in this paper is closely related to the use of Gaussian graphical models in time series analysis by Carvalho et al. (2007); Carvalho and West (2007); Wang and Li (2012); Dahlhaus and Eichler (2003). This paper is related to Eichler (2007) and Zou and Feng (2009), who present the network approach as a valid alternative to the Granger concept for causality identification and it extensions in the econometrics literature (Granger, 1969; Diks and Panchenko, 2005; Hoover, 2001).

We demonstrate the effectiveness of the Bayesian inference to network identification to analyze the interconnectedness of the return indexes of the major financial sectors considered by (Billio et al., 2012), and the volatility connectedness in the European stock market. Our result corroborate the findings of Billio et al. (2012), with evidence of a higher vulnerability in the financial system between 2001-2008 that amplified the crisis in 2007-2009. We also find evidence that Banks and Insurers are central in the "fear connectedness" (Diebold and Yilmaz, 2014) expressed by market participants in the financial sector of the Euro-area.

This paper proceed as follows. In Section 2.2, we survey the state of the arts on network models from a multidisciplinary perspective. In Section 2.3, we review the literature on networks in financial systemic risk. In Section 2.4, we relate network models to multivariate analysis with potential applications in econometrics and finance. In Section 2.5, we present the Bayesian approach to network inference. In Section 2.6 and 2.7, we illustrate applications of the Bayesian inference to network identification to analyze return and volatility connectedness respectively.

2.2 A Review of Graphical Models

Network (graphical) modeling is a class of multivariate analysis that uses graphs to represent statistical models. The graph displays explicitly the conditional independencies among variables. We introduce the essential concepts of graphical models, and review the state of the arts for statistical inference and application of networks from a multidisciplinary perspective.

Graphical models are represented by $(G, \theta) \in (\mathcal{G} \times \Theta)$, where G is a graph structure, θ is the model parameters, $\mathcal{G}$ is the space of the graphs and Θ is the parameter space. The graph consists of *nodes* (or vertices) as variables and *edges* (or links) as interactions.

A graph whose edges are without directions is an *undirected graph*. They are often referred to as *Markov networks*. They produce a class of models commonly known as undirected graphical models suitable for modeling similarity and correlated behaviors among variables.

A graph with directed edges is a *directed graph* and a directed graph without cycles is a *directed acyclic graph* (DAG). DAGs are typically based on the concept of family ordering. For instance in $A \to B \to C$, A is a *parent* of B and C is a *child* of B, A and B are *ancestors* of C, and B and C are *descendants* of A. $C \to A$ is illegal since a descendant cannot be his/her own ancestor. This type of ordering is

suitable for expressing causal relationships.

A *partially directed acyclic graph* is a type of DAG that allows for undirected or bi-directed edges. They are often referred to as essential or chain graphs and are suitable for applications where a unique direction of influence cannot be ascribed to interactions among some variables. They represent a class of Markov equivalent DAGs. Two or more DAGs are said to be *Markov equivalent* if they depict the same set of conditional independence relationships. For example, $A \rightarrow B \rightarrow C$, $A \leftarrow B \rightarrow C$ and $A \leftarrow B \leftarrow C$ are Markov equivalent, since they all represent A and C conditionally independent given B.

A *bipartite graph* is an undirected graph whose variable are categorized into two sets, such that nodes in one set can only interact with those in the other set, and no two nodes in the same set are connected. This type of graphs belongs to the class of *color graphs* where each variable is assigned a color such that no edge connects identically colored nodes. Thus a bipartite graph is equivalent to a two-colored graph. A *factor graph* is a bipartite graph whose nodes are categorized into variables and factors and represented by different shapes. Variables nodes are often represented by circles and factor nodes by squares.

A graph is *complete* if all vertices are connected. A *clique* is a subsect of vertices that are completely connected. Let V be the set of vertices and $V_A \subseteq V$, then G_A is defined as a *subgraph* on nodes in V_A. The triple $(V_A,\ V_B,\ V_C) \subseteq V$ forms a *decomposition* of a graph G if $V = V_A \cup V_B \cup V_C$ and $V_C = V_A \cap V_B$ such that G_C is complete and *separates* G_A and G_B. The subgraph G_C is called a *separator*. The decomposition is *proper* if $V_A = \varnothing$ and $V_B = \varnothing$. A sequence of subgraphs that cannot be further properly decomposed are the *prime components* of a graph. A graph is *decomposable* if it is complete or if every prime component is complete.

2.2.1 Statistical Inference

Statistical inference of the graph structure is central to the model estimation. The common methods are; (i) constraint-based, (ii) score-based and (iii) hybrid approach.

Constraint-based approach to graph selection involves the use of statistical tests to identify conditional independence relationships among variables. The outcome of these tests are used to constrain the graph selection process to identify the most plausible graph that is consistent with those constraints obtained. The widely applied constraint-based inference is the PC algorithm based on Fisher's Z-test (Spirtes et al., 2000).

Score-based approach to graph selection is typically based on assigning a score to each candidate graph. The score represents the goodness of fit of the graph given the data. This approach involves a search over the set of candidates which minimizes a penalized likelihood score. Another aspect of this approach is Bayesian in nature which usually involves priors and posterior computations taking advantage of model averaging to address the model uncertainty problem. Example of algorithms in this framework are greedy search, simulated annealing, Markov chain Monte Carlo, genetic algorithms (Friedman and Koller, 2003; Madigan and York, 1995; Giudici and Green, 1999).

The Hybrid approach combines techniques from constraint-based and score-based inference for graph selection. The hybrid method is designed to adopt the constraint-based reasoning as an initial step to restrict the search space for the application of the score-based scheme. The application of regularization methods such as the Least Absolute Shrinkage and Selection Operator (LASSO) and its variants can also be classified under this approach Tibshirani (1996); Meinshausen and Bühlmann (2006).

2.2.2 Applications and Developments

Graphical models have contributed to modeling challenging and complex real world phenomena in several fields. We review some of the recent developments.

They have been applied in forensics to provide tools to aid in reasoning under uncertainty. For example, Bayesian networks has been identified as a suitable tool for analyzing evidence in complex legal and crime cases (Wright, 2007; Dawid, 2003).

The study of gene interactions has become increasingly important because such information can be used as a basis for treating and diagnosing diseases which contributes to human understanding of biological processes. Several researchers have applied graphical models to model gene interactions to detect conditional dependencies (Friedman et al., 2000; Hensman et al., 2013). Many authors have studied the estimation of high-dimensional gene expressions by relaxing the assumption of normality. This is motivated by the observation that gene data are typically characterized with heavy tails or more outliers. The active research in this area focus on a generalized framework to account for non-normality and outliers (Miyamura and Kano, 2006; Sun and Li, 2012).

Several studies in biological network have revealed a complex hierarchical organization of cellular processes which poses great challenge to researchers. This has led to an active area of research with focus on designing algorithms to detect hierarchical modularity (Hao et al., 2012; Ravasz, 2009), latent variable models (Choi et al., 2011; Liu and Willsky, 2013), and hubs (Tan et al., 2014; Akavia et al., 2010).

Many real world phenomena, such as images, speech and human activities often exhibit regularities though characterized with uncertainty. Graphical models have been applied for structural representation of patterns in these phenomena. Many research works have advanced state of the arts algorithms to track structural patterns, image processing, speech and handwritten recognition and other wide variety of

applications. See Koller and Friedman (2009); Bishop (2006).

Graphical models have been advanced in multivariate analysis specifically in multiple regression problems. To deal with high dimensional problems, parsimony of the model is critical in achieving reasonable performances with limited sample size. Different research directions have been considered for building a parsimonious model. Several authors approach the problem by considering sparsity (Yuan and Lin, 2006; Fan and Peng, 2004). Others consider the reduced-rank approach (Bunea et al., 2011; Chen et al., 2013b) and for some other, the sparse reduced-rank approach (Chen and Huang, 2012; Lian et al., 2015).

Developing efficient large-scale algorithms for big data and high-dimensional problems has increasingly become an active area of in machine learning and statistics. A common approach to inference of graphical models is a centralized learning algorithm which is often hindered by restrictive resource constraints such as limited local computing, limited memory and expensive computational power especially when dealing with high dimensional problems. An active area of machine learning is the development of a decentralized system of distributed algorithms for high dimensional problems (see Liu and Ihler, 2012; Meng et al., 2013).

2.3 A Review of Network Aspect of Systemic Risk

Systemic risk as defined by Billio et al. (2012) is "any set of circumstances that threatens the stability or public confidence in the financial system". The European Central Bank (ECB) (2010) defines it as a risk of financial instability "so widespread that it impairs the functioning of a financial system to the point where economic growth and welfare suffer materially". A comprehensive review is discussed in (Brunnermeier and Oehmke, 2012; De Bandt and Hartmann, 2000; Acharya et al., 2010). Several authors have come to the same conclusion that the likelihood of major financial systemic crisis is related to the degree of correlation among the holdings of financial

institutions, how sensitive they are to changes in market prices and economic conditions, how concentrated the risks are among those financial institutions, and how closely linked they are with each other and the rest of the economy (Battiston et al., 2012; Brunnermeier and Pedersen, 2009).

Several systemic risk measures are discussed in the literature. Among them are: Banking System's Portfolio Multivariate Density (BSPMD) by Segoviano and Goodhart (2009); Conditional Value-at-Risk (CoVaR) by Adrian and Brunnermeier (2010); Absorption Ratio (AR) by Kritzman et al. (2010); Marginal Expected Shortfall (MES) by (Acharya et al., 2010; Brownlees and Engle, 2011); Distressed Insurance Premium (DIP) by Huang et al. (2012); Dynamic Causality Index (DCI) and Principal Component Analysis Systemic (PCAS) risk measures by Billio et al. (2012); Network Connectedness Measures (NCM) by Diebold and Yilmaz (2014). The BSPMD embed banks's distress inter-dependence structure, which captures distress dependencies among the banks in the system; CoVaR measures the value-at-risk (VaR) of the financial system conditional on an institution being under financial distress; AR measures the fraction of the total variance of a set of (N) financial institutions explained or "absorbed" by a finite number $(K < N)$ of eigenvectors; MES measures the exposure of each individual firm to shocks of the aggregate system; DIP measures the insurance premium required to cover distressed losses in the banking system; DCI capture how interconnected a set of financial institutions is by computing the fraction of significant Granger-causality relationships among their returns; PCAS captures the contribution of an institution to the multivariate tail dynamics of the system; NCM aggregates the contribution of each variable to the forecast error variance of other variables across multiple return series.

As to whether systemic risk can be reliably identified in advance, the chairman of Board of Governors of the Federal Reserve System, Ben Bernanke, distinguished between triggers and vulnerabilities as two set of factors that led to and amplified

the crisis of 2007-2009, and only by understanding the factors can we hope to guard against a repetition (Bernanke, 2013). The triggers are considered as the particular events or factors that touched off the crisis, i.e., the proximate causes. According to Bernanke (2010), the most prominent triggers of the crisis was the prospect of significant losses on residential mortgage loans to sub-prime borrowers that became apparent shortly after house prices began to decline. The vulnerabilities are attributed to structural weaknesses in the financial system and in regulation and supervision that amplified the initial shocks. Examples of such factors include high levels of leverage, interconnectedness, and complexity, all of which have the potential to magnify shocks to the financial system (Bernanke, 2010).

In the absence of vulnerabilities, triggers might produce sizable losses to certain firms, investors, or asset classes but would generally not lead to full-blown financial crises (Bernanke, 2013). For example, the collapse of the relatively small market for sub-prime mortgages would not have been nearly as consequential without pre-existing fragilities in securitization practices and short-term funding markets which greatly increased its impact. This assertion somehow corroborates the findings of Billio et al. (2012) who found that triggers of both the 1998 Long Term Capital Management (LTCM) crisis and the Financial Crisis of 2007-2009 were associated with liquidity and credit problems. In their analysis of the state of the financial system during the periods, the authors found that the 2007-2009 period experienced the largest number of interconnections compared other time periods. This suggests that the vulnerability of the system was much higher in 2007-2009 than in 1998 and hence the severity in the impact of the latter event which affected a much broader aspect of financial markets and threatened the viability of several important financial institutions, unlike the former.

To monitor and guard against a repetition of the 2007-2009 crisis, the attempt is to identify and address vulnerabilities in the financial environment to ensure ro-

bustness of the system. To this end, network analysis is a promising tool with the potential to help us better monitor the interconnectedness of financial institutions and markets (Bernanke, 2013). Many empirical studies have shown that networks reflect the architecture of interactions that arise among financial institutions and can provide insight into the structure and stability of the system. Some of the applications of the network tools include measuring the degree of connectivity of particular financial institutions to determine systemic importance, forecasting the likely contagion channels of institutional default or distress, and visualizing the "risk map" of exposure concentrations and imbalances in the system (Bisias et al., 2012). To ensure a robust system and to guard against risks that threatens the stability of the financial system, it is of crucial importance to: (a) identify systemically important institutions, (i.e. individual institutions posing threats to financial stability); (b) identify specific structural aspects of the financial system that are particularly vulnerable; (c) identify potential mechanisms for shock propagation in the financial system. These indicators can provide early signals to help regulators design appropriate policy responses.

The idea of network metrics for systemic risk has led to many discussions around the need for financial theory to absorb lessons from other disciplines like biostatistics, bio-informatics, machine learning, etc., to model the complex financial system. This has prompted many research with considerable increase in the number of empirical papers on the contribution of networks to existing theoretical result on systemic risk and contagion.

To learn networks from data, several statistical and econometric methods have been advanced to model interconnectedness and to represent sources of contagion and spillover effects. The commonly discussed approaches are: Granger-Causality Networks (Billio et al., 2012); Partial Correlation Networks (Barigozzi and Brownlees, 2014); Variance Decomposition Networks (Diebold and Yilmaz, 2014); and Tail Risk

Networks (Hautsch et al., 2014). Most of the above mentioned paper also consider the application of shrinkage methods like the LASSO for selecting relevant drivers to model sparsity in large datasets.

The different networks described in the above mentioned papers have contributed significantly to identifying the complex nature of interconnectedness in the financial system and possible mechanisms for risk propagation. The approaches of these papers however follow the standard practice of identifying a single model that summarize dependence in the observed data and often ignore the problem of model uncertainty. This can affect the estimated network and a cost in the generalization of the results.

In this paper, we discuss a Bayesian approach to network inference which allows us to address the problem of model uncertainty.

2.4 Graphical Models in Multivariate Analysis

Graphical models have become a center of attraction for modeling big data due to simplicity in presentation and ability to provide an intuitive way to visualize and interpret complex relationships. The graph visualization a deeper understanding of the relationships among variables by distinguishing direct and indirect interactions. The idea of connecting the multivariate time series literature (which has matured overtime) and graphical models is gradually becoming a vibrant field of research in econometrics, economics and finance. We relate graphical models to multivariate statistical analysis specifically in multivariate regression problems with potential applications in econometrics and finance.

2.4.1 Multivariate (Multiple) Regression

A typical multivariate multiple regression model is given by;

$$Y = BX + U \tag{2.1}$$

where X and Y are vector of exogenous and response variables respectively, B is a coefficient matrix and U is a vector of errors. The common approach in many empirical research is to fit the above model and to test for restrictions. In testing for the statistical significance of each of the estimated coefficients, we typically specify an acceptable maximum probability of rejecting the null hypothesis when it is true, i.e. committing a Type I error. In multiple hypothesis testing, the Type I errors committed increases with the number of hypotheses which may have serious consequences on the conclusions and generalization of the results. Several approaches have been proposed to deal with this problem. See Shaffer (1995); Drton and Perlman (2007) for review and discussion on multiple hypothesis testing.

Graphical models provide a convenient framework for exploring multivariate dependence patterns. By considering (2.1) as a causal (dependency) pattern of elements in Y on elements in X, the results of the coefficients matrix under a null hypothesis of single restrictions is, $B_{ij} = 0$ if y_i does not depend on x_j, and $B_{ij} \neq 0$ otherwise. Here we can define a binary connectivity matrix, G, such that $G_{ij} = 1$ implies $x_j \rightarrow y_i \iff B_{ij} \neq 0$. Thus G can be interpreted as a (directed) graph of the conditional dependences between elements in X and Y. From the above description, B can be represented as

$$B = (G \circ \Phi) \tag{2.2}$$

where Φ is a coefficients matrix and the operator $(\circ)$ is the element-by-element Hadamard's product (i.e., $B_{ij} = G_{ij}\ \Phi_{ij}$). There is a one-to-one correspondence between B and Φ conditional on G, such that $B_{ij} = \Phi_{ij}$, if $G_{ij} = 1$; and $B_{ij} = 0$, if $G_{ij} = 0$. Thus, Φ can be interpreted as the unconstrained regression coefficient matrix and B as the constrained regression coefficients matrix. For example:

$$\begin{array}{rcccc} y_1 = & 1.3x_1 & +0.5x_3 & +u_1 \\ y_2 = & 0.9x_1 & +0.5x_5 & +u_2 \\ y_3 = & x_2 & +0.7x_4 & +u_3 \end{array} \qquad B = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{pmatrix} 1.3 & a_1 & 0.5 & a_2 & a_3 \\ 0.9 & a_4 & a_5 & a_6 & 0.5 \\ a_7 & 1 & a_8 & 0.7 & a_9 \end{pmatrix} \end{array} \tag{2.3}$$

$$G = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{array} \qquad \Phi = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{pmatrix} 1.3 & a_1 & 0.5 & a_2 & a_3 \\ 0.9 & a_4 & a_5 & a_6 & 0.5 \\ a_7 & 1 & a_8 & 0.7 & a_9 \end{pmatrix} \end{array} \tag{2.4}$$

where $a_i \in \mathbb{R}, i = 1, \ldots, 9$, are expected to be statistically not different from zero by definition. The expression (2.4) presents some interesting alternative for modeling (2.3). Instead of estimating the unconstrained coefficients matrix, Φ, and performing multiple hypothesis testing, a more efficient alternative is to infer G as a variable selection matrix to estimate only the relevant coefficients in B. Thus, non-zero elements in B corresponds to non-zero elements in G. Inference of G taking into account all possible dependence configurations automatically handles the multiple testing problems in multivariate multiple regression models.

In most regression models, the parameters of the model are given by $\{B, \Sigma_u\}$, where Σ_u is the covariance matrix of the errors. In graphical models, inference of the graph structure is central to the model estimation, and the set of parameters, θ, consists the strength of the dependence among variables. Thus, applying graphical models in the context of regression models means that, θ must be equivalent to the regression parameters, i.e, $\theta \equiv \{B, \Sigma_u\}$.

Let Y denote a $n_y \times 1$ vector of dependent variables, X a $n_x \times 1$ vector of explanatory variables. Let $Z = (Y', X')'$ be $n = (n_y + n_x) \times 1$ vector of stacked Y and X. Suppose the joint, Z, follows the distribution, $Z \sim \mathcal{N}(0, \Omega^{-1})$, where $\Sigma = \Omega^{-1}$ is $n \times n$ covariance matrix. The joint distribution of Z can be summarized with a graphical model, (G, θ), where G is of dimension $n_y \times n_x$ and consists of

directed edges from elements in X to elements in Y. Thus, estimating the model parameters associated with G is equivalent to estimating Ω, i.e $\theta = \Omega$. Given Ω, the parameters of model (2.1) can be estimated from $\Sigma = \Omega^{-1}$ as

$$B = \Sigma_{yx}\Sigma_{xx}^{-1}, \qquad \Sigma_{\varepsilon} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \tag{2.5}$$

where Σ_{xy} is $n_x \times n_y$ the covariances between X and Y, Σ_{yy} is $n_y \times n_y$ covariances among Y and Σ_{xx} is $n_x \times n_x$ covariances among X. For further computational aspects of graphical models (see Lenkoski and Dobra, 2011; Heckerman and Geiger, 1994).

2.4.2 Applications in Econometrics and Finance

We present potential applications of graphical models in econometrics and finance.

Structural Model Estimation

VAR (Sims, 1980) have been widely used in macro-econometric models. It is widely discussed in the literature that such models do not have direct economic interpretations. However, due to their performance to forecast dynamics in macroeconomic variables, such limitation is overlooked. The structural VAR (SVAR) on the other hand is considered as a model with direct economic interpretation. They are however not directly estimable due to identification issues. The standard approach involves estimating a reduced form VAR, and identifying the relationships among shocks as a means of providing economic intuition on the structural dynamics. Based on (2.2), a SVAR model can be expressed in a graphical model form as

$$Y_t = \sum_{i=0}^{p} B_i Y_{t-i} + \varepsilon_t = \sum_{i=0}^{p} (G_i \circ \Phi_i) Y_{t-i} + \varepsilon_t \tag{2.6}$$

where $(G_0 \circ \Phi_0)$ and $(G_s \circ \Phi_s), s \geqslant 1$, are the graphical models representing the cross-sectional and temporal dependences, respectively. See Ahelegbey et al. (2015) for application of graphical models for identification restrictions in SVAR. See also

Corander and Villani (2006); Demiralp and Hoover (2003); Swanson and Granger (1997); Moneta (2008); Dahlhaus and Eichler (2003) for discussions on estimating causal structures in time series and VAR models.

Time Varying Model Estimation

A standard application in most empirical works is fixed/time varying parameter models. These approaches to modeling real world phenomena implicitly assumes that interactions among variables is stable over time and only the parameters are fixed or varying. This assumption may have consequences on the performance of estimated models. Some empirical works have shown that financial networks, especially, exhibit random fluctuations over various time scales, (see Billio et al., 2012), which must be incorporated in modeling dynamics in observed data. A typical time varying model of equation (2.1) is given by

$$Y_t = B_t X_{t-1} + U_t \tag{2.7}$$

Based on our expression in (2.2), the coefficients matrix of (2.7) can be expressed as

$$B_t = G \circ \Phi_t, \qquad B_t = G_t \circ \Phi, \qquad B_t = G_t \circ \Phi_t \tag{2.8}$$

The first expression of equation (2.8) follows the typical time varying parameter models commonly discussed in most empirical papers where the graph is invariant over time. The other two expression in equations (2.8) present cases of time varying structure of dependence. The state of the world is not constant overtime, hence, devoting much attention to modeling dynamics of the structure of interaction seems to be a more interesting line of research to understand the ever transforming modern economic and financial system. See Bianchi et al. (2014) for a new graphical factor model with Markov-switching graphs and parameters for modeling contagion and systemic risk. See Carvalho and West (2007) for graphical multivariate volatility modeling induced by time varying variances and covariances across series.

High Dimensional Model Estimation

There is an increasing interest in high-dimensional models and big data analysis. This has become necessary since many researchers have shown that information from large dataset enrich existing models and produces better forecasts for VAR models and also reveals the connectedness of the financial system. Graphical models are therefore relevant for high dimensional modeling by offering interpretation on information extracted from large datasets. See Ahelegbey et al. (2014) for discussions on modeling sparsity in graphical VAR models of large dimension and dealing also with uncertainty in the lag order. See also Jones et al. (2005); Scott and Carvalho (2008) for several discussions on approaches and priors to penalize globally or locally "dense" graphs when estimating high dimensional models.

The graphical approach can be used to build models alternative to the factor methods when dealing with large datasets. In factor augmented VAR models, information is extracted from a large number of variables to build factors to augment the VAR. Following the justification of this approach, (see Bernanke et al., 2005), graph search algorithms can be applied to select relevant predictors from a large set of exogenous variables which can be used to augment the VAR. This method will provide a more interpretable model than the factor approach.

CAPM-Like Model Estimation

A fundamental model in financial theory is the Capital Asset Pricing Model (CAPM) by Sharpe (1964) as an extension of Markowitz (1952) portfolio theory. This approach has received many criticism due to problems of empirical evidence. Fama and French (2004) summarize the popularity of the CAPM as follow: *The CAPM's empirical problems may reflect theoretical failings, the result of many simplifying assumptions. But they may also be caused by difficulties in implementing valid tests of the model.*

Graphical models can be applied to decompose asset return correlations into mar-

ket specific and idiosyncratic effects, as in the classical CAPM models. See Ahelegbey and Giudici (2014) for discussions on Bayesian hierarchical graphical models that allow correlations to be decomposed, into a country (market) effect plus a bank-specific (idiosyncratic) effect.

Portfolio Selection Problem

Portfolio risk analysis is typically based on the assumption that the securities in the portfolio are well diversified. A well-diversified portfolio is one that is exposed only to market risk within asset classes and includes a variety of significantly different asset classes. Precisely, asset classes that are not highly correlated and are thus considered to be complementary. The lesser the degree of correlation, the higher the degree of diversification and the lower the number of asset classes required to be well-diversified. Portfolios that contain securities with several correlated risk factors do not meet the well-diversified criteria. Despite an unprecedented access to information, it is evident that some portfolios by construction contain a predominant factor, and most risk modeling techniques are unable to capture these contagion.

To measure diversification more accurately, the graphical approach can be applied to study the structure of interconnection among the various asset classes. Since the Arbitrage Pricing Theory (APT) model of (Roll, 1977; Ross, 1976) and the return-based style model of Sharpe (1992) are regression models, the graphical approach provides useful technique for portfolio selection by modeling explicitly the dependence of the factors or asset classes in the form of networks. See Shenoy and Shenoy (2000) and Carvalho and West (2007) for applications of graphical models in the context of financial time series for predictive portfolio analysis.

Risk Management Style Assessment

Sharpe (1992) introduced a return-based analysis to measure management style and

performance. The analysis is based on the idea that a manager builds a portfolio according to a specific investment philosophy and investments reflect a style. The approach is based on a style regression model to determine the "effective mix", given by the estimated model which represents the return from the style and the residuals reflect the performance due to the "selection" (active management).

Suppose we are interested in modeling the annual return on an investment. Certainly, this will depend on factors like the amount invested in a fund, the annual stock market increase and the experience or style of the fund manager. A simple network to model this interaction is displayed in Figure 2.1, where $R = f(E, S, Y)$, the returns is a function of the amount invested (Y), the annual stock market increase (S) and the experience (or style) of the manager (E). This network can serve

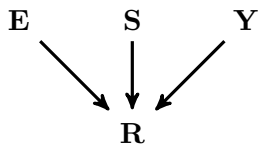


FIGURE2.1: Network for modeling returns on investment and style measurement. Y is the amount invested in the fund, S - the annual stock market increase, E - measures the experience (or style) of the fund manager and R - the annual return on the investment.

as a benchmark to assess the performance of fund managers. For instance, since the fund managers experience may be unknown and cannot be quantified, a qualitative measure can be applied to rank the experience, and a probabilistic inference can be obtained on the likely level of experience given information on the other variables in the network. See Ammann and Verhofen (2007) for application of networks to analyze the behavior of mutual fund managers.

2.5 Bayesian Inference Procedure

Graphical models presents a framework to deal with multiple testing problems in multivariate regression models. This involves exploring all candidate structures and

parameters of the model which poses a challenge since the space of possible structures increases super-exponentially (Chickering et al., 2004). The standard practice of identifying a single model that summarize the relationships often ignores the model uncertainty problem. The Bayesian approach however provides a way to handle this problem through prior information and model averaging. From a Bayesian perspective, the joint prior distribution over (Ω, G) is $P(G, \Omega) = P(G)P(\Omega|G)$. We focus on the inference of graph structure of the model.

2.5.1 Prior Distribution

In the absence of genuine prior knowledge on the dependence structure, the common approach is to assume the prior for G, i.e., $P(G) \propto 1$. For discussions on other graph priors, (see, Jones et al., 2005; Scott and Carvalho, 2008; Friedman and Koller, 2003).

There are two main approaches to define parameter priors for graphical models, however a common feature to these approaches is that both are graph conditional parameter priors. On one hand is a vast work on Gaussian DAG models discussing a list of conditions that permits an unconstrained precision matrix Ω (see, e.g. Heckerman and Geiger, 1994; Heckerman and Chickering, 1995; Geiger and Heckerman, 2002; Consonni and Rocca, 2012). On the other hand is a vast publication on Gaussian decomposable undirected graphical (UG) models with constraints on the precision matrix Ω, i.e. Ω_G (see, e.g. Roverato, 2002; Carvalho and Scott, 2009; Wang and Li, 2012; Lenkoski and Dobra, 2011). Note that, an unconstrained Ω characterizes a complete Gaussian DAG or UG model, i.e. a graph with no missing edges. The standard parameter prior for Gaussian DAG models with zero expectations is a Wishart distribution for Ω, whereas that of UG models is a G-Wishart distribution for Ω_G (or hyper-inverse Wishart distribution for Ω_G^{-1}).

2.5.2 Posterior Approximation

The likelihood function of a random sample $\mathcal{D} = (Z_1, \ldots, Z_T) \sim \mathcal{N}(0, \Omega^{-1})$ is

$$P(\mathcal{D}|\Omega, G) = (2\pi)^{-\frac{1}{2}nT}|\Omega|^{\frac{1}{2}T} \exp\Big\{ -\frac{1}{2}\langle \Omega, \hat{S}\rangle \Big\} \tag{2.9}$$

where $\hat{S} = \sum_{t=1}^{T} Z_t Z_t'$ is $n \times n$ sum of squared matrix, $\langle A, B\rangle = tr(A'B)$ denotes the trace inner product, $\underline{\nu} > n + 1$ is the degree of freedom parameter. Following Geiger and Heckerman (2002) that Ω conditional on G is Wishart distributed, we can integrate out Ω from (2.9)

$$P(\mathcal{D}|G) = \int P(\mathcal{D}|\Omega, G)\ P(\Omega|G) d\Omega = (2\pi)^{-\frac{1}{2}nT}\ \frac{K_n(\underline{\nu} + T, \underline{S} + \hat{S})}{K_n(\underline{\nu}, \underline{S})} \tag{2.10}$$

where $\underline{S}$ is a symmetric positive definite matrix, $K_n(\underline{\nu} + T, \underline{S} + \hat{S})$ and $K_n(\underline{\nu}, \underline{S})$ are the normalizing constants of the Wishart posterior and prior distribution on Ω respectively, with

$$K_n(\underline{\nu}, \underline{S}) = \int |\Omega|^{\frac{1}{2}(\underline{\nu}-n-1)} \exp\Big\{ -\frac{1}{2}\langle \Omega, \underline{S}\rangle \Big\}\ d\Omega\ =\ 2^{\frac{1}{2}\underline{\nu}n}|\underline{S}|^{-\frac{1}{2}\underline{\nu}}\Gamma_n\Big(\frac{1}{2}\underline{\nu}\Big) \tag{2.11}$$

with $\Gamma_n(a) = \pi^{\frac{1}{4}n(n-1)} \prod_{i=1}^{n} \Gamma\Big(a - \frac{i+1}{2}\Big)$ as the multivariate gamma function and $\Gamma(\cdot)$ the gamma function. The expression (2.10) is the marginal likelihood of the data, $\mathcal{D}$ given G expressed as ratio of the normalizing constants of the Wishart posterior and prior. Let $\overline{S} = \underline{S} + \hat{S}$ and $\overline{\nu} = \underline{\nu} + T$. The graph posterior distribution is therefore given by

$$P(G|\mathcal{D}) = P(G)P(\mathcal{D}|G) \approx (2\pi)^{-\frac{1}{2}nT}\frac{K_n(\overline{\nu}, \overline{S})}{K_n(\underline{\nu}, \underline{S})} \tag{2.12}$$

Under some conditions outlined by Geiger and Heckerman (2002), the marginal likelihood function (2.10) can be factorized into product of local terms as

$$P(\mathcal{D}|G) = \prod_{i=1}^{n_y} P(\mathcal{D}|G(y_i, \pi_i)) = \prod_{i=1}^{n_y} \frac{P(\mathcal{D}^{(y_i,\pi_i)}|G)}{P(\mathcal{D}^{(\pi_i)}|G)} \tag{2.13}$$

where $\pi_i = \{j = 1, \ldots, n_x : G_{ij} = 1\}$, n_y is the number of equations, $G(y_i, \pi_i)$ is the sub-graph with dependent variable y_i and explanatory variables π_i, $\mathcal{D}^{(y_i,\pi_i)}$ and $\mathcal{D}^{(\pi_i)}$ are sub-matrices of $\mathcal{D}$ consisting of (y_i, π_i) and π_i respectively. The closed form of (2.13) is:

$$P(\mathcal{D}^k|G) = (\pi)^{\frac{-n_k T}{2}} \frac{|\overline{S}_{k,k}|^{-\frac{1}{2}\overline{\nu}}}{|\underline{S}_{k,k}|^{-\frac{1}{2}\underline{\nu}}} \frac{\prod_{i=1}^{n_k} \Gamma\left(\frac{1}{2}(\overline{\nu} + 1 - i)\right)}{\prod_{i=1}^{n_k} \Gamma\left(\frac{1}{2}(\underline{\nu} + 1 - i)\right)} \tag{2.14}$$

where $k \in \left\{(y_i, \pi_i), \pi_i\right\}$ is of dimension n_k, $\mathcal{D}^k$ is a sub-matrix of $\mathcal{D}$ associated with k, $|\underline{S}_{k,k}|$ and $|\overline{S}_{k,k}|$ are the determinants of the prior and posterior sum of squared matrices of elements in k.

2.5.3 Graph Sampling Scheme

We sample the graph following the Markov Chain Monte Carlo (MCMC) algorithm in Madigan and York (1995). The scheme is such that, at the r-th iteration, given $G^{(r-1)}$, the sampler proposes a new graph $G^{(*)}$ based on a proposal distribution $Q(G^{(*)}|G^{(r-1)})$ with acceptance probability

$$A(G^{(*)}|G^{(r-1)}) = \min\left\{\frac{P(\mathcal{D}|G^{(*)})}{P(\mathcal{D}|G^{(r-1)})} \frac{P(G^{(*)})}{P(G^{(r-1)})} \frac{Q(G^{(r-1)}|G^{(*)})}{Q(G^{(*)}|G^{(r-1)})}, 1\right\} \tag{2.15}$$

where $Q(G^{(*)}|G^{(r-1))})$ and $Q(G^{(r-1)}|G^{(*)})$ are the forward and reverse proposal distribution respectively. For a single edge operation, the proposal distribution assigns a uniform probability to all possible graphs in the neighborhood of $G^{(r-1))}$, i.e., the set of graphs that can be reached from the current state $(G^{(r-1))})$ by adding or deleting a single edge. If the new graph $G^{(*)}$ is accepted, then the graph at the r-th iteration is set to $G^{(r)} = G^{(*)}$, otherwise $G^{(r)} = G^{(r-1)}$.

After R_T total iterations, we average the draws, $G^{(r)}$ and estimate the posterior probability of the edge by $\hat{e}_{ij} = \frac{1}{R_T - R_0} \sum_{r=R_0+1}^{R_T} G_{ij}^{(r)}$, where R_0 is the burn-in sample, $G_{ij}^{(r)}$ is the edge from x_j to y_i in the graph G at the r-th iteration. We define $\hat{G}_{ij}$

based on a one sided posterior credibility interval for the edge posterior distribution, and find the interval lower bound $\hat{G}_{ij} = 1$, if $\hat{e}_{ij} - z_{(1-\alpha)}\sqrt{\frac{\hat{e}_{ij}(1-\hat{e}_{ij})}{ESS}} > 0.5$, where $z_{(1-\alpha)}$ is the z-score of the normal distribution at the $(1-\alpha)$ significance level, ESS is the effective sample size obtained by $ESS = M/(1 + 2\sum_{l=1}^{\infty}\rho_l)$, where M is the number of MCMC samples of the graph, and ρ_l is the autocorrelation of lag l. We find a cutoff point s is obtained at which $\rho_s < 0.05$, and then sum all the ρ_s up to that point.

2.6 Financial System Interconnectedness

We illustrate the graphical approach for modeling financial system interconnectedness by considering the dataset of the four financial sectors in Billio et al. (2012). The data is monthly return indexes for hedge funds (HF), banks (BK), brokers (BR) and insurance (IN) companies in the US from January 1994 to December 2008. See the above paper for details on the construction of the indexes. We focus on the temporal dependence and consider a vector autoregressive (VAR) model with specification of the lag order ($p = 1$) based on testing the appropriate lag length using the BIC criterion. Following Billio et al. (2012), we consider two sample periods 1994-2000 and 2001-2008 for both network analysis.

Table 2.1 shows the posterior probabilities of the presence of edges in the four-variable network model of hedge funds, brokers, banks, and insurers. The left (right) panel shows the edge posterior probabilities in the network for the sample 1994-2000 (2001-2008). Bold values indicate links whose posterior probabilities are greater than 0.5 under a 95% credibility interval. The edges are directed from column labels (at $t-1$) to row labels (at t).

We see from the table that the probabilities in the first period (1994-2000) are very low compared to the second (2001-2008). The latter record six edges with bold values

and one of them indicates an autoregressive effect, i.e., $P(HF_{t-1} \rightarrow HF_t|\mathcal{D}) = 0.95$. On the interconnection among the institutions, we find in the second sample, strong evidence of the effect of Insurance companies on Brokers, $P(IN_{t-1} \rightarrow BR_t|\mathcal{D}) = 0.98$; on Banks, $P(IN_{t-1} \rightarrow BK_t|\mathcal{D}) = 0.82$; and on Hedge funds, $P(IN_{t-1} \rightarrow HF_t|\mathcal{D}) = 0.71$. Also, we find strong effects of Banks on Brokers, $P(BK_{t-1} \rightarrow BR_t|\mathcal{D}) = 0.75$, and effects of Brokers on Insurance companies, $P(BR_{t-1} \rightarrow IN_t|\mathcal{D}) = 0.64$.

	HF_{t-1}	BR_{t-1}	BK_{t-1}	IN_{t-1}	HF_{t-1}	BR_{t-1}	BK_{t-1}	IN_{t-1}
	Jan 1994 - Dec 2000				Jan 2001 - Dec 2008			
HF_t	0.23	0.21	0.19	0.19	**0.95**	0.23	0.29	**0.71**
BR_t	0.18	0.18	0.18	0.19	0.16	0.18	**0.75**	**0.98**
BK_t	0.26	0.20	0.30	0.19	0.33	0.31	0.50	**0.82**
IN_t	0.18	0.20	0.27	0.20	0.33	**0.64**	0.34	0.42

Table 2.1: Marginal posterior probabilities of connectedness of the institutions between 1994-2000 and 2001-2008. Bold values indicate links with probabilities greater than 0.5 under a 95% credibility interval.

(a) 1994-2000 (b) 2001-2008

FIGURE2.2: Network of hedge funds (HF), brokers (BR), banks (BK) and insurance (IN) between 1994-2000 and 2001-2008. Links are lagged dependencies and the blue (red) represent positive (negative) effects.

A representation of the results of Table 2.1 is shown in Figure 2.1. To visualize the dependencies, the edges of the network are presented in colors which depict the signs of the conditional dependencies. We find no connection in the first sample. In the second sample, we find a negative effect of Banks on Brokers, a bi-directional positive link between Insurance companies and Brokers, and a positive effect of Insurance companies on Banks and Hedge funds.

Our result corroborate the findings of Billio et al. (2012), with evidence of a higher vulnerability in the financial system between 2001-2008 which amplified the

2007-2009 crisis. We also find evidence that insurers played a central role in the interconnectedness leading to the crisis.

2.7 Volatility Connectedness in the Euro-Area

Volatility networks have gained attention lately due to their ability to track the fear of investors and to reflect the extent to which markets evaluate arrival of information. (Diebold and Yilmaz, 2014). In monitoring systemic risk, volatility connectedness referred to as "fear connectedness" has become increasingly important to identify risk transmission mechanisms in the financial system. As pointed out by Diebold and Yilmaz (2014), "the link between volatility connectedness and long-term rates is directly a result of the choices of investors. Rising long-term interest rates reflect optimism about future economic performance. As they expect the growth to pick up, investors sell more defensive stocks such as financial stocks and instead invest in manufacturing, energy and airlines sector stocks that are likely to benefit most from an economic recovery".

In this application, we analyze systemic risk in the European stock market by focusing on the volatility connectedness of the Euro Stoxx 600 super-sectors. The dataset consists of intra-day high-low price indexes from September 1, 2006 to September 19, 2014 from Datastream (see Table 2.2). The super-sectors represent the largest Euro area companies by the Industry Classification Benchmark (ICB) and covers countries like Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain.

Let $p^h_{i,t}$ and $p^l_{i,t}$ denotes the highest and lowest price of stock i on day t. Following Parkinson (1980) and Barigozzi and Brownlees (2014), we obtain the intra-day price range as a measure of the volatilities given by

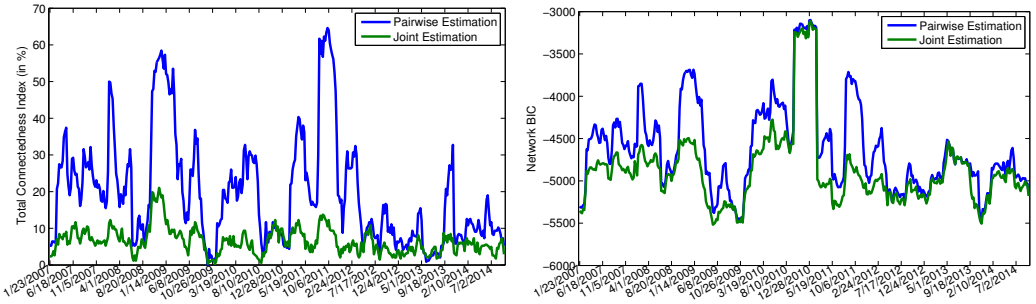
$$\tilde{\sigma}^2_{i,t} = \frac{1}{4\log(2)}\big(\ \log(p^h_{i,t}) - \log(p^l_{i,t})\ \big)^2 \tag{2.16}$$

Several estimators of volatility have been applied in the financial network literature. See for example Diebold and Yilmaz (2014) and Barigozzi and Brownlees (2014) for discussions and references of papers on other estimators of volatility in empirical financial models.

No.	Name	ID	No.	Name	ID
1	Banks*	BK	11	Media	MD
2	Insurance companies*	IN	12	Travel & Leisure	TL
3	Financial Services*	FS	13	Chemicals	CH
4	Real Estates* +	RE	14	Basic Resources	BR
5	Construction & Materials	CM	15	Oil & Gas	OG
6	Industrial goods & services	IGS	16	Telecommunication	TC
7	Automobiles & Parts	AP	17	Health Care	HC
8	Food & Beverage	FB	18	Technology	TG
9	Personal & Household Goods	PHG	19	Utilities	UT
10	Retail	RT			

Table 2.2: Description of Euro Stoxx 600 super-sectors. * - The financial sector variables.

We present the connectedness of the log volatilities of the super-sectors as the dependence pattern from a VAR(1) model, with lag order based on testing the appropriate lag length using the BIC criteria and the available dataset. We characterize the dynamics of the connectedness using a rolling estimation with window size of 100-days. We compare the pairwise (Granger-causality) estimation against a joint estimation based on the Bayesian graphical approach.



(a) Total Connectedness Index (b) Connectivity BIC scores

FIGURE2.3: Dynamics of total connectedness index and network BIC scores over the period 2007-2014 obtained from a rolling estimation with windows size of 100-days. The index of Granger-causality (pairwise estimation) is in blue and the Graphical model (joint estimation) is in green.

We present in Figure 2.3, the dynamics of the total connectedness index of the

Granger-causality (in blue) and the graphical model (in green) and their BIC scores over the sample period. See Billio et al. (2012) and Diebold and Yilmaz (2014) for discussions on network-based measures of connectedness. We see a significant difference in the total connectedness index of the two algorithms (see Figure 2.3a). We observe from the figure that the pairwise estimation procedure tends to over-estimate the linkages in periods of higher connectedness. This is typical of the standard Granger-causality approach. The BIC of the network (Figure 2.3b) favors the joint estimation approach over the pairwise scheme. This shows that the graphical approach performs relatively better than the Granger-causality at estimating the connectedness.

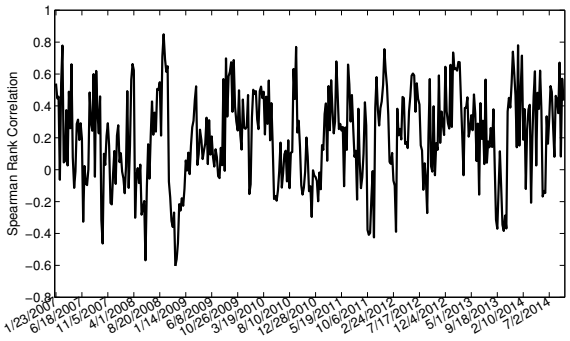
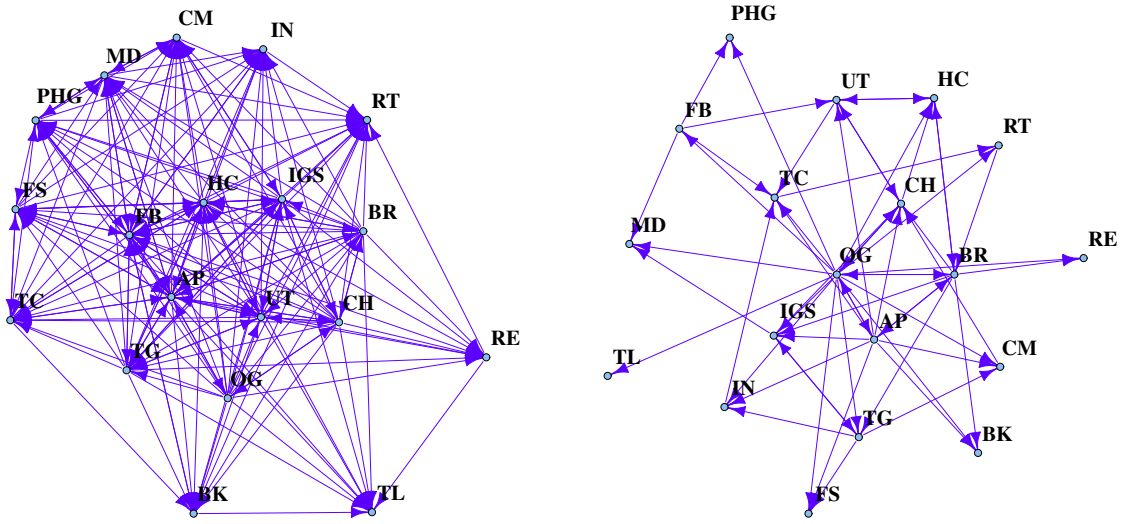


FIGURE2.4: Correlations of centrality rank metric from the pairwise and joint estimation networks.

Our main interest in this comparison is to focus on the centrality ranks of the two estimation procedures. To this end, we consider a correlation analysis of the eigenvector centrality ranks obtained by the two approaches over the sample period. The eigenvector centrality metric is a measure of the importance of a variable in a network. It assigns relative scores to all variables in the network, based on the principle that connections to few high scoring variables contribute more to the score of the variable in question than equal connections to low scoring variables. The eigenvector centrality measure has been extensively applied in financial network modeling to identify agents or institutions that are central to the spread of risk in the system. See Billio et al. (2012), Dungey et al. (2012), Barigozzi and Brownlees (2014) for

further discussions.

Figure 2.4 shows the dynamics of the Spearman rank correlations of the eigenvector centrality rank metrics of the two schemes. We see that on average, the rank by two schemes show a low positive correlation ($\rho = 0.203$). The distribution of the correlations seems negatively skewed with a mass of concentration on the positive. Furthermore, we notice that the extreme negatively correlated ranks is captured for the period ending November 12, 2008.



(a) Pairwise Estimation (BIC = -3716.2) (b) Joint Estimation (BIC = -4547.9)

FIGURE2.5: Volatility network for period ending November 12, 2008. Edges are lagged dependencies.

Next, we compares the predicted network of the two estimation approaches for the identified period with the extreme negative rank correlation (see Figure 2.5). We notice from the figure that the predicted network structure of the pairwise approach has more dense connectedness than the joint estimation approach. This is not surprising since (pairwise) Granger-causality deals only with bivariate time series and it is unable to distinguish between direct and mediated causal influences in the multivariate settings. The joint estimation through the graphical approach instead consider single and multiple testing possibilities to infer the dependence among the

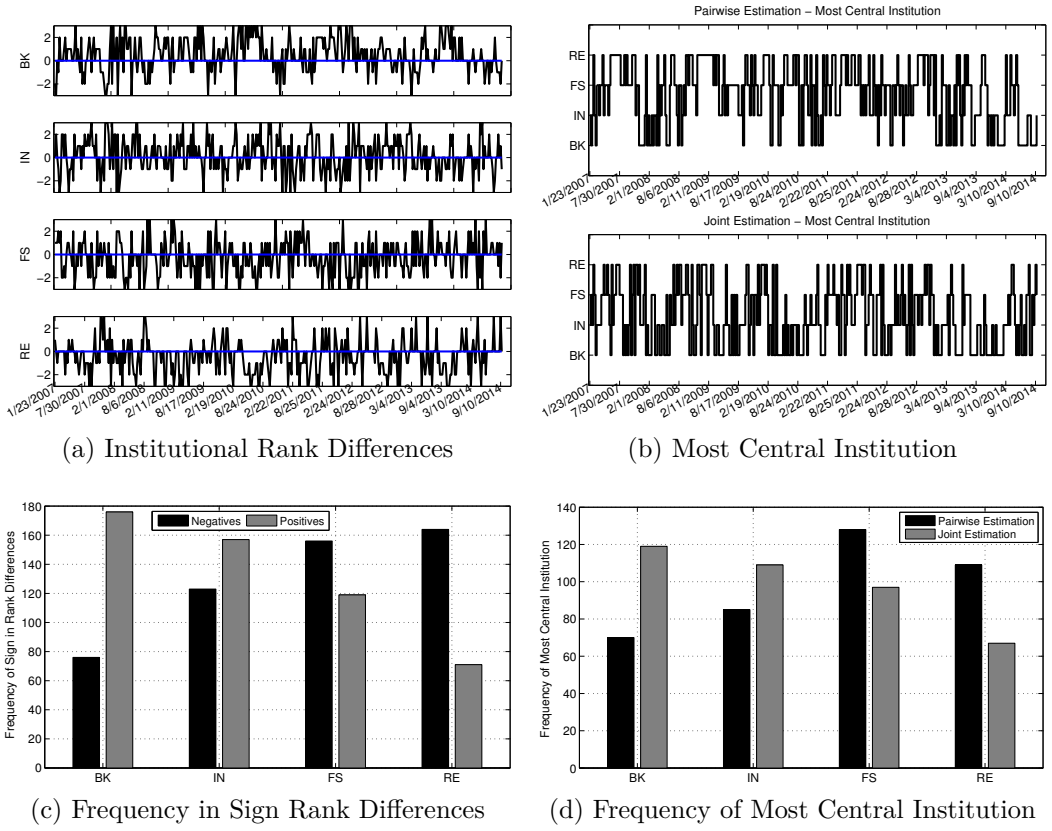
variables in a given data. Thus, the latter is able to distinguish between direct and mediated causal effects leading to a more sparse and parsimonious model. We notice from the figure that the BIC score favors the joint estimation over the pairwise scheme. Table 2.3 shows the top and bottom 5 super-sectors ranked by eigenvector centrality of the two estimation approaches for the period ending November 12, 2008. We see that the Utilities, Chemicals and Basic Resources which seems to be ranked low by the pairwise are rather ranked higher by the joint estimation. On the other hand Food & Beverage is highly ranked by the pairwise but lowly ranked by the joint estimation.

	Pairwise Estimation Centrality					Joint Estimation Centrality				
Rank	ID	Name	Eigen	In-D	Out-D	ID	Name	Eigen	In-D	Out-D
1	RT	Retail	0.3811	16	0	CH	Chemicals	0.4189	6	3
2	FB	Food Beverage	0.3219	14	15	HC	Health Care	0.3639	4	2
3	IN	Insurance	0.2940	13	2	UT	Utilities	0.2971	4	3
4	CM	Const & Mat	0.2940	13	4	TC	Telecom	0.2821	5	1
5	MD	Media	0.2688	12	7	BR	Basic Resource	0.2714	4	8
⋮	⋮					⋮				
15	TG	Technology	0.1591	8	15	AP	Auto. & Parts	0.1531	2	10
16	UT	Utilities	0.1591	8	16	RE	Real Estates	0.1531	2	0
17	CH	Chemicals	0.0859	5	18	PHG	Person H. Gds	0.0947	2	0
18	BR	Basic Resource	0.0711	4	17	FB	Food Beverage	0.0731	1	4
19	OG	Oil & Gas	0.0414	2	18	TL	Travel Leisure	0.0731	1	0

Table 2.3: Top and bottom 5 super-sectors ranked by eigenvector centrality of the pairwise and joint estimation approaches for the period ending November 12, 2008. Note: *Eigen* - eigenvector centrality, *In-D* means in-degree, *Out-D* means out-degree.

Finally, we focus our attention on the financial variables to analyze the dynamics of the centrality of the financial sector of the European stock market. We extract and analyze the differences between the eigenvector centrality rank of the individual financial super-sectors by the pairwise and the joint estimation schemes. We also compare the most central financial variable identified by the pairwise and the joint estimation approach . The dynamics of the rank differences and the most central financial institution is presented in Figure 2.6a and 2.6b respectively. Figures 2.6c and 2.6d depict the frequency of the sign of the institutional rank differences and the most central institution respectively over the sample period. We remind the reader that

in rank terms, 1 means a higher centrality and 4 means a lower centrality. Furthermore, by computing the rank difference as the pairwise minus the joint estimation rank of each institution, a negative rank difference connotes a higher centrality by the pairwise and a lower centrality by the joint estimation.



(a) Institutional Rank Differences (b) Most Central Institution

(c) Frequency in Sign Rank Differences (d) Frequency of Most Central Institution

FIGURE2.6: Centrality rank in the financial sector of European stock market over the sample period. Rank value 1 (4) means highest (lowest) centrality. A negative (positive) sign of rank difference means a higher (lower) centrality by the pairwise estimation and a lower (higher) centrality by the joint estimation.

The dynamics of the rank differences in Figure 2.6a shows quite many deviations from the reference line (in blue) for all four institutions. The reference line indicates equal ranks of institutions by the two estimations. We see from the figure what seems to be an opposite dynamics for Banks and Real Estates. Most periods of positive rank differences for Banks often experience negative rank differences for Real Estate and vice versa. The dynamics in Figure 2.6b also shows many periods of differences in the

most central financial institution predicted by the two estimations. For instance the pairwise approach found Real Estates as more central in the volatility connectedness than Banks in most part of 2007-2009. However, the joint estimation indicates the opposite.

The frequency of the signs from the rank differences as shown in Figure 2.6c indicates a completely opposite observation by the two estimation approaches. In most part of the sample period, the centrality rank of Banks and Insurance is lower in the pairwise estimation and higher in the joint estimation. The opposite is true for Financial services and Real Estates. For the most central financial institution, a researcher applying Granger-causality will identify Financial services as a major player in most part of the sample period, followed by Real Estates, with Banks as the least central. However, the graphical joint estimation shows Banks as more central, followed by Insurance, with Real Estates as the least.

In many real-world problems, interactions among random variables are more complex than pairwise. The graphical approach presented in this paper is designed to handle joint estimations and large scale multiple testing problems, and thus, more suitable to model the complex relationships in real world phenomena than the Granger-causality. Therefore, the estimated structure from the graphical approach seems more appropriate and reliable than that of the Granger-causality. It can be concluded from the above results that, over the sample period considered, Banks and Insurance companies are more central in the "fear connectedness" expressed by market participants in the financial sector of the Euro-area.

2.8 Conclusion

This paper surveys the state of the arts on graphical (network) models from a multidisciplinary perspective and in the context of systemic risk. It relates graphical models to multivariate analysis with potential applications in econometrics and fi-

nance. We demonstrate the effectiveness of the Bayesian approach to network inference to analyze vulnerabilities in the financial system. Using the return indexes of Billio et al. (2012), we find evidence of a higher vulnerability between 2001-2008 that amplified the 2007-2009 crisis. The evidence also suggests that insurers played a central role in the vulnerability of the system during the period. Using data on the European stock market, we find that Banks and Insurers are more central in the "fear connectedness" (Diebold and Yilmaz, 2014) expressed by market participants in the financial sector of the Euro-area. The result shows that the graphical (joint estimation) method allows us to extract a dependence patterns more suitable to model the complex relationships than the Granger-causality (pairwise estimation).

Acknowledgments

The author is thankful to Monica Billio, Roberto Casarin and Francesco Ravazzolo for their insightful comments on early versions of this paper. This research is supported by funding from the European Union, Seventh Framework Programme FP7/2007-2013 under grant agreement SYRTO-SSH-2012-320270 and by the Italian Ministry of Education, University and Research (MIUR) PRIN 2010-11 grant MISURA.

Chapter 3

Bayesian Graphical Models for SVAR Processes

3.1 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, 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.

This chapter is based on: Ahelegbey, D. F., Billio, M. and Casarin, R. (2015). Bayesian Graphical Models for Structural Vector Autoregressive Processes, *Journal of Applied Econometrics*, forthcoming.

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; 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 algorithm 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 3.2 presents BGVAR models. Section 3.3 discusses the model inference scheme. Section 3.4 provides an illustration of BGVAR on synthetic datasets, and a comparison with alternative approaches. Section 3.5 and 3.6 present macroeconomic and financial applications, respectively.

3.2 Bayesian Graphical Vector Autoregression

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

$$Y_t = B_0 Y_t + \sum_{i=1}^{p} B_i Y_{t-i} + \sum_{i=1}^{p} C_i Z_{t-i} + \varepsilon_t, \tag{3.1}$$

$t = 1, \ldots, T$, 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 $1 \leqslant i \leqslant p$, are $n_y \times n_y$ and $n_y \times n_z$ matrices of structural coefficients, respectively.

The identification problem of SVAR is that, (3.1) 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.

3.2.1 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 $B_i^* = (B_i, C_i), 1 \leqslant i \leqslant p$, the $n_y \times n$ matrices of unknown coefficients for the response and predictor variables. The reduced form of (3.1) can be expressed as

$$Y_t = \sum_{i=1}^{p} A_i X_{t-i} + A_0^{-1} \varepsilon_t \tag{3.2}$$

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, $A_i = A_0^{-1} B_i^*$, $1 \leqslant i \leqslant p$ are the reduced-form lag coefficient matrices such that A_i, $1 \leqslant i \leqslant p$, 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 (3.2) can be rewritten in a matrix form as:

$$Y = X'A'_{+} + U \tag{3.3}$$

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 $X'_{p+1-s}, \ldots, X'_{T-s}$, $1 \leqslant s \leqslant p$, 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 (3.3) 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 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 (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.

3.2.2 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 (3.3)

$$\Sigma_u \;=\; E(A_0^{-1}\varepsilon_t\varepsilon_t'(A_0^{-1})') \;=\; A_0^{-1}\Sigma_\varepsilon(A_0^{-1})' \tag{3.4}$$

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.

3.2.3 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 $P \rightarrow Q$ 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 realization of the i-th variable at time t. Equation (3.1) can be represented as a graphical model with a one-to-one correspondence between the coefficient matrices and a directed acyclic graph (DAG):

$$X_{t-s}^j \to X_t^i \iff B_{s,ij}^* \neq 0 \qquad 0 \leqslant s \leqslant p \tag{3.5}$$

where $B_0^* = B_0$, for $s = 0$, and $B_s^* = (B_s, C_s)$, for $1 \leqslant s \leqslant p$. By considering the structural dynamics as a causal dependence, the relationship in (3.5) for $1 \leqslant s \leqslant p$ 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 $\mu = 0$.

3.2.4 Bayesian Graphical VAR Models

Following the representation in equation (3.5), we define:

$$B_s^* = (G_s \circ \Phi_s), \qquad 0 \leqslant s \leqslant p \tag{3.6}$$

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 $1 \leqslant s \leqslant p$, $B_s^* = (B_s, C_s)$ 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, $1 \leqslant s \leqslant p$, as the matrix of the temporal dependence. Elements in G_s, $0 \leqslant s \leqslant p$, are such that $G_{s,ij} = 1 \iff X^j_{t-s} \to X^i_t$ and 0 otherwise. Elements in Φ_s, $0 \leqslant s \leqslant p$, are structural coefficients, such that $\Phi_{s,ij} \in \mathbb{R}$ represents the value of the effect of X^j_{t-s} on X^i_t. There is a one-to-one correspondence between Φ_s and B^*_s conditional on G_s:

$$B^*_{s,ij} = \begin{cases} \Phi_{s,ij} & if \qquad G_{s,ij} = 1 \\ 0 & if \qquad G_{s,ij} = 0 \end{cases} \tag{3.7}$$

Based on our notation in (3.6), equation (3.1) can be expressed as:

$$Y_t = (G_0 \circ \Phi_0)Y_t + \sum_{i=1}^{p}(G_i \circ \Phi_i)X_{t-1} + \varepsilon_t \tag{3.8}$$

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 (3.8) 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 A.1 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 (3.2), (3.3), (3.6), and (3.8), the reduced-form parameters of the standard VAR model can be mapped to that of the Bayesian graphical model as follows:

$$A_0 = \mathbf{I}_{n_y} - (G_0 \circ \Phi_0), \qquad A_i = (\mathbf{I}_{n_y} - G_0 \circ \Phi_0)^{-1}(G_i \circ \Phi_i), \qquad i = 1, \ldots, p \quad (3.9)$$

where A_0 is $n_y \times n_y$ coefficient matrix conditional on G_0, and A_i is a stacked reduced-form coefficient matrix. Let 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 the graph associated with A_+ in equation (3.3) 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 note 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 normalizing 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 restrictions allow us to avoid estimating all the coefficients in the VAR model. It presents an automatic way to achieve dimension 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 3.3.

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 responses, and n_z predictors, can be expressed as $G(p, n_y, n) = H(n_y)F(p, n_y, n)$. Figure 3.1 shows that the number of possible structures increases super-exponentially

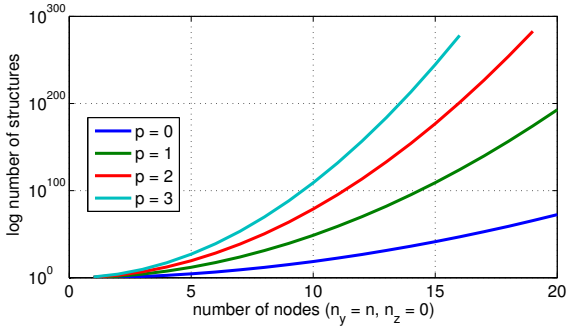


FIGURE3.1: Logarithmic estimates of the number of possible structures for a SVAR model of order p, $0 \leqslant p \leqslant 3$, with the number of response variables n_y equal to the number of nodes n, $(0 \leqslant n \leqslant 20)$, and no predictors $(n_z = 0)$.

with the number of variables, and lag order (different lines). This challenge has been

discussed extensively as a model determination problem (see Chickering et al., 2004; Corander and Villani, 2006). 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.

3.3 Efficient Model Inference Scheme

Under the Bayesian framework of Geiger and Heckerman (2002), 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 (G_0, G_+), from the joint distribution is computationally intensive, since the number of possible structures, as shown in Figure 3.1, 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) graph and to G_0 as the multivariate instantaneous (MIN) graph. 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 A.2 for further details of our MCMC scheme. In the following we describe the different steps of the Gibbs sampler.

3.3.1 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 remains in the current state, which satisfies aperiodicity. Hence, this proposal guarantees ergodicity of the MCMC chain. See A.2 for details.

3.3.2 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 contempora-

neous 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. $X_t^j \rightarrow X_t^i$ 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 $P \rightarrow \ldots \rightarrow Q \rightarrow \ldots \rightarrow R$, then P and Q are ancestors of R, and Q and R are descendants of P. We see that adding an edge $R \rightarrow P$ is illegal since it produces a cycle. Our proposal of a directed edge $X_t^j \rightarrow X_t^i$ 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 $X_t^j \rightarrow X_t^i$ only if a directed path from X_t^i to X_t^j, $(X_t^i \rightarrow \ldots \rightarrow X_t^j)$ 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 $X_t^j \rightarrow X_t^i$ and testing whether the resulting structure is a directed acyclic graph.

3.3.3 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 (3.9), 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. We estimate (A_+, Σ_u) by considering two typical prior distributions 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 variables that influence the response 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, σ_i^2, $i = 1, \ldots, n_y$, is the estimated variance of the residuals from a univariate AR 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 $\overline{\sigma}_i^2$, 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 $\overline{A}_i$, and the rest are restricted to zero.

3.4 Simulation Experiments

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

$$X_t = \sum_{i=0}^{p} (\mathbf{I}_k \otimes B_i) X_{t-i} + \varepsilon_t, \tag{3.10}$$

$t = 1, \ldots, T$, 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 3.1. 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 case, we replicate the estimation exercises 20 times, with random draws from the DGP. All the results reported in the following are averages over the replications.

We compare the MIN structure with the contemporaneous structure of the PC algorithm (see A.3.4) 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 condi-

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

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

Table 3.1: 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.

tional 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 Granger-causal variables at different lags (see A.3.5). As an alternative to the C-GC the Granger causal priority (see Sims, 2010, 1976; Qin, 2011) can be applied, which is an approach to discriminate 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 predictive AIC. See Appendix A.3 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 George et al. (2008), we run a total of 20,000 Gibbs iterations for the SSVS and exclude 10,000 burn-in samples.

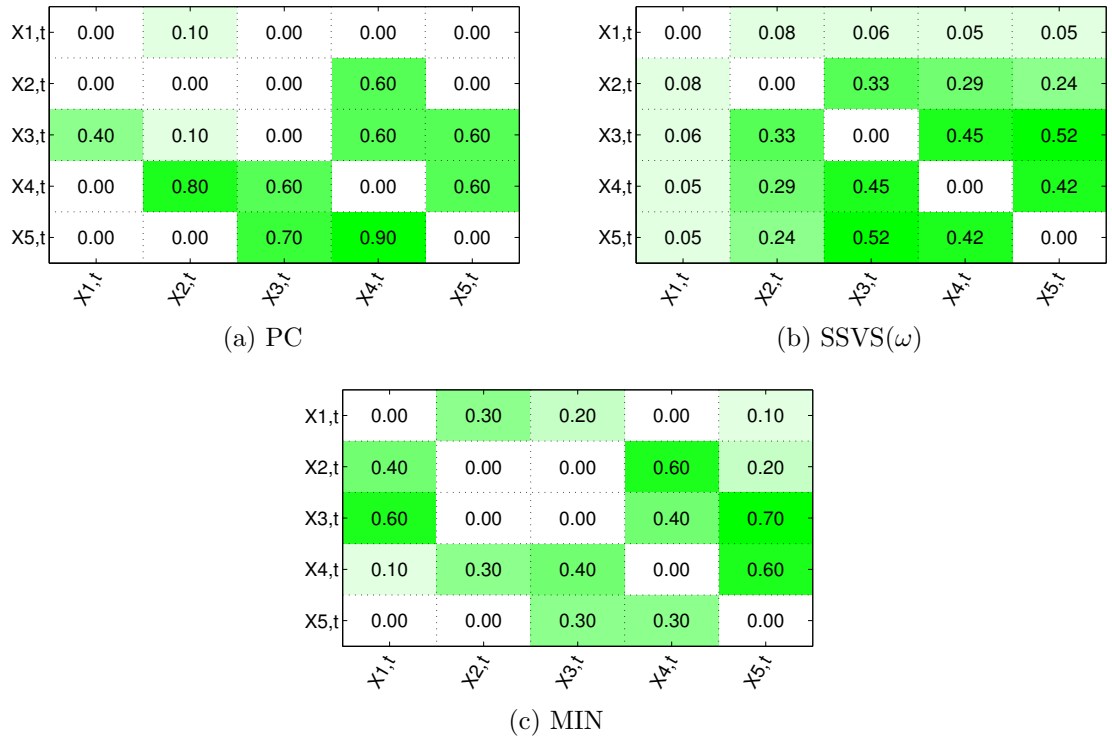
We report in Table 3.2 the results of the comparison between the different inference schemes. The top panel of the table shows that the MIN achieves a higher accuracy than the PC. In both small and moderate-dimension settings, we notice

Scheme	Small-size ($n = 5$) $p = 1$	$p = 2$	$p = 3$	Moderate-size ($n = 20$) $p = 1$	$p = 2$	$p = 3$
	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

Table 3.2: Metics on the 5- and 20-node models with lag = 1, 2, 3. *Note*: Comparison is 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).

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. In Figure 3.2, 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: $X_t^2 - X_t^4 - X_t^3 - X_t^5 - X_t^3$. The SSVS($\omega$) shows the following: $X_t^2 - X_t^3 - X_t^4 - X_t^5 - X_t^3$. The MIN structure indicates the following: $X_t^1 \rightarrow X_t^3$; $X_t^4 \rightarrow X_t^2$; $X_t^5 \rightarrow (X_t^3, X_t^4)$. 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. $P(X_t^4 \rightarrow X_t^5|\mathcal{X}) = 0.9 > P(X_t^5 \rightarrow X_t^4|\mathcal{X}) = 0.6$. 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: $X_t^2 \rightarrow X_t^4$; $X_t^3 \rightarrow X_t^1$; $X_t^5 \rightarrow X_t^3$, 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.

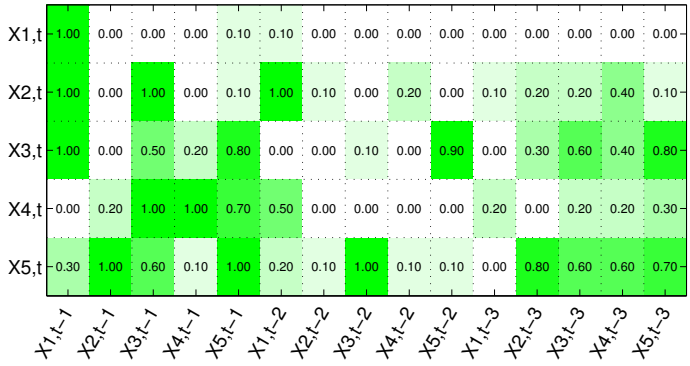


(a) PC (b) SSVS(ω)

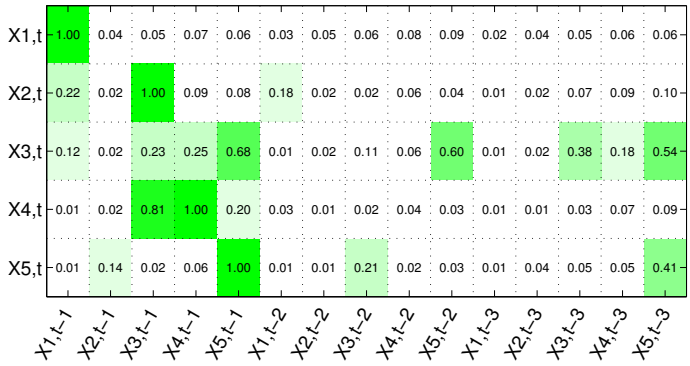
(c) MIN

FIGURE3.2: Structure of 5-node model ($p = 3$) averaged over 20 replications. Response (explanatory) variables are on rows (columns). The light (dark) green indicate weak (strong) dependence.

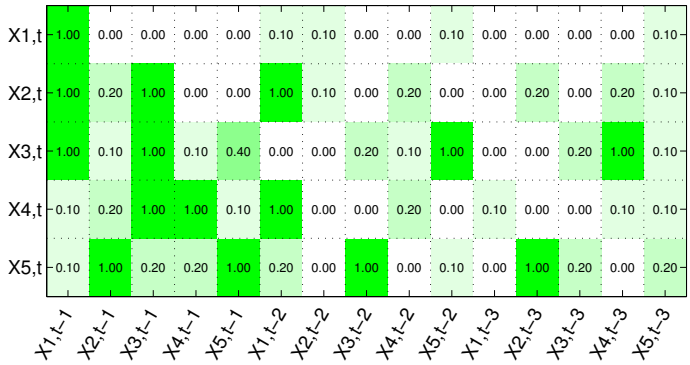
The second panel of Table 3.2 shows that both MAR and C-GC perform well



(a) C-GC



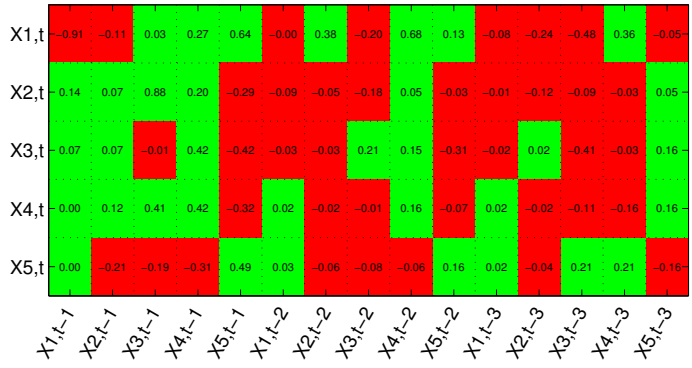
(b) SSVS(γ)



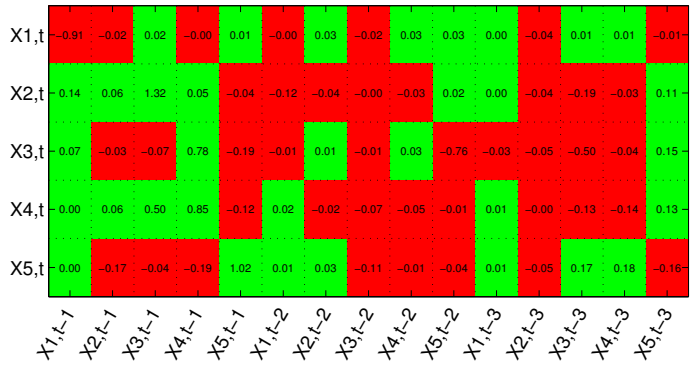
(c) MAR

FIGURE3.3: 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.

at inferring the temporal dependence relationships. 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. Figure 3.3 shows the temporal structure of C-GC,



(a) BVAR-NW Coefficient Matrix



(b) SSVS Coefficient Matrix

(c) BGVAR-NW Coefficient Matrix

FIGURE3.4: 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 for zeros.

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 than C-GC and SSVS(γ).

The third panel of Table 3.2 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.

In Figure 3.4, we report the estimates of the reduced-form coefficient matrices. For the sake of conciseness, a comparison between BVAR (normal-Wishart), SSVS and BGVAR (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.

The bottom panel of Table 3.2 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

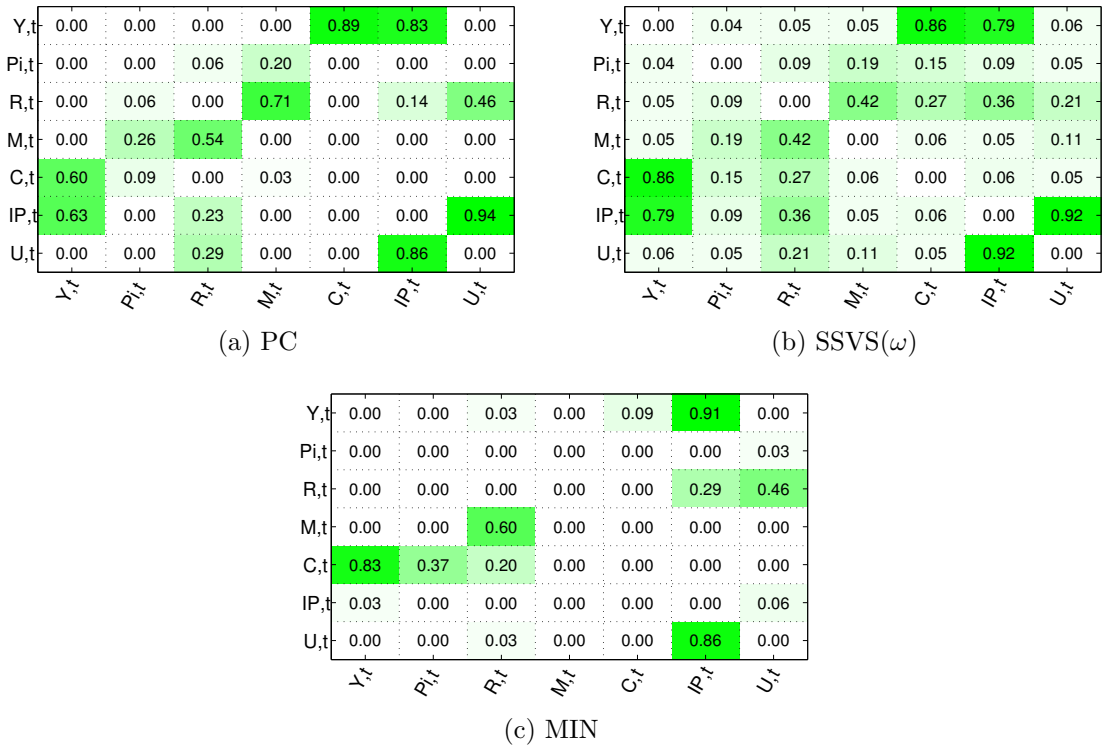
with the number of nodes and lags, for the moderate dimension DGP, we consider 40,000 iterations for BGVAR, and maintain 20,000 iterations for SVSS. Our results show that, in both small- and moderate-dimension settings, the collapsed Gibbs sampler used for the BGVAR posterior approximation is computationally 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.

3.5 Modeling and Forecasting Macroeconomic Time Series

The work by Banbura et al. (2010) has motivated an 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 models 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 the BVAR and SSVS models, and evaluate their predictive performance.

The dataset consists of quarterly series of 20 US-macroeconomic variables, from 1959Q1 to 2008Q4, used by Koop (2013). We transform the data as in Koop (2013). See A.4 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.

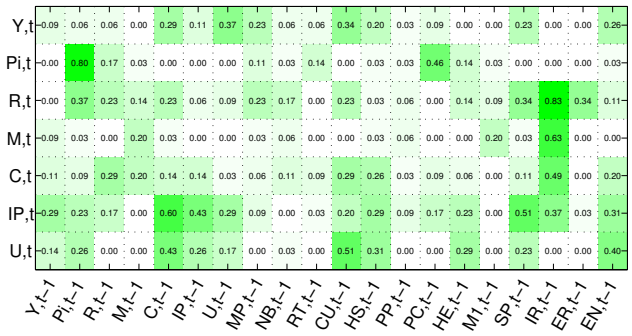


(a) PC (b) SSVS(ω)

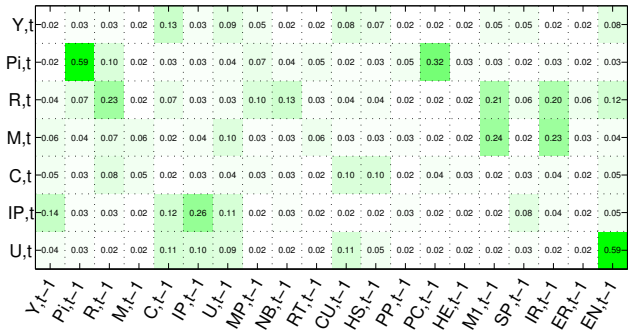
(c) MIN

FIGURE3.5: Contemporaneous structure of PC, SSVS and MIN averaged over the period $1960Q1-2006Q4$. 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.

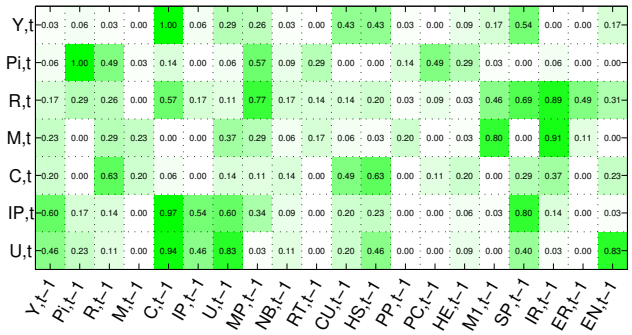
To allow for model changes, we apply a moving window to estimate, for each window, the structure and the associated predictive model. The moving window uses the most recent observations to estimate the model. Thus new data are added to the exiting dataset and the oldest observations are deleted. We initialize the first



(a) C-GC



(b) SSVS(γ)

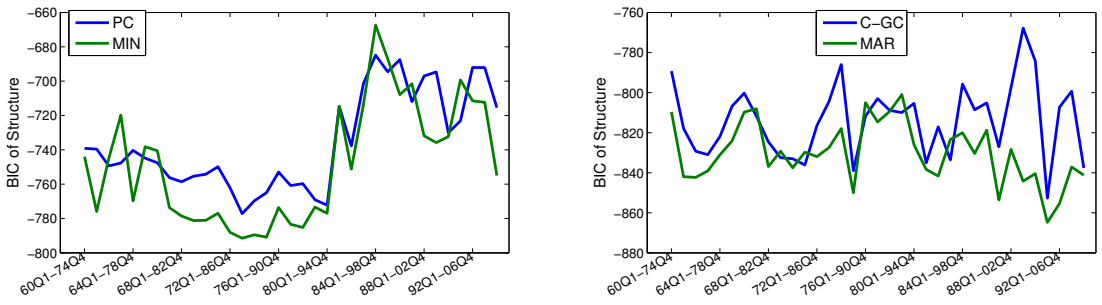


(c) MAR

FIGURE3.6: Temporal dynamic structure of C-GC, SSVS and MAR averaged $1960Q1-2006Q4$. The light (dark) green color indicates weak (strong) evidence of dependence. Response (explanatory) variables are on the rows (columns). The variables are: (Y) - real 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.

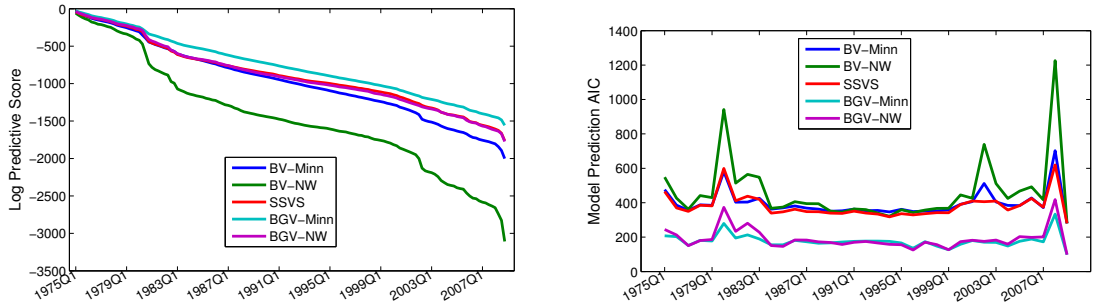
sample from $1960Q1 - 1974Q4$, and then move the window forward by 4-quarters.

We produce forecasts for all the horizons up to 4-quarters ahead. Our last sample is



(a) BIC of MIN (b) BIC of MAR

FIGURE3.7: The BIC of the contemporaneous and temporal dependence structures of the PC (blue), C-GC (blue) and the MAR (green) over the period $1960Q1 - 2006Q4$.

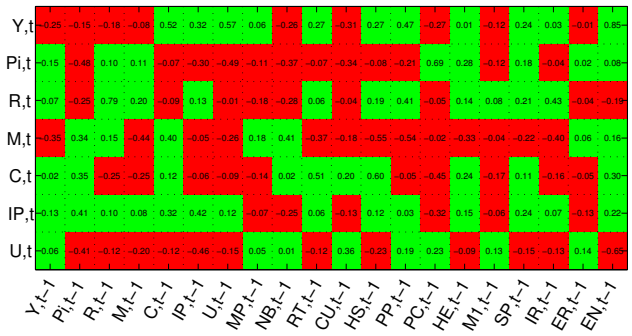


(a) Log predictive score (b) Predictive AIC

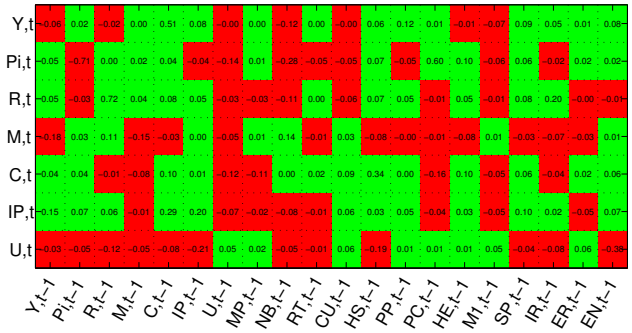
FIGURE3.8: Log predictive score and predictive AIC of models over $1975Q1 - 2008Q4$. BV-Minn (BVAR with Minnesota prior (blue)), BV-NW (BVAR with normal-Wishart (green)), SSVS (Stochastic Search Variable Selection (red)), BGV-Minn (BGVAR with Minnesota prior (cyan)), and BGV-NW (BGVAR with normal Wishart (pink)).

set to $1993Q1 - 2007Q4$. Thus the forecast period ranges from $1975Q1$ to $2008Q4$.

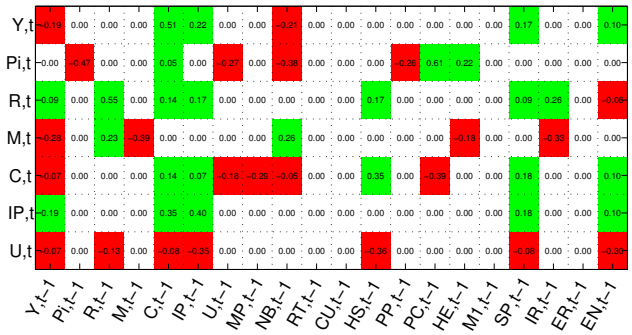
We report in Figure 3.5 the results of PC, SSVS(ω) and MIN for the macroeconomic application averaged over $1960Q1 - 2006Q4$. The PC algorithm identified strong evidence of the following relationships: $C_t - Y_t - IP_t - U_t$, and $M_t - R_t$. SSVS(ω) shows $C_t - Y_t - IP_t - U_t$, and MIN reveals: $Y_t \rightarrow C_t$, $IP_t \rightarrow (Y_t, U_t)$, and $R_t \rightarrow M_t$. Thus all the algorithms capture contemporaneous 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



(a) BVAR Coefficients



(b) SSVS Coefficients



(c) BGVAR Coefficients

FIGURE3.9: 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.

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 issues related to the exclusion of other relevant variables.

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 3.10 in 3.5.1). 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.

The results in 3.5.1 show that 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.

In Figure 3.6 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 3.6 may be close to zero since they are averages over a sequence of rolling estimates.

Figure 3.7a 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 3.7b 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 Figure 3.8a and 3.8b, we report the evolution of the log predictive scores and predictive AIC 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 predictive AIC 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 predictive AIC 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.

We compare the estimated coefficient matrices of the BVAR (normal-Wishart),

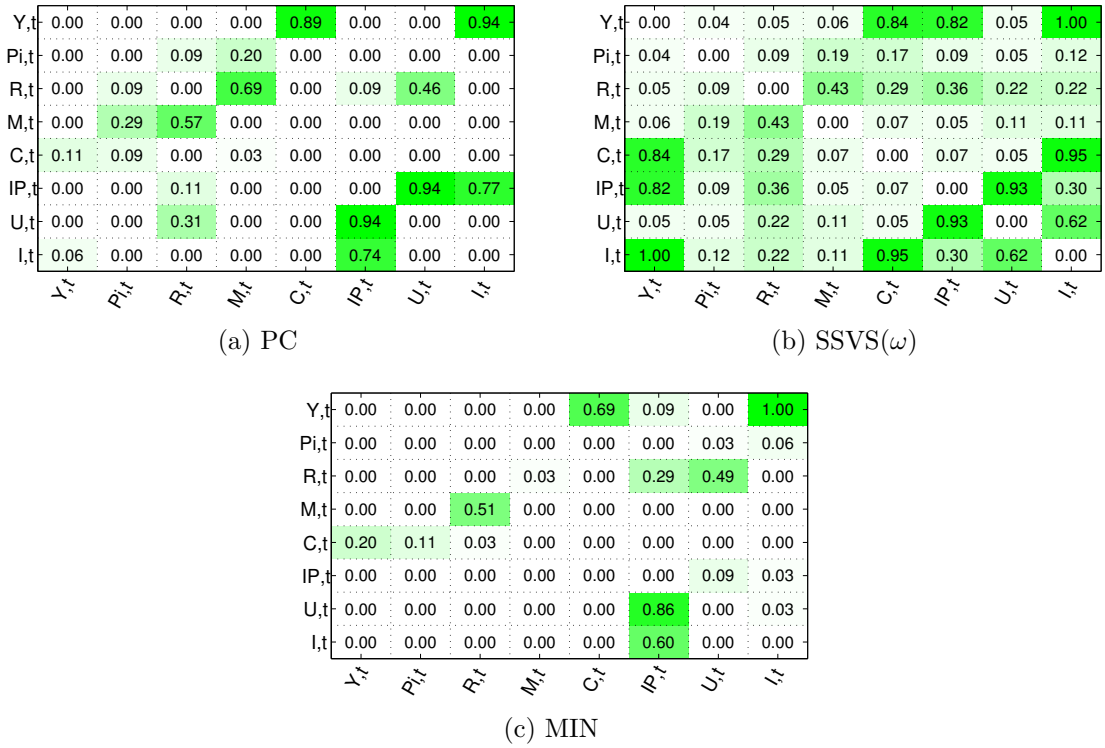
SSVS, and BGVAR (normal Wishart prior) for the last window ($1993Q1 - 2007Q4$) of the sample data. This is shown in Figure 3.9. Elements in red (green) represent negative (positive) coefficients and white elements represent zeros. Most of the BAVR coefficient values are slightly higher (in absolute values) 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 the figure, 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.

3.5.1 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 3.10 shows the results of PC, SSVS(ω) and MIN for the augmented model.

3.6 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



(a) PC (b) SSVS(ω)

(c) MIN

FIGURE3.10: 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.

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. (2014), 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 A.4). 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).

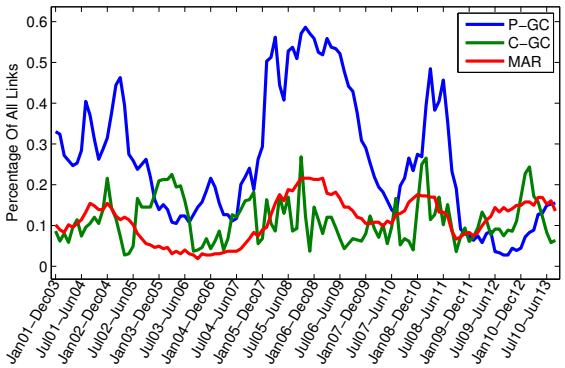
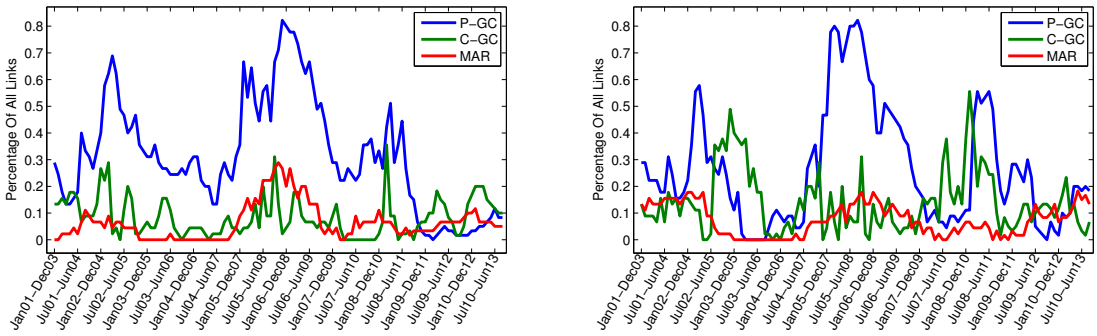


FIGURE3.11: Total connectivity index among super-sectors of Euro Stoxx 600 from January 2001 - August 2013 based on a 36-month moving window. P-GC (blue), C-GC (green) and MAR (red).

We compute the number of linkages by summing all the edges in the graph structure for each window. Figure 3.11 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 3.11, 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.



(a) Financial → Non Financial

(b) Non Financial → Financial

FIGURE3.12: Total connectivity index among (3.12a) from financial to non-financial and (3.12b) from non-financial to financial the super-sectors of Euro Stoxx 600 from January 2001 - August 2013. P-GC (blue), C-GC (green) and MAR (red).

From the MAR, we identify four periods of high interconnectedness. The periods are pre-2005, 2007-2009, 2010-2011, and 2011-2013. By matching these periods to notable global and European events, 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 3.12 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 Figure 3.11. 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.

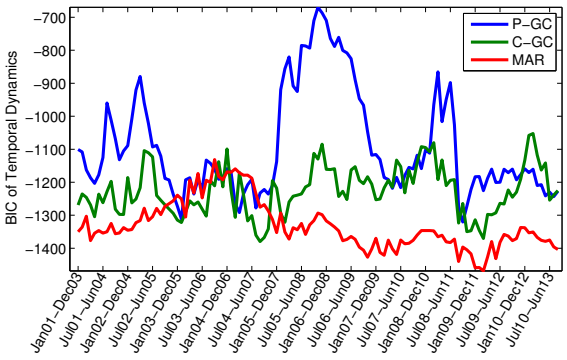


FIGURE3.13: The BIC score of the graph estimation performance from January 2001 - August 2013 based on a 36-month moving window. P-GC in blue, C-GC (green) and MAR (red).

The BIC scores (Figure 3.13, 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 confirms that BGVAR provides a more accurate linkages among the institutions than P-GC and C-GC, offering a better approach to studying systemic risk.

3.7 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 MCMC 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, results show that the BGVAR produce 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.

Chapter 4

Sparse Graphical VAR: A Bayesian Approach

4.1 Introduction

High dimensional modeling and large dataset handling have recently gain attention in several fields, particularly in economics and finance. This has become necessary since useful information is often scattered among large number of variables. Building models that allow for extraction of these information from large dataset enhances a better understanding of the modern economic and financial system. Many researchers have shown that combining financial and macroeconomic variables to estimate large vector autoregressive (VAR) models produces better forecasts than standard approaches (see, Banbura et al., 2010; Koop, 2013; Giannone et al., 2005; Stock and Watson, 2012; Carriero et al., 2013). Many others, using datasets of a large number of financial institutions, have shown that the financial system has become highly interconnected and thus, can be represented as a network where linkages among agents sharing common structures play a fundamental role in contagion and

This chapter is based on: Ahelegbey, D. F., Billio, M. and Casarin, R. (2015). Sparse Graphical Vector Autoregression: A Bayesian Approach, *Working Paper, Social Science Research Network*.(Submitted).

the spread of systemic risk (see, Huang et al., 2012; Billio et al., 2012; Diebold and Yilmaz, 2014; DasGupta and Kaligounder, 2014; Hautsch et al., 2014).

In this paper we propose a new Bayesian model for multivariate time series of large dimension by combining graph-based notion of causality (see Pearl, 1988; Lauritzen and Wermuth, 1989; Whittaker, 1990), with the concept of sparsity (see, e.g. Box and Meyer, 1986). Graphical models have been applied in time series analysis for estimating causal structures in VAR models (see Corander and Villani, 2006; Demiralp and Hoover, 2003; Swanson and Granger, 1997; Moneta, 2008) and identification restrictions in structural VAR (Ahelegbey et al., 2015). They have received increasing attention as tools to represent interconnectedness and sources of contagion among financial institutions (see Barigozzi and Brownlees, 2014; Billio et al., 2012; Diebold and Yilmaz, 2014; Ahelegbey and Giudici, 2014; Merton et al., 2013). As described in the following, we contribute to the literature in many ways.

One of the key challenges of high-dimensional models is the complex interactions among variables and the inferential difficulty associated with handling large datasets. For instance, in large VAR models, econometricians encounter the curse of dimensionality problem due to high number of variables relative to the number of data points. The standard Bayesian VAR approach to this problem is parameter shrinkage by applying the Minnesota prior (see, e.g. De Mol et al., 2008; Banbura et al., 2010; Doan et al., 1984). This approach is however inefficient to deal with the problem of indeterminacy (see Donoho, 2006), i.e. when the number of parameters to estimate in a system of equations exceeds the number of observations. Other common approaches discussed in the literature are dimension reduction and variable selection methods. For dimension reduction, there is a very large literature on dynamic factor models (DFM) (see, e.g. Geweke, 1977; Sargent and Sims, 1977; Bai and Ng, 2002; Stock and Watson, 2002; Forni et al., 2000, 2005, 2004), factor augmented VAR (FAVAR) models (Bernanke et al., 2005), and recent development in sparse (dynamic) factor

models (Kaufmann and Schumacher, 2013; Bhattacharya and Dunson, 2011; Carvalho et al., 2008). For variable selection methods, standard methods considered are Bayesian model averaging (see, e.g. Koop and Potter, 2004; Jacobson and Karlsson, 2004) and regularization methods like the Least Absolute Shrinkage and Selection Operator (LASSO) of Tibshirani (1996) and its variants, (see, e.g. Efron et al., 2004; Kyung et al., 2010; Park and Casella, 2008). The method considered in this paper is related to the latter, thus to variable selection.

Variable selection is a fundamental problem in high-dimensional models, and this is closely related to the possibility to describe the model with sparsity (Zhang et al., 2012). The idea of sparsity is associated with the notion that a large variation in the dependent variables is explained by a small proportion of predictors (Box and Meyer, 1986). Modeling sparsity has received attention in recent years in many fields, including econometrics, (see Korobilis, 2013; Elliott et al., 2013; Gefang, 2014). See also Belloni and Chernozhukov (2011) for an introduction to high-dimensional sparse econometric models.

This paper introduces and models sparsity in graphical VAR models of large dimension by dealing also with uncertainty in the lag order. It thus substantially extends the graphical VAR model, the inference approach and posterior approximation algorithm given in Ahelegbey et al. (2015). In most empirical analyses, researchers often overlook dependence among series when dealing with multi-equation regression models and large number of predictors, (see, e.g. Korobilis, 2013; Stock and Watson, 2014), since model selection is a difficult issue and such approach is often necessary to avoid the indeterminacy problem. However, this can be unsatisfactory in terms of interpretability and forecasting performance, since temporal dependence in the series is ignored. The graphical approach presented in this paper enables us to deal with this indeterminacy problem by exploiting sparsity to estimate the dynamic causal structure in large VAR models.

Many studies have considered several approaches to model sparse graphs (see, e.g. Jones et al., 2005; Carvalho et al., 2008; Scott and Carvalho, 2008; Shojaie and Michailidis, 2010). Also, there is an increasing interest in sparsity estimation for large VAR models (see, e.g. Davis et al., 2012; Gefang, 2014; Song and Bickel, 2011; Kock and Callot, 2015; De Mol et al., 2008; Medeiros and Mendes, 2012). We contribute to this literature by focusing on graphical VAR models from a Bayesian perspective with suitable prior specifications to deal with sparsity on the temporal dependence. More precisely, we build on the fan-in method of Friedman and Koller (2003) and propose a new approach to sparsity modeling. The idea of the fan-in is based on imposing a maximal number of predictors to ensure sparsity on the graph. Setting an a-priori hard fan-in might be too restrictive for large VAR applications. We therefore propose a prior distribution on the fan-in to allow for different prior information level about the maximal number of predictors for each equation of the VAR model. Thus, we allow for a random fan-in and adapt this prior distribution to the prior probability in variable selection problems. We show that this new prior distribution encourages sparsity on the graph taking into account the lag order. Since there is duality between prior and the penalty in the information criterion, our prior leads to a modified BIC for graphical model selection.

We also contribute to the literature on dynamic relationship identification. Here, we propose an efficient Markov Chain Monte Carlo (MCMC) algorithm to sample jointly, the graph structure, the lag order and the parameters of the VAR model. Due to the uncertainty on the lag order, we propose an efficient MCMC algorithm that takes advantage of computational power through parallel simulation of the graph and lag order. Inference of the graph and lag order allows us to estimate only the parameters of the relevant explanatory variables.

We show the efficiency and study the performance of our approach through simulation examples and an application to forecast macroeconomic times series with large

number of predictors. We consider the standard Lasso-type methods (i.e. LASSO and Elastic-Net) as benchmarks for comparing the identified causal structure and the forecast ability. We find evidence that our sparse graphical VAR model is more parsimonious than the LASSO and Elastic-Net. Furthermore, we find evidence of a gain in the predictive accuracy of our approach over the Lasso-type methods.

The paper is organized as follows: Section 4.2 presents the graphical concept for VAR models; Section 4.3 discusses prior distributions and focuses on the interaction between lag order and sparse graph prior distribution; Section 4.4 discusses the Bayesian inference scheme; Section 4.5 illustrates the simulation experiments; and Section 4.6 presents the empirical application.

4.2 Graphical VAR Models

Graphical models are statistical models that summarize the marginal and conditional independences among random variables by means of graphs (see Brillinger, 1996). The core of such models is representing relationships among variables in the form of graphs using nodes and edges, where nodes denote variables and edges show interactions. They can be represented in a more compact form by $(G, \theta) \in (\mathcal{G} \times \Theta)$, where G is a graph of relationships among variables, θ is the graphical model parameters, $\mathcal{G}$ is the space of the graphs and Θ is the parameter space.

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

$$Y_t = \sum_{i=1}^{p} B_i X_{t-i} + \varepsilon_t \tag{4.1}$$

$t = 1, \ldots, T$, where ε_t is $n_y \times 1$ vector of errors, independent and identically normal,

with mean zero and covariance matrix Σ_ε; p is the maximum lag order; B_i, $1 \leqslant i \leqslant p$ is $n_y \times n$ matrix of coefficients.

By interpreting (4.1) as a model with temporal dependence between explanatory and dependent variables, The VAR model can be expressed in a graphical framework (referred to as graphical VAR model), with a one-to-one correspondence between the coefficient matrices and a directed acyclic graph; if $B_{s,ij} \neq 0$ then there is a causal effect of X^j_{t-s} on Y^i_t, $1 \leqslant s \leqslant p$. Here we read X^i_t as realization of the i-th element of X at time t.

More formally, we define the relation $B_s = (G_s \circ \Phi_s)$, where G_s is a $n_y \times n$ binary connectivity matrix (also called adjacency matrix), Φ_s is a $n_y \times n$ coefficients matrix, and the operator $(\circ)$ is the element-by-element Hadamard's product (i.e., $B_{s,ij} = G_{s,ij}\ \Phi_{s,ij}$). Based on this definition, we identify a one-to-one correspondence between B_s and Φ_s conditional on G_s, such that $B_{s,ij} = \Phi_{s,ij}$, if $G_{s,ij} = 1$; and $B_{s,ij} = 0$, if $G_{s,ij} = 0$. The above relationship can be presented in a stacked matrix form as, $B = (G \circ \Phi)$, where $B = (B_1, \ldots, B_p)$, $G = (G_1, \ldots, G_p)$ and $\Phi = (\Phi_1, \ldots, \Phi_p)$, where each matrix is of dimension $n_y \times np$.

Let W_t be stacked lags of X_t, where $W_t = (X'_{t-1}, \ldots, X'_{t-p})'$ is of dimension $np \times 1$, with p as the lag order, and $V_t = (Y'_t, W'_t)'$ of dimension $(n_y + np) \times 1$. Suppose the joint, V_t, follows the distribution, $V_t \sim \mathcal{N}(0, \Omega^{-1})$, where $\Sigma = \Omega^{-1}$ is $(n_y + np) \times (n_y + np)$ is the covariance matrix. The joint distribution of the variables in V_t can be summarized with a graphical model, (G, θ), where $G \in \mathcal{G}$ consists of directed edges. In this paper, we focus on modeling directed edges from elements in W_t to elements in Y_t, capturing the temporal dependence among the observed variables. More specifically, $G_{ij} = 0$, means the i-th element of Y_t and j-th element of W_t are conditionally independent given the remaining variables in V_t, and $G_{ij} = 1$ corresponds to a conditional dependence between the i-th and j-th elements of Y_t and W_t respectively given the remaining variables in V_t. The graphical model

parameter, $\theta \in \Theta$, consist the coefficients, capturing the strength and sign of the temporal dependence relationship described by G. Based on the above assumption, estimating the model parameters associated with G is equivalent to estimating the precision matrix, Ω, i.e $\theta = \Omega$. The relationship between the parameters of the VAR, $\{B, \Sigma_\varepsilon\}$, and that of the graphical model, Ω, is as follows.

Proposition 1. Suppose the marginal distribution of $W_t \sim \mathcal{N}(0, \Sigma_{ww})$ and the conditional distribution of $Y_t|W_t \sim \mathcal{N}(BW_t, \Sigma_\varepsilon)$. There is a correspondence between $\{B, \Sigma_\varepsilon\}$ and Ω, such that given Ω, we obtain $\Sigma = \Omega^{-1}$ and $\{B, \Sigma_\varepsilon\}$ by

$$B = \Sigma_{yw}\Sigma_{ww}^{-1}, \qquad \Sigma_\varepsilon = \Sigma_{yy} - \Sigma_{yw}\Sigma_{ww}^{-1}\Sigma_{wy} \tag{4.2}$$

Also given $\{B, \Sigma_\varepsilon\}$ and Σ_{ww}, $\Omega = \Sigma^{-1}$ of (Y_t, W_t) is estimated by

$$\Omega = \begin{pmatrix} \Sigma_\varepsilon^{-1} & -\Sigma_\varepsilon^{-1}B \\ -B'\Sigma_\varepsilon^{-1} & \Sigma_{ww}^{-1} + B'\Sigma_\varepsilon^{-1}B \end{pmatrix} \tag{4.3}$$

Proof. See B.1.1. □

Following our definition, $B = (G \circ \Phi)$ and the results of Proposition 1, we identify the relationship between Ω and the dependence structure G, through the sub-matrix $(\Sigma_\varepsilon^{-1}B)$ of Ω. We denote $\Omega^G = \Sigma_\varepsilon^{-1}B$, defined on the space $\mathcal{M}(G)$, i.e. the set of precision matrices where elements of Ω^G corresponds to elements in G. Clearly, if the errors are assumed to be independent and normally distributed, Σ_ε is a diagonal matrix, which when normalized to identity leads to a one-to-one correspondence between B, Ω^G and G such that $B_{ij} = \Omega_{ij}^G = 0$ if $G_{ij} = 0$ and $B_{ij} = \Omega_{ij}^G \neq 0$ if $G_{ij} = 1$. In large VAR estimation, most empirical papers follow the above assumption on the errors to estimate the model, (see Stock and Watson, 2014; Kock and Callot, 2015).

In this paper we assume Σ_ε is a full matrix, i.e, the errors are correlated among the equations of the VAR. Estimating our graphical VAR model therefore involves: the temporal dependence graph, G, the maximum lag order, p, and the set of parameters

in Ω which related to $\{B, \Sigma_\varepsilon\}$. Estimating all these jointly is challenging. However, following the Bayesian paradigm allows us to take into account uncertainties on the quantities of interest through model averaging, (Madigan and York, 1995; Giudici and Green, 1999). The objective of this paper is to estimate jointly the lag order and graph from the observed time series, and to incorporate the inferred quantities to select the relevant variables to estimate the parameters of the model.

4.3 Sparse Bayesian Graphical VAR Models

In a system of linear equations where the number of parameters exceeds the number of observations, for instance in large VAR models, we face another problem referred to as indeterminacy, (see Donoho, 2006). Such systems can be modeled by exploiting sparsity. The description of our graphical VAR for high dimensional multivariate time series is completed with the elicitation of the prior distributions for the lag order p, a sparsity prior on the graph, and the prior on G and Ω.

4.3.1 Lag Order Prior Distribution

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

$$P(p_i) = \frac{1}{(\bar{p} - \underline{p} + 1)} \mathbb{I}_{\{\underline{p}, \ldots, \bar{p}\}}(p_i) \tag{4.4}$$

where $\mathbb{I}_A(x)$ is the indicator function, that is unity if $x \in A$ and zero otherwise. This is a standard choice in AR model selection problems (e.g.,Casarin et al. (2012)). Alternatively, the lag order can be assumed to follow a truncated Poisson distribution with mean λ and maximum $\bar{p}$ (Vermaak et al. (2004)), or a discretized Laplace distribution (Ehlers and Brooks (2004)). Our choice of discrete uniform distribution is fairly informative since $\underline{p}$ and $\bar{p}$ are defined a-priori following standard applications

and assigns equal weights to all possible values of p_i. For instance, in estimating VAR models with monthly (quarterly) time series, the standard lag selection procedure often consider $\underline{p} = 1$ and $\bar{p} = 12$ ($\bar{p} = 4$). The alternative lag order prior distributions are more informative and assigns different weights to the possible values of p_i.

4.3.2 Standard Graph Prior Distribution

Most of the literature on graphical models takes the prior for a graph G with n variables as uniform over all the relevant graphs, i.e., $P(G) = |\mathcal{G}|^{-1}$, where $|\mathcal{G}|$ is the cardinality of $\mathcal{G}$, (see Geiger and Heckerman, 2002; Giudici and Castelo, 2003). This can be attributed to the fact that the number of possible graphs increases super-exponentially with the number of variables, and there is difficulty in calculating the number of possible graphs. Assuming uniform probabilities for all graphs, however, does not ensure sparsity. Thus, many authors have discussed several approaches to penalize globally or locally "dense" graphs (see, e.g. Jones et al., 2005; Carvalho et al., 2008; Scott and Carvalho, 2008; Wang, 2010). See also Telesca et al. (2012) and Shojaie and Michailidis (2009) for the use of explicit information prior to improve the estimation of the graph structure.

Friedman and Koller (2003) proposed a factorization of the graph prior into equation specific terms for DAG models. As argued by the authors, setting an upper bound on the number of explanatory variables for each dependent variable encourages sparsity on the graph. This bound is referred to as the fan-in restriction in the graphical model literature. Let m be the maximum number of explanatory variables for each equation. Restricting the graph model selection to at most f explanatory variables instead of m, $f < m$, reduces the number of possible sets from $\mathcal{O}(2^m)$ to $\binom{m}{f}$, where $\binom{n}{k}$ is the binomial coefficient. A uniform choice on the latter set yields

a graph prior

$$P(G) = \prod_{i=1}^{n} P(\pi_i) \propto \prod_{i=1}^{n} \binom{n-1}{|\pi_i|}^{-1} \tag{4.5}$$

where $\pi_i = \{j = 1, \ldots, n : G_{ij} = 1\}$ is the set of explanatory variables of the i-th equation, $|\pi_i|$ is the number of elements in π_i, and $P(\pi_i)$ is the local graph prior of the i-th equation.

Jones et al. (2005) discussed a prior distribution for penalizing the inclusion of additional edges in dense graphs given by

$$P(G|\gamma) = \gamma^k (1-\gamma)^{m-k} \tag{4.6}$$

where m is the maximum number of edges and k represents the number of edges in the graph. In their application, the authors use a Bernoulli prior on each edge inclusion with parameter $\gamma = 2/(n-1)$ and set $m = \binom{n}{2}$.

For choices of the prior distribution on γ in the beta family, Scott and Carvalho (2008) showed that γ can be explicitly marginalized out. The uniform prior on γ gives a marginal prior inclusion probability of 1/2 for all edges and yields model probabilities

$$P(G) = \frac{1}{(m+1)} \binom{m}{k}^{-1} \tag{4.7}$$

4.3.3 Sparsity Prior Distribution

We build on the fan-in approach of Friedman and Koller (2003) by introducing a prior distribution on the fan-in to allow for different prior information level about the maximal number of predictors for each equation of the VAR model.

In a multivariate dynamic models with n variables and a lag order p, the number of possible predictors is np. Given that each series has T number of observations, then the number of observations of the model is $T-p$. In large VAR models, it is often natural that the number of predictors is greater than the number of observations,

i.e. $np > T - p$. When this happens, we expect that each equation has at most $T - p$ predictors to efficiently estimate the model. In cases where $T - p > np$, we expect that each equation has at most np predictors. Thus the maximal number of explanatory variables required to efficiently estimate a high dimensional model is given by $m_p = \min\{np, T - p\}$. Setting an a-priori hard fan-in (see Friedman and Koller, 2003) might be too restrictive for large VAR applications.

We denote with $\bar{\eta}$, $0 \leqslant \bar{\eta} \leqslant 1$, the measure of the maximal density, i.e. the fraction of the predictors that explains the dependent variables. Thus the level of sparsity is given by $(1 - \bar{\eta})$. We set the fan-in to $f = \lfloor \bar{\eta} m_p \rfloor$, where f is the largest integer less than $\bar{\eta} m_p$. To allow for different levels of sparsity for the equations in the VAR model, we assume a prior distribution on the maximal density for the different equations. We denote $\bar{\eta}_i$ the maximal density of the i-th equation and assume the prior on $\bar{\eta}_i$, given lag order p_i is beta distributed with parameters $a, b > 0$, $\bar{\eta}_i | p_i \sim \mathcal{B}e(a, b)$, on the interval $[0, 1]$

$$P(\bar{\eta}_i | p_i) = \frac{1}{B(a, b)} \bar{\eta}_i^{a-1} (1 - \bar{\eta}_i)^{b-1} \tag{4.8}$$

This leads to random fan-in's for the different equations in the VAR model. Note that the fan-in, f_i, must be directly related to the number of selected predictors in each equation and indirectly related to the number of observations of the model.

4.3.4 Our Graph Prior Distribution

We define the graph prior for each equation in the VAR model conditional on the sparsity prior. We refer to the prior on the graph of each equation as the local graph prior, denoted by $P(\pi_i | p_i, \gamma, \bar{\eta}_i)$. Following (Scott and Berger, 2010), we consider the inclusion of predictors in each equation as exchangeable Bernoulli trials with prior probability

$$P(\pi_i | p_i, \gamma, \bar{\eta}_i) = \gamma^{|\pi_i|} (1 - \gamma)^{np_i - |\pi_i|} \, \mathbb{I}_{\{0, \ldots, f_i\}}(|\pi_i|) \tag{4.9}$$

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

$$P(\pi_i|p_i, \bar{\eta}_i) = \frac{1}{2^{np_i}} \mathbb{I}_{\{0,\dots,f_i\}}(|\pi_i|) \tag{4.10}$$

Alternatively, a uniform prior on γ gives to each variable a marginal prior inclusion probability of 1/2 and a local graph prior (Foygel and Drton, 2011)

$$P(\pi_i|p_i, \bar{\eta}_i) = \binom{np_i}{|\pi_i|}^{-1} \mathbb{I}_{\{0,\dots,f_i\}}(|\pi_i|) \tag{4.11}$$

4.3.5 Parameter Prior Distribution

There are two main approaches to define parameter priors for graphical models, however a common feature to these approaches is that both are graph conditional parameter priors. On one hand is a vast work on Gaussian DAG models discussing a list of conditions that permits an unconstrained precision matrix Ω (see, e.g. Heckerman and Geiger, 1994; Heckerman and Chickering, 1995; Geiger and Heckerman, 2002; Consonni and Rocca, 2012). On the other hand is a vast publication on Gaussian decomposable undirected graphical (UG) models with constraints on the precision matrix Ω (see, e.g. Roverato, 2002; Carvalho and Scott, 2009; Wang and Li, 2012; Lenkoski and Dobra, 2011). Note that, an unconstrained Ω characterizes a complete Gaussian DAG or UG model, i.e. a graph with no missing edges. The standard parameter prior for Gaussian DAG models with zero expectations is a Wishart distribution, whereas that of UG models is a G-Wishart distribution. A consequence of the DAG conditional parameter prior is that, once we specify the parameter prior for one complete DAG model, all other priors can be generated automatically (see Consonni and Rocca, 2012). We follow the DAG framework since it allows us to

marginalize out Ω analytically and focusing on the structure inference, and then estimating the model parameters given the structure (see Section 4.4 for details).

Following Geiger and Heckerman (2002), we assume a prior distribution on the unconstrained precision matrix, Ω, conditional on any complete DAG, G, for a given lag order p, is Wishart distributed, with probability density function

$$P(\Omega|p,G) = \frac{1}{K_d(\nu, S_0)}|\Omega|^{\frac{(\nu-d-1)}{2}}\text{etr}\Big(-\frac{1}{2}\Omega S_0\Big) \tag{4.12}$$

where $\text{etr}(A) = \exp\{\text{tr}(A)\}$ and $\text{tr}(A)$ is the trace of a square matrix A, ν is the degree of freedom parameter, S_0 is a $d \times d$ symmetric positive definite matrix, with $d = n_y + np$, the size of the vector of stacked dependent and explanatory variables of the model. The normalizing constant is:

$$K_d(\nu, S_0) = 2^{\frac{\nu d}{2}}|S_0|^{-\frac{\nu}{2}}\Gamma_d\Big(\frac{\nu}{2}\Big) \tag{4.13}$$

where $\Gamma_d(a) = \pi^{\frac{d(d-1)}{4}}\prod_{i=1}^{d}\Gamma\Big(a + \frac{1-i}{2}\Big)$ is the multivariate gamma function and $\Gamma(\cdot)$ denotes the gamma function.

Based on the assumption that the conditional distribution of the dependent variables given the set of predictors, is described by equation (4.1), with parameters $\{B, \Sigma_\varepsilon\}$, we assume the prior distribution on $(B, \Sigma_\varepsilon|p, G)$ is an independent normal-Wishart (see, e.g. Heckerman and Geiger, 1994; Geiger and Heckerman, 2002). This is one of the prior distributions extensively applied in the Bayesian VAR literature. Specifically, we assumed the coefficients matrix B is independent and normally distributed, $B|p, G \sim \mathcal{N}(\underline{B}, \underline{V})$, and Σ_ε^{-1} is Wishart distributed, $\Sigma_\varepsilon^{-1} \sim \mathcal{W}(\underline{\nu}, \underline{S}/\underline{\nu})$. The prior expectation, $\underline{B} = \underline{0}_{n_y \times np}$, is a zero matrix, and the prior variance of the coefficient matrix, $\underline{V} = I_{np\times np}$, is an identity matrix. Also, the prior expectation of $\Sigma_\varepsilon = \frac{1}{\underline{\nu}}\underline{S}$ where is $\underline{S}$ is $n_y \times n_y$ positive semi-definite matrix and $\underline{\nu} > n_y - 1$ is the degrees of freedom.

4.4 Bayesian Inference

We define G_s as $n_y \times n$ binary connectivity matrix that captures the temporal relationship of variables at time $t-s$ with the variables at time t. We denote with $\vec{G}_p = (G_1, \ldots, G_p)$ as stacked $G_1, \ldots, G_p$, such that $\vec{G}_p$ is of dimension $n_y \times np$. We then define $\vec{G}_{p,i},\ i = 1, \ldots, n_y$ as the local graph associated with the i-th equation which is the i-th row of $\vec{G}_p$. The likelihood function of a random sample $\mathcal{X} = (V_1, \ldots, V_{T_0}) \sim \mathcal{N}(0, \Omega^{-1})$ is given by

$$P(\mathcal{X}|p, \vec{G}_p, \Omega) = (2\pi)^{-\frac{dT_0}{2}} |\Omega|^{\frac{T_0}{2}} \operatorname{etr}\Big(-\frac{1}{2}\Omega\hat{S}\Big) \tag{4.14}$$

where $T_0 = T - p$, T is the length of observation of the time series, $\hat{S} = \sum_{i=1}^{T_0} V_t V_t'$, sum of squares matrix of dimension $d \times d$. A nice feature of the unconstrained parameter prior in the DAG mode framework is that it allows for integrating out analytically, the precision matrix, Ω, with respect to its prior to obtain a marginal likelihood function for any DAG, $\vec{G}_p$ with lag p given by

$$P(\mathcal{X}|p, \vec{G}_p) = \int P(\mathcal{X}|p, \vec{G}_p, \Omega)\ P(\Omega|p, \vec{G}_p)\ d\Omega = \frac{K_d(\nu + T_0, S_0 + \hat{S})}{(2\pi)^{\frac{dT_0}{2}} K_d(\nu, S_0)} \tag{4.15}$$

where S_0 and $S_0 + \hat{S}$ are the prior and posterior sum of square matrices, which when normalized are $\underline{\Sigma} = \frac{1}{\nu} S_0$ and $\bar{\Sigma} = \frac{1}{\nu + T_0}(S_0 + \hat{S})$ respectively. Geiger and Heckerman (2002) outlined conditions for the integral in equation (4.15) to be analytically tractable and to have a close form expression that can be factorized into local marginal likelihoods. A key assumption is that the parameters must be independent within and across equations. In VAR models, the errors are correlated across equations which makes the factorization of (4.15) quite problematic. Following the idea of large-sample approximation by Kass et al. (1988) and Chickering and Heckerman (1997), we consider the errors of a large VAR model as unobserved variables which can be ignored when dealing with large datasets (see, e.g. Stock and Watson, 2014).

Based on this consideration and assumption that the coefficients are independent a-priori within and across equations, we approximate (4.15) with a pseudo-marginal likelihood given by the product of local densities

$$P(\mathcal{X}|p,\vec{G}_p) \approx \prod_{i=1}^{n_y} P(\mathcal{X}|p_i,\vec{G}_{p,i}(y_i,\pi_i)) = \prod_{i=1}^{n_y} \frac{P(\mathcal{X}^{(y_i,\pi_i)}|p_i,\vec{G}_{p,i})}{P(\mathcal{X}^{(\pi_i)}|p_i,\vec{G}_{p,i})} \tag{4.16}$$

where $\vec{G}_p(y_i,\pi_i)$ is the local graph of the i-th equation with y_i as dependent variable and π_i as the set of predictors; $\mathcal{X}^{(y_i,\pi_i)}$ is the sub-matrix of $\mathcal{X}$ consisting of y_i and π_i; and $\mathcal{X}^{(\pi_i)}$ is the sub-matrix of π_i. This approximation allows us to develop search algorithms to focus on local graph estimation. More specifically, a Markov chain Monte Carlo (MCMC) algorithm using the global score would be less efficient in exploration since the global score would be insensitive to the proposal of edge deletion or addition. Thus, the approximation allows the chain to explore the graph locally at equation level. The pseudo-likelihood has been used in MCMC by Zhou and Schmidler (2009) to circumvent the intractable normalizing constant problem in random fields. See also Andrieu and Roberts (2009); Maclaurin and Adams (2014) for one of the approximated likelihood in MCMC. The closed form of (4.16) is

$$P(\mathcal{X}^{d_i}|p_i,\vec{G}_{p,i}) = \pi^{\frac{-T_i|d_i|}{2}} \frac{|\bar{\Sigma}_{d_i}|^{-\frac{(\nu+T_i)}{2}}}{|\underline{\Sigma}_{d_i}|^{-\frac{\nu}{2}}} \prod_{i=1}^{|d_i|} \frac{\Gamma\left(\frac{\nu+T_i+1-i}{2}\right)}{\Gamma\left(\frac{\nu+1-i}{2}\right)} \tag{4.17}$$

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

4.4.1 Posterior Approximation

Inferring jointly the lag and the graph allows for selecting the relevant predictors to estimate the model parameters (B,Σ_ε). In order to approximate the posterior

distributions of the equations of interest, the standard approach is to consider a collapsed Gibbs sampling. At the j-th iteration, the sampler consists of the following:

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

2. Estimate $B^{(j)}$ and $\Sigma_\varepsilon^{(j)}$ directly from $P(B, \Sigma_\varepsilon|p^{(j)}, \vec{G}_p^{(j)}, \mathcal{X})$.

As regards to the first step, standard MCMC algorithms (Madigan and York, 1995) such as Gibbs sampling and Metropolis-Hastings (MH) apply only to model selection with fixed dimensions. In model selection problems with unknown lag order, the dimension of the model varies with the lag order. The algorithm extensively applied for this problem is the reversible jump (RJ) MCMC (Green, 1995). In graphical models especially, the space of possible graphs increases super-exponentially with the number of variables (Chickering et al., 2004). Therefore, sampling from a distribution on a union of varying graph dimension using the RJ algorithm will require a higher number of iterations to thoroughly search the space of all graphs. In our graphical VAR, the inferential difficulty increases due to the random fan-in restriction.

We propose an algorithm for sampling the graph taking into consideration the random fan-in and estimating the lag order. At the j-th iteration of the Gibbs, we consider for each equation $i = 1, \ldots, n_y$ and each lag order $p_i = \underline{p}, \ldots, \bar{p}$, a sample of $\bar{\eta}_i^{(j)}$ and $\vec{G}_{p,i}^{(j)}$ from $P(\bar{\eta}_i, \vec{G}_{p,i}|p_i, \mathcal{X}) \propto P(\bar{\eta}_i|p_i)P(\pi_i|p_i, \bar{\eta}_i)P(\mathcal{X}|p_i, \vec{G}_{p,i})$. By conditioning on each possible lag order, we are able to apply standard MCMC algorithm thereby avoiding movement between models of different dimensions since the dimension is fixed for each lag. After J iterations, we average the draws, $\vec{G}_{p,i}^{(j)}$, over J and estimate $\hat{\vec{G}}_{p,i}$, for each $p_i = \underline{p}, \ldots, \bar{p}$, using the criterion discussed in B.2.2. This procedure estimates the local graph for the possible lags of $p_i \in \{\underline{p}, \ldots, \bar{p}\}$. Next, we find $(\hat{p}_i, \hat{G}_{\hat{p},i})$ which minimizes a penalized local log-likelihood (BIC) score given in (4.24). Given the estimated graph and lag order, $(\hat{p}_i, \hat{G}_{\hat{p},i})$, we select the relevant predictors

for each equation to estimate the model parameters (B, Σ_ε).

4.4.2 Graphical Model Selection

Graphical model selection is a challenge since the dimension of the graph space to explore increases super-exponentially with the number of variables. In this paper we apply MCMC and build on the MCMC algorithm described in Madigan and York (1995); Grzegorczyk and Husmeier (2011). Our algorithm differs from that of the above mentioned papers in two aspects: the initialization and the inclusion of the random fan-in restriction.

As regards to the initialization, we propose a strategy which improves the mixing of the chain. In MCMC search algorithms the space exploration crucially depends on the choice of the starting point of the chain. A set of burn-in chain iterations is often used to have a good starting point. However, Brooks et al. (2011) pointed out that any sample that is believed to be representative of the equilibrium distribution is an equally good starting point. In view of this, we propose an initialization which extracts variables (and their lags) with reliable information to improve predictions of the dependent variables. Let $\vec{G}_{p,i}$ denote the local graph of the i-th equation, $\mathbf{V}^i_{p,x}$, the vector of all possible explanatory variables with lags up to p for each equation, with $p \in [\underline{p}, \ldots, \bar{p}]$, and $\mathbf{V}_y$, the vector of dependent variables. We run the following

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

2. For each equation, $i = 1, \ldots, n_y$:

    2a. Test whether or not predictions of $y_i \in \mathbf{V}_y$ is improved by incorporating information from each $x_k \in \mathbf{V}^i_{p,x}$, i.e, $P(y_i|x_k) > P(y_i)$. Following a Minnesota type of prior, we assume recent lags (specifically lag 1) of dependent variables are more reliable to influence current realizations. Based on this idea, set $\vec{G}_p(y_i, x_k) = 1$ if x_k is equal to lag 1 of y_i, and retain x_k in $\mathbf{V}^i_{p,x}$.

2b. For x_k not equal to lag 1 of y_i, we compare the probability of the null hypothesis, $H_0 = P(\mathcal{X}|p_i, \vec{G}_p(y_i, \varnothing))$, where $\varnothing$ denote the empty set, against the probability of the alternative, $H_1 = P(\mathcal{X}|p_i, \vec{G}_p(y_i, \{x_k\}))$. If $H_1 > H_0$, we reject the null, set $\vec{G}_p(y_i, x_k) = 1$ and retain x_k in $\mathbf{V}^i_{p,x}$. If $H_1 \leqslant H_0$, we set $\vec{G}_p(y_i, x_k) = 0$ and remove x_k from $\mathbf{V}^i_{p,x}$.

3. We then denote $N_p(\pi_i)$ as the set of variables, $x'_k s$, retained in $\mathbf{V}^i_{p,x}$.

In our experience, the above initialization provides a good starting point for graphical model selection. See Figure B.1 for a comparison of the convergence diagnostics of a random initialization MCMC for the graph simulation. Using the set $N_p(\pi_i)$ of candidate predictors of the dependent variable of the i-th equation, we start our MCMC search. We proceed with the local causal search by investigating the combination of variables in $N_p(\pi_i)$ that produces the highest scoring local graph(s).

As regards the inclusion of the random fan-in restriction, we denote with $m_p = \min\{np, T - p\}$, the maximal number of predictors required to efficiently estimate the model, for $p \in [\underline{p}, \ldots, \bar{p}]$. At the j-th iteration, let $\vec{G}^{(j-1)}_{p,i}$ be the current local graph and $\pi^{(j-1)}_i$, the current set of predictors in $\vec{G}^{(j-1)}_{p,i}$, then for each equation, $i = 1, \ldots, n_y$, the Gibbs iterates the following steps:

1. Draw the sparsity parameter for the forward proposal probability, $\bar{\eta}^{(*)}_i$ from a $\mathcal{B}e(a, b)$ and set the fan-in $f^{(*)}_i = \lfloor m_p \bar{\eta}^{(*)}_i \rfloor$.

2. If the number of elements in $\pi^{(j-1)}_i$ is less than the fan-in, i.e. $|\pi^{(j-1)}_i| < f^{(*)}_i$, then randomly draw a x_k from the set of candidate predictors, $N_p(\pi_i)$, and add or remove the edge between y_i and x_k, i.e. $\vec{G}^{(*)}_p(y_i, x_k) = 1 - \vec{G}^{(j-1)}_p(y_i, x_k)$. The forward proposal probability, $Q(\vec{G}^{(*)}_{p,i}|\vec{G}^{(j-1)}_{p,i}, \bar{\eta}^{(*)}_i) = 1/|N_p(\pi_i)|$. If $|\pi^{(j-1)}_i| \geqslant f^{(*)}_i$, then randomly draw a variable, x_k, from the current set of predictors,

$\pi_i^{(j-1)}$, and remove the edge between y_i and x_k, i.e. $\vec{G}_p^{(*)}(y_i, x_k) = 0$. In this case, the forward proposal probability is $Q(\vec{G}_{p,i}^{(*)}|\vec{G}_{p,i}^{(j-1)}, \bar{\eta}_i^{(*)}) = 1/|\pi_i^{(j-1)}|$.

3. To obtain the reverse proposal probability, we denote $\pi_i^{(*)}$, the set of predictors in $\vec{G}_{p,i}^{(*)}$ taking into consideration the changes made in step 2.

4. Next, we draw the sparsity parameter for the reverse proposal probability, $\bar{\eta}_i^{(**)}$ from a $\mathcal{B}e(a, b)$ and set $f_i^{(**)} = \lfloor m_p \bar{\eta}_i^{(**)} \rfloor$.

5. If $|\pi_i^{(*)}| < f_i^{(**)}$, the reverse move will involve a random draw of a variable from $N_p(\pi_i)$ to add or delete from $\vec{G}_{p,i}^{(*)}$. The reverse proposal probability is given by $Q(\vec{G}_{p,i}^{(j-1)}|\vec{G}_{p,i}^{(*)}, \bar{\eta}_i^{(**)}) = 1/|N_p(\pi_i)|$. If $|\pi_i^{(*)}| \geqslant f_i^{(**)}$, the reverse will randomly draw a variable from $\pi_i^{(*)}$ to delete from $\vec{G}_{p,i}^{(*)}$. The reverse proposal probability in this case is given by $Q(\vec{G}_{p,i}^{(j-1)}|\vec{G}_{p,i}^{(*)}, \bar{\eta}_i^{(**)}) = 1/|\pi_i^{(*)}|$.

6. From (4.10), the ratio of the local graph priors simplifies to 1 and the acceptance probability is $A(\vec{G}_{p,i}^{(*)}, \bar{\eta}_i^{(*)}|\vec{G}_{p,i}^{(j-1)}, \bar{\eta}_i^{(**)}) = \min\{1,\ R_A\}$ where

$$R_A = \left\{ \frac{P(\mathcal{X}|p_i, \vec{G}_{p,i}^{(*)})}{P(\mathcal{X}|p_i, \vec{G}_{p,i}^{(j-1)})} \ \frac{Q(\vec{G}_{p,i}^{(j-1)}|\vec{G}_{p,i}^{(*)}, \bar{\eta}_i^{(**)})}{Q(\vec{G}_{p,i}^{(*)}|\vec{G}_{p,i}^{(j-1)}, \bar{\eta}_i^{(*)})} \right\} \tag{4.18}$$

where $P(\mathcal{X}|p_i, \vec{G}_{p,i}) = P(\mathcal{X}|p_i, \vec{G}_p(y_i, \pi_i))$, and can be computed from (4.16) and (4.17). Without the fan-in restriction, the proposal distribution is symmetric, thus, the prior and inverse proposal ratio in (4.18) simplifies to 1.

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

A description of the pseudo-code for the graph selection is given in B.3.

4.4.3 Duality between Priors and Penalties

Thanks to the duality between prior distributions and the penalization of likelihood functions, we define an information criterion for choosing $\hat{p}_i$ and $\hat{\vec{G}}_{p,i}$ by solving

$$(\hat{p}_i, \hat{\vec{G}}_{p,i}) = \arg\max_{p_i, \vec{G}_{p,i}} \quad P(p_i)P(\pi_i|p_i)P(\mathcal{X}|p_i, \vec{G}_{p,i}) \tag{4.19}$$

$$\text{where} \qquad P(\pi_i|p_i) = \int P(\pi_i|p_i, \bar{\eta}_i)P(\bar{\eta}_i|p_i)d\bar{\eta}_i \tag{4.20}$$

Following the standard BIC for graphical models, the logarithm of the marginal likelihood function of the local graph can be approximated by

$$\log P(\mathcal{X}|p_i, \vec{G}_{p,i}) \;\approx\; \log P(\mathcal{X}|p_i, \hat{\vec{G}}_{p,i}) - \frac{1}{2}|\hat{\pi}_i| \log T \tag{4.21}$$

where $\hat{\vec{G}}_{p,i}$ and $\hat{\pi}_i$ are the estimated local graph and set of relevant predictors of the dependent variable of the i-th equation.

Proposition 2. For choices of $P(\bar{\eta}_i|p_i)$ in (4.8), as beta distributed and $P(\pi_i|p_i, \bar{\eta}_i)$ according to (4.10), $\bar{\eta}_i$ can be integrated out of (4.20) to obtain

$$P(\pi_i|p_i) = \frac{1}{2^{np_i}} \sum_{j=0}^{m_p-1} \mathbb{I}_{\{0,\dots,j\}}(|\pi_i|)\Big(I_{\frac{j+1}{m_p}}(a,b) - I_{\frac{j}{m_p}}(a,b)\Big) \tag{4.22}$$

where $I_z(a,b) = \int_0^z (B(a,b))^{-1}(\bar{\eta}_i)^{a-1}(1-\bar{\eta}_i)^{b-1}\, d\bar{\eta}_i$, is the incomplete beta function (Abramowitz and Stegun, 1964, p. 263).

Proof. See B.1.2. □

Corollary 3. A uniform prior on $\bar{\eta}_i$, means $a = b = 1$ and yields

$$P(\pi_i|p_i) = \frac{1}{2^{np_i}} \left(1 - \frac{|\pi_i|}{m_p}\right) \tag{4.23}$$

Proof. See B.1.3. □

Proposition 4. Let $P(\pi_i|p_i)$ as shown in (4.23) be evaluated at the values of π_i such that $|\pi_i| = k$. If $\varphi(k) = -\log P(\pi_i|p_i)$ is considered a function of π_i, with $|\pi_i| = k$, $k = 0, \ldots, np_i$, then $\varphi(k)$ is a convex function given $p_i > 0$ and $n > 0$.

Proof. See B.1.4. □

From the marginalized local graph prior in (4.23), it follows that $P(\pi_i|p_i) \leqslant \frac{1}{2^{np_i}}, \forall |\pi_i| \leqslant m_p$. For choices of discrete uniform prior on the lag order according to (4.4) means $P(p_i)$ is constant, $\forall p_i \in \{\underline{p}, \ldots, \bar{p}\}$. Following (4.21), we define a modified BIC for local graph and lag selection as

$$BIC(p_i, \vec{G}_{p,i}) = -2 \log P(\mathcal{X}|p_i, \hat{\vec{G}}_{p,i}) + |\hat{\pi}_i| \log T + 2np_i \log 2 \tag{4.24}$$

Following a similar approach proposed by Chib and Greenberg (1995), we use the estimated local graph $\hat{\vec{G}}_{p,i}$ to evaluate the score and to select the lag order $\hat{p}_i$. Selecting the local graph and lag order for each equation may automatically produce asymmetric lags for the different equations. Closely related to our modified BIC is the extended BIC discussed by (see Bogdan et al., 2004; Chen and Chen, 2008; Foygel and Drton, 2011). These papers show that the standard BIC has the tendency to select large size models when dealing with high-dimensional data. A significant difference between our modified BIC and the ones discussed in the above papers is that the additional penalty term depends only on the number of possible links (np_i) for each lag order and independent of the number of estimated links ($|\hat{\pi}_i|$). Thus in comparing graphs with the same lag order but different configuration, our modified criterion coincides with the standard BIC with an additional constant term.

4.4.4 Model Estimation

The posterior of Ω conditional on the lag order $\hat{p}$ and a given graph $\hat{G}_p$ is Wishart distributed (see Geiger and Heckerman, 2002). Since Ω is directly related to B and

Σ_ε, (see Proposition 1), we proceed with the posterior estimation of the model parameters focusing on B and Σ_ε. Thus we estimate $\hat{B}$ and $\hat{\Sigma}_\varepsilon$ from $P(B, \Sigma_\varepsilon|\hat{p}, \hat{G}_p, \mathcal{X})$ with an independent normal-Wishart. By conditioning on $\hat{G}_p$, we estimate the parameters $\{\bar{B}_{G,i}, \bar{V}_{G,i}\}$ that corresponds to the non-zero elements of the $i-$th equation graph $\hat{G}_{p,i}$. We define the selection matrix $E_i = (e_{j_1}, \ldots, e_{j_{|\pi_i|}})$, where E_i is of dimension $np \times |\pi_i|$, $j_k \in \pi_i$ is an element of the set of predictors of the i-th equation, and e_k is the standard orthonormal basis of the set of real np-dimensional vectors. The posterior mean and variance of $\{\bar{B}_{G,i}, \bar{V}_{G,i}\}$ is

$$\bar{B}_{G,i} = \bar{V}_{G,i}(\underline{V}_{G,i}^{-1}\underline{B}_{G,i} + \bar{\sigma}_i^{-2}W'_{G,i}Y_i) \tag{4.25}$$

$$\bar{V}_{G,i} = (\underline{V}_{G,i}^{-1} + \bar{\sigma}_i^{-2}W'_{G,i}W_{G,i})^{-1} \tag{4.26}$$

with

$$W_{G,i} = WE_i, \qquad \underline{B}_{G,i} = \underline{B}_i E_i, \qquad V_{G,i} = E_i'\underline{V}_i E_i \tag{4.27}$$

where $W_{G,i} \in W'$, is the set of selected predictors of the i-th equation; W' is stacked $W'_1, \ldots, W'_{T_0}$, such that W' is of dimension $T_0 \times np$; Y is stacked $Y'_1, \ldots, Y'_{T_0}$, such that Y is of dimension $T_0 \times n_y$; Y_i is the i-th column of Y; $\underline{B}_{G,i}$ and $\underline{V}_{G,i}$, are the prior mean and variance of $W_{G,i}$ respectively; $\bar{\sigma}_i^2, i = 1, \ldots, n_y$, is the variance of residuals from the posterior of Σ_ε, where the posterior of Σ_ε^{-1} is Wishart distributed with scale matrix

$$\bar{S} = \underline{S} + (Y' - \bar{B}W')'(Y - W\bar{B}') \tag{4.28}$$

and degrees of freedom $\bar{\nu} = \underline{\nu} + T_0$, where $\bar{B} = (\bar{B}_{G,1}, \ldots, \bar{B}_{G,n_y})$, the stacked posterior mean of the coefficients, such that $\bar{B}$ is of dimension $n_y \times np$ with positions of non-zero elements corresponding to non-zero elements in $\hat{G}_p$.

4.5 Simulation Study

4.5.1 Metrics for Performance Evaluation

We investigate the effectiveness of our graphical approach with our new prior distribution against one without sparsity restriction together with the LASSO of (Tibshirani, 1996) and the Elastic-net (ENET) of Zou and Hastie (2005). We evaluate the efficiency of the algorithms in terms of the estimated graph, the predictive performance of the estimated models on out-of-sample observations and computational cost in terms of run time.

Given the graph of the data generating process (DGP), we extract from the estimated graph the number of true links correctly predicted as TP; FP as number of true zero edges predicted as positives; TN as number of true zero edges correctly predicted; and FN as number of true links unidentified. We evaluate the graph estimation performance based on the number of predicted positive links ($PP = TP + FP$), the graph accuracy (ACC) and precision (PRC) given as

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \qquad PRC = \frac{TP}{TP + FP} \tag{4.29}$$

Furthermore, we evaluate the graph estimation performance in terms of log-likelihood and BIC scores. Following (4.24), the graph BIC is obtained as

$$BIC_G = \sum_{i=1}^{n_y} BIC(p_i, \vec{G}_{p,i}) = -2L_G + \sum_{i=1}^{n_y} (|\hat{\pi}_i| \log T + 2n\hat{p}_i \log 2) \tag{4.30}$$

where $L_G = \sum_{i=1}^{n_y} L_i$, with L_G is the log-likelihood of the estimated graph and $L_i = \log P(\mathcal{X}|p_i, \vec{G}_p(y_i, \pi_i))$ is the log-likelihood of the local graph of the i-th equation.

We evaluate the model estimation performance based on the out-of-sample joint density and point forecasts. The log-predictive score (LPS) is the most common measure of the joint predictive density discussed in the literature. Since the competing models might have different number of variables and lags across the equations,

the predictive AIC presents a meaningful comparison for purposes of parsimony and is given by

$$AIC_M \;=\; -\,2\log P(Y_{\tau_1}|X_{\tau_0};\hat{B},\hat{\Sigma}_\varepsilon) \;+\; 2|\hat{B}| \tag{4.31}$$

for $\tau_1 = \tau_0+1,\ldots,T$, where τ_0 is the number of observations for the training sample, X_{τ_0} is the training sample dataset; Y_{τ_1} is the out-of-sample observations of the dependent variables; $|\hat{B}|$ is the number of non-zero coefficients in $\hat{B}$; $\hat{\Sigma}_\varepsilon$ is the estimated error covariance matrix; and $\log P(Y_{\tau_1}|X_{\tau_0};\hat{B},\hat{\Sigma}_\varepsilon)$ is the log predictive score.

For point forecast, the mean squared forecast error (MSFE) is the most common measure discussed in the literature. To compare the joint point forecasts, we compute the mean MSFE (MMSFE) following

$$MMSFE = \frac{1}{T-\tau_0-1}\sum_{\tau_1=\tau_0+1}^{T}\Big(\frac{1}{n_y}\sum_{i=1}^{n_y}(Y_{\tau_1}^i-\hat{Y}_{\tau_1}^i)^2\Big) \tag{4.32}$$

where $Y_{\tau_1}^i$ and $\hat{Y}_{\tau_1}^i$ are the out-of-sample observed and predicted values of the i-th dependent variable respectively.

4.5.2 Simulation Study Set-up and Results

The data generating process (DGP) of the simulated study is as follows

$$Y_t = BX_{t-1}+\varepsilon_t, \qquad\qquad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,\Sigma_\varepsilon) \tag{4.33}$$

$t=1,\ldots,T$, where $\Sigma_\varepsilon \sim \mathcal{IW}(b,I_{n_y})$, is a full matrix drawn from an inverse-Wishart distribution with degree of freedom $b=n_y+1$ and an identity scale matrix, I_{n_y}, B is $n_y\times n$ coefficient matrix, Y_t and X_t are is a $n_y\times 1$ and $n\times 1$ respectively. To analyze different sparsity levels, we generate the coefficients matrix B such that, the number of non-zero coefficients for each equation is drawn from a uniform on $\{0,\ldots,40\}$. We considered a large dimensional model by setting $n_y=10, n=100$. We replicate the simulation and estimation exercises 100 times. The 100 replicatons

have been conducted on a cluster multiprocessor system which consists of 4 nodes; each comprises four Xeon E5-4610 v2 2.3GHz CPUs, with 8 cores, 256GB ECC PC3-12800R RAM, Ethernet 10Gbit, 20TB hard disk system with Linux. The simulation study in Table 4.1 takes about 14 minutes of CPU time. For each replication, we generate a sample size, $T = 60$ and use $T_0 = 50$ for model estimation and 10 for out-sample forecast analysis.

We run 20,000 Gibbs iterations for the graph estimation and 2000 iterations for parameter estimations. We applied the standard approach of Tibshirani (1996) and Zou and Hastie (2005) for the LASSO and ENET estimation respectively. We set $\underline{p} = 1$ and $\bar{p} = 4$ and implement a parallel estimation for the LASSO and ENET. For each $p \in [\underline{p}, \bar{p}]$, we sequentially use one variable as the dependent variable and the remaining as the predictors. We apply a five-fold cross validation to select the regularization parameter λ with minimal plus one standard error point (index1SE).

Figure B.2 shows the convergence diagnostics of the graph simulation and the local graph BIC for the lags. The figure of the PSRF indicates convergence of the chain. We also notice from Figure B.2d that the posterior distribution on the lag order for each equation of the simulation experiment using our modified BIC favors lag order $p = 1$. We report in Table 4.1, the performance of the LASSO, ENET, BGVAR and SBGVAR for the inference of the DGP in (4.33).

We proceed by comparing the effectiveness of the algorithms in estimating the graph of the true DGP, when the DGP average links number is 201.5. Table 4.1 shows that without the sparsity restriction, the BGVAR overestimates the number of links compared to the other algorithms. The Lasso-type methods (LASSO and Elastic-Net) fall in the middle with a lower number of links compared to that of the DGP. The SBGVAR on the other hand recorded the least number of edges. This is quite expected since the idea is to select the subset of the explanatory variables that explains a large variation in the dependent variables.

	LASSO	ENET	BGVAR	SBGVAR
DGP Average number of links = 201.5				
PP	108.40	131.30	249.42	69.88
TP	62.04	67.92	**97.76**	48.62
ACC	95.35	95.08	93.61	**95.65**
PRC	58.06	51.85	39.31	**69.58**
L_G	145.28	169.83	**242.35**	160.71
BIC_G	1457.40	1509.23	1837.14	**1307.80**
LPS	-243.07	-304.32	-236.79	**-166.49**
AIC_M	702.94	871.24	972.42	**472.74**
MMSFE	0.67	0.69	0.62	**0.59**
Time (in seconds)	55.76	50.18	162.52	**42.47**

Table 4.1: Average graph and model performance of algorithms over 100 replications. PP - number of predicted positive links; TP - number of true positive links; ACC - graph accuracy; PRC - graph precision; L_G - graph log-likelihood; BIC_G - graph BIC; LPS - log predictive score; AIC_M - predictive AIC; $MMSFE$ - Mean of MSFE. Bold values indicate the best choice for each metric.

By including more edges than the true DGP, the graphical search algorithm without sparsity restriction (BGVAR) records the highest true positive links but relatively low accuracy and precision compared to the other algorithms. Again the Lasso-type methods fall in the middle, recording a lower number of true positive links but with a higher accuracy and precision than the BGVAR. The graphical approach with sparsity restriction instead had the least number of true positive edges but tends to be more accurate and much precise than the other algorithms. The log-likelihood score of the graph favored the BGVAR but the graph BIC score favored the SBGVAR. Thus the BIC score confirms the outcome of the graph accuracy metric which shows that though the SBGVAR records the least edges, it produced a better representation of the temporal dependence in the simulated dataset than the Lasso-type methods and the BGVAR.

The log predictive score, predictive AIC and the MMSFE all favor the SBGVAR over the other competitors. One would expect the Lasso-type methods to perform better than the graphical VAR, however this is not the case according to the above simulation results. This is attributable to the fact that the Lasso-type techniques perform both model selection and parameter estimation simultaneously. This may

seem to be an advantage but on the other hand it affects the estimated parameters, since it shrinks all coefficients at the same rate (see Gefang, 2014). In addition, the Lasso-type methods only focus on estimating the coefficients in each equation neglecting the interaction among the errors across the different equations. The graphical approach instead focus on selecting and estimating only the coefficients of relevant variables taking into consideration the interaction among the errors across the different equations. Thus the latter achieves better parameter estimation efficiency than the Lasso-type models. The result shows that the sparsity restriction on the graph enables us to identify the small set of the most influential explanatory variables that explains a large variation in the dependent variables. Also, the SBGVAR produce a more parsimonious model with better out-of-sample forecasts than the Lasso-type methods.

On the computational intensity, the SBGVAR spends less time than the other algorithms. Interestingly, it records about one-fourth of the run time of the BGVAR. Thus, the sparsity restriction helps to reduce the run time by considering a relatively lower search space in terms of the number of combinations of explanatory variables. The higher run time of the Lasso-type methods is due to the cost of cross-validation to select the regularization parameter.

4.5.3 Sparsity and Indeterminacy Evaluation

A system of linear equations is said to be under-determined when the number of parameters to estimate exceeds the number of observations (see Donoho, 2006). Such systems can be modeled by exploiting sparsity. We investigate the performance of the graphical model approaches against the standard Lasso-type methods for different level of indeterminacy and sparsity of the DGP.

For a VAR model with n_y dependent variables and n explanatory variables for each equation, and a lag order p, we have a total of $n_y np$ number of coefficients

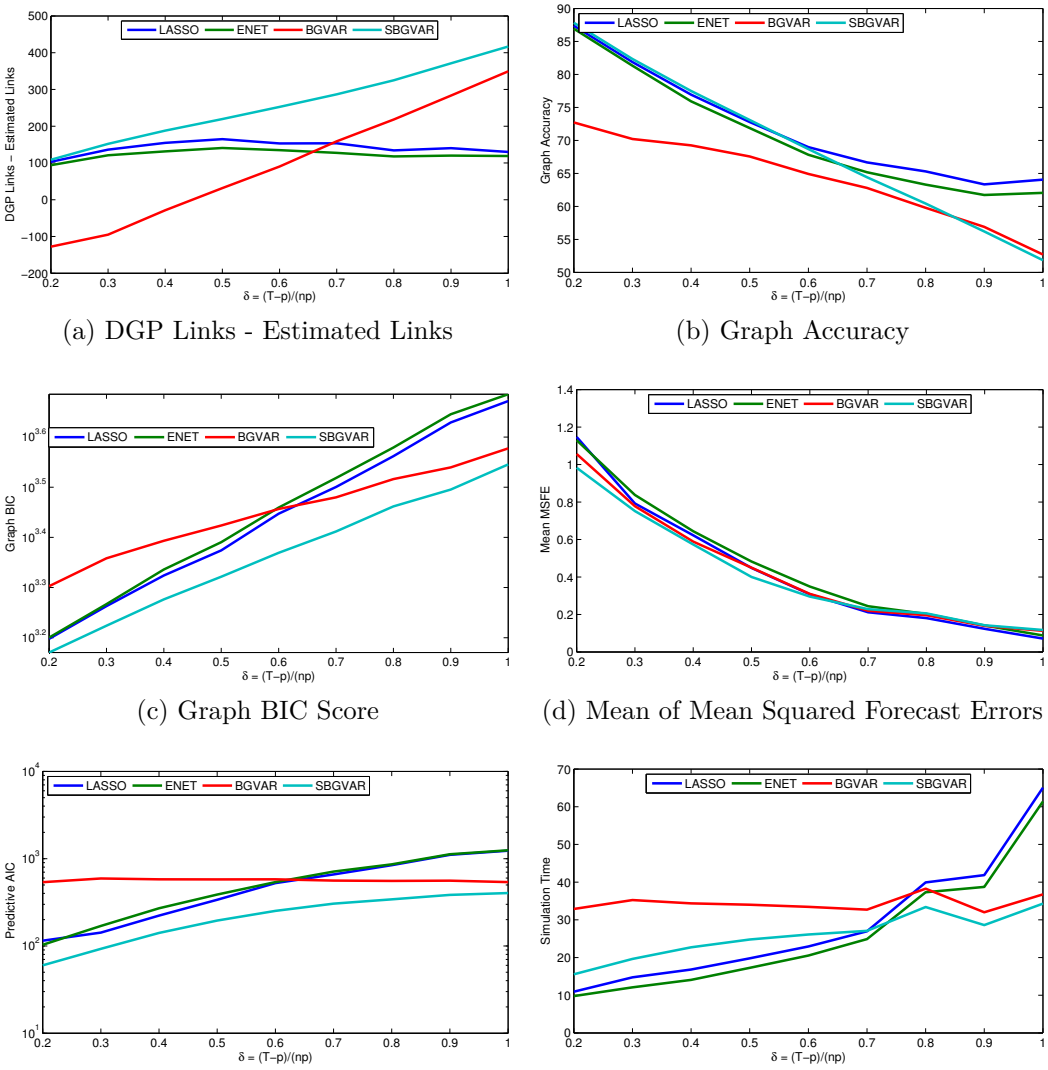
to estimate. Given a multiple time series with T observations, the total number of observations of the dependent variables is given by $(T-p)n_y$. Following Donoho (2006), we measure the level of indeterminacy by $\delta = (T-p)n_y/n_y np = (T-p)/np$, and the level of sparsity by $\rho = kn_y/(T-p)n_y = k/(T-p)$, where k is the number of non-zero coefficients in each equation of the DGP.

Following Donoho and Stodden (2006), we formulate our experiment by setting the DGP to generate a VAR model with $n_y = 10$, $n = 100$ and lag order $p = 1$. For different level of indeterminacy, we set $T-p$ to take values $\{20, \ldots, 100\}$. For each $T-p$, we generate for each equation, $k = \lceil \rho(T-p) \rceil$, where ρ takes values $\{0.2, 0.3, \ldots, 1\}$. This is to allow for different sparsity levels for each level of indeterminacy.

We proceed by comparing the effectiveness of the LASSO, ENET, BGVAR and SBGVAR in estimating the true DGP by setting $\underline{p} = \bar{p} = 1$. For each T and k, we replicate the simulation and estimation exercise 10 times with the magnitude of the coefficients drawn from a uniform on $[-1, 1]$. In each replication, we estimate the model and perform a 1-step ahead forecast. Figure 4.1 shows the estimation performance of the algorithms for the different levels of indeterminacy averaged over the different levels of sparsity.

Figure 4.1a shows the difference between the average DGP number of links and the estimated links of the algorithms. Except for the BGVAR, all the other algorithms estimated a lower number of links compared to that of the DGP. More specifically, the BGVAR seems to overestimate the number of DGP links for lower under-determined models, whereas the SBGVAR underestimates the number of DGP links regardless of the level of indeterminacy. We also see that, the difference between the DGP and the estimated links of the BGVAR and SBGVAR increases overtime while the Lasso-type methods are relatively stable regardless of the level of indeterminacy.

The graph accuracy in Figure 4.1b shows that all the algorithms experienced a



(a) DGP Links - Estimated Links

(b) Graph Accuracy

(c) Graph BIC Score

(d) Mean of Mean Squared Forecast Errors

(e) Predictive AIC

(f) Simulation Time (in seconds)

FIGURE4.1: Estimation performance of the algorithms for different level of indeterminacy averaged over different level of sparsity. The LASSO is in green, ENET in blue, BGVAR in red and SBGVAR in cyan.

deterioration in the accuracy of the prediction of the graph associated with the DGP. However, on average the SGBVAR performs slightly better at the graph estimation for lower under-determined models than the Lasso-type methods.

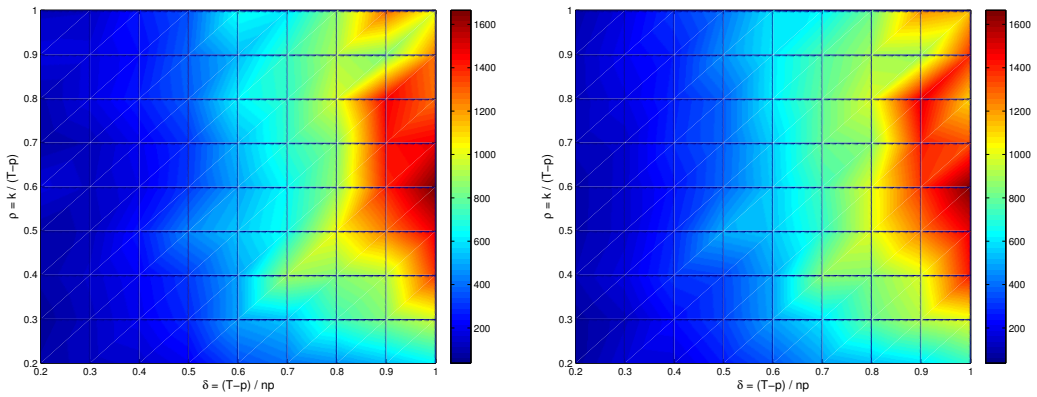
In Figure 4.1c, the graph BIC of the algorithms increases with the level of indeterminacy. This is not surprising since the BIC is a direct function of the number of observations which increases with the level of indeterminacy. Again we observe

that the graph BIC score favors the graph estimated by the SBGVAR over the other competing algorithms. This shows that though the SBGVAR recorded the minimum number of links, it produce a better representation of the graph associated with the DGP.

For model estimation performance, Figure 4.1d shows that all the algorithms perform better at out-of-sample point forecasts for higher under-determined models. The MMSFE of the algorithms are not significantly different though we find that it favors the SBGVAR for lower under-determined models. The predictive AIC (in Figure 4.1e) on the other strongly favors the SBGVAR for all level of indeterminacy.

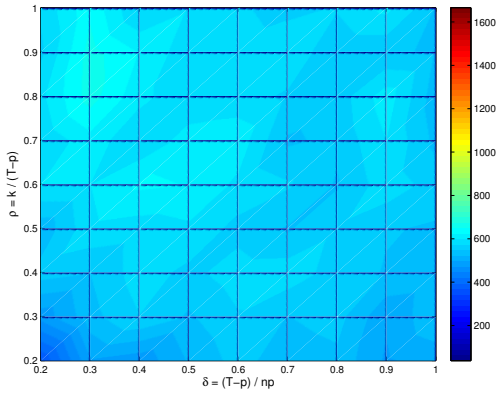
On the computational intensity, we notice (from Figure 4.1f) an increase in run time with the level of indeterminacy for all algorithms except the BGVAR which seems slightly constant over time. Overall, the Lasso-type methods achieve a lower run time for lower under-determined models whiles the SBGVAR achieves lower run time for higher under-determined models.

We focus attention on the model estimation performance of the algorithms for the different levels of indeterminacy and sparsity. Figure 4.2 shows the heatmap of the predictive AIC of the models of the algorithms estimated over the levels of sparsity and indeterminacy of the DGP. The color bar shows the different range of values of the predictive AIC, where blue represents lower AIC, and red for highest AIC. Clearly, we notice a significant difference between the results of the Lasso-type methods and that of the graphical model approaches. Thus the LASSO and ENET are not significantly different from each other, whiles the BGVAR and SBGVAR are quite different, dominated by cyan and blue respectively. The figure shows that the predictive AIC favor the SBGVAR over all levels of sparsity and indeterminacy. The Lasso-type methods only performs better than the BGVAR for lower under-determined models with different level of sparsity, whiles the BGVAR dominates in higher under-determined models.

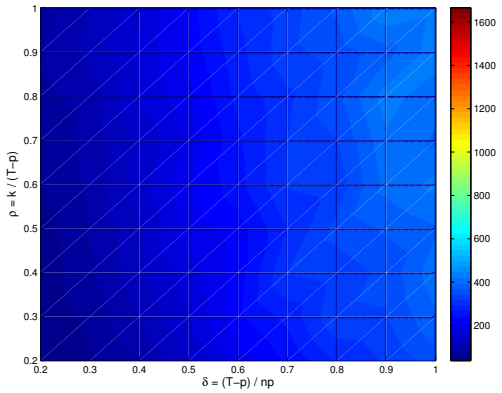


(a) LASSO Predictive AIC

(b) ENET Predictive AIC



(c) BGVAR Predictive AIC



(d) SBGVAR Predictive AIC

FIGURE4.2: Heatmap of the predictive AIC of the models estimated by the four algorithms over the different levels of indeterminacy and sparsity in the data generating process. The result is an average of 10 replication exercises for each δ and ρ. The color bar shows the different range of values of the predictive AIC, where blue represents lower AIC, and red for highest AIC.

The results of this exercise confirm that of our first simulation experiment. Firstly, the sparsity restriction on the graph space induces sparsity on the estimated graph of the temporal relationship among the variables. Secondly, the random fan-in restriction helps to reduce the computational complexity by considering a relatively lower search space in terms of the number of combinations of explanatory variables. Thirdly, though the SBGVAR under-estimates the number of links compared to the DGP and other algorithms, it is able to identify the small set of the most influential explanatory variables that explains a large variation in the dependent variables.

Thus, the SBGVAR produces a more parsimonious model with competitive out-of-sample joint point forecasts and better density forecasts than the competing models.

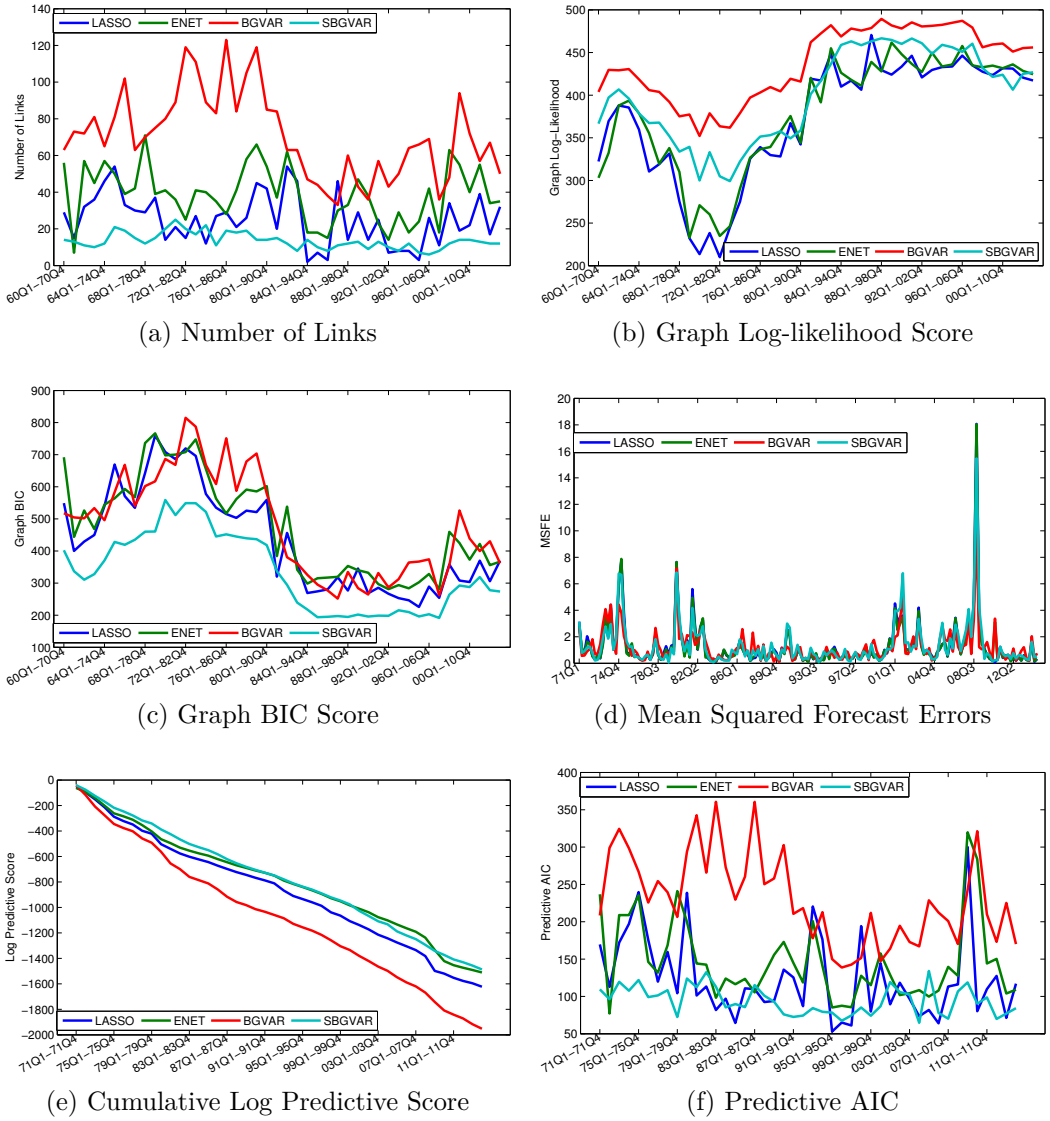
4.6 Forecasting VAR with Many Predictors

Several studies have shown empirically that applying large VAR models for macroeconomic time series produces better forecasts than standard approaches (see Banbura et al., 2010; Koop, 2013; Giannone et al., 2005; Stock and Watson, 2012; Carriero et al., 2013). In the literature, researchers typically work with a single model with fixed or time varying coefficients (see Koop and Korobilis, 2013). It is therefore important to allow for changes in structure and/or parameters to understand the dynamic evolution of the relationship among variables. As part of our contribution, we apply our graphical scheme to model and forecast selected macroeconomic variables with large number of predictors.

The dataset is quarterly observations of 130 US-macroeconomic variables. All series were downloaded from St. Louis' FRED database and cover the quarters from $1959Q1$ to $2014Q3$. Some series had missing observations which are completed with earlier version of the database used by Korobilis (2013). We follow the adjustment codes of De Mol et al. (2008); Stock and Watson (2012) and Korobilis (2013) to transform all the series into stationarity. See B.4 for the list of series and adjustment codes. We consider 6 series as dependent variables and the remaining 124 as predictors. The dependent variables are: consumer price index (CPIAUCSL), Federal funds rate (FEDFUNDS), real gross domestic product (GDPC96), real gross private domestic investment (GPDIC96), industrial production index (INDPRO) and real personal consumption expenditure (PCECC96).

We set the minimum and maximum lag order equal to $\underline{p} = 1$ and $\bar{p} = 4$ respectively according to the literature. We consider a moving window with a starting sample from $1960Q1$ to $1970Q4$ to estimate the model and to forecast 1 to 4-quarters

ahead. We then move the window forward by 4-quarters. Our last sample covers $2003Q1$ to $2013Q4$, and the final forecast is up to $2014Q3$.



(a) Number of Links (b) Graph Log-likelihood Score

(c) Graph BIC Score (d) Mean Squared Forecast Errors

(e) Cumulative Log Predictive Score (f) Predictive AIC

FIGURE4.3: Performance of the algorithms in modeling and forecasting selected macroeconomic variables with many predictors over the sample period $1960Q1 - 2014Q3$. Figures 4.3a - 4.3c show the graph estimation performance, whilst 4.3d - 4.3f depict the model estimation performance.

Figures B.3 in B.2 show the convergence diagnostics of the graph simulation and the local graph BIC for the lags for the macroeconomic application. The figure of the PSRF indicates convergence of the chain. Clearly, the global log score of the graph seems to increase with the lag order in Figure B.3b whereas the total number

of links of the different lags seems quite close as displayed in Figure B.3a. However, we notice from Figure B.3d that the posterior distribution on the lag order for each equation of the macroeconomic application using our modified BIC score favors lag order $p = 1$.

We report in Figure 4.3, the graph and model estimation performance of the Lasso-type methods and the graphical VAR approaches in modeling and forecasting the selected macroeconomic variables over the sample period $1960Q1 - 2014Q3$. The graph estimation performance is compared in terms of the number of link (PP - predicted positive edges), the log-likelihood of the graph (L_G) and the BIC score of the graph (BIC_G). The model estimation performance is compared in terms of the log predictive score (LPS), the predictive AIC (AIC_M) and the average of the mean squared forecast errors (MMSFE). Table 4.2 presents the averages of the graph and model estimation performance of the algorithms including the computational time over the sample period.

From Figure 4.3, we observe that the BGVAR estimated more edges than the other algorithms in a greater part of the sample period. This is followed by the Lasso-type methods, (ENET, then the LASSO) and the SBGVAR records the least number of links over the entire sample period. In scoring the estimated graphs, the BGVAR obtained the highest log likelihood over the entire period whiles the SBGVAR records the minimum at the beginning but showed significant improvement over the rest of the sample period. The BIC score of the graph however favored the SBGVAR over the other algorithms. The summary of the averages in Table 4.2 shows that the by including more edges than the other algorithms the BGVAR records the highest log likelihood of the graph whilst the SBGVAR with the least number of links obtained the minimum BIC score indicating that the SGBVAR graph presents a better representation of the temporal dependence in the macroeconomic application than the Lasso-type methods and the BGVAR.

Figure 4.3d shows the evolution of the out-of-sample joint point forecasts of the models estimated by the algorithms. We observe from the plot that the MSFE are not very different from each other. However, in terms of the out-of-sample joint density forecasts, the BGVAR model presents the minimum cumulative log predictive score and that of the SBGVAR model dominates the LASSO but is very competitive against the ENET model. When adjusted for the number of selected variables used for the forecasting analysis, the SBGVAR model obtain the minimum predictive AIC whilst the BGVAR model performed worst than the Lasso-type models. From Table 4.2, we see that on average the Lasso-type models obtain the minimum MMSFE and this indicate that they produce slightly better point forecasts than the graphical approaches. The average log predictive score and AIC on the other hand are in favor of the SGBVAR over the Lasso-type models.

	LASSO	ENET	BGVAR	SBGVAR
PP	25.14	39.09	70.82	13.50
L_G	372.26	379.45	**437.21**	400.05
BIC_G	434.09	473.71	481.01	**333.46**
LPS	-36.87	-34.33	-44.36	**-33.79**
AIC_M	124.01	146.84	230.36	**94.57**
MMSFE	1.30	1.26	**1.24**	1.32
Time (in seconds)	57.93	42.26	65.74	**23.77**

Table 4.2: Average graph and model estimation performance of algorithms in modeling and forecasting selected macroeconomic series from $1960Q1 - 2014Q3$. PP - number of predicted positive edges; L_G - graph log-likelihood; BIC_G - graph BIC; LPS - log predictive score; AIC_M - predictive AIC; and $MMSFE$ - mean of MSFE. Bold values indicate the best choice for each metric.

On the computational intensity, the result shows that on average the SBGVAR spends less simulation time on graph sampling and parameter estimation than the other algorithms. Interestingly, it records about one-fourth of the run time of the BGVAR.

In summary, we find evidence that the graphical VAR approach with our new graph prior distribution induces sparsity on the graph structure. In modeling and

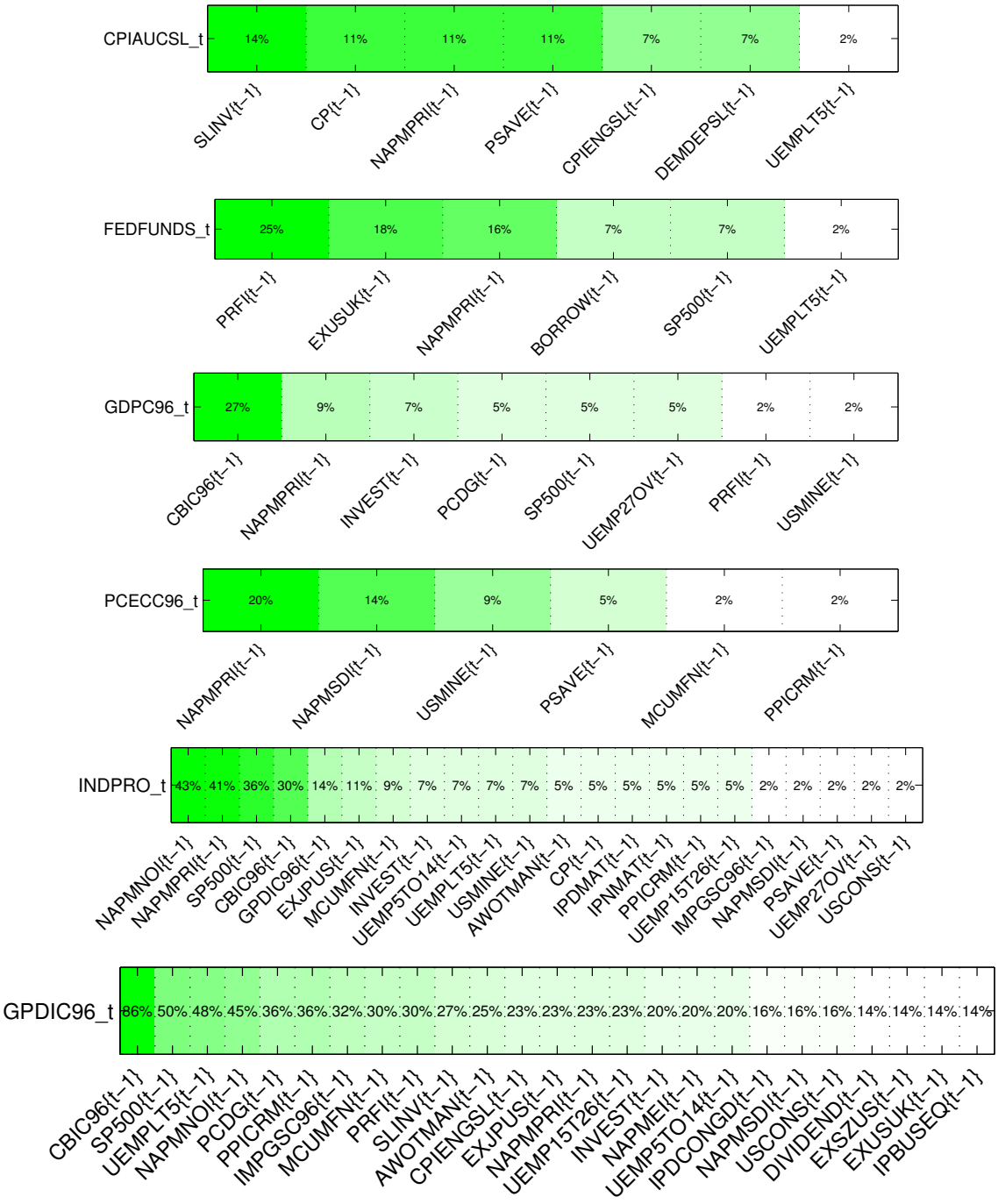


FIGURE4.4: Frequency of inclusion of the most influential variables that explain a large variation in the dependent variables of the macroeconomic application averaged over the sample period $1960Q1-2014Q3$. CPIAUCSL is consumer price index, FEDFUNDS - Federal funds rate, GDPC96 - real gross domestic product, GPDIC96 - real gross private domestic investment, INDPRO - industrial production index and PCECC96 - real personal consumption expenditure.

forecasting our selected macroeconomic series, the result shows that the SBGVAR better represents the temporal dependence, since it is more parsimonious than the competitors. Furthermore, we find evidence of a gain in the predictive performance of the SBGVAR approach over the Lasso-type methods. It is also less computationally intensive compared to the graphical approach without sparsity restriction and the Lasso-type methods.

In Figure 4.4, we report the frequency of inclusion of the relevant predictors for the variables of interest in our macroeconomic application, averaged over the sample period $1960Q1 - 2014Q3$. For convenience, we report only the top explanatory variables of real investment (GPDIC96) with frequency up to 14%. We notice from the figure that all the variables of interest are not explained by the same set of predictors. Thus, apart from real investment and industrial production index, which have a higher number of predictors over the sample period, the rest can be predicted by a handful of macroeconomic variables.

We find evidence supporting the effect of financial variables on the real sector of the US economy. More specifically, we find that over the sample period, S&P 500 and exchange rates with Japan, Switzerland and UK, have strong effects on real investment (GPDIC96) and industrial production index (INDPRO), and weak effects on real gross domestic product (GDPC96) and on the Federal funds rate (FEDFUNDS). The results are in line with Diebold and Yilmaz (2015) suggesting the importance of monitoring the connectedness between real activity and stock returns (or financial variables). Furthermore, it offers some insight for further evaluation of macro-financial linkages which have long been at the core of the IMF's mandate to oversee the stability of the global financial system.

4.7 Conclusion

In recent years, there has been increasingly available rich datasets with large number of variables relative to the length of observations. Standard estimation techniques often encounter problems since there are too many parameters to estimate. This paper contributes to solve this problem by developing a Bayesian graphical approach to model dependence and to address over-parametrization in high-dimensional multivariate time series models. The novelty of the approach of this paper is proposing a new (sparsity) prior limiting the number of predictors for each equation of a large vector autoregressive (VAR) model and developing a new Markov chain Monte Carlo (MCMC) algorithm to determine the relevant links and lag order for each equation, and the parameter estimates.

We show through simulation and real data macroeconomic application that our proposed method is a valid alternative to the standard Lasso-type methods - LASSO and Elastic-Net (ENET). In both applications, we find evidence that our new prior distribution induces sparsity on the graph of temporal dependence among variables. The comparison with the standard LASSO and ENET as benchmarks show that our model perform well in identifying the relevant predictors and forecasting the variables of interest. Furthermore, the simulation exercise shows that our model performs comparatively better when indeterminacy is high. On the macroeconomic application, we find evidence that financial variables helps in predicting economic activity. Thus, our proposed method can be considered for empirical evaluation of macro-financial linkages.

Acknowledgments

Earlier versions of this work have been presented at the 34th International Symposium on Forecasting (ISF 2014), Rotterdam; Workshop on Networks in Economics

and Finance, Louvain-la-Neuve, 2014; the 8th International Conference on Computational and Financial Econometrics, (CFE 2014), Pisa; the European Winter Meetings of the Econometric Society, (EWM 2014), Madrid; the 6th Italian Congress of Econometrics and Empirical Economics (ICEEE 2015), Salerno; and the BOMOPAV 2015 Economics Meeting, Modena. We are grateful for useful comments from the participants in these conferences. This research used the SCSCF multiprocessor cluster system at University Ca' Foscari of Venice. We would like to thank Kamil Yilmaz, Mario Forni, Lorenzo Frattarolo and Fulvio Corsi for their comments. Author's 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.

Chapter 5

Bayesian Selection of Systemic Risk Networks

5.1 Introduction

In the latest financial crisis, started in 2007, the core capital of banks has proved to be insufficient to cover impairment losses arising from loans and security portfolios. Consequently, several banks have strengthened their capital base or reduced their asset exposure. Other banks have been bailed out by state aids or have defaulted. To reduce the risk of similar crises in the future and to enhance the resilience of the banking sector, a new regulatory framework, the so-called Basel III package, has been proposed, implying more stringent capital requirements for financial institutions (Basel Committee on Banking Supervision, 2011). The effectiveness of the new regulatory framework to prevent banking default and financial crisis is an open problem, particularly as regulations themselves are still in progress, and may thus benefit from the results of research findings in the field.

The adoption of a robust financial network approach for systemic risk is recommended not only because of its proper emphasis on financial interdependencies,

This chapter is based on: Ahelegbey, D. F. and Giudici, P. (2014). Bayesian Selection of Systemic Risk Networks, *Advances in Econometrics: Bayesian Model Comparison*, vol. 34, 117–153.

but also for its possibility to describe how the structure of these interdependencies evolves in time. If this is achieved, we would be able to address systemic financial risks in two directions: on one hand, to understand the role that a network structure plays in the spread of financial shocks; on the other hand, to understand the impact of simulated stress events on a network of interdependencies .

To learn financial networks from the data we propose empowering network models with multivariate graphical models. Graphical models embed the idea that interactions among random variables in a system can be represented in the form of graphs, whose nodes represents the variables and whose edges shows their interactions. For an introduction to graphical models see, for example, Pearl (1988); Lauritzen and Wermuth (1989); Whittaker (1990); Wermuth and Lauritzen (1990); Edwards (1990); Lauritzen (1996).

Graphical models can be employed to accurately estimate the adjacency matrix, aimed at measuring interconnectedness between different financial institutions and, in particular, to assess central ones that may be the most contagious or the strongest source of contagion (as in Billio et al., 2012). Network models use the correlation matrix estimated from the data to derive the adjacency matrix. Although useful, this approach takes into account only the marginal (indirect) effect of a variable on another, without looking at the (direct) effect of other variables. In our context, it does not distinguish between the direct and the indirect effect of a country on other ones. Graphical models, instead, focus on the partial correlation matrix, that is obtained by measuring only the direct correlation between two variables. A partial correlation coefficient can express the change in the expected value of a dependent variable, caused by a unitary change of the independent variables, when the remaining variables are held constant. In so doing, the effect of a bank on another is split into a direct effect (estimated by the partial correlation) and an indirect effect (what is left in the marginal correlation). Here we follow this approach and derive the adja-

cency matrix, the main input of a financial network model, not from the correlation matrix but, rather, from the partial correlation matrix obtained from the application of graphical models to the available data.

To achieve this aim we consider multivariate Gaussian graphical models, defined in terms of Markov properties, that is, conditional independencies associated with the underlying graph (Lauritzen, 1996; Whittaker, 1990). While traditional network models assume fixed graphical structures (such as fully connected graphs), the structure of a graphical model is typically random, and can be learned from the data, as a good fitting structure. Such a model selection can be performed by testing, in a stepwise procedure, the statistical significance of conditional independencies, which are equivalent to specified zeroes among certain partial correlation coefficients which, in turn, are equivalent to missing edges in the network representation.

The use of graphical models can thus help to have a deeper understanding of the relationships between variables, by distinguishing direct from indirect relationships. From their appearance in the 90's, several methodological advances have been made for graphical models. Less so in terms of applications, especially in financial economics. In our opinion, this requires to solve two main problems.

First, the assumption of a random graph may be an important added value in situations where little a priori knowledge is present, as is the case for systemic risk. In addition, results should not be conditioned on single models, but, rather, should be model averaged, to avoid suboptimal inferences. Second, graphical models do not allow to decompose asset return correlations into market specific and idiosyncratic effects, as in the classical CAPM models (Sharpe, 1964). This assumption seems to be restrictive in finance.

The first problem can be solved with the use of more advanced, Bayesian, graphical models, as shown in Madigan and York (1995), Giudici and Green (1999) and, more recently, Ahelegbey et al. (2015). In particular, Madigan and York (1995) and

Giudici and Green (1999) propose a Bayesian model able to consider all possible graphical structures, choose the best fitting ones and, if necessary, average inferential results over the set of all models. The methodological contribution of this paper can solve the second problem as well. We propose Bayesian hierarchical graphical models, that allow correlations to be decomposed, as in a CAPM-like model, into a country (market) effect plus a bank-specific (idiosyncratic) effect.

The applied contribution of this paper is in the understanding of whether and how a distress probability is transmitted between different banks, that belong to different countries, with different regulatory systems. The world financial market is not perfect: many frictions exists between different countries, mainly due to different regulations, given the fundamental relevance of banks for the economies to which they belong. A very interesting case study, in this respect, is the Eurozone, where the European Central Bank (ECB) has recently taken over the supervision of the largest banks (with total assets greater than 30bn euro) in each of the member states. Thus, eventually, the euro banking market will evolve into a single market but, at the time being, it is still fragmented. It thus becomes timely and rather interesting to study the degree of convergence towards a European banking union, looking at the comovements between stock returns of the banks in the area, that may give important insights.

Finally, we remark that the paper also contributes to computational statistics as, in order to apply the proposed hierarchical model to the large p/small n database at hand, a novel Markov Chain Monte Carlo algortihm, based on Bayes factor thresholding, has been developed and implemented in a Matlab routine, available upon request.

The paper is organized as follows. In Section 5.2 we review the relevant literature, both from a methodological and an applied viewpoint. In Section 5.3, we introduce Bayesian graphical models methodology, and show their theoretical implications. In

Section 5.4 we introduce a simulated data application that is helpful for the set-up of our computational methodology. Section 5.5 describes the empirical results obtained with the application of the Bayesian hierarchical graphical models to analyze European Banks Risk Network.

5.2 Literature Review

The study of bank failures is important for two reasons. First, an understanding of the factors related to bank failure enables regulatory authorities to supervise banks more efficiently. In other words, if supervisors can detect problems early enough, regulatory actions can be taken, to prevent a bank from failing and, therefore, to reduce the costs of its bail-in, faced by shareholders, bondholders and depositors; as well as those of its bail-out, faced by the governments and, therefore, by the taxpayers. Second, the failure of a bank very likely induces failures of other banks or of parts of the financial system as a whole. Understanding the determinants of a single bank failure may thus help to understand the determinants of financial systemic risks, which could be due to microeconomic, idiosyncratic factors or to macroeconomic imbalances. When problems are detected, their causes can be removed or isolated, to limit "contagion effects".

The literature on predictive models for single bank failures is relatively recent: until the 1990s, most authors emphasize the absence of default risk of a bank (see e.g. Gup, 1998; Roth, 1994), in the presence of a generalized expectation of state interventions. However, in the last years we have witnessed the emergence of financial crisis in different areas of the world, and a correlated emphasis on systemic financial risks. In addition, government themselves are less willing than before to save banks, partly for their financial shortages and partly for a growing negative sentiment from the public opinion. As a consequence of all of these, recent research trends have seen a growing body of literature on bank failures, and systemic risks originating from

the above mentioned sentiments and actions.

The studies on bank failures can be classified into three main streams: financial market models, scoring models and macroeconomic models.

Financial market models originate from the seminal paper of Merton (Merton, 1974), in which the market value of a bank's asset, typically modeled as a diffusion process, is insufficient to meet its liabilities. Due to its practical limitations, Merton's model has been evolved into a reduced form (see e.g. Vasicek, 1984), leading to a widespread diffusion of the resulting model, and the related implementation in Basel II credit portfolio model. For a review of this evolution see, for example, the book by Resti and Sironi (2007). In order to implement market models, diffusion process parameters and, therefore, bank default probabilities can be obtained on the basis of share price data that can be collected almost in real time from financial markets. Market data are public, relatively easy to collect and are quite objective. On the other hand, they may not reflect the true fundamentals of the underlying financial institutions, and may lead to a biased estimation of the probability of failure. Indeed, the recent paper by Idier et al. (2013) and Fantazzini and Maggi (2012) show that market models may be good in very short-term predictions, but not in medium and long-term ones, where the importance of fundamental financial data emerge.

Scoring models are based only on financial fundamentals, taken from the publicly available balance sheet information. Their diffusion followed the seminal paper by Altman (Altman, 1968), which has induced the production of scoring models for banks themselves: notable examples are Sinkey (1975), Tam and Kiang (1992), Rose and Kolari (1985), Cole and Gunther (1998). The development of the Basel regulation and the recent financial crisis have further boosted the literature on scoring models for banking failure predictions. Recent examples include Arena (2005), Davis and Karim (2008) who use logit models; Vazquez and Federico (2012) who use a probit model and Klomp and Haan (2012) who use a principal component factor approach.

Systemic risk can indeed also be defined as the risk that the failure of one significant financial institution can cause or significantly contribute to the failure of other significant financial institutions, as a result of their linkages to each other (see e.g. Billio et al., 2012). Trying to address this aspect of systemic risks, researchers have recently proposed network models, that can help model the systemic risk in financial systems which display complex degrees of connectedness. In particular, Billio et al. (2012) propose several econometric measures of connectedness based on principal component analysis and Granger-causality networks. They find that hedge funds, banks and insurance companies have become highly interrelated over the past decade, likely increasing the level of systemic risk through a complex and time-varying network of relationships. Chen et al. (2013a) and Barigozzi and Brownlees (2014) follow similar approaches.

The emergence of systemic risks has also directed the attention on macroeconomic models to predict bank failures, especially for those countries whose economies are heavily dependent on banks. As in Merton's reduced form model, the main intuition behind these models is to decompose failure risk into an idiosyncratic component, that can be studied using microeconomic data, and a systematic component. Scoring models have been extended in different ways: interesting developments include the incorporation of macroeconomic components (see e.g., Koopman et al., 2011; Mare, 2012; Kanno, 2013; Kenny et al., 2013) ; and the explicit consideration of the credit portfolio, as in the symbol model of De Lisa et al. (2011) - that allows stress tests of banking asset quality and capital, as emphasized in the recent paper by Halaj (2013). The problem with scoring models is that they are mostly based on balance sheet data, which are different from the market and has a low frequency of update (annual or, at best, quarterly) and depends on subjective strategic choices. They may thus be good to predict defaults (especially in the medium term) but not in the assessment of systemic risks, which occur very dynamically and with short notice.

Our focus here will not be on the prediction of single defaults but rather, on how such prediction are correlated with each other, in a systemic perspective. The research literature on systemic risk is very recent, and follows closely the developments of the recent financial crisis. A comprehensive review is discussed in Brunnermeier and Oehmke (2012) who also provide a historical comparison of different crisis. Specific measures of systemic risk have been proposed, in particular, by (Adrian and Brunnermeier, 2010; Acharya et al., 2010; Brownlees and Engle, 2011; Huang et al., 2012; Billio et al., 2012; Segoviano and Goodhart, 2009). All of these approaches are built on financial market price information, on the basis of which they lead to the estimation of appropriate quantiles of the estimated loss probability distribution of a financial institution, conditional on a crash event on the financial market. These literature developments have led to and are still contributing to the identification of the Systemically Important Financial Institutions (SIFIs), at the global and regional level. They however do not address the issue of how risks are transmitted between different institutions.

In this paper we aim to provide a model for the estimation of contagion between banks supported by the data using hierarchical graphical models.

5.3 Hierarchical Bayesian Graphical Models

In this section we review the basics of Bayesian graphical models and, then, describe the methodology and the computational approach we propose.

From a statistical viewpoint, while correlations can be estimated, on the basis of the N observed time series of data, assuming that, at each time point, observations follow a multivariate Gaussian model, with unknown variance-covariance matrix Σ, partial correlations can be estimated assuming that the same observations follow a graphical Gaussian model, in which Σ is constrained by the conditional independence described by a graph.

Let $X = (X_1, ..., X_N) \in R^N$ be a random vector distributed according to a multivariate normal distribution $\mathcal{N}_N(\mu, \Sigma)$. In this paper, without loss of generality, we will assume that the data are generated by a stationary process, and, therefore, $\mu = 0$. In addition, we will assume throughout that the covariance matrix Σ is non singular.

Let $G = (V, E)$ be an undirected graph, with vertex set $V = \{1, ..., N\}$, and edge set $E = V \times V$, a binary matrix, with elements e_{ij}, that describes whether pairs of vertices are (symmetrically) linked between each other ($e_{ij} = 1$), or not ($e_{ij} = 0$). If the vertices V of this graph are put in correspondence with the random variables $X_1, ..., X_N$, the edge set E induces conditional independence on X via the so-called Markov properties (see e.g. Lauritzen (1996), Whittaker (1990) or, from an econometric viewpoint, Carvalho and West (2007) and Corander and Villani (2006)). More precisely, the pairwise Markov property determined by G states that, for all $1 \leqslant i < j \leqslant N$,

$$e_{ij} = 0 \Longleftrightarrow X_i \perp X_j | X_{V \setminus \{i,j\}}$$

that is, the absence of an edge between vertices i and j is equivalent to independence between the random variables X_i and X_j, conditionally on all other variables $X_{V \setminus \{i,j\}}$. In our context, all random variables are continuous and it is assumed that $X \sim \mathcal{N}_N(0, \Sigma)$. Let the elements of Σ^{-1}, the inverse of the variance-covariance matrix, be indicated as $\{\sigma^{ij}\}$.It can be shown (see e.g. Whittaker (1990)) that the following equivalence also holds:

$$X_i \perp X_j | X_{V \setminus \{i,j\}} \Longleftrightarrow \rho_{ijV} = 0, \qquad \rho_{ijV} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}} \tag{5.1}$$

where ρ_{ijV} denotes the ij-th partial correlation, that is, the correlation between X_i and X_j conditionally on the remaining variables $X_{V \setminus \{i,j\}}$.

Therefore, by means of the pairwise Markov property, given an undirected graph $G = (V, E)$, a graphical Gaussian model can be defined as the family of all N-variate

normal distributions $\mathcal{N}_N(0, \Sigma)$ that satisfy the constraints induced by the graph on the partial correlations, as follows:

$$e_{ij} = 0 \Longleftrightarrow \rho_{ijV} = 0, \qquad \forall 1 \leqslant i < j \leqslant N \tag{5.2}$$

Graphical model uncertainty can be taken into account, along with parameter uncertainty, within a Bayesian approach, whose main practical advantage is that inferences on quantities of interest can be averaged over different models, each of which has a weight that corresponds to its Bayesian posterior probability. See, for example, Madigan and York (1995), Giudici and Green (1999) and Giudici and Castelo (2003).

To achieve this aim, the first task is to recall the expression of the marginal likelihood of a graphical Gaussian model, and specify prior distributions over the parameter Σ as well as on the graphical structures G. For a given graph G, consider a sample x of size n from $P = N_N(0, \Sigma)$, and let S_n be the corresponding observed variance-covariance matrix. For a subset of vertices $A \subset N$, let Σ_A denote the variance-covariance matrix of the variables in X_A, and define Σ_A the corresponding observed matrix. When the graph G is decomposable the likelihood of a graphical gausssian model specified by P nicely decomposes as follows (see e.g. Dawid and Lauritzen (1993)):

$$p(x|\Sigma, G) = \frac{\prod_{C \in \mathcal{C}} p(x_C|\Sigma_C)}{\prod_{S \in \mathcal{S}} p(x_S|\Sigma_S)} \tag{5.3}$$

where $\mathcal{C}$ and $\mathcal{S}$ respectively denote the set of cliques and separators of the graph G, and:

$$P(x_C|\Sigma_C) = (2\pi)^{-\frac{n|C|}{2}} |\Sigma_C|^{-\frac{n}{2}} \exp\left[-\frac{1}{2} tr\{S_C \Sigma_C^{-1}\} \right] \tag{5.4}$$

and similarly for $P(x_S|\Sigma_S)$.

Dawid and Lauritzen (1993) propose a convenient prior for the parameters of the above likelihood, which is named hyper Wishart. It can be obtained from a collection

of clique specific marginal Wishart as follows:

$$\frac{\prod_{C\in\mathcal{C}} l(\Sigma_C)}{\prod_{S\in\mathcal{S}} l(\Sigma_S)} \tag{5.5}$$

where $l(\Sigma_C)$ is the density of a Wishart distribution, with hyper-parameters T_C and α, and similarly for $l(\Sigma_S)$. For the definition of the hyper-parameters, we follow Giudici and Green (1999) and let T_{0C} and T_{0S} be the sub-matrices of a large matrix T_0 of dimension $N \times N$, and choose $\alpha > N$. To complete prior specification, for $P(G)$ we assume a uniform prior over all possible graphical structures.

Dawid and Lauritzen (1993) show that, under the previous assumptions, the posterior distribution of the variance-covariance matrix Σ, is a hyper-Wishart distribution with $\alpha + N$ degrees of freedom and a scale matrix given by:

$$T_n = T_0 + S_n \tag{5.6}$$

where S_n is the sample variance-covariance matrix.

The proposed prior distributions can also be used to integrate the likelihood with respect to the unknown random parameters, obtaining the so-called marginal likelihood of a graph, which will be the main metric for model selection. This follows from

$$P(X|G) = P(X|\Sigma, G)P(\Sigma) \tag{5.7}$$

Giudici and Green (1999) show that such marginal likelihood is equal to:

$$p(x|G) = \frac{\prod_{C\in\mathcal{C}} p(x_C)}{\prod_{S\in\mathcal{S}} p(x_S)} \tag{5.8}$$

where

$$P(x_C) = (\pi)^{-\frac{n|C|}{2}} \frac{k(|C|, \alpha + n)}{k(|C|, \alpha)} \frac{det(T_{0C})^{\alpha/2}}{\det(T_{nC})^{(\alpha+n)/2}} \tag{5.9}$$

where $k(\cdot)$ is a normalization constant, given by: $k(x, y) = \prod_{j=1}^{x} \Gamma(\frac{y+1-j}{2})$, T_{0C} and T_{nC} are sub-matrices corresponding to x_C of a large matrix T_0 and T_n. The

metric expressed by the above marginal likelihood is the basic ingredient for Gaussian graphical model selection and averaging, as will now be shown.

According to the conventional Bayesian paradigm, being the model space discrete, the best graphical model will be that with the highest a posteriori probability. By Bayes rule, the posterior probability of a graph is given by:

$$P(G|X) \propto P(X|G)P(G) \tag{5.10}$$

and, therefore, since we assumed a uniform prior over the graph structures, maximizing the posterior probability is equivalent to maximizing the marginal likelihood metric. For graphical model selection purposes we shall thus search in the space of all possible graphs for the structure such that

$$G^* = \arg\max_G P(G|X) \propto P(X|G) \tag{5.11}$$

The Bayesian paradigm does not force conditioning inferences on the (best) model chosen. The assumption of G being random, with a prior distribution on it, allows any inference on quantitative parameters to be model averaged with respect to all possible graphical structures, with weights that correspond to the posterior probabilities of each graph. This because, by Bayes' theorem:

$$P(\Sigma|X) = P(\Sigma|X, G)P(G|X) \tag{5.12}$$

The above allows to overcome the main drawback of non-Bayesian graphical models, namely, the fact that, once a model is chosen, all inferences will be conditional on that model, even if it has a little support from the data (although maximal).

However, in real situations, the number of possible graphical structures may be very large and we may need to restrict the number of models to be averaged. This can be done efficiently, for example, following a simulation-based procedure for model search, such as Markov Chain Monte Carlo (MCMC) sampling, described in Madigan and York (1995). One of the standard MCMC methods is the Metropolis-Hastings (MH) algorithm, which is based on an acceptance-rejection scheme. In our

context, given an initial graph, the algorithm samples a new graph using a proposal distribution.

We remark that, to guarantee irreducibility of the Markov chain, we follow Giudici and Castelo (2003) and test whether the proposal is a decomposable graph. Following Giudici and Green (1999) and Frydenberg and Lauritzen (1989) we apply the concept of junction tree (see Lauritzen (1996) or Cowell et al. (1999)). More specifically, the addition of an edge between two nodes is allowed if the two nodes belong to two different connected components or if the cliques they belong to are connected by a separator. On the other hand, the removal of an edge is allowed if such edge belongs to a single clique. After each iteration the junction forest is updated according to the new organization of the cliques. The newly sampled graph is compared with the old graph, with a decision rule to either reject or accept the proposed sample. The proposed graph G_{new} is either accepted, or rejected in which case the previous graph is maintained as G_{old}. The decision to accept or reject a proposed graph depends on an acceptance probability. By assuming a uniform graph prior, the log acceptance probability is given by:

$$\log(Ac) = \max\{\log(P(X|G_{new})) - \log(P(X|G_{old})), 0\} \tag{5.13}$$

5.3.1 Hierarchical Graphical Models

One of the appealing features of graphical models for multivariate time series analysis is to represent graphically the logical implications as well as the conditional independence relationships among the variables. In high dimensional settings, it is extremely hard to extract meaningful information from the complex interaction among the variables. In addition, learning such complex interactions is computationally intensive when using standard structure learning schemes. There is a high chance that the learning algorithm gets trapped, spending much time to learn local optimum structures which might not be representative A possible solution to the

trap problem is to add more structure to the graphical model. For example, in the Bayesian setting, the prior distribution over the graph space can be enriched and used to penalize complex structures. Another approach, particularly suited for the applied problem considered in this paper, is to employ hierarchical graphical models: this will allow structural learning to be localized to groups of variables which however are not independent and, thus, can borrow inferential strength from each other.

Several researchers have discussed and applied hierarchies in graphical models (see e.g. Hensman et al., 2013; Gyftodimos and Flach, 2002; Zhang et al., 2005). The approach proposed in this paper is in the spirit of Guo et al. (2011, 2013) but from a Bayesian perspective. Suppose we have a large dataset of n random variables, $X = (X_1, \ldots, X_n)$, that can be grouped into k categories, $(Z_1, \ldots, Z_k)$, such that the kth category contains nk variables $Z_k = (Y_1, \ldots, Y_{nk})$. Without loss of generality, we assume that each variable in X belongs to exactly one of the categories, i.e $\sum_{i=1}^{k} n_i = n$, and that the observations in the same category are centered along each variable. To exemplify, the random variables in X can be banks that can be grouped into k different countries (represented as Z) with each country consisting of a number of banks (represented as Y), so that $X = (Y, Z)$.

In this context, we consider Z as a compressed representation of the banking system in different countries. Based on the classification, we build a two-level hierarchical model composed of: a country specific component, that explains relationships between banking systems of different countries, and an idiosyncratic component that models relationships between banks in the same country. More formally

$$P(X) = P(Y, Z) = P(Z)P(Y|Z)$$

To achieve this aim, we introduce a one-way decomposition of the covariance between any two banks, X_i, $X_j \in X$, that belong, respectively, to countries Z_a and Z_b. Let G_z be a $k \times k$, $0\backslash 1$ matrix, representing the structure of between country relationships.

In our context, all random variables are continuous and it is assumed that $X \sim \mathcal{N}_n(0, \Sigma)$, $Z \sim \mathcal{N}_k(0, \Phi)$ and $Y \sim \mathcal{N}_n(0, \Psi)$. We assume that:

$$\sigma_{i,j} = g^z_{a,b} \circ (\phi_{a,b} \circ \psi_{i,j}) + (1 - g^z_{a,b}) \circ \psi_{i,j}, \qquad X_i \in Z_a, X_j \in Z_b$$

$$\sigma_{i,j} = (\phi_{a,a} \circ \psi_{i,j}), \qquad X_i, X_j \in Z_a, X_i \neq X_j$$

where $(\circ)$ is the Hadamard product, $g^z_{a,b} = G_z(a, b)$ captures the link between country Z_a and Z_b, $\phi_{a,b}$ is the covariance between Z_a and Z_b, and $\psi_{i,j}$ is the covariance between banks Y_i and Y_j. This decomposition treats $\phi_{a,b}$ as a common factor controlling any linkage between two banks Y_i and Y_j from different countries. On the other hand, $\psi_{i,j}$ measures the covariance between banks Y_i and Y_j, independent of a country specific effect.

A nice aspect of the above model is that it can be further decomposed, to accommodate for further grouping effects. For instance, in our specific application, the term $\phi_{a,b}$ may be further be decomposed into a macroeconomic country effect plus a banking sector specific ones, calculating the difference between macroeconomic log returns (e.g. of the Gross Domestic Product of a country) and capitalization weighted log returns of the banking sector of a country. In addition, the term ψ_{y_i,y_j} can also be similarly decomposed in groups, according to bank-specific balance sheet variables (such as asset size, leverage, etc.).

5.3.2 Efficient Structural Inference Scheme

We now focus on sampling the multivariate instantaneous relationships among random variables by allowing for simultaneous interactions. We will do so by extending the work by Ahelegbey et al. (2015). In large dimensional settings, a common drawback of the classical Metropolis-Hastings sampling scheme is the likelihood of spending much time to learn local optimum structures which might not be representative of the global optimum structure. Secondly, sampling structures with simultaneous interactions (undirected links) leads to difficulties in diagnosing convergence of the

chain. According to Madigan and York (1995), irreducibility is guaranteed on condition that the structures satisfy acyclic constraints. This makes the standard scheme not feasible in sampling undirected structures. In this paper, we propose a Markov chain Monte Carlo scheme that employs a Bayes factor thresholding.

A sketch of the idea is as follows: We initialize the algorithm by assuming a fully connected structure (G_{old}). At each iteration, we delete the edge between two variables X_i and X_j to produce a new structure (G_{new}). We then compute the posterior of the two structures and compute the Bayes factor. By assuming a uniform graph prior, it can be shown that the log Bayes factor is given by:

$$\log(BF) = \log(P(\mathcal{X}|G_{new})) - \log(P(\mathcal{X}|G_{old}))$$

If $\log(BF) > \tau$, where $\tau \geqslant 0$, then the model of the new structure is much preferred which means that the edge between X_i and X_j must be deleted. However, if $\log(BF) \leqslant \tau$, then the edge between X_i and X_j can be retained. Thus the mechanism automatically accepts edges between variables leading to improvements in structural learning.

To sample our two-level hierarchical model, we marginalize out analytically the parameters of the structural model to obtain an efficient Gibbs sampling algorithm (e.g., see Casella and Robert, 2004). That is, we sample the parameters and the relationships in blocks (e.g., see Roberts and Sahu, 1997). Let G_z and G_y be the structure of the between country and between bank relationships. Thus, our approach involves sampling from the posterior distribution of the between country structural relationships (G_z) given Ψ, then updating Ψ given (G_z). The next step involves sampling from the posterior distribution of the between banks structural relationships (G_y) given Ψ, Φ and G_z. Let $\mathcal{X} = (\mathcal{Z}, \mathcal{Y})$ where $\mathcal{X}$ is the observed dataset of the variables arranged into $\mathcal{Z}$ and $\mathcal{Y}$ representing the country specific and bank specific dataset respectively. The resulting collapsed Gibbs sampler (Liu, 1994)

consists of the following steps:

1. Sample between country structural relationships: $(G_z|\mathcal{Z})$,
2. Update between country parameters: $(\Phi|\mathcal{Z}, G_z)$,
3. Sample between banks structural relationships: $(G_y|\mathcal{Y}, \mathcal{Z}, \Phi, G_z)$,
4. Update between banks parameters: $(\Psi|\mathcal{Y}, \mathcal{Z}, \Phi, G_z, G_y)$,

See C.2 for details on the MCMC for sampling the relationships and estimating the parameters.

5.3.3 Centrality Measures

Let A be $n \times n$, $0\backslash 1$ matrix, which we will call the adjacency matrix. Network statistics can be derived using A: in particular, meaningful summary measures can be obtained using an appropriate singular value decomposition of such matrix. The simplest measure of node centrality is the degree which measures the number of connection of a given node in a network. In the framework of undirected network, the degree (d_i) of node i is simply:

$$d_i = \sum_{j=1}^{n} a_{i,j}$$

where j represents all other nodes, n is the total number of nodes, and a_{ij} is the (i, j) element of the adjacency matrix of the network, which is defined as 1 if node i is connected to node j, and 0 otherwise. The centrality measure that has been proposed in financial network modeling to explain the capacity of an agent to cause systemic risk, that is, a large contagion loss on other agents, is the eigenvector centrality (see e.g. Furfine, 2003; Billio et al., 2012). The eigenvector centrality measure is a measure of the importance of a node in a network. It assigns relative scores to all

nodes in the network, based on the principle that connections to few high scoring nodes contribute more to the score of the node in question than equal connections to low scoring nodes.

More formally, for the i-th node, the centrality score is proportional to the sum of the scores of all nodes which are connected to it, as in the following equation:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} a_{i,j} x_j,$$

where x_j is the score of a node j, $a_{i,j}$ is the (i, j) element of the adjacency matrix of the network, λ is a constant and n is the number of nodes of the network. The previous equation can be rewritten for all nodes, more compactly, as:

$$Ax = \lambda x,$$

where A is the adjacency matrix, λ is the eigenvalue of the matrix A, with associated eigenvector x, an n-vector of scores (one for each node). Note that, in general, there will be many different eigenvalues λ for which a solution to the previous equation exists. However, the additional requirement that all the elements of the eigenvector be positive (a natural request in our context) implies (by the Perron-Frobenius theorem) that only the eigenvector corresponding to the largest eigenvalue provides the desired centrality measures. Therefore, once an estimate of A is provided, network centrality scores can be obtained from the previous equation, as elements of the eigenvector associated to the largest eigenvalue.

In our two-level hierarchical model, it follows that the centrality measure decomposes into a country specific component, and a between banks component. In addition to centrality measures we will consider, as a summary measure of systemic risk of a bank, the number of its links with other banks in the systems.

5.4 Simulation Experiment

To illustrate how our methodology and computational algorithm work, In this Section we illustrate it with a 6-node simulation experiment of Wang and Li (2012). The experiment considers a model of $n = 6$ variables, with $T = 18$ observations, and an estimated sum of squares $S = T\hat{\Sigma} = TA^{-1}$, where the upper triangular matrix of A is defined as:

$$A = \begin{pmatrix} 1 & 0.5 & 0 & 0 & 0 & 0.4 \\ & 1 & 0.5 & 0 & 0 & 0 \\ & & 1 & 0.5 & 0 & 0 \\ & & & 1 & 0.5 & 0 \\ & & & & 1 & 0.5 \\ & & & & & 1 \end{pmatrix} \tag{5.14}$$

Using our proposed algorithm, we run 2000 Gibbs iterations with 200 burn-in ones and considered three thresholds $\tau = 0, 1, 2$. The result of the posterior probabilities of edge presence in the graph are shown in Table 5.1:

$(\tau = 0)$	$(\tau = 1)$	$(\tau = 2)$
$\begin{pmatrix} 1 & 0.94 & 0.02 & 0.01 & 0.02 & 0.83 \\ & 1 & 0.94 & 0.02 & 0.01 & 0.02 \\ & & 1 & 0.93 & 0.02 & 0.01 \\ & & & 1 & 0.93 & 0.01 \\ & & & & 1 & 0.94 \\ & & & & & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.96 & 0.05 & 0.03 & 0.04 & 0.92 \\ & 1 & 0.96 & 0.04 & 0.02 & 0.04 \\ & & 1 & 0.96 & 0.04 & 0.03 \\ & & & 1 & 0.96 & 0.04 \\ & & & & 1 & 0.96 \\ & & & & & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.95 & 0.76 & 0.76 & 0.76 & 0.92 \\ & 1 & 0.95 & 0.77 & 0.76 & 0.76 \\ & & 1 & 0.94 & 0.78 & 0.77 \\ & & & 1 & 0.95 & 0.77 \\ & & & & 1 & 0.96 \\ & & & & & 1 \end{pmatrix}$

Table 5.1: Marginal posterior probabilities of edges for threshold $\tau = 0, 1, 2$.

Following the idea of testing significance of regression coefficients, we apply a test for significance of an edge between two variables. We test if the posterior edge probability is greater than 0.5 under a 95% credibility interval. If this is true, then the edge is statistically significant, otherwise there is no link between the variables.

We monitor convergence of the MCMC using potential scale reduction factor (PSRF) of Gelman and Rubin (1992). See C.2 for computational details and convergence diagnostics of both simulated and empirical application of our inference approach. By comparing our graph estimation performance with the data gener-

ating process, we observe that fixing a threshold $\tau = 0$ or 1 produces an accurate description of the data generating process. According to the scale provided by Kass and Raftery (1995), using $\tau = 0$ leads to accepting edges that are not worth more than a bare mention. However, fixing $\tau = 1$ produces strong evidence against the null network. When $\tau = 2$, we observe from the results that the algorithm overestimates the number of links in the network therefore yielding an unsatisfactory representation of the links. Based on these results, we consider $\tau = 1$ as a robust logarithm Bayes factor threshold for our empirical application, that will be presented in the next Section.

5.5 European Banks Risk Network

In this section we apply our proposed model to the estimation of the systemic credit risk of the largest listed banks of the Eurozone. We consider the Eurozone because of the changes in progress in this area, where the European central bank (ECB) start the take over of the supervision of the largest banks on November 1st, 2014. It is a gradual process, that aims at replacing the previous fragmented supervision and regulation (between 17 member states) into a unified one, with common rules and practices. It thus seems timely and important to focus the analysis on the banks that belong to this area, with the aim of contributing to identifying the most contagious institutions, at the super-national level.

In this work we consider only large banks, whose total assets are greater than $30b$ euros, and that are included in the ECB comprehensive assessment review. As we aim for an approach that integrates market information with bank-specific data, we consider only publicly listed banks, for which market data is available. Finally, in the case of a banking group with more entities that satisfy the above criteria, we consider only the controlling entity. The complete list of the 45 considered financial institutions is in Table C.1, with the corresponding ticker code acronyms. To com-

plete the information, Table C.1 contains, besides bank names and their codes, their prevalent country, whether they are Systematically Important Financial Institutions (SIFI) either at the Global or at the Domestic level, their Total Assets from the last available balance sheet (in thousands of euro), the Year Over Year variation of their equity returns (in percentage points), and the corresponding standard deviation of the returns, also in percentage points.

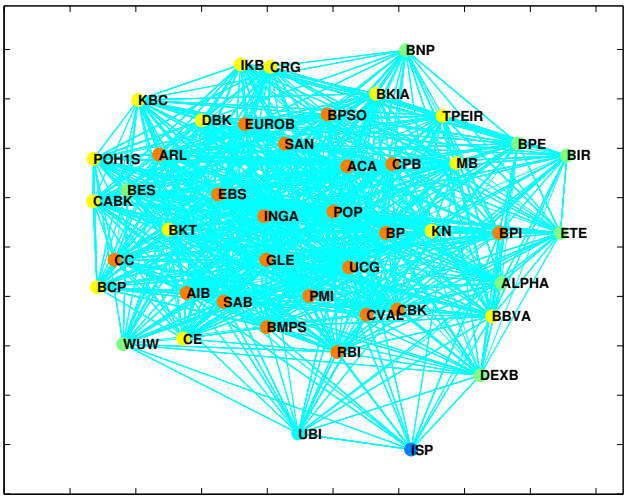
From Table C.1, note that 12 of the 17 eurozone countries have at least one bank with total assets greater than $30bn$ euro. The five missing countries are Estonia, Luxembourg, Malta, Slovakia and Slovenia. Apart from the missing (small) countries, Table C.1 underlines a remarkable difference in the banking structure of the different countries, that is to be remembered when interpreting the obtained results. Some countries have a large number of banks in the sample: Italy, the third populated country, has 12, Spain, the fourth populated country, has 7. Others have instead a limited number of banks: Germany, the most populated country, has only 5, and so does France, the second populated country. Note that Greece, a small country, with the most troubled eurozone economy, has 4. For comparison purposes, the ECB assessment sample (that includes unlisted banks) contains a total of 130 banks in the Eurozone, among which 24 German banks, 13 French, 15 Italian, 16 Spanish and 4 Greek ones. Looking at country representation note that our sample of banks, which does represent about a third of all banks in the ECB list, is overrepresented in terms of Greece banks (it contains them all) and Italian banks (has about 80% of them), and underrepresented in terms of German banks (has about 20% of them). The above bias is the result of the structure of the banking systems in the different countries, and especially of medium sized banks, that typically operate at a local (subnational) level: while most such banks, in Italy (and in Greece), are listed, in Germany this is rarely the case, with other countries lying in between.

The market and financial data that we considered is from Bankscope, a com-

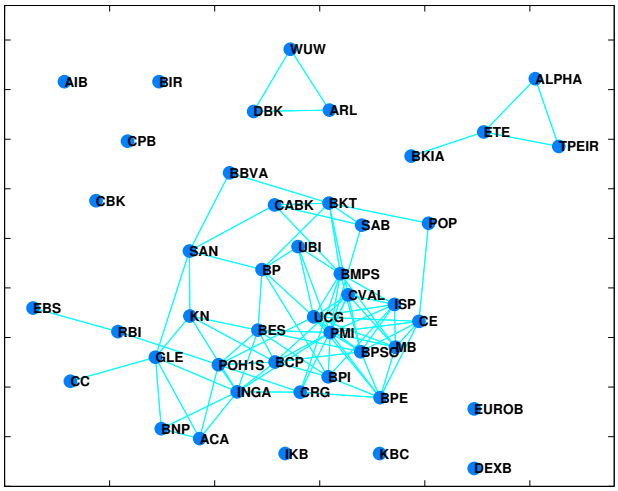
prehensive database for individual banks across the world provided by the private company Bureau Van Dijk. In particular, the information that we use covers a year of price stock data, just before the start of the ECB supervision, with observations on a weekly basis. Precisely, our observation period goes from October 22, 2012, to October 18, 2013. From the Bankscope database, we have extracted the weekly closing stock prices for each bank, P_t and, then, transformed them into weekly returns, defined as: $R_t = log(P_t/P_{t-1})$, where t is a week in the last year and $t-1$ the week that precedes it.

We now present the results from the application of Bayesian graphical models to the weekly log returns of the 45 considered banks. We applied to such data both a (non-hierarchical) Bayesian graphical model, and our proposed Bayesian hierarchical models that enlarges the state space of each bank log return variables with country specific log return variables, obtained averaging the returns of the banks in each country. This inclusion, if significant, is expected to lower the connections between banks, and especially cross-border ones, as connections between banks may be replaced by connections between countries.

Figure 5.1a shows the unconditional graphical network model between banks with the highest a posteriori probability, in the non hierarchical model, and Figure 5.1b the same best model, conditional on the country to country relationships, in the hierarchical model. From Figure 5.1a note that the model is highly interconnected, and it is rather difficult to interpret. The only apparent message conveyed is about banks that appear "peripheral", with a lower degree and a lower centrality measures. Such banks range from the troubled Greek banks Alpha bank and Piraeus Bank, and Dexia bank, to medium size banks such as Wustenrot and Wurttenbergische in Germany, Pohjola Bank in Finland, UBI in Italy, and two internationally oriented SIFIs such as Banco de Bilbao and BNP. These banks, because of their lower centrality, seem to appear less contagious/subject to contagion.



(a) Non-Hierarchical Network (Total Links : 711)



(b) Hierarchical Network (Total Links : 95)

FIGURE5.1: Comparison of Inter-Bank connections for banks in the Euro area from October, 2012, to October, 2013. (5.1a): a Non-Hierarchical Network and (5.1b): a Hierarchical Network. Banks are represented using their Bank Codes (See Table C.1 for Bank Names).

Figure 5.1b, instead presents a much clearer picture: the total number of links drop from 711 to 95. There is a remarkable difference between the unconditional results in Figure 5.1a and those obtained conditionally on a country effect, in Figure 5.1b. The most evident difference is the lower degree of connectivity between banks, especially at the cross-border level. This means that the country effect explains a lot of the co-movements between the returns of the banks. Table 5.3, that contains, for

each bank, the centrality measure and the total number of connections, gives further evidence in this direction: indeed, most banks have connections mainly with banks of their country, and this is a further confirmation of the fact that share prices of banks heavily depend on the risk of the countries in which they operate. This also explains why the most central and connected banks in Table 5.3 are Italian banks, which are more represented in the sample. To the other side of the spectrum, German banks, that appear disconnected from other countries' banks. Similarly, Greek banks form their own circle of troubled banks, together with the Spanish Bankia.

Besides the existence of a country effect, Table 5.3 emphasizes that some Systematically Important Financial Institutions (SIFI in Table 1) are potentially more contagious than others, as they are more connected. This is the case of Unicredit, heavily linked to most italian banks; of ING, strongly interconnected at an international level; and, although to a lesser extent, for the French banks BPCE, Credit Agricole and Societe General and the Spanish Banco Santander. Other SIFIs, including the two German ones Deutsche bank and Commerzbank, as well as BNP Paribas, are less central.

We now examine in detail the inter country linkages estimated by the hierarchical model. Figure 5.2 shows the between country graphical network that is learned (model averaged) from the data, and Table 5.2 gives the corresponding centrality measures. To aid interpretation, we have added a marker for the estimated sign of the partial correlation found: positive or negative.

From Figure 5.2 and Table 5.2, notice that country financial systems are not much connected, confirming the image of Europe as a "market with financial frictions". In particular, due to the still persisting strong crisis, that generates high credit loss impairments, southern European countries, such as Greece, Spain and Italy are connected with each other. Similarly, so are stronger economies such as Germany, Austria and Finland. France act as gates between troubled southern countries and

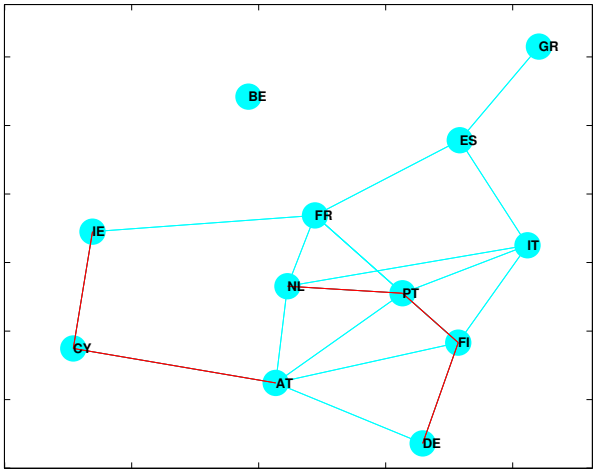


FIGURE5.2: Inter-Country connections in the Euro area from October, 2012, to October, 2013. AT stands for Austria, BE for Belgium, CY for Cyprus, DE for Germany, ES for Spain, FI for Finland, FR for France, GR for Greece, IE for Ireland, IT for Italy, NL for the Netherlands and PT for Portugal. The color of links indicate the signs of the partial correlations: red for negative and blue for positive.

Rank	Code	Country	Eig.V	Degree	Num. Banks	YOY	Stdev
1	PT	Portugal	0.4497	5	5	5.24	4.64
2	FI	Finland	0.4148	5	1	24.34	4.52
3	AT	Austria	0.3714	5	7	15.45	3.80
4	FR	France	0.3695	5	3	11.21	6.68
5	NL	Netherland	0.3692	4	1	59.85	8.36
6	IT	Italy	0.3412	4	12	-10.39	9.49
7	DE	Germany	0.1895	2	5	16.04	4.92
8	ES	Spain	0.1818	3	2	-61.51	10.98
9	CY	Cyprus	0.1178	2	1	44.17	4.96
10	IE	Ireland	0.1174	2	2	16.18	4.63
11	GR	Greece	0.0438	1	4	-24.67	11.65
12	BE	Belgium	0	0	2	24.16	4.00

Table 5.2: Inter-Country centrality measures from October, 2012, to October, 2013.

stronger ones, along with the Netherlands and Portugal whose banks have a high international exposure. Finally, Belgium, Ireland and Cyprus, smaller economies, follow very specific paths: Belgium, disconnected, because of the very different behaviour of its two banks: Dexia, under restructuring, and KBC, healthier; Cyprus, that went through a dramatic financial crisis, in the last year; Ireland, whose banks have gone through a year of gains, that followed years of crisis.

Indeed, looking at Table 5.2, the most contagious countries seem France, Portu-

gal, Finland, Austria and Netherlands. While the French economy is indeed a gate between southern and Northern european economies, the other centralities can be explained by the multinational activities of the banks of the related countries.

Having estimated the country-to-country connections, we now look in detail the interdependencies between banks, that emerge from Figure 5.1b. As already remarked, Table 5.3 gives the corresponding numeric description. In particular, it gives, for each bank, its eigenvalue centrality measure, and the total number of links in the conditional model. Such total number is split between the within country links and the cross-border links of each bank.

We first focus on the within country links that are estimated by the Bayesian model averaged hierarchical graphical model. Figure 5.3 extracts from Figure 5.1b the bank relationships that are learned from the data, within each country. A first remark is that, conditionally on the country-to country relationships, the number of links between banks, within country, decreases from 96 to 58. The results in Figure 5.3 can be read off jointly with the column of Table 5.3 that shows the Within Country links, country by country. We focus on the largest ones.

Italy

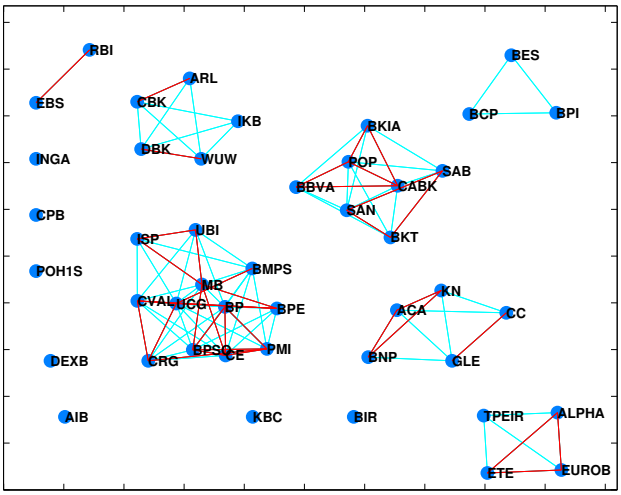
Figure 5.4 shows that Italian banks seem to rotate around Unicredit, and Intesa San Paolo, as well as with the investment bank Mediobanca. Among the other banks the most central ones are the cooperative territorial banks such as Popolare di Milano, Banca Popolare di Sondrio, Credito Valtellinese and Banca Popolare Emilia Romagna, indeed linked with each other. Note that the two most troubled banks, Monte dei Paschi di Siena and Banca Carige, have different centrality: the former is more connected and, therefore, can spread its distress among more neighbours.

We also underline that a remarkable aspect of the model is that it seems to be able to capture even regional (subnational) links between banks, expression of the

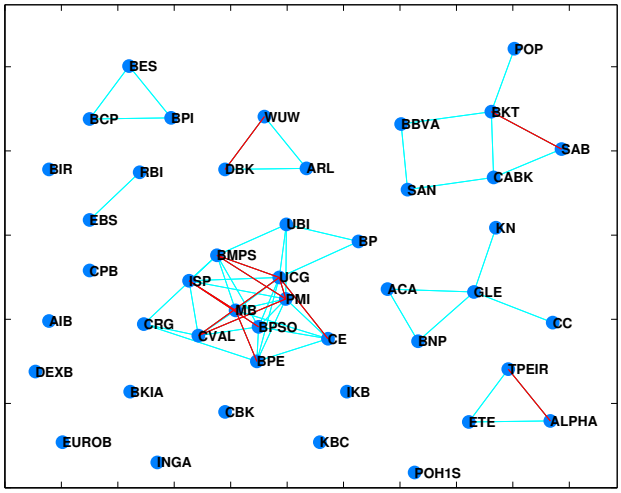
	Code	Bank Name	Country	Eig.V	Degree		
					Tot	W-C	A-C
1	PMI	Banca Popolare di Milano	Italy	0.3995	14	9	5
2	UCG	UniCredit	Italy	0.3314	11	10	1
3	BPSO	Banca Popolare di Sondrio	Italy	0.3052	9	7	2
4	CVAL	Credito Valtellinese	Italy	0.2587	8	6	2
5	BMPS	Banca Monte dei Paschi di Siena	Italy	0.2557	9	7	2
6	ISP	Intesa Sanpaolo	Italy	0.2552	7	6	1
7	MB	Mediobanca	Italy	0.2482	7	7	0
8	BPE	Banca popolare Emilia Romagna	Italy	0.2324	7	6	1
9	BCP	Banco Comercial Portugues	Portugal	0.2289	8	2	6
10	BES	Banco Espirito Santo	Portugal	0.2145	8	2	6
11	CE	Credito Emiliano	Italy	0.1999	6	5	1
12	BPI	Banco BPI	Portugal	0.1879	6	2	4
13	INGA	ING Groep	Netherland	0.1751	9	0	9
14	POH1S	Pohjola Bank Oyj	Finland	0.1490	7	0	7
15	UBI	Unione di Banche Italiane	Italy	0.1437	4	4	0
16	CRG	Banca Carige	Italy	0.1375	5	3	2
17	BP	Banco Popolare	Italy	0.1325	6	2	4
18	BKT	Bankinter	Spain	0.1081	7	4	3
19	KN	Natixis-BPCE Group	France	0.0921	5	1	4
20	SAB	Banco de Sabadell	Spain	0.0732	3	2	1
21	CABK	Caixabank	Spain	0.0620	4	3	1
22	ACA	Credit Agricole	France	0.0530	4	2	2
23	GLE	Societe Generale	France	0.0525	6	4	2
24	SAN	Banco Santander	Spain	0.0461	5	2	3
25	POP	Banco Popular Espanol	Spain	0.0396	2	1	1
26	BNP	BNP Paribas	France	0.0360	3	2	1
27	BBVA	Banco Bilbao Vizcaya Argentaria	Spain	0.0198	2	2	0
28	RBI	Raiffeisen Bank International	Austria	0.0195	2	1	1
29	CC	CIC Credit Mutuel Group	France	0.0067	1	1	0
30	EBS	Erste Group Bank	Austria	0.0025	1	1	0
31	BIR	Bank of Ireland	Ireland	0	0	0	0
32	ALPHA	Alpha Bank	Greece	0	2	2	0
33	EUROB	Eurobank Ergasias	Greece	0	0	0	0
34	TPEIR	Piraeus Bank	Greece	0	2	2	0
35	ETE	National Bank of Greece	Greece	0	3	2	1
36	BKIA	Bankia	Spain	0	1	0	1
37	DBK	Deutsche Bank	Germany	0	2	2	0
38	CBK	Commerzbank	Germany	0	0	0	0
39	DEXB	Dexia	Belgium	0	0	0	0
40	KBC	KBC Group	Belgium	0	0	0	0
41	CPB	Bank of Cyprus	Cyprus	0	0	0	0
42	WUW	Wustenrot & Wurttembergische	Germany	0	2	2	0
43	ARL	Aareal Bank	Germany	0	2	2	0
44	IKB	IKB Deutsche Industriebank	Germany	0	0	0	0
45	AIB	Allied Irish Banks	Ireland	0	0	0	0

Table 5.3: Inter-Bank centrality measures from the Hierarchical Network for banks in the Euro area over the period October, 2012, to October, 2013. Tot : Total Degree, W-C : Within Country Degree, and A-C : Across Country Degree.

fact that banks may share the same lenders and, therefore, similar risks. This is the case for: CVAL and BPSO , both operating mainly in the Lombardy region; BP and UBI, both operating in the north center-east region; BPE and CE both operating in the Emilia Romagna region.



(a) Non-Hierarchical Network (Total Links : 96)



(b) Hierarchical Network (Total Links : 58)

FIGURE5.3: Comparison of Within-Country connections over the period October, 2012, to October, 2013. (5.3a): a non-Hierarchical Network and (5.3b): a Hierarchical Network. Banks are represented using their Bank Codes (See Table C.1 for Bank Names). The color of links indicate the signs of the partial correlations: red for negative and blue for positive.

Spain

In the case of Spain, the territorial banks Bankinter and Caixabank are the most central, within country. The two large banks Santander and Banco de Bilbao are less central, but related to each other. Finally, Bankia is disconnected from the others, due to its state led restructuring. Considerations on regional dependences can be

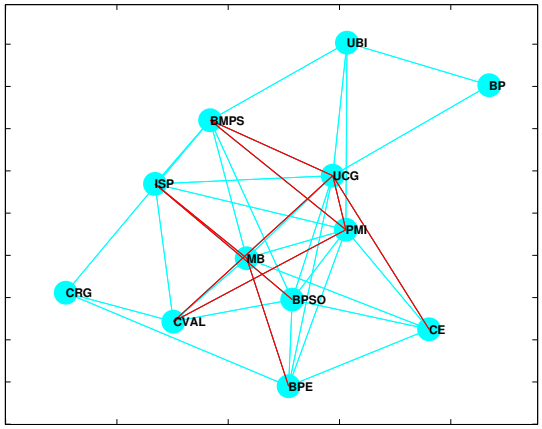


FIGURE5.4: Inter-Bank connections in Italy over the period October, 2012, to October, 2013. UniCredit (UCG), Intesa Sanpaolo (ISP), Banca Monte dei Paschi di Siena ($BMPS$), Unione di Banche Italiane (UBI), Banco Popolare (BP), Mediobanca (MB), Banca popolare Emilia Romagna (BPE), Banca Popolare di Milano (PMI), Banca Carige (CRG), Banca Popolare di Sondrio ($BPSO$), Credito Emiliano (CE) and Credito Valtellinese ($CVAL$).

drawn for spanish banks in a similar fashion to Italiian ones.

France

In France the Systematically Important Institution Societe Generale acts as a gate between the territorial bank groups BPCE and Credit Mutuel and the larger groups: BNP Paribas and Credit Agricole.

Germany

In Germany the Systematically Important Deutsche Bank is connected to both Aareal Bank and Wustenrot, while Commerzbank and IKB are isolated, with their respectively bad and good performances.

Greece

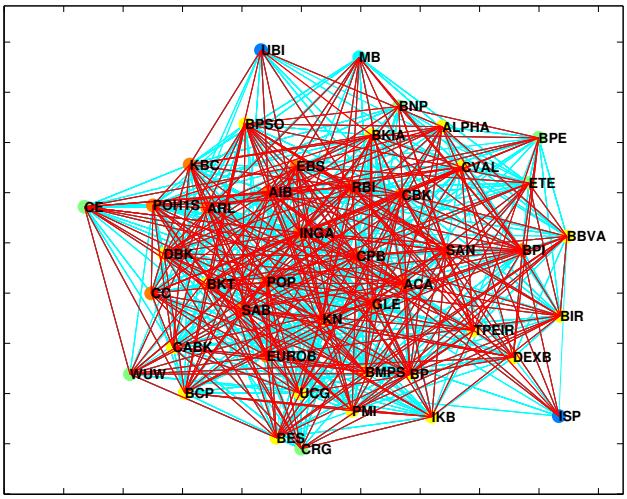
All Greek banks but Eurobank (involved in a restructuring plan) are related with each other, and this reflects the fact that they reflect a common deep financial crisis of the country.

5.5.1 Discussion

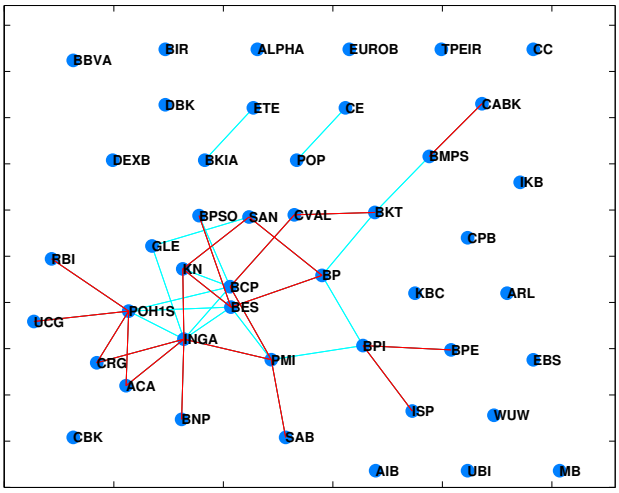
From the above discussion, the existence of a country effect, that determines statistical correlations and dependencies between banks, is clear. From an economical viewpoint such effect can be explained as follows. On the asset side, banks acts as lenders and, thus, if the economy is doing badly, so do banks, as enterprises do not repay the given credit and/or go default. In addition, countries with high public debts typically persuade banks to buy the related government securities. When the economy is badly performing, the value of such bond decreases. On the liability side, depositors from troubled countries typically put less money in banks or withdraw it. Furthermore, when a bank does badly, the government may intervene, and capitalize it buying a relevant portion of shares. Because of this inter-linkage between banks and their countries, we expect stock prices of banks to be highly determined by a country risk effect, which may increase the interconnections between banks in troubled countries, siuch as southern European ones.

We now look at bank to bank connections, that are estimated by our model, across countries, rather than within each country. Figure 5.5 shows the across country between bank network that is learned from the data, extracted from 5.1b. The results in Figure 5.5 can be read off jointly with the column of Table 5.3 that shows the Across Country links.

We remind the reader that, on the basis of the assumptions of the proposed hierarchical model, the less a country is connected, the less likely the presence of significant cross border relationships for its banks. To exemplify, a cross border correlation for a German (or an Irish) bank is less likely to remain in the model than a similar (in size) cross border correlation for a French (or a Portoguese) bank, as France (and Portugal) are much more central than Germany (and Belgium) in the country-to-country network.



(a) Non-Hierarchical Network (Total Links : 615)



(b) Hierarchical Network (Total Links : 37)

FIGURE5.5: Comparison of Inter-Bank connections across countries in the Euro area over the period October, 2012, to October, 2013. (5.5a): a Non-Hierarchical Network and (5.5b): a Hierarchical Network. Banks are represented using their Bank Codes (See Table C.1 for Bank Names). The color of links indicate the signs of the partial correlations: red for negative and blue for positive.

From Figure 5.5, notice that the number of connections significantly decreases, from 615 to 37, much more than what do the within country connections. From the figure and Table 5.3, the most connected banks, at the cross border level, are either banks in small countries, such as ING, Pohjola, Banco Commercial Portoguese and Banco Espirito Santo, or multinational SIFIs such as Unicredit and Societe Generale. Notice also the high centrality of territorial banks such as Banca Popolare di Milano,

Banco Popolare, Banco BPI and BPCE, presumably as they have relevant portions of shares in the hand of international investment funds, that look at them with common portfolio based strategies.

Another interesting remark that can be drawn from Figure 5.5 concerns the banks with no cross-border connections: the Greek and the German banks, whose country, as already seen, is disconnected from the others; the Irish banks which have no specific connections beyond the aggregate country ones and, finally, the Bank of Cyprus, Bankia and the Belgian banks.

A final remark concerns the comparison between SIFIs, whose contagion is the most dangerous beacuse of their asset size. The most central ones appear Unicredit, ING, and the French banks Societe Generale, BPCE and Credit Agricole. Banco Santander, BNP Paribas and Banco Bilbao are connected, but less then the previous ones; Commerzbank, Dexia, and Deutsche Bank are the least connected.

For sensitivity analysis, we run our algorithm with different log Bayes factor thresholds, $\tau = 0$ and $\tau = 2$, (see C.3). Our results, summarised in Table C.2 show that the results do not differ considerably from previous ones but the corresponding networks are, respectively, less or more interconnected.

5.6 Conclusions

Financial network models are a useful tool to model interconnectedness and systemic risks in financial systems. Such models are essentially descriptive, and based on highly correlated networks. The motivation of this paper is to provide a stochastic framework for financial network models, aimed at a more parsimonious and more realistic representation. The paper contains two main research contributions in this direction. First we have introduced Bayesian Gaussian graphical models in the field of systemic risk modelling, thus estimating the adjacency matrix of a network in a robust and coherent way. Following this approach, (Bayesian) confidence inter-

vals for the centrality measures can be easily obtained, from the simulated output. Second, we have proposed a hierarchical Bayesian graphical model that can usefully decompose dependencies between financial institutions into correlations into between countries and correlations between institutions, within and/or across countries.

We have applied our proposed methods to the largest Eurozone banks, with the aim of identifying central institutions, more subject to contagion of or, conversely, whose failure could result in further distress or breakdowns in the whole system. Our results show that, in the transmission of the perceived default risk, there is a strong country effect, that reflects the weakness and the strenth of the underlying economies. Besides the country effect, the most central banks appear the large international ones, especially if from a relatively small country.

From an applied viewpoint, in the paper we have considered only stock market data and not balance sheet data. This because the latter is typically backward-looking, whereas market data incorporate future expectations. We believe, however, that the two sources of information are complementary and should be combined in a methodological extension of our work, possibly also considering "soft" information, that may be "forward looking", such as that present in social network data.

Acknowledgement

We acknowledge financial support from the Italian ministry PRIN MISURA project: Multivariate models for risk assessment. We also thank the referee for useful comments and discussion that have helped improving the content of the paper.

Chapter 6

Conclusion

6.1 Summary

This thesis is motivated by the increase in attention and application of econometric methods for systemic risk analysis mainly through network models. More specifically, empirical studies on systemic risk have shown that networks reflect the architecture of interactions that arise among financial institutions and can provide insight into the structure and stability of the financial system. The network approach has shown to be very useful in understanding the global financial system as a web of interconnected institutions, where linkages among institutions helps to analyze contagion, spillover effects and mechanisms for propagating risk. The thesis contributes to the literature on network econometrics by providing new techniques for building models that enhance our understanding of the economic and/or financial system.

In Chapter 2, we review the literature on graphical models from a multidisciplinary perspective and their application in the context of systemic risk and financial contagion. We discussed potential applications of such models in econometrics and finance. We also demonstrate empirically the application of the Bayesian approach to graph structural inference on financial time series. We consider the dataset of Billio et al. (2012) on the monthly return indexes of hedge funds, brokers, banks

and insurance companies over two samples: 1994-2000 and 2001-2008. Our result corroborate the findings of Billio et al. (2012), with evidence of a higher interconnection between 2001-2008 leading to the global financial crisis between 2007-2009. The evidence also suggests that insurance companies played a central role in the interconnectedness among the financial institutions during the period. We extended the application to analyze the volatility connectedness in the European stock market. The result shows that, the graphical (joint estimation) method allows us to extract a dependence patterns more suitable to model the complex relationships in the European stock market than the Granger-causality (pairwise estimation). We find evidence that Banks and Insurance companies are more central in the "fear connectedness" (Diebold and Yilmaz, 2014) expressed by market participants in the financial sector of the Euro-area.

In Chapter 3, we propose 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.

In Chapter 4, we propose a model selection approach to multivariate time series of large dimension by combining graph-based notion of causality with the concept of sparsity on the structure of dependence among the variables. In particular, we build on the application of fan-in restriction for graphical models by proposing a sparsity-inducing prior distribution that allows for different prior information level about the maximal number of predictors for each equation of a VAR model. We discuss the joint inference of the temporal dependence in the observed series and the maximum lag order of the process, with the parameter estimation of the model. The applied contribution focuses on modeling and forecasting selected macroeconomic and financial time series with many predictors. Our result shows a gain in predictive performance using our sparse graphical VAR.

In Chapter 5, we develop a stochastic framework for financial network models, aimed at a more parsimonious and more realistic representation. We introduce Bayesian Gaussian graphical models in the field of systemic risk modeling, by estimating the adjacency matrix of a network of financial institutions in a robust and coherent way. In addition, we propose a hierarchical Bayesian graphical model that can usefully decompose dependencies between financial institutions into linkages between different countries financial systems and linkages between banking institutions, within and/or across countries. This is applied to study the largest banks in the Euro-area with the aim of identifying central institutions. More specifically to understand how linkages of inter-country systems contribute to the contagion process among banks. Our results show that, in the transmission of default risk, there is a strong country effect, that reflects the weakness/strength of the underlying economies. Besides the country effects, we observe that banks that are central to the transmission of default risk are the large international banks.

6.2 Extensions and Further Research

The Bayesian graphical approach as presented in this work has a broad range of real world applications, especially in macro-econometrics as well as other fields in empirical economics and finance. Further research with this approach will be to develop the research lines discussed in Chapter 2 with methodological and/or application contributions. Given the promising results of the subsequent chapters, the graphical approach can be extended to other applications of VAR models as well to Dynamic Stochastic General Equilibrium (DSGE) models.

In Chapter 2, one of the research lines discussed is time varying model estimation. In the light of this, Ahelegbey and Koopman (working paper, 2014) adopt the generalized autoregressive score (GAS) methodology to model varying dependence in multivariate time series. The GAS approach allows the dynamics of the graph to be modeled as an observation-driven time varying parameter model where the graphs are used as parameters and the mechanism to update them is the scaled score of the marginal likelihood function.

In Chapter 3, we pointed out that though our graphical approach provides a solution to the identification problem in structural VAR (SVAR) models, the result is not a unique solution to the problem. This is so because cross-sectional dependence graphs belonging to the same equivalence class have the same score and therefore cannot be distinguished on the basis of the probability distribution of the data. A possible extension is to combine our approach with other sources of information, by incorporating prior experts knowledge into the inference of the cross-sectional dependence from the equivalence class of graphs. If external interventions are not possible and experts knowledge are not available to disentangle and identify the contemporaneous causal ordering in SVAR, then a possible alternative is to allow for partially directed graphs. In this case, only directed edges can be assigned a causal

interpretation. Undirected edges on the other hand can be considered as simultaneous interactions among variables. This notion is in line with Hoover (2001), who argues that an adequate account of causality must permit simultaneity, especially in economics since much of economics is represented in simultaneous systems.

In Chapter 4, we found strong evidence supporting the effect of financial variables on the real sector of the US economy. Further research in this area (under preparation) is to consider the application of graphical models for empirical evaluation of macro-financial linkages to understand how policy instruments targeted at financial institutions might have direct or indirect effect on the macroeconomy through the connectedness between the financial sector and the macroeconomy.

A methodological extension of Chapter 5 is to consider a fairly informative (or sparsity) prior over the different graphical structures. This will lead to penalizing overly complicated models. Although this may introduce a selection bias, it may be help model selection when large networks are being considered. A further extension of the paper involves considered directed, rather than undirected, graphical Gaussian models. Doing so we may be able to provide "causal" directions of systemic risks and not only symmetric ones.

Bibliography

Abramowitz, M. and Stegun, I. A. (1964), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Courier Dover Publications.

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2010), "Measuring Systemic Risk," Working Paper 1002, Federal Reserve Bank of Cleveland.

Adrian, T. and Brunnermeier, M. K. (2010), "CoVaR," Staff Reports 348, Federal Reserve Bank of New York.

Ahelegbey, D. F. (2015), "The Econometrics Aspects of Networks: A Review," *Working Paper*.

Ahelegbey, D. F. and Giudici, P. (2014), "Bayesian Selection of Systemic Risk Networks," *Advances in Econometrics: Bayesian Model Comparison*, 34, 117–153.

Ahelegbey, D. F., Billio, M., and Casarin, R. (2014), "Sparse Graphical Vector Autoregression: A Bayesian Approach," *Working Paper, Social Science Research Network*.

Ahelegbey, D. F., Billio, M., and Casarin, R. (2015), "Bayesian Graphical Models for Structural Vector Autoregressive Processes," *Journal of Applied Econometrics (forthcoming)*.

Akavia, U. D., Litvin, O., Kim, J., Sanchez-Garcia, F., Kotliar, D., Causton, H. C., Pochanard, P., Mozes, E., Garraway, L. A., and Pe'er, D. (2010), "An Integrated Approach to Uncover Drivers of Cancer," *Cell*, 143, 1005–1017.

Altman, E. I. (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *The journal of finance*, 23, 589–609.

Ammann, M. and Verhofen, M. (2007), "Prior Performance and Risk-Taking of Mutual Fund Managers: A Dynamic Bayesian Network Approach," *The Journal of Behavioral Finance*, 8, 20–34.

Andersson, S. A., Madigan, D., and Perlman, M. D. (1997), "A Characterization of Markov Equivalence classes for Acyclic Digraphs," *The Annals of Statistics*, 25, 505–541.

Andrieu, C. and Roberts, G. O. (2009), "The Pseudo-Marginal Approach for Efficient Monte Carlo Computations," *The Annals of Statistics*, pp. 697–725.

Arena, M. (2005), "Bank Failures and Bank Fundamentals: A Comparative Analysis of Latin America and East Asia During the Nineties Using Bank-Level Data," *Journal of Banking and Finance*, 32, 299–310.

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

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

Barigozzi, M. and Brownlees, C. (2014), "Nets: Network Estimation for Time Series," *Working Paper, Social Science Research Network*.

Basel Committee on Banking Supervision (2011), "Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems," Tech. rep., Bank for International Settlements.

Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B., and Stiglitz, J. E. (2012), "Liaisons Dangereuses: Increasing Connectivity, Risk Sharing, and Systemic Risk," *Journal of Economic Dynamics and Control*, 36, 1121–1141.

Belloni, A. and Chernozhukov, V. (2011), *High Dimensional Sparse Econometric Models: An Introduction*, Springer.

Berger, J. O. and Pericchi, L. R. (1998), "On Criticisms and Comparisons of Default Bayes Factors for Model Selection and Hypothesis Testing," in *Proceedings of the Workshop on Model Selection*, Edizioni Pitagora, Bologna, Italy.

Bernanke, B. (2013), "Monitoring the Financial System," in *Intervento alla 49th Annual Conference Bank Structure and Competition, Chicago*, vol. 10.

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

Bernanke, B. S. (1986), "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49–100.

Bernanke, B. S. (2010), "Causes of the Recent Financial and Economic Crisis," *Statement before the Financial Crisis Inquiry Commission, Washington, September*, 2.

Bhattacharya, A. and Dunson, D. B. (2011), "Sparse Bayesian Infinite Factor Models," *Biometrika*, 98, 291–306.

Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2014), "Modeling Contagion and Systemic Risk," *Working Paper, Social Science Research Network*.

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

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, vol. 4, Springer New York.

Bisias, D., Flood, M. D., Lo, A. W., and Valavanis, S. (2012), "A Survey of Systemic Risk Analytics," *US Department of Treasury, Office of Financial Research*.

Bogdan, M., Ghosh, J. K., and Doerge, R. (2004), "Modifying the Schwarz Bayesian Information Criterion to Locate Multiple Interacting Quantitative Trait Loci," *Genetics*, 167, 989–999.

Box, G. E. and Meyer, R. D. (1986), "An Analysis for Unreplicated Fractional Factorials," *Technometrics*, 28, 11–18.

Brillinger, D. R. (1996), "Remarks Concerning Graphical Models for Time Series and Point Processes," *Revista de Econometria*, 16, 1–23.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011), *Handbook of Markov Chain Monte Carlo*, CRC press.

Brooks, S. P. and Gelman, A. (1998), "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.

Brownlees, C. T. and Engle, R. (2011), "Volatility, Correlation and Tails for Systemic Risk Measurement," Technical report, New York University.

Brunnermeier, M. K. and Oehmke, M. (2012), "Bubbles, Financial Crises, and Systemic Risk," NBER Working Papers 18398, National Bureau of Economic Research, Inc, Cambridge, MA.

Brunnermeier, M. K. and Pedersen, L. H. (2009), "Market Liquidity and Funding Liquidity," *Review of Financial studies*, 22, 2201–2238.

Bunea, F., She, Y., Wegkamp, M. H., et al. (2011), "Optimal Selection of Reduced Rank Estimators of High-dimensional Matrices," *The Annals of Statistics*, 39, 1282–1309.

Carriero, A., Clark, T. E., and Marcellino, M. (2013), "Bayesian VARs: Specification Choices and Forecast Accuracy," *Journal of Applied Econometrics*, 25, 400–417.

Carvalho, C., Massam, H., and West, M. (2007), "Simulation of Hyper-inverse Wishart Distributions in Graphical Models," *Biometrika*, 94, 647–659.

Carvalho, C. M. and Scott, J. G. (2009), "Objective Bayesian Model Selection in Gaussian Graphical Models," *Biometrika*, 96, 497–512.

Carvalho, C. M. and West, M. (2007), "Dynamic Matrix-Variate Graphical Models," *Bayesian Analysis*, 2, 69–98.

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and West, M. (2008), "High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics," *Journal of the American Statistical Association*, 103.

Casarin, R., Dalla Valle, L., and Leisen, F. (2012), "Bayesian Model Selection for Beta Autoregressive Processes," *Bayesian Analysis*, 7, 385–410.

Casella, G. and Robert, C. P. (2004), *Monte Carlo Statistical Methods*, Springer Verlag, New York.

Chen, H., Cummins, J. D., Viswanathan, K. S., and Weiss, M. A. (2013a), "Systemic Risk and the Interconnectedness between Banks and Insurers: an Econometric Analysis," *Journal of Risk and Insurance*.

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

Chen, K., Dong, H., and Chan, K.-S. (2013b), "Reduced Rank Regression via Adaptive Nuclear Norm Penalization," *Biometrika*, 100, 901–920.

Chen, L. and Huang, J. Z. (2012), "Sparse Reduced-Rank Regression for Simultaneous Dimension Reduction and Variable Selection," *Journal of the American Statistical Association*, 107, 1533–1545.

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

Chickering, D. M. and Heckerman, D. (1997), "Efficient Approximations for the Marginal Likelihood of Bayesian Networks with Hidden Variables," *Machine Learning*, 29, 181–212.

Chickering, D. M., Heckerman, D., and Meek, C. (2004), "Large-Sample Learning of Bayesian Networks is NP-Hard," *Journal of Machine Learning Research*, 5, 1287–1330.

Choi, M. J., Tan, V. Y., Anandkumar, A., and Willsky, A. S. (2011), "Learning Latent Tree Graphical Models," *The Journal of Machine Learning Research*, 12, 1771–1812.

Cole, R. A. and Gunther, J. W. (1998), "Predicting Bank Failures: A Comparison of on-and off-site Monitoring Systems," *Journal of Financial Services Research*, 13, 103–117.

Coleman, J. S. (1986), "Social Theory, Social Research, and a Theory of Action," *American journal of Sociology*, pp. 1309–1335.

Consonni, G. and Rocca, L. L. (2012), "Objective Bayes Factors for Gaussian Directed Acyclic Graphical Models," *Scandinavian Journal of Statistics*, 39, 743–756.

Cooley, T. F. and Leroy, S. F. (1985), "A Theoretical Macroeconometrics: A Critique," *Journal of Monetary Economics, Elsevier,*, 16, 283–308.

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

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999), *Probabilistic Networks and Expert Systems: Statistics for Engineering and Information Science*, Springer Verlag New York, Inc.

Dahlhaus, R. and Eichler, M. (2003), "Causality and Graphical Models for Time Series," *n: P. Green, N. Hjort, and S. Richardson (eds.): Highly structured stochastic systems. University Press, Oxford.*

DasGupta, B. and Kaligounder, L. (2014), "On Global Stability of Financial Networks," *Journal of Complex Networks*, pp. 1–59.

Davis, E. P. and Karim, D. (2008), "Comparing Early Warning Systems for Banking Crises," *Journal of Financial stability*, 4, 89–120.

Davis, R. A., Zang, P., and Zheng, T. (2012), "Sparse Vector Autoregressive Modeling," *arXiv preprint arXiv:1207.0520.*

Dawid, A. (2003), "An Object-Oriented Bayesian Network for Estimating Mutation Rates," in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, pp. 3–6.

Dawid, A. P. and Lauritzen, S. L. (1993), "Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models," *The Annals of Statistics*, 21, 1272–1317.

Dawid, A. P. and Lauritzen, S. L. (2001), "Compatible Prior Distributions," *In Bayesian Methods with Application to Science, Policy and Official Statistics; Selected Papers from ISBA 2000*.

De Bandt, O. and Hartmann, P. (2000), "Systemic Risk: A Survey," .

De Lisa, R., Zedda, S., Vallascas, F., Campolongo, F., and Marchesi, M. (2011), "Modelling Deposit Insurance Scheme Losses in a Basel 2 Framework," *Journal of Financial Services Research*, 40, 123–141.

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

Demiralp, S. and Hoover, K. D. (2003), "Searching for the Causal Structure of a Vector Autoregression," *Oxford Bulletin of Economics and Statistics*, 65, 745–767.

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

Diks, C. and Panchenko, V. (2005), "A Note on the Hiemstra-Jones Test for Granger Non-Causality," *Studies in Nonlinear Dynamics & Econometrics*, 9.

Ding, M., Chen, Y., and Bressler, S. L. (2006), "Granger Causality: Basic Theory and Applications to Neuroscience," in *Schelter B., Winterhalder M., Timmer J. (eds) Handbook of time series analysis*, pp. 437–460, Wiley, Weinheim.

Doan, T., Litterman, R., and Sims, C. (1984), "Forecasting and Conditional Projection using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.

Donoho, D. and Stodden, V. (2006), "Breakdown Point of Model Selection when the Number of Variables Exceeds the Number of Observations," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pp. 1916–1921, IEEE.

Donoho, D. L. (2006), "High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension," *Discrete Computational Geometry*, 35, 617–652.

Drton, M. and Perlman, M. D. (2007), "Multiple Testing and Error Control in Gaussian Graphical Model Selection," *Statistical Science*, pp. 430–449.

Dungey, M., Luciani, M., and Veredas, D. (2012), "Ranking Systemically Important Financial Institutions," Tech. rep., Tinbergen Institute.

Edwards, D. (1990), "Hierarchical Interaction Models (with discussion)," *Journal of the Royal Statistical Society, Series B*, 52, 3–20 and 51–72.

Edwards, D. (2000), *Introduction to Graphical Modelling*, Springer Science & Business Media.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), "Least Angle Regression," *The Annals of Statistics*, 32, 407–499.

Ehlers, R. S. and Brooks, S. P. (2004), "Bayesian Analysis of order Uncertainty in ARIMA Models," *Federal University of Parana, Tech. Rep.*

Eichler, M. (2007), "Granger Causality and Path Diagrams for Multivariate Time Series," *Journal of Econometrics*, 137, 334–353.

Elliott, G., Gargano, A., and Timmermann, A. (2013), "Complete Subset Regressions," *Journal of Econometrics*, 177, 357–373.

European Central Bank (ECB) (2010), "Financial Networks and Financial Stability," *Financial Stability Review*, 2010, 155–160.

Fama, E. F. and French, K. R. (2004), "The Capital Asset Pricing Model: Theory and Evidence," *Journal of Economic Perspectives*, 18, 25–46.

Fan, J. and Peng, H. (2004), "Nonconcave Penalized Likelihood with a Diverging Number of Parameters," *The Annals of Statistics*, 32, 928–961.

Fantazzini, D. and Maggi, M. (2012), "Computing Reliable Default Probabilities in Turbulent Times," *Rethinking Valuation and Pricing Models*, pp. 241–255.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalized Factor Model: Identification and Estimation," *Review of Economics and statistics*, 82, 540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2004), "The Generalized Factor Model: Consistency and Rates," *Journal of Econometrics*, 119, 231–255.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005), "The Generalized Dynamic Factor Model," *Journal of the American Statistical Association*, 100.

Foygel, R. and Drton, M. (2011), "Bayesian Model Choice and Information Criteria in Sparse Generalized Linear Models," *arXiv preprint arXiv:1112.5635*.

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

Friedman, N., Murphy, K., and Russell, S. (1998), "Learning the Structure of Dynamic Probabilistic Networks," in *Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence*, pp. 139–147, Morgan Kaufmann Publishers Inc.

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000), "Using Bayesian Networks to Analyze Expression Data," *Journal of Computational Biology*, 7, 601–620.

Frydenberg, M. and Lauritzen, S. L. (1989), "Decomposition of Maximum Likelihood in Mixed Graphical Interaction Models," *Biometrika*, 76, 539–555.

Furfine, C. H. (2003), "Interbank Exposures: Quantifying the Risk of Contagion," *Journal of Money, Credit and Banking*, 35, 111–128.

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

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

Gelman, A. and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences, (with discussion)," in *Statistical Science*, vol. 7, pp. 457–511.

George, E. I., Sun, D., and Ni, S. (2008), "Bayesian Stochastic Search for VAR Model Restrictions," *Journal of Econometrics*, 142, 553–580.

Geweke, J. (1977), *The Dynamic Factor Analysis of Economic Time-Series Models.*, SSRI workshop series, Social Systems Research Institute, University of Wisconsin-Madison.

Giannone, D., Reichlin, L., and Sala, L. (2005), "Monetary Policy in Real Time," in *NBER Macroeconomics Annual 2004*, vol. 19, pp. 161–224, MIT Press.

Giudici, P. and Castelo, R. (2003), "Improving Markov Chain Monte Carlo Model Search for Data Mining," *Machine Learning*, 50, 127–158.

Giudici, P. and Green, P. J. (1999), "Decomposable Graphical Gaussian Model Determination," *Biometrika*, 86, 785–801.

Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, 37, 424–438.

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

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

Grzegorczyk, M., Husmeier, D., and Rahnenführer, J. (2010), "Modelling Nonstationary Gene Regulatory Processes," *Advances in Bioinformatics*, pp. 1–17.

Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011), "Joint Estimation of Multiple Graphical Models," *Biometrika*, 98, 1–15.

Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2013), "Estimating Heterogeneous Graphical Models for Discrete Data with an Application to Roll Call Voting," *Annals of Applied Statistics*.

Gup, B. E. (1998), *Bank Failures in the Major Trading Countries of the World: Causes and Estimation*, Westport, CT: Quorum Books.

Gyftodimos, E. and Flach, P. A. (2002), "Hierarchical Bayesian Networks: A Probabilistic Reasoning Model for Structured Domains," in *Proceedings of the ICML-2002 Workshop on Development of Representations. University of New South Wales*, pp. 23–30.

Halaj, G. (2013), "Optimal Asset Structure of a Bank - Bank Reactions to Stressful Market Conditions," Working Paper Series 1533, European Central Bank, Frankfurt, Germany.

Hao, D., Ren, C., and Li, C. (2012), "Revisiting the Variation of Clustering Coefficient of Biological Networks Suggests New Modular Structure," *BMC Systems Biology*, 6, 34.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014), "Financial Network Systemic Risk Contributions," *Review of Finance*.

Heckerman, D. and Chickering, D. M. (1995), "Learning Bayesian Networks: The Combination of Knowledge and Statistical Data," in *Machine Learning*, pp. 20–197.

Heckerman, D. and Geiger, D. (1994), "Learning Gaussian Networks," in *Uncertainty in Artificial Intelligence*, pp. 235–243.

Hensman, J., Lawrence, N. D., and Rattray, M. (2013), "Hierarchical Bayesian Modelling of Gene Expression Time Series Across Irregularly Sampled Replicates and Clusters," *BMC Bioinformatics*, 14, 1–12.

Hoover, K. D. (2001), *Causality in Macroeconomics*, Cambridge University Press, Cambridge.

Huang, X., Zhou, H., and Zhu, H. (2012), "Systemic Risk Contributions," *Journal of financial services research*, 42, 55–83.

Idier, J., Lamé, G., and Mésonnier, J. S. (2013), "How Useful is the Marginal Expected Shortfall for the Measurement of Systemic Exposure? A Practical Assessment," Working papers 348, Banque de France, Paris, France.

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

Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., and West, M. (2005), "Experiments in Stochastic Computation for High-dimensional Graphical Models," *Statistical Science*, pp. 388–400.

Kadiyala, K. R. and Karlsson, S. (1993), "Forecasting with Generalized Bayesian Vector Autoregressions," *Journal of Forecasting*, 12, 365–378.

Kanno, M. (2013), "Credit Migration Forecasting and Correlation between Business and Credit Cycles," Technical report, Kanagawa University.

Karlsson, S. (2013), "Forecasting with Bayesian Vector Autoregressions," in *Handbook of Economic Forecasting, G. Elliott and A. Timmermann, (eds.)*, vol. 2, pp. 689–1324, Elsevier.

Kass, R., Tierney, L., and Kadane, J. (1988), "Asymptotics in Bayesian Computation," *Bayesian Statistics*, 3, 261–278.

Kass, R. E. and Raftery, A. E. (1995), "Bayes Factors," *Journal of the american statistical association*, 90, 773–795.

Kaufmann, S. and Schumacher, C. (2013), *Bayesian Estimation of Sparse Dynamic Factor Models with Order-Independent Identification*, Studienzentrum Gerzensee.

Kenny, G., Kostka, T., and Masera, F. (2013), "Can Macroeconomists Forecast Risk? Event-based Evidence from the Euro Area SPF," Working Paper Series 1540, European Central Bank.

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

King, R. G., Plosser, C. I., Stock, J. H., and Watson, M. W. (1991), "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819–840.

Klomp, J. and Haan, J. d. (2012), "Banking Risk and Regulation: Does one size fit all?" *Journal of Banking and Finance*, 36, 3197–3212.

Kock, A. B. and Callot, L. (2015), "Oracle Inequalities for High Dimensional Vector Autoregressions," *Journal of Econometrics, (Forthcoming)*.

Koller, D. and Friedman, N. (2009), *Probabilistic Graphical Models: Principles and Techniques*, MIT press.

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

Koop, G. and Korobilis, D. (2013), "Large Time-Varying Parameter VARs ," *Journal of Econometrics*, 177, 185 – 198.

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

Koopman, S. J., Lucas, A., and Schwaab, B. (2011), "Modeling Frailty-correlated Defaults Using Many Macroeconomic Covariates," *Journal of Econometrics*, 162, 312–325.

Korobilis, D. (2013), "Hierarchical Shrinkage Priors for Dynamic Regressions with Many Predictors," *International Journal of Forecasting*, 29, 43–59.

Kritzman, M., Li, Y., Page, S., and Rigobon, R. (2010), "Principal Components as a Measure of Systemic Risk," .

Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010), "Penalized Regression, Standard Errors, and Bayesian Lassos," *Bayesian Analysis*, 5, 369–411.

Lauritzen, S. L. (1996), *Graphical Models*, Oxford University Press, Oxford.

Lauritzen, S. L. and Wermuth, N. (1989), "Graphical Models for Associations Between Variables, Some of Which Are Qualitative and Some Quantitative," *Annals of Statistics*, 17, 31–57.

Lenkoski, A. and Dobra, A. (2011), "Computational Aspects Related to Inference in Gaussian Graphical Models with the G-Wishart Prior," *Journal of Computational and Graphical Statistics*, 20.

Lian, H., Feng, S., and Zhao, K. (2015), "Parametric and Semiparametric Reduced-rank Regression with Flexible Sparsity," *Journal of Multivariate Analysis*, 136, 163–174.

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.

Liu, Q. and Ihler, A. T. (2012), "Distributed Parameter Estimation via Pseudo-likelihood," in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pp. 1487–1494.

Liu, Y. and Willsky, A. (2013), "Learning Gaussian Graphical Models with Observed or Latent FVSs," in *Advances in Neural Information Processing Systems*, pp. 1833–1841.

Maclaurin, D. and Adams, R. P. (2014), "Firefly Monte Carlo: Exact MCMC with Subsets of Data," *arXiv preprint arXiv:1403.5693*.

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

Mare, D. S. (2012), "Contribution of Macroeconomic Factors to the Prediction of Small Bank Failures," *Working Paper, Social Science Research Network*.

Markowitz, H. (1952), "Portfolio Selection," *The Journal of Finance*, 7, 77–91.

McNees, S. (1986), "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts," *Journal of Business and Economic Statistics*, 4, 5–15.

Medeiros, M. C. and Mendes, E. (2012), "Estimating High-dimensional Time Series Models," *CREATES Research Paper*, 37.

Meinshausen, N. and Bühlmann, P. (2006), "High Dimensional Graphs and Variable Selection with the Lasso," *The Annals of Statistics*, pp. 1436–1462.

Meng, Z., Wei, D., and Wiesel, A. (2013), "Distributed Learning of Gaussian Graphical Models via Marginal Likelihoods," in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, pp. 39–47.

Merton, R. C. (1974), "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance*, 2, 449 – 471.

Merton, R. C., Billio, M., Getmansky, M., Gray, D., Lo, A. W., and Pelizzon, L. (2013), "On a New Approach for Analyzing and Managing Macrofinancial Risks," *Financial Analysts Journal*, 69.

Miyamura, M. and Kano, Y. (2006), "Robust Gaussian Graphical Modeling," *Journal of Multivariate Analysis*, 97, 1525–1550.

Moneta, A. (2008), "Graphical Causal Models and VARs: An Empirical Assessment of the Real Business Cycles Hypothesis," *Empirical Economics*, 35, 275–300.

Park, T. and Casella, G. (2008), "The Bayesian LASSO," *Journal of the American Statistical Association*, 103, 681–686.

Parkinson, M. (1980), "The Extreme Value Method for Estimating the Variance of the Rate of Return," *Journal of Business*, pp. 61–65.

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann publishers, San Mateo, CA, USA,.

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

Qin, D. (2011), "Rise of VAR Modelling Approach," *Journal of Economic Surveys*, 25, 156–174.

Ravasz, E. (2009), "Detecting Hierarchical Modularity in Biological Networks," in *Computational Systems Biology*, pp. 145–160, Springer.

Resti, A. and Sironi, A. (2007), *Risk Management and Shareholders' Value in Banking: from Risk Measurement Models to Capital Allocation Policies*, vol. 417, Wiley, London.

Roberts, G. and Sahu, S. (1997), "Updating Schemes, Covariance Structure, Blocking and Parametrisation for the Gibbs Sampler," *Journal of the Royal Statistical Society*, 59, 291 – 318.

Robinson, R. W. (1977), "Counting Unlabeled Acyclic Digraphs," in *Combinatorial mathematics V*, pp. 28–43, Springer.

Roll, R. (1977), "A Critique of the Asset Pricing Theory's tests Part I: On Past and Potential Testability of the Theory," *Journal of financial economics*, 4, 129–176.

Rose, P. S. and Kolari, J. W. (1985), "Early Warning Systems as a Monitoring Device for Bank Condition," *Quarterly Journal of Business and Economics*, 24, 43 – 60.

Ross, S. A. (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of economic theory*, 13, 341–360.

Roth, M. (1994), "Too Big to Fail and the Instability of the Banking System: Some Insights from Foreign Countries," *Business Economics*, 4, 43–49.

Roverato, A. (2002), "Hyper Inverse Wishart Distribution for Non-decomposable Graphs and its Application to Bayesian Inference for Gaussian Graphical Models," *Scandinavian Journal of Statistics*, 29, 391–411.

Rubio-Ramirez, J. F., Waggoner, D. F., and Zha, T. (2010), "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, 77, 665–696.

Santis, F. and Spezzaferri, F. (1999), "Methods for Default and Robust Bayesian Model Comparison: The Fractional Bayes Factor Approach," *International Statistical Review*, 67, 267–286.

Sargent, T. J. and Sims, C. A. (1977), "Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory," *New methods in business cycle research*, 1, 145–168.

Scott, J. G. and Berger, J. O. (2010), "Bayes and Empirical-Bayes Multiplicity Adjustment in the Variable-Selection Problem," *The Annals of Statistics*, 38, 2587–2619.

Scott, J. G. and Carvalho, C. M. (2008), "Feature-inclusion Stochastic Search for Gaussian Graphical Models," *Journal of Computational and Graphical Statistics*, 17.

Segoviano, M. A. and Goodhart, C. A. (2009), "Banking Stability Measures," IMF Working Papers 09/4, International Monetary Fund, Washington, DC.

Shaffer, J. P. (1995), "Multiple Hypothesis Testing," *Annual Review of Psychology*, 46, 561–584.

Sharpe, W. F. (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425–442.

Sharpe, W. F. (1992), "Asset Allocation: Management Style and Performance Measurement," *The Journal of Portfolio Management*, 18, 7–19.

Shenoy, C. and Shenoy, P. P. (2000), "Bayesian Network Models of Portfolio Risk and Return," *Computational Finance*, 1999, 87.

Shojaie, A. and Michailidis, G. (2009), "Analysis of Gene Sets Based on the Underlying Regulatory Network," *Journal of Computational Biology*, 16, 407–426.

Shojaie, A. and Michailidis, G. (2010), "Penalized Likelihood Methods for Estimation of Sparse High Dimensional Directed Acyclic Graphs," *Biometrika*, 97, 519–538.

Sims, C. A. (1976), "Exogeneity and Causal Ordering in Macroeconomic Models," .

Sims, C. A. (1980), "Macroeconomics and Reality," *Econometrica, Econometric Society,*, 48, 1–48.

Sims, C. A. (2010), "Lecture Notes: Causal Orderings and Exogeneity," .

Sinkey, J. F. (1975), "A Multivariate Statistical Analysis of the Characteristics of Problem Banks," *The Journal of Finance*, 30, 21–36.

Song, S. and Bickel, P. J. (2011), "Large Vector Auto Regressions," Tech. rep., Humboldt University, Collaborative Research Center 649.

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

Stock, J. H. and Watson, M. W. (2001), "Vector Autoregressions," *Journal of Economic Perspectives*, 15, 101–115.

Stock, J. H. and Watson, M. W. (2002), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American statistical Association*, 97, 1167–1179.

Stock, J. H. and Watson, M. W. (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)*, pp. 173 – 205, Oxford University Press: Oxford.

Stock, J. H. and Watson, M. W. (2012), "Generalized Shrinkage Methods for Forecasting using Many Predictors," *Journal of Business & Economic Statistics*, 30, 481–493.

Stock, J. H. and Watson, M. W. (2014), "Estimating Turning Points using Large Datasets," *Journal of Econometrics*, 178, 368–381.

Sun, H. and Li, H. (2012), "Robust Gaussian Graphical Modeling via l_1 Penalization," *Biometrics*, 68, 1197–1206.

Swanson, N. R. and Granger, C. W. J. (1997), "Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions," *Journal of the American Statistical Association*, 92, 357–367.

Tam, K. Y. and Kiang, M. Y. (1992), "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions," *Management Science*, 38, 926–947.

Tan, K. M., London, P., Mohan, K., Lee, S.-I., Fazel, M., and Witten, D. (2014), "Learning Graphical Models with Hubs," *The Journal of Machine Learning Research*, 15, 3297–3331.

Telesca, D., Müller, P., Kornblau, S. M., Suchard, M. A., and Ji, Y. (2012), "Modeling Protein Expression and Protein Signaling Pathways," *Journal of the American Statistical Association*, 107, 1372–1384.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the LASSO," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

Vasicek, O. A. (1984), "Credit Valuation," Tech. rep., KMV corporation.

Vazquez, F. and Federico, P. (2012), "Bank Funding Structures and Risk: Evidence from the Global Financial Crisis," *Working Paper, Social Science Research Network*.

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S. (2004), "Reversible Jump Markov Chain Monte Carlo Strategies for Bayesian Model Selection in Autoregressive Processes," *Journal of Time Series Analysis*, 25, 785–809.

Wang, H. (2010), "Sparse Seemingly Unrelated Regression Modelling: Applications in Econometrics and Finance," *Computational Statistics & Data Analysis*, 54, 2866–2877.

Wang, H. and Li, S. Z. (2012), "Efficient Gaussian Graphical Model Determination Under G-Wishart Prior Distributions," *Electronic Journal of Statistics*, 6, 168–198.

Wasserman, S. (1994), *Social Network Analysis: Methods and Applications*, vol. 8, Cambridge University press.

Wermuth, N. and Lauritzen, S. L. (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*.

Woodbury, M. A. (1950), "Inverting Modified Matrices," *Memorandum Report, 42, Statistical Research Group, Princeton University*.

Wright, L. (2007), "Bayesian Networks and Probabilistic Inference in Forensic Science," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170, 1187–1187.

Yuan, M. and Lin, Y. (2006), "Model Selection and Estimation in Regression with Grouped Variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 49–67.

Zhang, D., Wells, M. T., Turnbull, B. W., Sparrow, D., and Cassano, P. A. (2005), "Hierarchical Graphical Models," *Journal of the American Statistical Association*, 100, 719–727.

Zhang, Z., Wang, S., Liu, D., and Jordan, M. I. (2012), "EP-GIG Priors and Applications in Bayesian Sparse Learning," *The Journal of Machine Learning Research*, 13, 2031–2061.

Zhou, X. and Schmidler, S. C. (2009), "Bayesian Parameter Estimation in Ising and Potts Models: A Comparative StudyWith Applications to Protein Modeling," Tech. rep., Duke University.

Zou, C. and Feng, J. (2009), "Granger Causality vs. Dynamic Bayesian Network Inference: A Comparative Study," *BMC Bioinformatics*, 10, 122.

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

Appendix A

Technical Details of Chapter 3

A.1 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

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

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

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

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

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

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 (2002), 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

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

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 (Y^i) and its set of explanatory variables (π_i)

$$P(\mathcal{X}|G) \quad = \prod_{i=1}^{n_y} P(\mathcal{X}|G(Y^i, \pi_i)) = \prod_{i=1}^{n_y} \frac{P(\mathcal{X}^{(Y^i,\pi_i)}|G)}{P(\mathcal{X}^{(\pi_i)}|G)} \tag{A.3}$$

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

$$P(\mathcal{X}^D|G) = (\pi)^{\frac{-n_d T}{2}} \; \frac{|\overline{\Sigma}_{x,d}|^{-\frac{(\nu+T)}{2}}}{|\underline{\Sigma}_{x,d}|^{-\frac{\nu}{2}}} \; \prod_{i=1}^{n_d} \frac{\Gamma\left(\frac{\nu+T+1-i}{2}\right)}{\Gamma\left(\frac{\nu+1-i}{2}\right)} \tag{A.4}$$

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 (MP) and the normal-Wishart (NW) 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 Y^i. Under MP, $\overline{\sigma}_i = \sigma_i$, are diagonal elements of $\Sigma_u = \text{diag}(\sigma_1^2, \ldots, \sigma_{n_y}^2)$. Under NW, $\overline{\sigma}_i^2$, is the variance of residuals from the posterior of Σ_u, which is assumed to be inverse-Wishart distributed, $\Sigma_u \sim \mathcal{IW}(\underline{\nu}, \underline{S})$.

A.2 MCMC Sampling

Given a graph G, the algorithm samples a new graph G^* based on a proposal distribution. The new graph is accepted with probability

$$A(G^*|G) \quad = \min\left\{\frac{P(\mathcal{X}|G^*)}{P(\mathcal{X}|G)}\frac{P(G^*)}{P(G)}\frac{Q(G|G^*)}{Q(G^*|G)}, 1\right\} \tag{A.5}$$

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, based on 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 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

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 $G^{(t+1)} = G^{(t)}$

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 $G^{(t+1)}(\tilde{y}_i, \pi_i) = G^{(t)}(\tilde{y}_i, \pi_i)$

restriction. However, to reduce the computing time, we pre-compute $\overline{\Sigma}_x$, $\underline{\Sigma}_x$ and the metric in (A.4). 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.3) and (A.4) 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 $X'_{p+1-s}, \ldots, X'_{T-s}$, $1 \leqslant s \leqslant p$, such that X is of dimension

$np \times (T-p)$, $\mathcal{X}(i) = (X', (Y^i)')'$, $\mathcal{X}(i)$ is $(np+1) \times (T-p)$. Let G_+ be a stacked graph of the MAR relationships. Then marginal likelihood of the MAR structure is

$$P(\mathcal{X}|G_+) \quad = \prod_{i=1}^{n_y} \frac{P\big(\mathcal{X}(i)^{(Y^i,\pi_+^i)}\big|G_+\big)}{P\big(\mathcal{X}(i)^{(\pi_+^i)}\big|G_+\big)} \tag{A.6}$$

where π_+^i is the candidate explanatory variables of Y^i, consisting of the lagged variables in X_+. We score the MAR by evaluating (A.3) and (A.4), and by replacing G with G_+, and $\underline{\Sigma_x}$ and $\overline{\Sigma_x}$ with the prior and posterior covariance matrices of $\mathcal{X}(i)$, $i = 1, \ldots, n_y$.

A.3 Graphical Model Evaluation

A.3.1 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 $PSRF$ $(MPSRF) \leqslant 1.2$. 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.

A.3.2 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

$$q_{(1-\alpha)} = \hat{e}_{ij} - z_{(1-\alpha)}\sqrt{\frac{\hat{e}_{ij}(1-\hat{e}_{ij})}{n_{eff}}} \tag{A.7}$$

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

$$\hat{G}_{ij} = \begin{cases} 1 & \text{if} \quad q_{(1-\alpha)} > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{A.8}$$

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 (A.7), 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 A.1. TP indicates the number of real

Table A.1: Classification table for graph-predictive accuracy evaluation.

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

positives correctly predicted as positives, FP 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. We measure the graph inference accuracy by

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

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

$$BIC(\hat{G}) = -2\log(P(\mathcal{X}|\hat{G})) + |E|\log(T) \tag{A.9}$$

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.

A.3.3 Model Prediction Accuracy

We evaluate the estimated model based on the predictive AIC given by

$$AIC(\hat{M}) = -2\log(P(\mathcal{X}|\hat{M})) + 2|\hat{A}_M| \tag{A.10}$$

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.

A.3.4 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.

A.3.5 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

$$X_t^i = \sum_{s=1}^{p} \beta_{s,1} X_{t-s}^1 + \ldots + \sum_{s=1}^{p} \beta_{s,i} X_{t-s}^i + \ldots + \sum_{s=1}^{p} \beta_{s,j} X_{t-s}^j + \ldots + \sum_{s=1}^{p} \beta_{s,n} X_{t-s}^n + u_t^i \quad \text{(A.11)}$$

where $i, j \in \{1, \ldots, n\}$, $i \neq j$, and test whether $X_{t-s}^j \rightarrow X_t^i$, 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.

A.4 Macroeconomic and Financial Data Description

Table A.2 gives the data description and transformation code from Koop (2013) used for our macroeconomic application in Section 3.5. Table A.3 presents the description of the Euro Stoxx 600 super-sectors from Datastream for our financial application in Section 3.6.

Short ID	Mnemonic	Code	Description
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$)
C	GDP252	5	Real Personal Cons. Exp., Quantity Index
IP	IPS10	5	Industrial production index: total
U	LHUR	2	Unemp. rate: All workers, 16 and over (%)
I^*	GDPIC96	5	Real gross domestic private investment
MP	PSCCOMR	5	Real spot market price index: all commodities
NB	FMRNBA	3	Depository inst reserves: non-borrowed (mil$)
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 A.2: Data description and transformation code. 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.

No.	Name	ID	No.	Name	ID
1	Banks*	BK	11	Media	MD
2	Insurance companies*	IN	12	Travel & Leisure	TL
3	Financial Services*	FS	13	Chemicals	CH
4	Real Estates* $^+$	RE	14	Basic Resources	BR
5	Construction & Materials	CM	15	Oil & Gas	OG
6	Industrial goods & services	IGS	16	Telecommunication	TC
7	Automobiles & Parts	AP	17	Health Care	HC
8	Food & Beverage	FB	18	Technology	TG
9	Personal & Household Goods	PHG	19	Utilities	UT
10	Retail	RT			

Table A.3: Description of Euro Stoxx 600 super-sectors. * - The financial sector variables. ($^+$) from September 2008.

Appendix B

Technical Details of Chapter 4

B.1 Proofs

B.1.1 Proof of Proposition 1

Proof. Let X_t be $n \times 1$ vector of observations at time t, $Y_t \subseteq X_t$ a $n_y \times 1$ vector of dependent variables, W_t the stacked lags of X_t, where $W_t = (X'_{t-1}, \ldots, X'_{t-p})'$ is of $np \times 1$ dimension, with p as the maximum lag order. Suppose the joint distribution of $(Y'_t, W'_t) \sim \mathcal{N}(\mu, \Omega^{-1})$, where μ is the $((np + n_y) \times 1)$ vector of means and Ω^{-1} is $(np + n_y) \times (np + n_y)$ matrix of covariances. Without loss of generalization, we assume μ is a zero vector.

Suppose the marginal distribution of $W_t \sim \mathcal{N}(0, \Sigma_{ww})$ and the conditional distribution of $Y_t|W_t \sim \mathcal{N}(BW_t, \Sigma_\varepsilon)$, where B is $n_y \times np$ matrix of coefficients and Σ_ε is $n_y \times n_y$ covariance matrix of the errors. Then given Ω as the precision matrix of (Y_t, W_t), we can obtain $\Sigma = \Omega^{-1}$ which can be expressed as

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yw} \\ \Sigma_{wy} & \Sigma_{ww} \end{pmatrix} \tag{B.1}$$

where Σ_{wy} is $np \times n_y$ the covariances between W_t and Y_t, and Σ_{yy} is $n_y \times n_y$ covariances among Y_t. Then B and Σ_ε can be obtained from Σ by

$$B = \Sigma_{yw}\Sigma_{ww}^{-1}, \qquad \Sigma_\varepsilon = \Sigma_{yy} - \Sigma_{yw}\Sigma_{ww}^{-1}\Sigma_{wy} \tag{B.2}$$

Now given Σ as in (B.1), $\Omega = \Sigma^{-1}$ can be obtained as:

$$\Omega = \begin{pmatrix} (\Sigma_{yy} - \Sigma_{yw}\Sigma_{ww}^{-1}\Sigma_{wy})^{-1} & -(\Sigma_{yy} - \Sigma_{yw}\Sigma_{ww}^{-1}\Sigma_{wy})^{-1}\Sigma_{yw}\Sigma_{ww}^{-1} \\ -(\Sigma_{ww} - \Sigma_{wy}\Sigma_{yy}^{-1}\Sigma_{yw})^{-1}\Sigma_{wy}\Sigma_{yy}^{-1} & (\Sigma_{ww} - \Sigma_{wy}\Sigma_{yy}^{-1}\Sigma_{yw})^{-1} \end{pmatrix} \quad \text{(B.3)}$$

To complete the proof, we report the well-known Sherman-Morrison-Woodbury formula (see Woodbury, 1950). The inverse of a partitioned symmetric matrix is

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}\Big(A_{22} - A_{21}A_{11}^{-1}A_{12}\Big)^{-1}A_{21}A_{11}^{-1}$$

$$-(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} = -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \quad \text{(B.4)}$$

Following (B.4) and the expressions in (B.2), (B.3) can be simplified as

$$\Omega = \begin{pmatrix} \Sigma_{\varepsilon}^{-1} & -\Sigma_{\varepsilon}^{-1}B \\ -B'\Sigma_{\varepsilon}^{-1} & \Sigma_{ww}^{-1} + B'\Sigma_{\varepsilon}^{-1}B \end{pmatrix} \quad \text{(B.5)}$$

□

B.1.2 Proof of Proposition 2

Proof. From the prior distributions in (4.8) and (4.10), we marginalize out $\bar{\eta}_i$

$$P(\pi_i|p_i) = \frac{1}{2^{np_i}} \int_0^1 \mathbb{I}_{\{0,\dots,f_i\}}(|\pi_i|) \, \frac{1}{B(a,b)}(\bar{\eta}_i)^{a-1}(1-\bar{\eta}_i)^{b-1} \, d\bar{\eta}_i \quad \text{(B.6)}$$

where $f_i = \lfloor \bar{\eta}_i m_p \rfloor$ with $m_p = \min\{np_i, T - p_i\}$, $\mathbb{I}_{\{0,\dots,f_i\}}(|\pi_i|)$ is the indicator function

$$\mathbb{I}_{\{0,\dots,f_i\}}(|\pi_i|) = \begin{cases} \mathbb{I}_{\{0\}}(|\pi_i|), & 0 \leqslant \bar{\eta}_i < \frac{1}{m_p} \\ \quad \vdots & \cdots \\ \mathbb{I}_{\{0,\dots,m_p-1\}}(|\pi_i|), & \frac{m_p-1}{m_p} \leqslant \bar{\eta}_i < 1 \\ \mathbb{I}_{\{0,\dots,m_p\}}(|\pi_i|), & \bar{\eta}_i = 1 \end{cases} \quad \text{(B.7)}$$

Let $f(\bar{\eta}_i) = (B(a,b))^{-1}\ \bar{\eta}_i^{a-1}(1-\bar{\eta}_i)^{b-1}$. From (B.6)

$$P(\pi_i|p_i) = \frac{1}{2^{np_i}}\left[\mathbb{I}_{\{0\}}(|\pi_i|)\int_0^{\frac{1}{m_p}} f(\bar{\eta}_i)\ d\bar{\eta}_i + \ldots + \mathbb{I}_{\{0,\ldots,m_p-1\}}(|\pi_i|)\int_{\frac{m_p-1}{m_p}}^{1} f(\bar{\eta}_i)\ d\bar{\eta}_i\right]$$

$$= \frac{1}{2^{np_i}}\left[\sum_{j=0}^{m_p-1}\mathbb{I}_{\{0,\ldots,j\}}(|\pi_i|)\ \left(I_{\frac{j+1}{m_p}}(a,b) - I_{\frac{j}{m_p}}(a,b)\right)\right] \tag{B.8}$$

where $I_z(a,b) = \int_0^z f(\bar{\eta}_i)d\bar{\eta}_i$ is the incomplete beta function (Abramowitz and Stegun, 1964, p. 263). □

B.1.3 Proof of Corollary 3

Proof. By assuming a uniform prior on $\bar{\eta}_i$, $f(\bar{\eta}_i) = 1$. Furthermore, the difference between the incomplete beta functions in (B.8) is $I_{\frac{j+1}{m_p}}(a,b) - I_{\frac{j}{m_p}}(a,b) = \frac{1}{m_p}$. Thus

$$P(\pi_i|p_i) = \frac{1}{2^{np_i}}\ \frac{1}{m_p}\ \sum_{j=0}^{m_p-1}\mathbb{I}_{\{0,\ldots,j\}}(|\pi_i|)\ = \frac{1}{2^{np_i}}\ \left(1 - \frac{|\pi_i|}{m_p}\right) \tag{B.9}$$

□

B.1.4 Proof of Proposition 4

Proof. The function $\varphi(k)$ is convex if and only if $\varphi''(k) > 0$, $\forall k$. By defining $\varphi(k) = -\log P(\pi_i|p_i) = np_i\log(2) + \log(m_p) - \log(m_p - k)$, for $|\pi_i| = k$, it can be shown that

$$\varphi''(k) = \frac{1}{(m_p-k)^2} > 0 \tag{B.10}$$

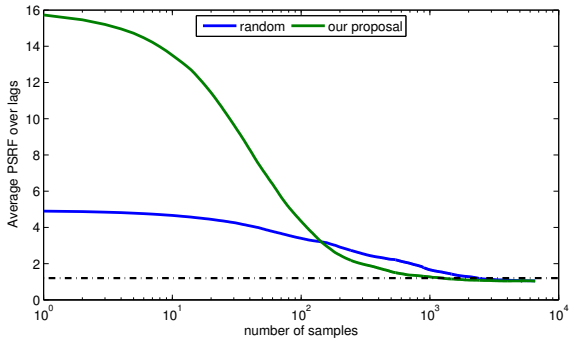
□

B.2 Convergence Diagnostics and Posterior Approximation

B.2.1 Convergence Diagnostics

For our graphical approach, we monitor the convergence of the MCMC chain using the potential scale reduction factor (PSRF), see Gelman and Rubin (1992). See also

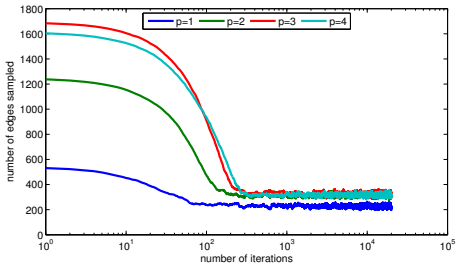
Casella and Robert (2004), ch. 12, for a review on methods for convergence monitoring in MCMC. The PSRF monitors the within-chain and between-chain covariances of the global log posterior densities of the sampled structures to test whether The chain is said to have properly converged if $PSRF \leqslant 1.2$. Figure B.1 display a comparison of the MCMC convergence diagnostics for a random initialization and our initialization procedure of the graph averaged over lags. Figures B.2 and B.3 shows plots of links and graph score at each MCMC iteration, the local graph BIC for the lags for the simulation experiments and the macroeconomic application respectively.



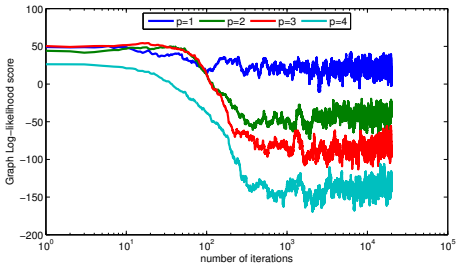
FIGUREB.1: Comparison of the MCMC convergence diagnostics for a random initialization (in blue) and our initialization (in green) procedure of the graph averaged over lags. The black dashed line is 1.2, and colored lines close to this line indicate convergence of the chain.

B.2.2 Edge Posterior Approximation

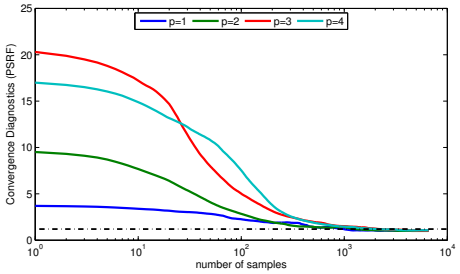
We estimate the posterior probability of the edge by $\hat{e}_{ij}$, which is the average of the MCMC samples from the G_{ij} posterior distribution. For variable selection purposes, we define the estimator G^*_{ij} of the edge from X^j to X^i based on a one sided posterior credibility interval for the edge posterior distribution, and find the interval lower bound $G^*_{ij} = 1$ if $\hat{e}_{ij} - z_{(1-\alpha)}\sqrt{\frac{\hat{e}_{ij}(1-\hat{e}_{ij})}{n_{eff}}} > 0.5$, where n_{eff} is the effective sample size 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.



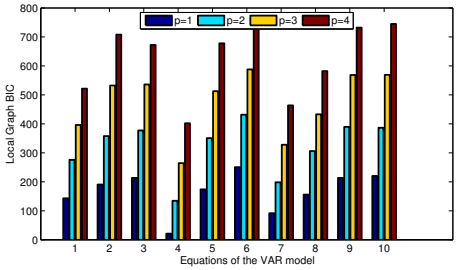
(a) Number of links and iterations



(b) Log score and iterations

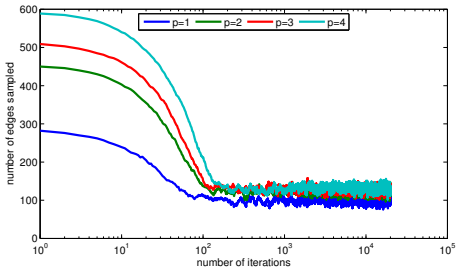


(c) Convergence diagnostics (PSRF)

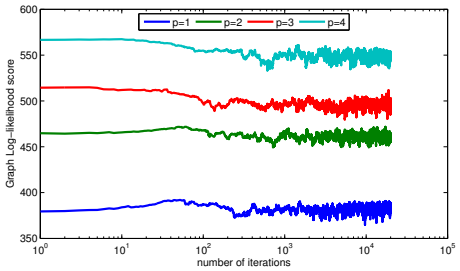


(d) Local graph BIC for the lags

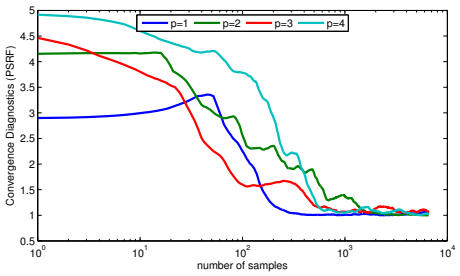
FIGUREB.2: Links (B.2a) and graph score (B.2b)at each MCMC iteration, with convergence diagnostics (B.2c) and local graph BIC (B.2d) for lags for each simulation experiment equation.



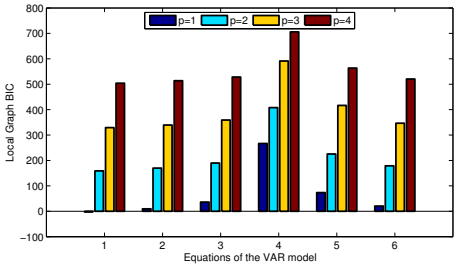
(a) Number of links and iterations



(b) Log score and iterations



(c) Convergence diagnostics (PSRF)



(d) Local graph BIC for the lags

FIGUREB.3: Links (B.3a) and graph score (B.3b) at each MCMC iteration, with convergence diagnostics (B.3c) and local graph BIC (B.3d) for lags for each macroeconomic equation.

B.3 Pseudo-Code for Sparse Graph Selection

Algorithm 3 presents a description of the pseudo-code for the sparse graph selection.

Algorithm 3 Graphical VAR Model Selection Algorithm

for $p \in [\underline{p}, \ldots, \bar{p}]$ **do**
  Initialize $\vec{G}_p$ as $(n_y \times np)$ zero matrix, set $m_p = \min\{np, T-p\}$
  $\mathbf{V}^i_{p,x}$ the vector of all possible explanatory variables up to lag p
  $\mathbf{V}_y$ the vector of dependent variables
  for each $y_i \in \mathbf{V}_y$ **do**
    for each $x_k \in \mathbf{V}^i_{p,x}$ **do**
      if x_k is equal to lag 1 of y_i **then**
        Set $\vec{G}_p(y_i, x_k) = 1$
      else
        Compute $H_0 = P(\mathcal{X}|p_i, \vec{G}_p(y_i, \varnothing))$ and $H_1 = P(\mathcal{X}|p_i, \vec{G}_p(y_i, \{x_k\}))$
        if $H_1 > H_0$ **then**
          Set $\vec{G}_p(y_i, x_k) = 1$ and retain x_k in $\mathbf{V}^i_{p,x}$
        else
          Set $\vec{G}_p(y_i, x_k) = 0$ and remove x_k from $\mathbf{V}^i_{p,x}$
  Set $N_p(\pi_i)$ the set of variables, $x'_k s$, retained in $\mathbf{V}^i_{p,x}$
  for $j = 1 \rightarrow J$, the total iterations **do**
    for each $y_i \in \mathbf{V}_y$ **do**
      Set $\pi_i^{(j-1)}$ = the set explanatory variables of y_i in $\vec{G}^{(j-1)}_{p,i}$
      Draw $\bar{\eta}_i^{(*)}$ from a $\mathcal{B}e(a,b)$ and set $f_i^{(*)} = \lfloor m_p \bar{\eta}_i^{(*)} \rfloor$
      if $|\pi_i^{(j-1)}| < f_i^{(*)}$ **then**
        Set $Q(\vec{G}^{(*)}_{p,i} | \vec{G}^{(j-1)}_{p,i}, \bar{\eta}_i^{(*)}) = 1/|N_p(\pi_i)|$, Draw $x_k \in N_p(\pi_i)$
        Add/remove edge; i.e. $\vec{G}_p^{(*)}(y_i, x_k) = 1 - \vec{G}_p^{(j-1)}(y_i, x_k)$
      else
        Set $Q(\vec{G}^{(*)}_{p,i} | \vec{G}^{(j-1)}_{p,i}, \bar{\eta}_i^{(*)}) = 1/|\pi_i^{(j-1)}|$, Draw $x_k \in \pi_i^{(j-1)}$
        Remove edge; i.e. $\vec{G}_p^{(*)}(y_i, x_k) = 0$
      Set $\pi_i^{(*)}$ = the set explanatory variables of y_i in $\vec{G}^{(*)}_{p,i}$
      Draw $\bar{\eta}_i^{(**)}$ from a $\mathcal{B}e(a,b)$ and set $f_i^{(**)} = \lfloor m_p \bar{\eta}_i^{(**)} \rfloor$
      if $|\pi_i^{(*)}| < f_i^{(**)}$ **then**
        Set $Q(\vec{G}^{(j-1)}_{p,i} | \vec{G}^{(*)}_{p,i}, \bar{\eta}_i^{(**)}) = 1/|N_p(\pi_i)|$
      else
        Set $Q(\vec{G}^{(j-1)}_{p,i} | \vec{G}^{(*)}_{p,i}, \bar{\eta}_i^{(**)}) = 1/|\pi_i^{(*)}|$
      Sample $u \sim \mathcal{U}_{[0,1]}$ and Compute R_A following equation (4.18)
      if $u < \min\{1, R_A\}$ **then** $\vec{G}^{(j)}_{p,i} = \vec{G}^{(*)}_{p,i}$
      else $\vec{G}^{(j)}_{p,i} = \vec{G}^{(j-1)}_{p,i}$

B.4 Data Description - Large Macroeconomic Application

Table B.1 provides a description and stationarity transformation codes used for our large macroeconomic application in Section 4.6.

No.	Mnemonic	Description	Tcode
1	CPIAUCSL*	Consumer Price Index for All Urban Consumers: All Items	6
2	FEDFUNDS*	Effective Federal Funds Rate	2
3	GDPC96*	Real Gross Domestic Product, 3 Decimal	5
4	GPDIC96*	Real Gross Private Domestic Investment, 3 decimal	5
5	INDPRO*	Industrial Production Index	5
6	PCECC96*	Real Personal Consumption Expenditures	5
7	AAA	Moody's Seasoned Aaa Corporate Bond Yield	2
8	AHECONS	Ave Hr Earnings Of Prod & Nonsupervisory Employees: Construction	6
9	AHEMAN	Ave Hr Earnings of Prod & Nonsupervisory Empl: Manufacturing	6
10	AWHMAN	Ave Wkly Hr of Prod & Nonsupervisory Empl: Manufacturing	5
11	AWOTMAN	Ave Wkly Overtime Hrs of Prod & Nonsup. Empl: Manufacturing	5
12	BAA	Moody's Seasoned Baa Corporate Bond Yield	2
13	BORROW	Total Borrowings of Depository Institutions from the Federal Reserve	6
14	BUSLOANS	Commercial and Industrial Loans, All Commercial Banks	6
15	CBIC96	Real Change in Private Inventories	1
16	CCFC	Corporate: Consumption of Fixed Capital	6
17	CIVPART	Civilian Labor Force Participation Rate	5
18	CONSUMER	Consumer Loans at All Commercial Banks	5
19	CP	Corporate Profits After Tax (without IVA and CCAdj)	6
20	CPIAPPSL	Consumer Price Index for All Urban Consumers: Apparel	6
21	CPIENGSL	Consumer Price Index for All Urban Consumers: Energy	6
22	CPILEGSL	Consumer Price Index for All Urban Consumers: All Items Less Energy	6
23	CPIMEDSL	Consumer Price Index for All Urban Consumers: Medical Care	6
24	CPITRNSL	Consumer Price Index for All Urban Consumers: Transportation	6
25	CPIUFDSL	Consumer Price Index for All Urban Consumers: Food	6
26	CPIULFSL	Consumer Price Index for All Urban Consumers: All Items Less Food	6
27	CURRCIR	Currency in Circulation	6
28	CURRSL	Currency Component of M1	6
29	DEMDEPSL	Demand Deposits at Commercial Banks	6
30	DIVIDEND	Corporate Profits after tax with IVA and CCAdj: Net Dividends	6
31	DMANEMP	All Employees: Durable goods	5
32	DPIC96	Real Disposable Personal Income	6
33	EMRATIO	Civilian Employment-Population Ratio	5
34	EXCAUS	Canada / U.S. Foreign Exchange Rate	5
35	EXJPUS	Japan / U.S. Foreign Exchange Rate	5
36	EXPGSC96	Real Exports of Goods & Services, 3 Decimal	5
37	EXSZUS	Switzerland / U.S. Foreign Exchange Rate	5
38	EXUSUK	U.S. / U.K. Foreign Exchange Rate	5
39	FINSLC96	Real Final Sales of Domestic Product	5
40	GCEC96	Real Government Consumption Expenditures & Gross Investment	5
41	GDPDEF	Gross Domestic Product: Implicit Price Deflator	5
42	GPDICTPI	Gross Private Domestic Investment: Chain-type Price Index	6
43	GS1	1-Year Treasury Constant Maturity Rate	2
44	GS10	10-Year Treasury Constant Maturity Rate	2
45	GS3	3-Year Treasury Constant Maturity Rate	2
46	GS5	5-Year Treasury Constant Maturity Rate	2
47	GSAVE	Gross Saving	5
48	HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	4
49	HOUST1F	Privately Owned Housing Starts: 1-Unit Structures	4
50	HOUST5F	Privately Owned Housing Starts: 5-Unit Structures or More	4

Table B.1: Variables and transformation codes, 1 = no transformation, 2 = first difference, 4 = log, 5 = 100×(first difference of log), 6 = 100×(second difference of log). *- The dependent variables.

No.	Mnemonic	Description	Tcode
51	HOUSTMW	Housing Starts in Midwest Census Region	4
52	HOUSTNE	Housing Starts in Northeast Census Region	4
53	HOUSTS	Housing Starts in South Census Region	4
54	HOUSTW	Housing Starts in West Census Region	4
55	IMPGSC96	Real Imports of Goods & Services, 3 Decimal	5
56	INVEST	Securities in Bank Credit at All Commercial Banks	5
57	IPBUSEQ	Industrial Production: Business Equipment	5
58	IPCONGD	Industrial Production: Consumer Goods	5
59	IPDCONGD	Industrial Production: Durable Consumer Goods	5
60	IPDMAT	Industrial Production: Durable Materials	5
61	IPFINAL	Industrial Production: Final Products (Market Group)	5
62	IPMAT	Industrial Production: Materials	5
63	IPNCONGD	Industrial Production: Nondurable Consumer Goods	5
64	IPNMAT	Industrial Production: nondurable Materials	5
65	LOANS	Loans and Leases in Bank Credit, All Commercial Banks	6
66	M1SL	M1 Money Stock	6
67	M1V	Velocity of M1 Money Stock	5
68	M2SL	M2 Money Stock	6
69	M2V	Velocity of M2 Money Stock	5
70	MCUMFN	Capacity Utilization: Manufacturing (NAICS)	1
71	MPRIME	Bank Prime Loan Rate	2
72	MZMSL	MZM Money Stock	6
73	NAPM	ISM Manufacturing: PMI Composite Index	1
74	NAPMEI	ISM Manufacturing: Employment Index	1
75	NAPMII	ISM Manufacturing: Inventories Index	1
76	NAPMNOI	ISM Manufacturing: New Orders Index	1
77	NAPMPI	ISM Manufacturing: Production Index	1
78	NAPMPRI	ISM Manufacturing: Prices Index	1
79	NAPMSDI	ISM Manufacturing: Supplier Deliveries Index	1
80	NDMANEMP	All Employees: Nondurable goods	5
81	NONREVSL	Total Nonrevolving Credit Owned and Securitized, Outstanding	6
82	OPHPBS	Business Sector: Real Output Per Hour of All Persons	5
83	PAYEMS	All Employees: Total nonfarm	5
84	PCDG	Personal Consumption Expenditures: Durable Goods	5
85	PCECTPI	Personal Consumption Expenditures: Chain-type Price Index	5
86	PCESV	Personal Consumption Expenditures: Services	5
87	PCND	Personal Consumption Expenditures: Nondurable Goods	5
88	PFCGEF	Producer Price Index: Finished Consumer Goods Excluding Foods	6
89	PINCOME	Personal Income	6
90	PNFI	Private Nonresidential Fixed Investment	6
91	PPIACO	Producer Price Index: All Commodities	6
92	PPICPE	Producer Price Index: Finished Goods: Capital Equipment	6
93	PPICRM	Producer Price Index: Crude Materials for Further Processing	6
94	PPIFCF	Producer Price Index: Finished Consumer Foods	6
95	PPIFCG	Producer Price Index: Finished Consumer Goods	6
96	PPIFGS	Producer Price Index: Finished Goods	6
97	PPIITM	Producer Price Index: Intermediate Materials: Supplies & Components	6
98	PRFI	Private Residential Fixed Investment	6
99	PSAVE	Personal Saving	5
100	REALLN	Real Estate Loans, All Commercial Banks	6
101	SAVINGSL	Savings Deposits - Total	6
102	SLEXPND	State & Local Government Current Expenditures	6
103	SLINV	State & Local Government Gross Investment	6
104	SP500	S&P 500	5
105	SRVPRD	All Employees: Service-Providing Industries	5
106	TB3MS	3-Month Treasury Bill: Secondary Market Rate	2
107	TB6MS	6-Month Treasury Bill: Secondary Market Rate	2
108	TCDSL	Total Checkable Deposits	6
109	TOTALSL	Total Consumer Credit Owned and Securitized, Outstanding	6
110	TVCKSSL	Travelers Checks Outstanding	6
111	UEMP15T26	Number of Civilians Unemployed for 15 to 26 Weeks	5
112	UEMP27OV	Number of Civilians Unemployed for 27 Weeks and Over	5

No.	Mnemonic	Description	Tcode
113	UEMP5TO14	Number of Civilians Unemployed for 5 to 14 Weeks	5
114	UEMPLT5	Number of Civilians Unemployed - Less Than 5 Weeks	5
115	ULCNFB	Nonfarm Business Sector: Unit Labor Cost	5
116	UNRATE	Civilian Unemployment Rate	2
117	USCONS	All Employees: Construction	5
118	USEHS	All Employees: Education & Health Services	5
119	USFIRE	All Employees: Financial Activities	5
120	USGOOD	All Employees: Goods-Producing Industries	5
121	USGOVT	All Employees: Government	5
122	USINFO	All Employees: Information Services	5
123	USLAH	All Employees: Leisure & Hospitality	5
124	USMINE	All Employees: Mining and logging	5
125	USPBS	All Employees: Professional & Business Services	5
126	USPRIV	All Employees: Total Private Industries	5
127	USTPU	All Employees: Trade, Transportation & Utilities	5
128	USTRADE	All Employees: Retail Trade	5
129	USWTRADE	All Employees: Wholesale Trade	5
130	WASCUR	Compensation of Employees: Wages & Salary Accruals	6

Appendix C

Technical Details of Chapter 5

C.1 Data Description For European Banks Risk Network

Table C.1 provides a description of the dataset used for our analysis European banks risk network in Section 5.5.

C.2 Computational Details and Convergence Diagnostics

Following the idea by Wang and Li (2012), we implement the algorithm such that steps 1 and 2 can be combined and step 3 and 4 as well. A pseudo presentation of the algorithm is shown in Algorithm (4):

Algorithm 4 Sampling Network

Initialize T_0 to be identity matrix
Compute $T_n = T_0 + S_n$
Initial with a fully connected graph (upper triangle) G
for $t = 1$ to Total iterations **do**
    for each variable x_i **do**
        extract the clique x_{iC} centered around x_i
        for each variable $x_j \in x_{iC} \backslash x_i$ **do**
            extract the clique $x_{iC\backslash j} = x_{iC} \backslash x_j$
            Compute log Bayes factor, $\log(BF) = \log(P(x_{iC\backslash j})) - \log(P(x_{iC}))$
            if $\log(BF) > \tau$ **then**
                $x_j \notin x_{iC}$, set $G_{ij} = 0$ and update $T_{0,ii}$
            else
                $x_j \in x_{iC}$, set $G_{ij} = 1$ and update $T_{0,ii}$ and $T_{0,ij}$
            Update T_n following Carvalho et al. (2007)

As discussed in Section 5.4, we choose $\tau = 1$ as a threshold value for the Bayes

Bank Name	Code	Country	T.Assets	SIFI	YOY	Stdev
Deutsche Bank	DBK	Germany	2012329	G	3.96	4.00
BNP Paribas	BNP	France	1907290	G	26.94	3.59
Credit Agricole	ACA	France	1842361	G	37.42	4.39
Banco Santander	SAN	Spain	1269628	G	10.90	3.60
Societe Generale	GLE	France	1250696	G	44.17	4.96
ING Groep	INGA	Netherland	1168632	G	24.86	4.08
UniCredit	UCG	Italy	926827	G	41.21	4.99
Intesa Sanpaolo	ISP	Italy	673472	D	30.72	4.45
Banco Bilbao Vizcaya Argentaria	BBVA	Spain	637785	G	32.08	3.82
Commerzbank	CBK	Germany	635878	G	-48.65	7.28
Natixis-BPCE Group	KN	France	528370	G	37.39	4.74
Dexia	DEXB	Belgium	357210	G	-133.50	22.46
Caixabank	CABK	Spain	348294	-	25.86	5.08
Bankia	BKIA	Spain	282310	-	-469.33	32.13
KBC Group	KBC	Belgium	256886	D	72.92	4.16
CIC Credit Mutuel Group	CC	France	235732	-	33.69	2.02
Banca Monte dei Paschi di Siena	BMPS	Italy	218882	-	2.39	6.29
Erste Group Bank	EBS	Austria	213824	D	24.34	4.52
Banco de Sabadell	SAB	Spain	161547	-	-3.57	6.21
Banco Popular Espanol	POP	Spain	157618	-	-44.69	12.19
Bank of Ireland	BIR	Ireland	148146	D	91.20	5.80
Raiffeisen Bank International	RBI	Austria	136116	D	-25.59	3.91
Unione di Banche Italiane	UBI	Italy	132433	-	38.72	5.30
Banco Popolare	BP	Italy	131921	-	7.77	6.13
Allied Irish Banks	AIB	Ireland	122516	D	53.60	8.68
National Bank of Greece	ETE	Greece	104798	D	-186.01	25.86
Banco Comercial Portugues	BCP	Portugal	89744	D	25.95	5.92
Banco Espirito Santo	BES	Portugal	83690	-	37.05	7.37
Wustenrot & Wurttembergische	WUW	Germany	77192	-	-4.68	1.89
Mediobanca	MB	Italy	72841	-	29.96	5.55
Piraeus Bank	TPEIR	Greece	70406	D	-146.26	21.33
Eurobank Ergasias	EUROB	Greece	67653	D	-7.49	19.93
Banca popolare Emilia Romagna	BPE	Italy	61637	-	30.46	5.77
Alpha Bank	ALPHA	Greece	58357	D	-121.81	18.04
Bankinter	BKT	Spain	58165	-	83.17	5.68
Banca Popolare di Milano	PMI	Italy	52475	-	10.74	6.06
Banca Carige	CRG	Italy	49325	-	-18.23	7.07
Aareal Bank	ARL	Germany	45734	-	37.33	4.34
Pohjola Bank Oyj	POH1S	Finland	44623	-	14.87	3.38
Banco BPI	BPI	Portugal	44564	D	22.54	6.02
Bank of Cyprus	CPB	Cyprus	33762	-	-60.01	10.77
IKB Deutsche Industriebank	IKB	Germany	31593	-	59.85	8.36
Banca Popolare di Sondrio	BPSO	Italy	32349	-	-3.35	3.74
Credito Emiliano	CE	Italy	30748	-	30.57	4.50
Credito Valtellinese	CVAL	Italy	29896	-	-11.50	5.68

Table C.1: Banks listed and supervised by the European Central Bank from October 2012, to October 2013. G: Global Systemically Important, D: Domestically Important, T.Asset: Total Assets, YOY: Year on Year Returns, Stdev: standard deviations

factor. We run 2200 Gibbs iterations with 200 as burn-in. Our MCMC simulations are implemented in Matlab. We have verified posterior convergence of the algorithm by diagnosing the complexity of the network.

We monitor convergence of the chain using potential scale reduction factor of Gelman and Rubin (1992). The approach divides the sampled chains into three

parts to compute within-chain and between-chain covariances to test whether all the chains converge to the same posterior distribution. The chain is said to have properly converged if $PSRF \leqslant 1.2$. The results from our convergence diagnostics, indicates that all our simulations converged with a PSRF smaller than 1.02.

C.3 Sensitivity Analysis

For sensitivity analysis purposes, we have run our computational algorithm with different log Bayes factor thresholds, $\tau = 0$ and $\tau = 2$. Table C.2 shows the result of the total links of both hierarchical and non-hierarchical networks for the different Bayes factor thresholds.

Type	Network	Baye's Factor Threshold and Total Links		
		$\tau = 0$	$\tau = 1$	$\tau = 2$
Non-Hierarchical	Over-all Inter Banks	611	711	867
	Within-Country	85	96	106
	Inter-Bank Across Country	526	615	761
	SIFI	44	45	51
Hierarchical	Over-all Inter Banks	59	95	343
	Within-Country	41	58	93
	Inter-Bank Across Country	18	37	250
	SIFI	8	11	28
	Inter-Countries	12	19	46

Table C.2: Sensitivity Analysis with different log Bayes factor thresholds

Our results, summarized in Table C.2 show that the results do not differ considerably from previous ones but the corresponding networks are, respectively, less or more interconnected. For example, in terms of the between country network, the estimated present links drop from 19 to 12 for $\tau = 0$, and increase to 46 for $\tau = 2$. In terms of specific relationships, we remark that the substance of what we found and discussed for the threshold of $\tau = 1$, chosen according to what presented in the Simulation section, remains stable with other choices of the threshold.