The Econometrics of Bayesian Graphical Models: A Review With Financial Application

Daniel Felix Ahelegbey

Boston University - Department of Mathematics and Statistics

28 May 2015

Journal of Network Theory in Finance **2**(2), 1–33

Research Paper

The econometrics of Bayesian graphical models: a review with financial application

Daniel Felix Ahelegbey

Department of Mathematics and Statistics, Boston University, 111 Cummington Mall, Boston, MA 02115, USA; email: dfkahey@bu.edu

(Received February 18, 2016; revised April 25, 2016; accepted May 16, 2016)

ABSTRACT

Recent advances in empirical finance have shown that the adoption of network theory is critical in order to understand contagion and systemic vulnerabilities. While interdependencies among financial markets have been widely examined, only a few studies review networks, and they do not focus on the econometrics aspects. This paper presents a state-of-the-art review on the interface between statistics and econometrics in the inference and application of Bayesian graphical models. We specifically highlight the connections and possible applications of network models in financial econometrics in the context of systemic risk.

Keywords: Bayesian inference; graphical models; model selection; systemic risk; financial crisis.

1 INTRODUCTION

The global financial crisis has prompted new research interests to understand the structure of the financial system and risk propagation channels. In order to provide a framework for strengthening financial stability, policymakers are currently not only refining the regulatory and institutional setup, but also looking for analytical tools to identify, monitor and address systemic risk better. According to Bernanke (2010) and

Print ISSN 2055-7795 | Online ISSN 2055-7809

the Financial Crisis Inquiry Commission (2011), the global financial crisis was triggered by losses suffered by holders of subprime mortgages, and amplified by financial system vulnerabilities. They pointed out that in the absence of the systemic vulnerabilities, the triggers might produce sizable losses to certain firms, investors or asset classes, but generally they would not lead to a global recession. Many authors have found the complex connections in financial markets to have been the vulnerabilities that magnified the initial shocks of the crisis (see, for example, Acemoglu *et al* 2015; Battiston *et al* 2012; Billio *et al* 2012; Diebold and Yilmaz 2014; Gai *et al* 2011).

Network analysis has proven to be a promising tool for understanding the systemic vulnerabilities. It studies the interconnectedness among financial markets and the macroeconomy. The novelty of network analyses in this domain is crucial to model the complex connections in the financial system through information from network structure, density, homophily and centralities (Jackson 2014). While interconnectedness and contagion have been widely studied, only a few studies review networks, and they do not focus on the econometrics aspects (see Dornbusch *et al* (2000), Pericoli and Sbracia (2003) and Dungey *et al* (2005) for a review of financial contagion models). This study contributes to the literature on network econometrics. In particular, it presents a state-of-the-art review on the inference and possible applications of networks in financial econometrics in the context of systemic risk.

Many techniques for network modeling have been developed in statistics and social network literature. The common class of models includes the following: exponential random graph models (Frank and Strauss 1986; Holland and Leinhardt 1981), stochastic block models (Nowicki and Snijders 2001; Wang and Wong 1987) and latent space models (Handcock *et al* 2007; Hoff *et al* 2002). For a review of statistical models for social networks, see Kolaczyk (2009), Goldenberg *et al* (2010), Snijders (2011) and Kolaczyk and Csárdi (2014).

In most applications of social networks, the network is assumed to be known and is considered as the observed data. However, in the systemic risk literature, the role of interconnectedness in the risk-propagation process crucially depends on the network structure, which is generally unknown. Much of the earlier work on contagion has focused on interconnectedness arising from actual exposures among institutions, based on either balance sheet information or other financial market data. There is relatively little empirical work on the former, largely because of problems of balance sheet data accessibility. However, several studies have focused on the latter in order to understand sources of contagion and spillovers (Ahelegbey *et al* 2016; Barigozzi and Brownlees 2016; Billio *et al* 2012; Diebold and Yilmaz 2014). Although market data is readily available these days, inferring networks from observed data is characterized by uncertainty and highly complex structures.

Relationships among many real-world phenomenons are often more complex than pairwise. It is well known that interring networks from data is a model determination problem in which the number of candidates increases super-exponentially with the number of variables (Chickering *et al* 2004). Therefore, applying standard techniques, such as Granger causality (Granger 1969), to identify a single model from this set of candidates often ignores the problem of model uncertainty. Also, such techniques are unable to distinguish between direct and mediated causes. Here, we present an overview of the Bayesian approach to network identification. This method takes into account network uncertainty by allowing us to incorporate prior information, where necessary, and perform model averaging (see Heckerman *et al* 1995). The approach discussed in this paper is closely related to the literature on Gaussian graphical models for time series (Carvalho and West 2007; Carvalho *et al* 2007; Dahlhaus and Eichler 2003). It is also related to Eichler (2007) and Zou and Feng (2009), who present the network techniques as a valid alternative to the Granger concept for causal identification and its extensions in the econometrics literature (Diks and Panchenko 2005; Hoover 2001).

We demonstrate the effectiveness of the Bayesian method in identifying interconnectedness among both the major financial sectors in the US, using the monthly returns indexes of Billio *et al* (2012), and the daily volatilities of the super-sectors of the European stock market. The results on the returns network corroborate the findings of Billio *et al* (2012), with evidence of higher connectedness between 2001 and 2008 and insurance companies being central to the spread of risk in the US financial market during the subperiod leading to the global financial crisis. The volatility network shows that banks and insurance companies are central to the spread of the "fear connectedness" (Diebold and Yilmaz 2014) expressed by market participants in the financial sector of the euro area.

This paper proceeds as follows. We review network applications from a statistical perspective in Section 2 and the literature on financial networks for systemic risk in Section 3. In Section 4, we relate network models to multivariate analysis and present possible applications in financial econometrics. We then discuss the Bayesian network inference in Section 5. In Section 6, we illustrate the effectiveness of the Bayesian network inference in analyzing the return and volatility connectedness of financial time series.

2 A REVIEW OF GRAPHICAL MODELS

Graphical modeling is a class of multivariate analysis that uses graphs to represent statistical models. They are represented by $(G, \theta) \in (\mathcal{G} \times \Theta)$, where G is a network, θ is the model parameters, $\mathcal{G}$ is the space of graphs and Θ is the parameter space. A graph $G = (V, E)$ is defined in terms of a set V of vertices or nodes (variables) joined by a set E of edges or links (interactions). We introduce the essential concepts of graphical models and review the developments in statistical inference and application of Bayesian network models.

2.1 Basic terminologies

A graph with undirected edge interactions between variables is an "undirected graph" (or "Markov network"). These graphs produce a class of models commonly known as undirected graphical models, which are more suitable for analyzing similarity and correlated behaviors among variables (see Koller and Friedman 2009; Meinshausen and Bühlmann 2006; Wainwright and Jordan 2008).

A graph with directed edge interactions between variables 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 $\rightarrow$ B $\rightarrow$ 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 $\rightarrow$ A is illegal, since a descendant cannot be his or her own ancestor. This type of ordering is suitable for expressing causal relationships. These graphs produce a class of models referred to as Bayesian networks (see, for example, Ghahramani 1998; Heckerman *et al* 1995; Neapolitan 2004).

A partially directed acyclic graph (or chain graph) is a type of DAG that allows bidirected edges. These graphs are suitable for applications in which 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 the conditional independence of A and C, given B (see Andersson *et al* 1997; Gillispie and Perlman 2001; Pearl 2000).

A bipartite graph is an undirected graph in which variables 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 graph belongs to the class of color graphs, in which 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 with nodes categorized into variables and factors, which are represented by different shapes. Variable nodes are often represented by circles, and factor nodes are often represented by squares (see Asratian 1998; Kschischang *et al* 2001; Zha *et al* 2001).

A weighted graph is one that has a numeric value (weights) associated with each edge. A graph is complete if all vertices are connected. A clique is a subset of vertices that are completely connected. Let V be the set of vertices and $V_{\mathrm{A}} \subseteq V$; then, G_{A} is defined as a subgraph on nodes in V_{A}. The triple $(V_{\mathrm{A}}, V_{\mathrm{B}}, V_{\mathrm{C}}) \subseteq V$ forms a decomposition of a graph G if $V = V_{\mathrm{A}} \cup V_{\mathrm{B}} \cup V_{\mathrm{C}}$ and $V_{\mathrm{C}} = V_{\mathrm{A}} \cap V_{\mathrm{B}}$, 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 $V_A \neq \emptyset$ and $V_B \neq \emptyset$. 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 (see Giudici and Green 1999; Koller and Friedman 2009; Wainwright and Jordan 2008).

2.2 Statistical inference

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

2.2.1 Constraint-based approach

The constraint-based approach to graph selection involves the use of statistical tests to identify conditional independence relationships among variables. The outcomes of these tests are used to constrain the graph selection process to estimate the most plausible graph that is consistent with those constraints obtained. The most widely applied constraint-based inference is the PC algorithm, based on Fisher's z-transform (see Spirtes *et al* 2000; Verma and Pearl 1991).

2.2.2 Score-based approach

Score-based approaches are 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 that minimizes a penalized likelihood score. Another aspect of this approach is Bayesian in nature: it usually involves priors and posterior computations, taking advantage of model averaging to address the model uncertainty problem. Examples of such algorithms are greedy search, simulated annealing and Markov chain Monte Carlo (MCMC) (Friedman and Koller 2003; Giudici and Green 1999; Madigan and York 1995).

2.2.3 Hybrid approach

The hybrid approach combines techniques from the constraint-based and score-based inferences for graph selection. Methods based on this approach are designed to adopt constraint-based reasoning as an initial step to restrict the search space for the application of the score-based scheme. An example of this is the Max-Min Hill-Climbing algorithm (Tsamardinos *et al* 2006). Regularization methods such as the least absolute shrinkage and selection operator (LASSO) and its variants are also hybrid methods (Banerjee *et al* 2008; Friedman *et al* 2008; Meinshausen and Bühlmann 2006; Tibshirani 1996).

2.3 Applications and developments in Bayesian network models

Graphical models have contributed to modeling challenging and complex real-world phenomenons in several fields. They have been applied in forensic science as tools to aid in reasoning under uncertainty. For example, Bayesian networks have been identified as a suitable tool for analyzing evidence in complex legal and criminal cases (Dawid 2003; Wright 2007).

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 our understanding of biological processes. Several researchers have applied graphical models to analyze gene interactions in detecting conditional dependencies (Friedman *et al* 2000; Hensman *et al* 2013). Several authors studied change points and time-varying interactions (see Grzegorczyk *et al* 2011; Lebre *et al* 2010; Robinson and Hartemink 2009).

The observation that gene data is typically characterized by heavy tails or outliers has motivated research on high-dimensional gene expressions by relaxing the assumption of normality. The active research in this area focuses on non-normality and outliers (Miyamura and Kano 2006; Vogel and Fried 2011) and non-paranormal distributions (nonparametric or semi-parametric models) to allow for mixed data (binary, ordinal or continuous) (see Teramoto *et al* 2014; Zhao *et al* 2012). Many studies have considered applications in which conditional distributions assume different probability models, such as Bernoulli, multinomial, Poisson and exponential families (Höfling and Tibshirani 2009; Ravikumar *et al* 2010).

Studies in biological networks have revealed the complex hierarchical structures of cellular processes, which pose a challenge to researchers. An active focus in this area is designing algorithms to detect hierarchical modularity (Hao *et al* 2012; Ravasz 2009), latent variables (Choi *et al* 2011; Liu and Willsky 2013) and hubs (Akavia *et al* 2010; Tan *et al* 2014).

Graphical models have been applied in areas such as image processing, speech and handwriting recognition, which often exhibit regularities, even though they are characterized by uncertainty. Such models have been applied to acquire a structural representation of the patterns in these phenomenons. Many researchers have developed state-of-the-art algorithms to track structural patterns in these domains (Bishop 2006; Koller and Friedman 2009; Murphy 2012).

Many real-world systems are too complex and complicated for humans to learn from. Machine learning has therefore become necessary to help identify the patterns in such systems. Graphical models provide a suitable framework to represent the relationships in such patterns. Developing large-scale algorithms for big data and high-dimensional problems has increasingly become a great concern in machine learning and statistics. A common approach to network inference is a centralized learning algorithm, which is often hindered by restrictive resource constraints such as limited local computing, limited memory and expensive computational power. An active area of research in this domain is development of a decentralized system of distributed algorithms for high-dimensional problems (Liu and Ihler 2012; Meng *et al* 2013).

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

3 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) 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 of systemic risk is given in De Bandt and Hartmann (2000), Acharya *et al* (2010) and Brunnermeier and Oehmke (2012). Several authors have come to the same conclusion that the likelihood of major systemic crisis is related to

(1) the degree of correlation among the holdings of financial institutions,

(2) how sensitive they are to changes in market prices and economic conditions,

(3) how concentrated the risks are among those financial institutions,

(4) how closely linked they are with each other and the rest of the economy (Acharya and Richardson 2009; Brunnermeier and Pedersen 2009; Diebold and Yilmaz 2015).

Several systemic risk measures are discussed in the literature. Among them are Banking System's (Portfolio) Multivariate Density (BSPMD; Segoviano and Goodhart 2009), conditional value-at-risk (CoVaR; Adrian and Brunnermeier 2010), absorption ratio (AR; Kritzman *et al* 2011), marginal expected shortfall (MES; Acharya *et al* 2010; Brownlees and Engle 2011), distressed insurance premium (DIP; Huang *et al* 2012), dynamic causality index (DCI) and principal component analysis systemic (PCAS) risk measures (Billio *et al* 2012) and network connectedness measures (NCMs; Diebold and Yilmaz 2014). The BSPMD embeds banks' distress interdependence structures, which capture 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; the DCI captures 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; and NCMs aggregate the contribution of each variable to the forecast error variance of other variables across multiple return series.

Systemic risks are the major contributors to financial crises. Since the 1990s, it has been observed that financial crises appear in clusters (for instance, the East Asian crisis in 1997, the Russia and Long-Term Capital Management (LTCM) crisis in 1998, the Brazil crisis in 1999, the dot-com crisis in 2000, the Argentina crisis in 2001, the Iceland and Turkey crises in 2006, the China crisis in 2007, the global financial crisis in 2007–9 and the European crisis in 2010–13). As to whether systemic risks can be reliably identified in advance, Bernanke (2013) analyzed the global financial crisis by distinguishing between triggers and vulnerabilities of the system that caused the event. This distinction is helpful and allows us to identify which factors to focus on to guard against a repetition of the global financial crisis. The triggers are the events that began the crisis. One prominent trigger of the global financial crisis were the losses suffered by holders of subprime mortgages. Tang *et al* (2010) discussed how triggers for crises differ, and examples include sovereign debt default, risk management strategies, sudden stops in capital flows, collapses of speculative bubbles, inconsistencies between fundamentals and policy settings and a liquidity squeeze. Billio *et al* (2012) found liquidity and credit problems to be the triggers of both the LTCM crisis and the global financial crisis.

The vulnerabilities are the preexisting structural weaknesses of the financial system that amplified the initial shocks (Bernanke 2013). Examples of such factors include a lack of macroprudential focus in regulation and supervision, high levels of leverage and complex interconnectedness. In the absence of these vulnerabilities, the triggers might produce sizable losses to certain firms, investors or asset classes, but generally they would not lead to a global recession. Many authors have found complex financial interdependencies to be channels that magnify initial shocks to the system (Acemoglu *et al* 2015; Billio *et al* 2012; DasGupta and Kaligounder 2014; Diebold and Yilmaz 2014, 2015; Gai *et al* 2011; Tang *et al* 2010). Tang *et al* (2010) showed that financial crises are indeed alike, as all linkages are statistically important across all crises. They identified three potential channels for contagion effects:

(1) idiosyncratic channels, which provide a direct link from the source asset market to international asset markets;

(2) market channels, which operate through either the bond or stock markets;

(3) country channels, which operate through the asset markets of a country jointly.

Billio *et al* (2012) found that the 2007–9 period experienced a higher level of interconnectedness and systemic vulnerability than the LTCM. This allowed the authors to explain why the impact of the global financial crisis affected a much broader segment of financial markets and threatened the viability of several important financial institutions to a greater extent than the LTCM. Acemoglu *et al* (2015) showed that, beyond a certain point, dense interconnections serve as a mechanism for the propagation of shocks, leading to a more fragile financial system. Diebold and Yilmaz (2015) showed that the impact of systemic risk depends on the collective action and connectedness of financial institutions as well as interaction between financial markets and the macroeconomy.

Bernanke (2013) argued that shocks are inevitable, but identifying and addressing vulnerabilities is key to ensuring a robust financial system. To understand vulnerabilities, researchers and regulators are currently focusing on network analyses to identify complex connectivities in financial markets. Some of the applications of 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 financial system, it is of crucial importance to

(a) identify systemically important institutions,

(b) identify specific structural aspects of the system that are particularly vulnerable,

(c) identify potential mechanisms for shock propagation in the system.

Early studies on systemic risk networks focused on linkages arising from actual exposures based on balance sheet information. However, due to data accessibility issues, little empirical work has been done in this area (Cont *et al* 2013; Georg 2013). Meanwhile, several statistical and econometric methods have been advanced to study interdependencies, contagion and spillover effects from observed market data. The commonly discussed approaches in recent literature include the following: Granger causality (Billio *et al* 2012), variance decomposition (Diebold and Yilmaz 2014), tail risk (Hautsch *et al* 2015) and partial correlation (Barigozzi and Brownlees 2016). Most of these approaches often ignore the problem of network uncertainty and highly complex structures. In this paper, we discuss a Bayesian approach to network inference that helps to address the above problems.

4 GRAPHICAL MODELS IN MULTIVARIATE ANALYSIS

Graphical models have become popular for modeling patterns in complex systems due to their ability to provide an intuitive interpretation of the interactions. The graphs present a way of visualizing the relationships between variables in order to distinguish between direct and indirect interactions. The idea of connecting the multivariate time series literature and graphical models is gradually becoming a vibrant field of research in economics and finance. We relate graphical models to multivariate statistical analysis, specifically in multivariate regression problems, with possible applications in financial econometrics.

4.1 Multivariate (multiple) regression

A typical multivariate multiple regression model is given by

$$Y = BX + U, \tag{4.1}$$

where X and Y are vectors of exogenous and response variables, respectively, B is a coefficient matrix, and U is a vector of errors. The common approach in much empirical research is to fit the above model and 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, ie, committing a type I error. In multiple hypothesis testing, the type I errors committed increase with the number of hypotheses, which may have serious consequences for the conclusions and generalization of the results. Several approaches have been proposed to deal with this problem: see Shaffer (1995) and Drton and Perlman (2007) for a review and discussion of multiple hypothesis testing.

Graphical models provide a convenient framework for exploring multivariate dependence patterns. By considering (4.1) as a causal (dependency) pattern of elements in Y on elements in X, the coefficient 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. We define a binary connectivity matrix, G, such that $G_{ij} = 1$ implies $x_j \to y_i \iff B_{ij} \neq 0$. Thus, G can be interpreted as a (directed) graph of the conditional dependencies between elements in X and Y. B can be represented as

$$B = (G \circ \Phi), \tag{4.2}$$

where Φ is a coefficient matrix and the operator $(\circ)$ is the element-by-element Hadamard product (ie, $B_{ij} = G_{ij}\Phi_{ij}$). 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, we interpret Φ (B) as the unconstrained (respectively, constrained) regression coefficient matrix. For example,

$$\begin{aligned} y_1 &= 1.3x_1 + 0.5x_3 + u_1 \\ y_2 &= 0.9x_1 + 0.5x_5 + u_2 \\ y_3 &= x_2 + 0.7x_4 + u_3 \end{aligned} \qquad B = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{pmatrix} 1.3 & 0 & 0.5 & 0 & 0 \\ 0.9 & 0 & 0 & 0 & 0.5 \\ 0 & 1 & 0 & 0.7 & 0 \end{pmatrix} \end{array} \tag{4.3}$$

$$G = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{array} \quad \Phi = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{pmatrix} 1.3 & a_1 & 0.5 & a_2 & a_3 \\ 0.9 & a_4 & a_5 & a_6 & 0.5 \\ a_7 & 1 & a_8 & 0.7 & a_9 \end{pmatrix} \end{array}, \tag{4.4}$$

where $a_i \in \mathbb{R}$, $i = 1, \dots, 9$, are expected to be statistically not different from zero by definition. The expression (4.4) presents an interesting alternative for modeling (4.3). Instead of estimating the unconstrained coefficient matrix, Φ, and performing multiple hypothesis tests, a more efficient alternative is to infer G as a variable selection matrix to estimate only the relevant coefficients in B. Thus, nonzero elements in B correspond to nonzero elements in G. Inference of G, taking into account all possible dependence configurations, automatically handles the multiple testing problem in multivariate multiple regression models.

In most regression models, the parameters to be determined are $\{B, \Sigma_u\}$, where Σ_u is the covariance matrix of U. In graphical models, inference of the graph structure is central to the model estimation, and the set of parameters, θ, describes the strength of the dependence among variables. Hence, by relating graphical models to multivariate regressions, θ must be equivalent to the regression parameters, ie, $\theta \equiv \{B, \Sigma_u\}$.

Let Y and X denote an $n_y \times 1$ and an $n_x \times 1$ vector of dependent and explanatory variables, respectively. Let $Z = (Y', X')'$ be the $n = (n_y + n_x) \times 1$ vector of stacked Y and X. Suppose the joint distribution, Z, follows the distribution $Z \sim \mathcal{N}_n(0, \Omega^{-1})$, where $\Sigma = \Omega^{-1}$ is an $n \times n$ 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. Therefore, estimating the model parameters associated with G is equivalent to estimating Ω, ie, $\theta = \Omega$. Given Ω, the parameters of model (4.1) can be found from $\Sigma = \Omega^{-1}$ as

$$B = \Sigma_{yx}\Sigma_{xx}^{-1}, \qquad \Sigma_\varepsilon = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}, \tag{4.5}$$

where Σ_{xy} is $n_x \times n_y$ 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 Heckerman and Geiger (1994) and Lenkoski and Dobra (2011).

4.2 Applications in econometrics and finance

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

4.2.1 Structural model estimation

Vector autoregressive (VAR) models are widely used to estimate and forecast multivariate time series in macroeconomics. It is generally known that such models do not have direct economic interpretations. However, due to their ability to forecast dynamics in macroeconomic variables, such a limitation is overlooked. Structural VARs (SVARs), however, have direct economic interpretations, but these are not estimable due to identification issues. For SVAR identification, the standard approach relies on a reduced-form VAR to determine the relationships among shocks as a means of providing economic intuition about the structural dynamics. To achieve this, some researchers impose structures provided by a specific economic model, in which case the empirical results will only be as credible as the underlying theory (Kilian 2013). Moreover, in many cases, there are not enough credible exclusion restrictions to achieve identification.

Following (4.2), the SVAR model can be expressed in a graphical model form as

$$Y_t = \sum_{i=0}^{p} B_i Y_{t-i} + \varepsilon_t = \sum_{i=0}^{p} (G_i \circ \Phi_i) Y_{t-i} + \varepsilon_t, \tag{4.6}$$

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.[1]

4.2.2 Time-varying model estimation

Fixed or time-varying parameter models are standard applications in most empirical works. These approaches to modeling real-world phenomenons implicitly assume that interactions between variables are stable over time, and only the parameters vary or are fixed. This assumption may have consequences for the performance of estimated models. Some empirical works have shown that financial networks especially exhibit

[1] See Ahelegbey *et al* (2016) for an application of graphical models to identifying restrictions in SVAR. See also Swanson and Granger (1997), Dahlhaus and Eichler (2003), Demiralp and Hoover (2003), Corander and Villani (2006) and Moneta (2008) for estimation of causal structures in time series and VAR models.

random fluctuations over various time scales (Billio *et al* 2012), which must be incorporated into modeling dynamics in observed data. A typical time-varying model of (4.1) is given by

$$Y_t = B_t X_{t-1} + U_t. \tag{4.7}$$

Following the expression in (4.2), the coefficient matrix of (4.7) can be expressed as

$$B_t = G \circ \Phi_t, \qquad B_t = G_t \circ \Phi, \qquad B_t = G_t \circ \Phi_t. \tag{4.8}$$

The first expression of (4.8) follows the typical time-varying parameter models commonly discussed in most empirical papers, where the graph is invariant over time. The other two expressions (4.8) present cases of time-varying structures of dependence. The state of the world is not constant over time, so devoting our attention to modeling dynamics of the structure of interaction seems to be a more interesting line of research in order to understand the ever-transforming modern economic and financial system.[2]

4.2.3 High-dimensional model estimation

There is an increasing interest in high-dimensional models and big data analysis. This has become necessary, as many studies have shown that information from large data sets enriches existing models, produces better forecasts for VAR models and also reveals the connectedness of the financial system. Graphical models are therefore relevant for high-dimensional modeling, as they offer interpretations of information extracted from large data sets.[3]

The graphical approach can be used to build models that serve as alternatives to the factor methods when dealing with large data sets. 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 (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.

4.2.4 CAPM-like model estimation

A fundamental model in financial theory is the capital asset pricing model (CAPM; Sharpe 1964). This is an extension of the portfolio theory of Markowitz (1952). This

---

[2] See Bianchi *et al* (2014) for an example of a graphical factor model with Markov-switching graphs and parameters for modeling contagion. See also Carvalho and West (2007) for graphical multivariate volatility modeling induced by time-varying covariances across series.

[3] See Ahelegbey *et al* (2014) for a discussion on modeling sparsity in large graphical VAR models with uncertainty on the lag order. See also Jones *et al* (2005) and Scott and Carvalho (2008) for discussions on approaches for penalizing globally or locally "dense" graphs when estimating high-dimensional models.

approach has received much criticism due to problems of empirical evidence. Fama and French (2004) summarized the popularity of the CAPM as follows:

> 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 market specific and idiosyncratic effects, as in the classical CAPM models.[4]

4.2.5 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, a well-diversified portfolio is made up of asset classes that are highly uncorrelated and are 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. Portfolios that contain securities with several correlated risk factors do not meet this diversification criteria. Despite unprecedented access to information, some portfolios, by construction, contain a predominant factor, and most risk-modeling techniques are unable to capture this contagion.

To measure diversification more accurately, the graphical approach can be applied to study the structure of interconnection among the asset classes. Since the arbitrage pricing theory (APT) model of Roll (1977) and Ross (1976) and the return-based style model of Sharpe (1992) are regression models, the graphical approach provides a useful technique for portfolio selection by explicitly modeling the dependence of the factors or asset classes.[5]

4.2.6 Risk-management-style assessment

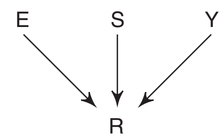
Sharpe (1992) introduced a return-based analysis to measure management style and performance. The analysis is based on the idea that a manager builds a portfolio according to a specific investment philosophy, and investments reflect a style. The approach is based on a style-regression model to determine the "effective mix", given by the estimated model, which represents the return from the style, while the residuals reflect the performance due to the "selection" (active management).

---

[4] See Ahelegbey and Giudici (2014) for a discussion on Bayesian hierarchical graphical models, which allow correlations to be decomposed into a country (market) effect plus a bank-specific (idiosyncratic) effect.

[5] See Shenoy and Shenoy (2000) and Carvalho and West (2007) for an application of graphical models in the context of financial time series for predictive portfolio analysis.

FIGURE 1 Network model for returns on investment and style measurement.



Y is the amount invested, S is the annual increase in market stocks, E is a measure of the fund manager's style and R is the return.

Suppose we are interested in modeling the annual return on an investment. This will depend on factors such as the amount invested (Y), the annual stock market increase (S) and the experience or style of the fund manager (E). A simple network to model this interaction is displayed in Figure 1, where $R = f(E, S, Y)$, and the returns are a function of Y, S and E.

This can serve as a benchmark to assess the performance of fund managers. For instance, since the fund manager's level of experience may be unknown, a qualitative measure can be applied to rank their experience, and a probabilistic inference can be obtained on the likely level of experience, given information on the other variables in the network.[6]

5 BAYESIAN INFERENCE PROCEDURE FOR STRUCTURE LEARNING

Network structure learning using standard frequentist techniques presents multiple testing problems. It also involves exploring all candidate structures 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 summarizes these relationships often ignores the uncertainty problem. The Bayesian methodology has proved to be more efficient in addressing uncertainties and complexities in the inference problem than standard frequentist techniques (Heckerman *et al* 1995). The Bayesian approach to network modeling is a class of probabilistic graphical models, where each node represents a random variable, and the links express probabilistic relationships between these variables. The network captures the way in which the joint distribution over all the random variables can be decomposed into a product of factors, each depending only on a subset of the variables.

[6] See Ammann and Verhofen (2007) for a network application for mutual fund managers behavior analysis.

The description of the Bayesian inference is completed with the elicitation of prior distributions on the graphs and parameters, the posterior approximations and graph estimation.

5.1 Prior distribution

Modeling the joint distribution of (G, Ω) from a Bayesian perspective can be expressed in a natural hierarchical structure $P(G, \Omega) = P(G)P(\Omega \mid G)$. In the absence of genuine prior knowledge of the network of interactions, the common approach is to assume a uniform prior for G, ie, $P(G) \propto 1$. (For discussions on other graph priors, see Friedman and Koller (2003), Jones *et al* (2005) and Scott and Carvalho (2008).)

The standard parameter prior for graphical models is often conditioned on the graphs. Two main classes of such priors are commonly discussed in the literature. The first is based on DAG models that permit an unconstrained precision matrix Ω (see Consonni and Rocca 2012; Geiger and Heckerman 2002; Grzegorczyk 2010; Heckerman *et al* 1995). The second is based on decomposable undirected graph (UG) models, which allow constraints or no constrains on Ω (see Carvalho and Scott 2009; Roverato 2002; Wang and Li 2012). An unconstrained Ω often characterizes a complete 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 or hyper-inverse Wishart distribution.

We follow the standard parameter prior for Gaussian DAG models, with density given by

$$P(\Omega \mid G) = \frac{1}{K_n(\nu, \underline{S})} |\Omega|^{(\nu-n-1)/2} \exp\{-\tfrac{1}{2}\langle \Omega, \underline{S}\rangle\}, \tag{5.1}$$

where $\langle A, B\rangle = \mathrm{tr}(A'B)$ denotes the trace inner product, $\nu > n+1$ is the degrees of freedom parameter, $\underline{S}$ is a prior sum of squares matrix and $K_n(\nu, \underline{S})$ is the normalizing constant:

$$\begin{aligned} K_n(\nu, \underline{S}) &= 2^{\nu n/2} |\underline{S}|^{-\nu/2} \Gamma_n\left(\frac{\nu}{2}\right), \\ \Gamma_b(a) &= \pi^{b(b-1)/4} \prod_{i=1}^{b} \Gamma\left(a + \frac{1-i}{2}\right), \end{aligned} \tag{5.2}$$

where $\Gamma_b(a)$ is the multivariate generalization of the Gamma function, $\Gamma(\cdot)$. In this application, we set $\nu = n+2$ and $\underline{S} = \nu I_n$, where I_n is the n-dimensional identity matrix.

5.2 Posterior approximation

Let $Z_t = (Y_t', X_t')'$, where Y_t represents the set of dependent variables at time t and X_t is the set of explanatory variables. Suppose that $Z_t \sim \mathcal{N}_n(\mathbf{0}, \Omega^{-1})$, where n is the number of elements in Z_t. Let $\mathcal{D} = (Z_1, \dots, Z_T)$ be a complete series of the observed variables. Then, the likelihood function $P(\mathcal{D} \mid \Omega, G)$ is multivariate Gaussian with density

$$P(\mathcal{D} \mid \Omega, G) = (2\pi)^{-nT/2} |\Omega|^{T/2} \exp\{-\tfrac{1}{2}\langle \Omega, \hat{S} \rangle\}, \tag{5.3}$$

where $\hat{S} = \sum_{t=1}^{T} Z_t Z_t'$ is the $n \times n$ sum of squares matrix. Since G is unknown, we estimate the graph by integrating out Ω analytically to produce a marginal likelihood function

$$P(\mathcal{D} \mid G) = \int P(\mathcal{D} \mid \Omega, G) P(\Omega \mid G) \,\mathrm{d}\Omega = (2\pi)^{-nT/2} \frac{K_n(\nu + T, \overline{S})}{K_n(\nu, \underline{S})}, \tag{5.4}$$

where $\overline{S} = \underline{S} + \hat{S}$ is the posterior sum of squares matrix. The graph posterior is given by $P(G \mid \mathcal{D}) = P(G) P(\mathcal{D} \mid G)$. Based on the uniform prior assumption over G, maximizing $P(G \mid \mathcal{D})$ is equivalent to maximizing $P(\mathcal{D} \mid G)$. Following the standard Bayesian paradigm, (5.4) can be factorized into a product of local terms, each involving a response variable (y_i) and its set of selected predictors (π_i):

$$P(\mathcal{D} \mid G) = \prod_{i=1}^{n_y} P(\mathcal{D} \mid G(y_i, \pi_i)) = \prod_{i=1}^{n_y} \frac{P(\mathcal{D}^{(y_i, \pi_i)} \mid G)}{P(\mathcal{D}^{(\pi_i)} \mid G)}, \tag{5.5}$$

where $\pi_i = \{j = 1, \dots, n_x \colon G_{ij} = 1\}$, n_y is the number of equations and $G(y_i, \pi_i)$ is the subgraph of G, with links from π_i to y_i. $\mathcal{D}^{(y_i, \pi_i)}$ and $\mathcal{D}^{(\pi_i)}$ are submatrixes of $\mathcal{D}$, consisting of (y_i, π_i) and π_i, respectively. The closed form of (5.5) is

$$P(\mathcal{D}^k \mid G) = (\pi)^{-n_k T/2} \frac{|\underline{S}_{k,k}|^{\nu/2}}{|\overline{S}_{k,k}|^{(\nu+T)/2}} \prod_{i=1}^{n_k} \frac{\Gamma((\nu + T + 1 - i)/2)}{\Gamma((\nu + 1 - i)/2)}, \tag{5.6}$$

where $k \in \{(y_i, \pi_i), \pi_i\}$ is of dimension n_k, $\mathcal{D}^k$ is a submatrix of $\mathcal{D}$ associated with k, and $|\underline{S}_{k,k}|$ and $|\overline{S}_{k,k}|$ are the determinants of the prior and posterior sum of the square matrixes of $\mathcal{D}^k$.

5.3 Graph estimation

We sample the graph following the MCMC algorithm in Madigan and York (1995). The scheme is such that, at the rth iteration, given $G^{(r-1)}$, the sampler proposes a new graph $G^{(*)}$ with acceptance probability

$$A(G^{(*)} \mid G^{(r-1)}) = \min\left\{ \frac{P(\mathcal{D} \mid G^{(*)})}{P(\mathcal{D} \mid G^{(r-1)})} \frac{P(G^{(*)})}{P(G^{(r-1)})} \frac{Q(G^{(r-1)} \mid G^{(*)})}{Q(G^{(*)} \mid G^{(r-1)})}, 1 \right\}, \tag{5.7}$$

where $Q(G^{(*)} \mid G^{(r-1)})$ and $Q(G^{(r-1)} \mid G^{(*)})$ are the forward and reverse proposal distribution, respectively. The proposal distribution assigns a uniform probability to all possible edges 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 rth iteration is set to $G^{(r)} = G^{(*)}$; otherwise, $G^{(r)} = G^{(r-1)}$.[7]

6 FINANCIAL APPLICATION

We now illustrate the effectiveness of the Bayesian approach to network inference in analyzing return and volatility connectedness of financial time series.

6.1 Financial system interconnectedness

We study the structure of interconnectedness among the four major sectors of the US financial system, following Billio *et al* (2012). The data consists of monthly return indexes for hedge funds (HF), banks (BK), brokers (BR) and insurance companies (IN) in the United States from January 1994 to December 2008. Single hedge-fund data was obtained from the Lipper TASS database. Data for individual banks, brokers and insurers was obtained from the Center for Research in Security Prices database. The monthly returns of all companies with standard industrial classification (SIC) codes 6000–6199, 6200–6299 and 6300–6499 were used to construct value-weighted indexes for banks, brokers and insurers, respectively.

We estimate the temporal dependence pattern in a VaR with lag order ($p = 1$), chosen according to a Bayesian information criterion (BIC). We examine the network for two sample periods (1994–2000 and 2001–8). Table 1 shows the posterior probabilities of the presence of edges in the network. The left (respectively, right) panel shows the edge posterior probabilities in the network for the period 1994–2000 (respectively, 2001–8). Bold values indicate links for which 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). The edge posterior probabilities in the first period (1994–2000) are very low compared with the second (2001–8). Thus, we find no evidence of significant linkages in the first period. However, in the second period, we find strong interconnectedness among the institutions. More specifically, we find evidence of an autoregressive effect among hedge funds, ie, $P(\mathrm{HF}_{t-1} \to \mathrm{HF}_t \mid \mathcal{D}) = 0.95$, and a strong effect of insurers on brokers, $P(\mathrm{IN}_{t-1} \to \mathrm{BR}_t \mid \mathcal{D}) = 0.98$; insurers on banks, $P(\mathrm{IN}_{t-1} \to \mathrm{BK}_t \mid \mathcal{D}) = 0.82$; insurers on hedge funds, $P(\mathrm{IN}_{t-1} \to \mathrm{HF}_t \mid \mathcal{D}) = 0.71$; banks on brokers, $P(\mathrm{BK}_{t-1} \to \mathrm{BR}_t \mid \mathcal{D}) = 0.75$; and brokers on insurers, $P(\mathrm{BR}_{t-1} \to \mathrm{IN}_t \mid \mathcal{D}) = 0.64$.

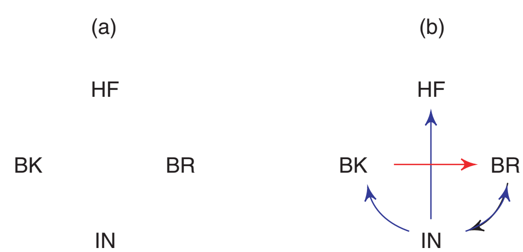
[7] See Ahelegbey *et al* (2016) for discussions on the convergence diagnostics of the MCMC, estimation of the edge posterior probabilities and a pseudo-code of the algorithm.

TABLE 1 Marginal posterior probabilities of linkages among institutions between 1994–2000 and 2001–8.

	January 1994–December 2000				January 2001–December 2008			
	$\mathbf{HF}_{t-1}$	$\mathbf{BR}_{t-1}$	$\mathbf{BK}_{t-1}$	$\mathbf{IN}_{t-1}$	$\mathbf{HF}_{t-1}$	$\mathbf{BR}_{t-1}$	$\mathbf{BK}_{t-1}$	$\mathbf{IN}_{t-1}$
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

Bold values indicate links with probabilities greater than 0.5 under a 95% credibility interval.

FIGURE 2 Network of hedge funds (HF), brokers (BR), banks (BK) and insurance (IN) between (a) 1994–2000 and (b) 2001–8, without self-loops.



The blue (red) links are lagged positive (negative) effects.

The network representation of the results of Table 1 is shown in Figure 2. This figure reveals the direction and sign (represented by the colored arrows) of the linkages among the institutions. The blue (respectively, red) links are lagged positive (respectively, negative) effects. We see no links in the first period. In the second period, we find a negative effect of banks on brokers, a bidirectional positive link between insurers and brokers and a positive effect of insurers on banks and hedge funds. These results corroborate the findings of Billio *et al* (2012), providing evidence of a higher vulnerability in the system between 2001 and 2008. We also find that insurers are central to the spread of risk in the system during the 2001–8 period.

6.2 Volatility connectedness in the euro area

Volatility networks (also referred to as "fear connectedness") have become increasingly important due to their ability to track the fear of investors and identify risk transmission mechanisms in the financial system (Diebold and Yilmaz 2014). We

TABLE 2 Description of EURO STOXX 600 super-sectors.

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 estate*	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			

* The financial sector variables.

analyze the "fear connectedness" in the European stock market using intraday high–low price indexes of the nineteen super-sectors of EURO STOXX 600, obtained from Datastream and covering the period September 1, 2006 to September 19, 2014. See Table 2 for a list of the super-sectors that represent the largest euro area companies by the Industry Classification Benchmark (ICB). The institutions cover countries such as Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain.

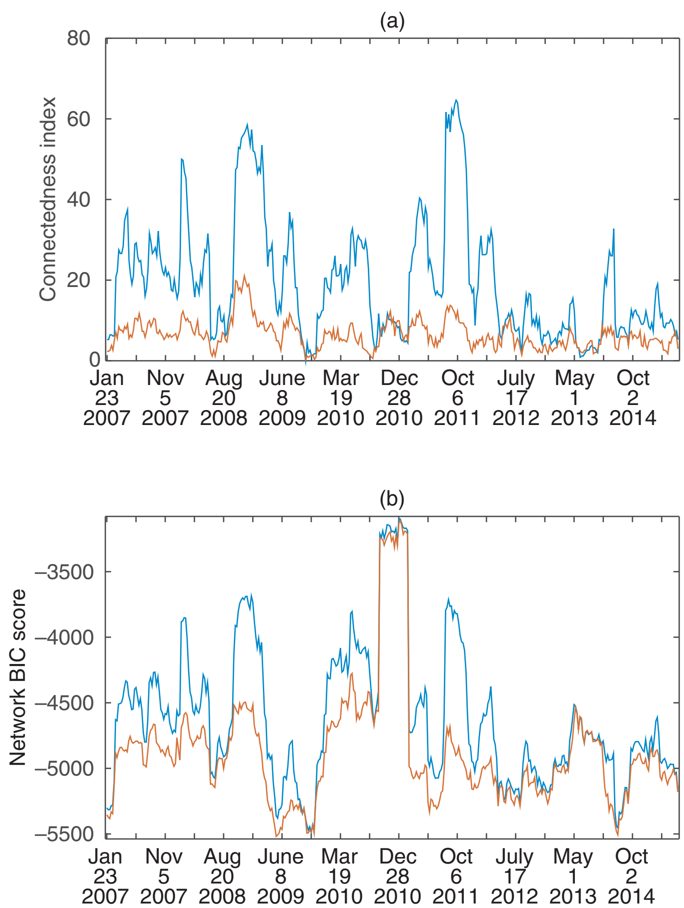
Let $p^{\mathrm{h}}_{i,t}$ and $p^{\mathrm{l}}_{i,t}$ denote the highest and lowest prices of stock i on day t. We obtain the intraday price range as a measure of the realized volatilities

$$\mathrm{RV}_{i,t} = \frac{1}{4\log(2)}(\log p^{\mathrm{h}}_{i,t} - \log p^{\mathrm{l}}_{i,t})^2. \tag{6.1}$$

We study the connectedness of the log volatilities as the dependence pattern in a VaR(1). We characterize the dynamics of the connectedness using a rolling estimation with a window size of 100 days. We compare the Granger causality network (henceforth GCnet) with that of the Bayesian method (henceforth BGnet).

We present in Figure 3 the dynamics of the total connectedness index of the GCnet (in blue) and the BGnet (in red) with their respective BIC scores over the sample period. We notice a significant difference in the total number of interconnections among the institutions (see Figure 3(a)). We also see that the GCnet records higher linkages than the BGnet. This is not surprising, since the Granger causality approach deals only with bivariate time series and is unable to distinguish between direct and mediated causal influences in multivariate settings. The Bayesian method, however, considers single and multiple testing possibilities, and it is therefore able to identify direct and mediated causal effects. The BIC score (see Figure 3(b)) shows that the BGnet produces relatively better structures than the GCnet.

FIGURE 3 Dynamics of total connectedness index and network BIC scores over the period 2007–2014, obtained from a rolling estimation with a window size of 100 days.



(a) Total connectedness index. (b) BIC of connectedness. The index of Granger causality is in blue and the Bayesian method is in red.

We compare the rank positions of the eigenvector centrality on the estimated networks. Eigenvector centrality is a measure of the importance of a variable in the transmission of systemic information and the spread of risks. (See Billio *et al* (2012) and Dungey *et al* (2012) for further discussions.) Figure 4 shows the evolution of the Spearman correlations over the sample period. The distribution of the correlations seems negatively skewed, with the extremely negative correlated rank recorded for the window ending November 12, 2008. Figure 5 shows the network of the window

FIGURE 4 Spearman correlations of centrality rank on the estimated networks over the sample period.

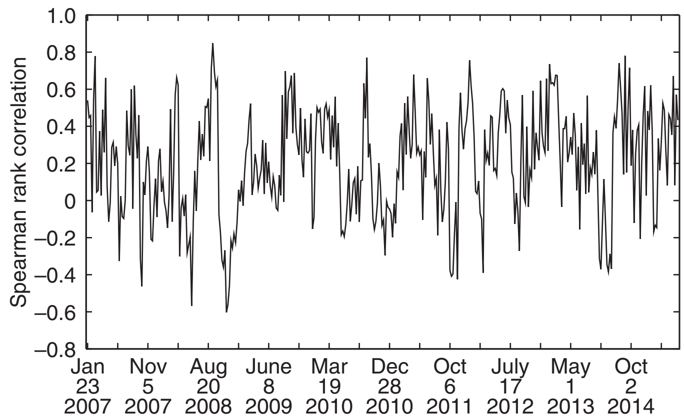
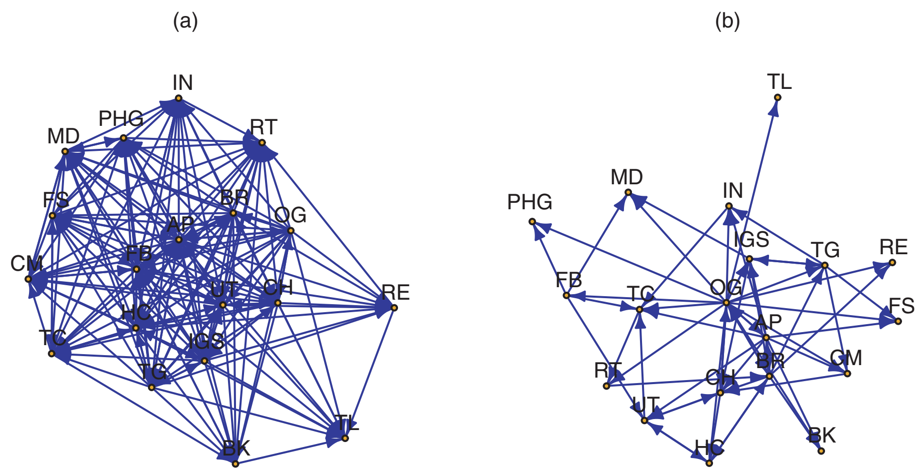


FIGURE 5 Volatility network for period ending November 12, 2008.



(a) GCnet (BIC = −3716.2). (b) BGnet (BIC = −4547.9). Edges are lagged dependencies.

ending November 12, 2008. The GCnet looks more connected than the BGnet; however, the associated BIC score favors the latter. Table 3 shows the top and bottom five ranked institutions from the two networks. We notice that utilities, chemicals and

TABLE 3 The top and bottom five super-sectors ranked by eigenvector centrality (EC) of the GCnet and BGnet for the period ending November 12, 2008.

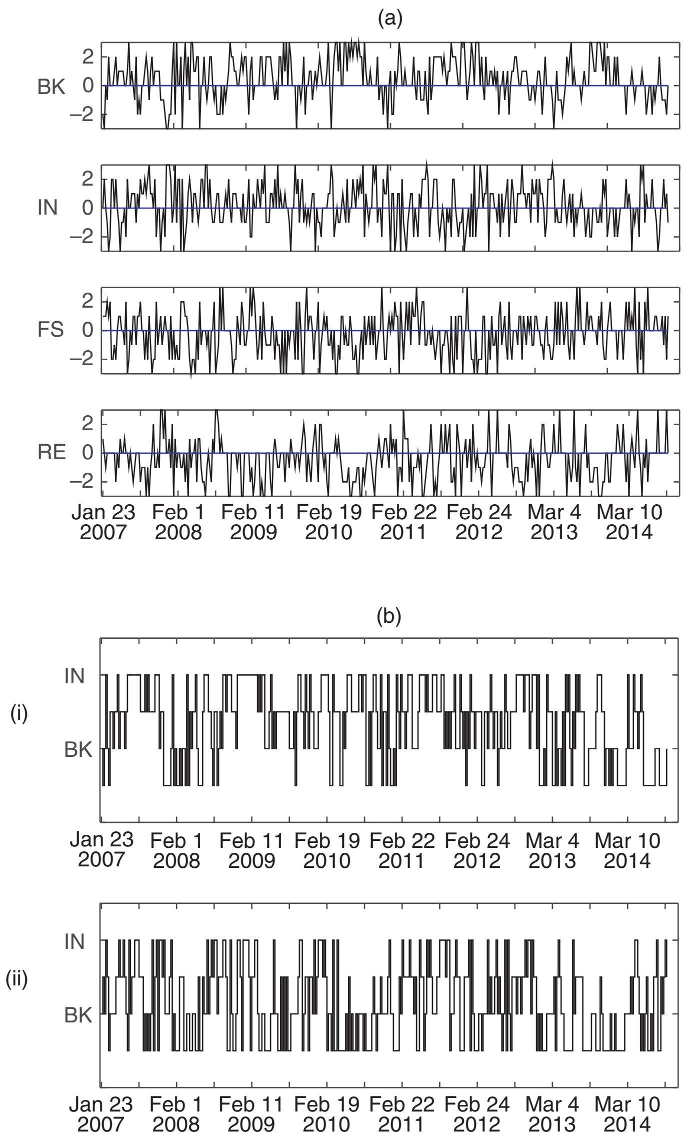
	Granger causality (EC)			Bayesian method (EC)		
Rank	**ID**	**Name**	**EC**	**ID**	**Name**	**EC**
1	RT	Retail	0.3811	CH	Chemicals	0.4189
2	FB	Food & beverage	0.3219	HC	Health care	0.3639
3	IN	Insurance	0.2940	UT	Utilities	0.2971
4	CM	Construction & materials	0.2940	TC	Telecom	0.2821
5	MD	Media	0.2688	BR	Basic resources	0.2714
15	TG	Technology	0.1591	AP	Automobiles & parts	0.1531
16	UT	Utilities	0.1591	RE	Real estate	0.1531
17	CH	Chemicals	0.0859	PHG	Personal & household goods	0.0947
18	BR	Basic resources	0.0711	FB	Food & beverage	0.0731
19	OG	Oil & gas	0.0414	TL	Travel & leisure	0.0731

basic resources are ranked low in the GCnet but high in the BGnet. Also, food & beverage ranks highly in the GCnet but lowly in the BGnet.

Finally, we focus on the centrality of the financial sector of the market. Figures 6(a) and 6(b) show the dynamics of the rank differences and the most central institutions. Figures 6(c) and 6(d) also show the frequency of the sign rank differences and the frequency of the most central institution, respectively. We remind the reader that in rank terms, 1 means higher centrality and 4 means lower centrality. Also, the rank difference is the difference between the rank on the GCnet and that on the BGnet. Thus, a negative rank difference denotes a higher (respectively, lower) centrality on the GCnet (respectively, BGnet). The evolution of the rank differences in Figure 6(a) shows many deviations from the reference line (in blue) for all four institutions. The reference line indicates equal centrality ranks of institutions by the two estimations.

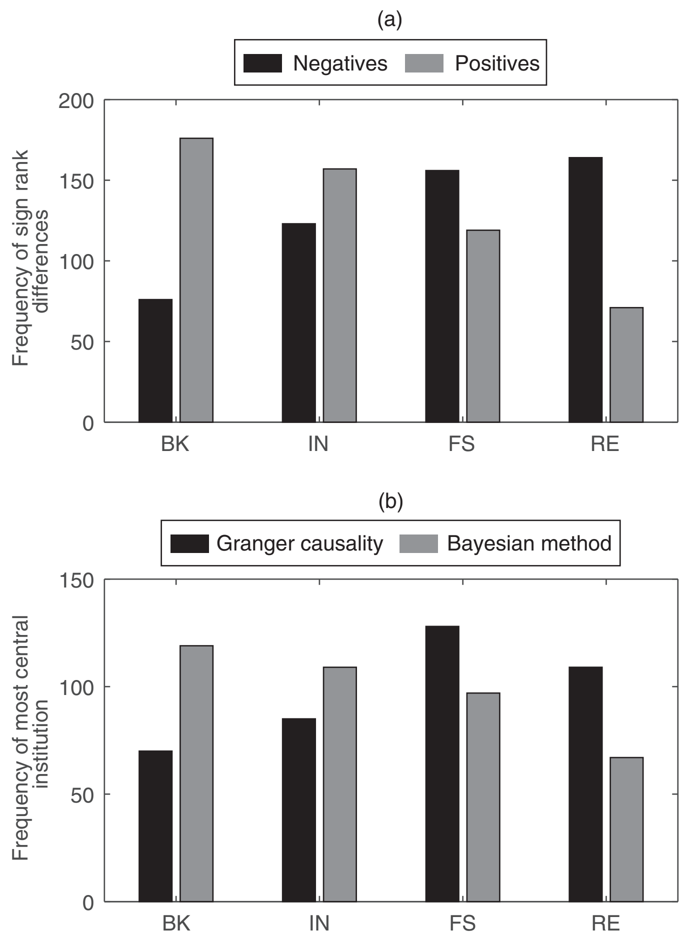
Figure 6(b) shows many periods of differences in the most central institution predicted by the two methods. For instance, GCnet found real estate to be more central than banks for most of 2007 to 2009, while the BGnet shows the opposite. The frequency of the sign rank differences (in Figure 6(c)) indicates results that are the complete opposite, such that, in most of the sample, banks and insurers are ranked relatively low on GCnet and high on the BGnet. Similarly, financial services and real estate are ranked relatively high on GCnet and low on the BGnet. Thus, a researcher using Granger causality will identify financial services to be most central in the spread

FIGURE 6 Centrality in the financial sector of European stock market. Rank value 1 (4) means highest (lowest) centrality. [Figure continues on next page.]



(a) Institutional rank differences. (b) Most central institution: (i) Granger causality; (ii) Bayesian method. A negative (positive) sign of rank difference means a higher (lower) centrality by the GCnet and a lower (higher) centrality by the BGnet.

FIGURE 6 Continued.



(c) Frequency in sign rank differences. (d) Frequency of most central institution. A negative (positive) sign of rank difference means a higher (lower) centrality by the GCnet and a lower (higher) centrality by the BGnet.

of risk, and banks the least, while a researcher applying the Bayesian method will find banks to be the most central, and real estate the least.

In many real-world interactions, dependencies among random variables are more complex than pairwise. The Bayesian approach to network selection discussed in this paper is designed to handle joint estimation and large-scale multiple testing problems. Thus, it produces more suitable graphs than Granger causality to analyze complex interactions. From the BGnet, we find evidence that banks and insurers are more central in the “fear connectedness” expressed by market participants in the financial sector of the euro area.

7 CONCLUSION

This paper presents a state-of-the-art review of the interface between statistics and econometrics in the inference and application of Bayesian graphical models. We specifically highlight connections and possible applications of network models in financial econometrics in the context of systemic risk. Using the monthly return indexes of Billio *et al* (2012) for hedge funds, banks, brokers and insurers, we find evidence of a higher connectedness between 2001 and 2008 in the US financial system. We also find evidence that insurers play a central role in the vulnerability of the system, which amplified the global financial crisis. Further empirical study on the financial super-sectors of the European stock market reveals banks and insurers to be central figures in the “fear connectedness” (Diebold and Yilmaz 2014) expressed by market participants in the euro area. A comparison of the estimated networks shows that the Bayesian method produces dependence patterns that are more suitable than those produced by Granger causality to capture complex interdependencies.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. The authors’ research is supported by funding from the European Union, Seventh Framework Programme FP7/2007-2013 under grant agreement SYRTO-SSH-2012-320270; the Italian Ministry of Education; University and Research (MIUR) PRIN 2010-11 grant MISURA; and the Air Force Office of Scientific Research (AFOSR) grant.

ACKNOWLEDGEMENTS

I would like to thank the editor, the anonymous referees, and Monica Billio, Roberto Casarin, Francesco Ravazzolo and Fabrizio Leisen for their comments on an earlier version of this paper.

REFERENCES

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review* **105**, 564–608 (http://doi.org/bhxf).

Acharya, V. V., and Richardson, M. (2009). *Restoring Financial Stability: How to Repair a Failed System*. Wiley (http://doi.org/fzbfh4).

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., and Giudici, P. (2014). Bayesian selection of systemic risk networks. *Advances in Econometrics: Bayesian Model Comparison* **34**, 117–153 (http://doi.org/bhxg).

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. (2016). Bayesian graphical models for structural vector autoregressive processes. *Journal of Applied Econometrics* **31**, 357–386 (http://doi.org/bhxh).

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 (http://doi.org/fdbtv4).

Ammann, M., and Verhofen, M. (2007). Prior performance and risk-taking of mutual fund managers: a dynamic Bayesian network approach. *Journal of Behavioral Finance* **8**, 20–34 (http://doi.org/dq2mcq).

Andersson, S. A., Madigan, D., and Perlman, M. D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Annals of Statistics* **25**, 505–541 (http://doi.org/dpf8zw).

Asratian, A. S. (1998), *Bipartite Graphs and Their Applications*, Volume 131. Cambridge University Press.

Banerjee, O., El Ghaoui, L., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research* **9**, 485–516.

Barigozzi, M., and Brownlees, C. (2016). NETS: network estimation for time series. Working Paper, Social Science Research Network.

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 (http://doi.org/bhxj).

Bernanke, B. (2010). Causes of the recent financial and economic crisis. Testimony, Statement Before the Financial Crisis Inquiry Commission, Washington, September.

Bernanke, B. (2013). Monitoring the financial system. Speech, 49th Annual Conference Bank Structure and Competition, May 10, Federal Reserve Bank of Chicago.

Bernanke, B., Boivin, J., and Eliasz, P. S. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* **120**, 387–422.

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 (http://doi.org/bhxm).

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

Bisias, D., Flood, M. D., Lo, A. W., and Valavanis, S. (2012). A survey of systemic risk analytics. Working Paper 1, US Department of Treasury, Office of Financial Research.

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. Working Paper 18398, NBER.

Brunnermeier, M. K., and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial Studies* **22**, 2201–2238 (http://doi.org/drd3sm).

Bunea, F., She, Y., and Wegkamp, M. H. (2011). Optimal selection of reduced rank estimators of high-dimensional matrices. *Annals of Statistics* **39**, 1282–1309 (http://doi.org/d8cg8b).

Carvalho, C. M., and Scott, J. G. (2009). Objective Bayesian model selection in Gaussian graphical models. *Biometrika* **96**, 497–512 (http://doi.org/bkd7v2).

Carvalho, C. M., and West, M. (2007). Dynamic matrix-variate graphical models. *Bayesian Analysis* **2**, 69–98 (http://doi.org/cqdc38).

Carvalho, C. M., Massam, H., and West, M. (2007). Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika* **94**, 647–659 (http://doi.org/bd9tpp).

Chen, K., Dong, H., and Chan, K.-S. (2013). Reduced rank regression via adaptive nuclear norm penalization. *Biometrika* **100**, 901–920 (http://doi.org/bhxp).

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 (http://doi.org/bhxq).

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. *Journal of Machine Learning Research* **12**, 1771–1812.

Consonni, G., and Rocca, L. L. (2012). Objective Bayes factors for Gaussian directed acyclic graphical models. *Scandinavian Journal of Statistics* **39**, 743–756 (http://doi.org/bhxr).

Cont, R., Santos, E., and Moussa, A. (2013). Network structure and systemic risk in banking systems. In *Handbook of Systemic Risk*, pp. 327–368. Cambridge University Press.

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

Dahlhaus, R., and Eichler, M. (2003). Causality and graphical models for time series. In *Highly Structured Stochastic Systems*, Green, P. J., Hjort, N., and Richardson, S. (eds). Oxford University Press.

DasGupta, B., and Kaligounder, L. (2014). On global stability of financial networks. *Journal of Complex Networks* **2**, 313–354 (http://doi.org/bhxs).

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. Society for Artificial Intelligence and Statistics.

De Bandt, O., and Hartmann, P. (2000). Systemic risk: a survey. Working Paper, European Central Bank.

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 (http://doi.org/d4qgmm).

Diebold, F., and Yilmaz, K. (2014). On the network topology of variance decompositions: measuring the connectedness of financial firms. *Journal of Econometrics* **182**, 119–134 (http://doi.org/v9t).

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

Diks, C., and Panchenko, V. (2005). A note on the Hiemstra–Jones test for Granger noncausality. *Studies in Nonlinear Dynamics & Econometrics* **9**, 1–9 (http://doi.org/dsnsvf).

Dornbusch, R., Park, Y. C., and Claessens, S. (2000). Contagion: understanding how it spreads. *World Bank Research Observer* **15**, 177–197 (http://doi.org/bhxw).

Drton, M., and Perlman, M. D. (2007). Multiple testing and error control in Gaussian graphical model selection. *Statistical Science* **22**, 430–449 (http://doi.org/b8sskz).

Dungey, M., Fry, R., González-Hermosillo, B., and Martin, V. L. (2005). Empirical modelling of contagion: a review of methodologies. *Quantitative Finance* **5**, 9–24 (http://doi.org/dgcnpf).

Dungey, M., Luciani, M., and Veredas, D. (2012). Ranking systemically important financial institutions. Discussion Paper, Tinbergen Institute.

Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics* **137**, 334–353 (http://doi.org/dttdkh).

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 (http://doi.org/bq8s9c).

Fan, J., and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Annals of Statistics* **32**, 928–961 (http://doi.org/fp8q45).

Financial Crisis Inquiry Commission (2011). The financial crisis inquiry report. Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States, Public Affairs.

Frank, O., and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association* **81**, 832–842 (http://doi.org/bhxz).

Friedman, N., and Koller, D. (2003). Being Bayesian about network structure. *Journal of Machine Learning* **50**, 95–125 (http://doi.org/dg8mjh).

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 (http://doi.org/dzt3cq).

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical LASSO. *Biostatistics* **9**, 432–441 (http://doi.org/db7svr).

Gai, P., Haldane, A., and Kapadia, S. (2011). Complexity, concentration and contagion. *Journal of Monetary Economics* **58**, 453–470 (http://doi.org/bkw8vg).

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 (http://doi.org/fkxnrp).

Georg, C.-P. (2013). The effect of the interbank network structure on contagion and common shocks. *Journal of Banking & Finance* **37**, 2216–2228 (http://doi.org/bhx3).

Ghahramani, Z. (1998). Learning dynamic Bayesian networks. In *Adaptive Processing of Sequences and Data Structures*, pp. 168–197. Springer (http://doi.org/chctf5).

Gillispie, S. B., and Perlman, M. D. (2001). Enumerating Markov equivalence classes of acyclic digraph models. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 171–177. Morgan Kaufmann.

Giudici, P., and Green, P. J. (1999). Decomposable graphical Gaussian model determination. *Biometrika* **86**, 785–801 (http://doi.org/bp2wbn).

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning* **2**, 129–233 (http://doi.org/fkz6s4).

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (http://doi.org/cr52sb).

Grzegorczyk, M. (2010). An introduction to Gaussian Bayesian networks. In *Systems Biology in Drug Discovery and Development*, Yan, Q. (ed.). Methods in Molecular Biology, Volume 662. Springer (http://doi.org/cs3kgs).

Grzegorczyk, M., Husmeier, D., and Rahnenführer, J. (2011). Modelling non-stationary dynamic gene regulatory processes with the BGM model. *Computational Statistics* **26**, 199–218 (http://doi.org/cjcgzn).

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society* A **170**, 301–354 (http://doi.org/fnm3cz).

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) (http://doi.org/bhx4).

Hautsch, N., Schaumburg, J., and Schienle, M. (2015). Financial network systemic risk contributions. *Review of Finance* **19**, 685–738 (http://doi.org/bhx5).

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

Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning* **20**, 197–243 (http://doi.org/drgw8b).

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 (http://doi.org/bhx6).

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97**, 1090–1098 (http://doi.org/dqmnzj).

Höfling, H., and Tibshirani, R. (2009). Estimation of sparse binary pairwise Markov networks using pseudo-likelihoods. *Journal of Machine Learning Research* **10**, 883–906.

Holland, P. W., and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association* **76**, 33–50 (http://doi.org/bhx7).

Hoover, K. D. (2001). *Causality in Macroeconomics*. Cambridge University Press (http://doi.org/fh6kdz).

Huang, X., Zhou, H., and Zhu, H. (2012). Systemic risk contributions. *Journal of Financial Services Research* **42**, 55–83 (http://doi.org/frpbnf).

Jackson, M. O. (2014). Networks in the understanding of economic behaviors. *Journal of Economic Perspectives* **28**, 3–22 (http://doi.org/bhx8).

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* **20**, 388–400 (http://doi.org/dmm6v6).

Kilian, L. (2013). Structural vector autoregressions. In *Handbook of Research Methods and Applications in Empirical Macroeconomics*, pp. 515–554. Edward Elgar, Cheltenham (http://doi.org/bhx9).

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer (http://doi.org/cz9zfh).

Kolaczyk, E. D., and Csárdi, G. (2014). *Statistical Analysis of Network Data with R*, Volume 65. Springer (http://doi.org/bhzb).

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

Kritzman, M., Li, Y., Page, S., and Rigobon, R. (2011). Principal components as a measure of systemic risk. *Journal of Portfolio Management* **37**, 112 (http://doi.org/c6b52c).

Kschischang, F. R., Frey, B. J., and Loeliger, H.-A. (2001). Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory* **47**, 498–519 (http://doi.org/b2w8tp).

Lebre, S., Becq, J., Devaux, F., Stumpf, M. P., and Lelandais, G. (2010). Statistical inference of the time-varying structure of gene regulation networks. *BMC Systems Biology* **4**(130) (http://doi.org/cpbbd3).

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**, 140–157 (http://doi.org/dj652w).

Lian, H., Feng, S., and Zhao, K. (2015). Parametric and semiparametric reduced-rank regression with flexible sparsity. *Journal of Multivariate Analysis* **136**, 163–174 (http://doi.org/bhzc).

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. International Machine Learning Society.

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. Neural Information Processing Systems Foundation.

Madigan, D., and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review* **63**, 215–232 (http://doi.org/bjp899).

Markowitz, H. (1952). Portfolio selection. *Journal of Finance* **7**, 77–91 (http://doi.org/bhzd).

Meinshausen, N., and Bühlmann, P. (2006). High dimensional graphs and variable selection with the lasso. *Annals of Statistics* **34**, 1436–1462 (http://doi.org/fwt5kt).

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. Society for Artificial Intelligence and Statistics.

Miyamura, M., and Kano, Y. (2006). Robust Gaussian graphical modeling. *Journal of Multivariate Analysis* **97**, 1525–1550 (http://doi.org/fq4dz4).

Moneta, A. (2008). Graphical causal models and VaRs: an empirical assessment of the real business cycles hypothesis. *Empirical Economics* **35**, 275–300 (http://doi.org/bx7wkz).

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA.

Neapolitan, R. E. (2004). *Learning Bayesian Networks*, Volume 38. Prentice Hall, Upper Saddle River, NJ.

Nowicki, K., and Snijders, T. A. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* **96**, 1077–1087 (http://doi.org/bchhqk).

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

Pericoli, M., and Sbracia, M. (2003). A primer on financial contagion. *Journal of Economic Surveys* **17**, 571–608 (http://doi.org/chhw3z).

Ravasz, E. (2009). Detecting hierarchical modularity in biological networks. In *Computational Systems Biology*, pp. 145–160. Springer.

Ravikumar, P., Wainwright, M. J., and Lafferty, J. D. (2010). High-dimensional Ising model selection using l_1-regularized logistic regression. *Annals of Statistics* **38**, 1287–1319 (http://doi.org/fdq6hd).

Robinson, J. W., and Hartemink, A. J. (2009). Non-stationary dynamic Bayesian networks. In *Advances in Neural Information Processing Systems (NIPS)*, Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L. (eds), Volume 21, pp. 1369–1376. Neural Information Processing Systems Foundation.

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 (http://doi.org/ccmg77).

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory* **13**, 341–360 (http://doi.org/crw95p).

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 (http://doi.org/dg5j64).

Scott, J. G., and Carvalho, C. M. (2008). Feature-inclusion stochastic search for Gaussian graphical models. *Journal of Computational and Graphical Statistics* **17**, 790–808 (http://doi.org/bndsk4).

Segoviano, M. A., and Goodhart, C. A. (2009). Banking stability measures. Working Paper 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 (http://doi.org/bhzm).

Sharpe, W. F. (1992). Asset allocation: management style and performance measurement. *Journal of Portfolio Management* **18**, 7–19 (http://doi.org/fjqs2p).

Shenoy, C., and Shenoy, P. P. (2000). Bayesian network models of portfolio risk and return. *Computational Finance* **1999**, 87–106.

Snijders, T. A. (2011). Statistical models for social networks. *Annual Review of Sociology* **37**, 131–153 (http://doi.org/bc9qm2).

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

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 (http://doi.org/bhzn).

Tan, K. M., London, P., Mohan, K., Lee, S.-I., Fazel, M., and Witten, D. (2014). Learning graphical models with hubs. *Journal of Machine Learning Research* **15**, 3297–3331.

Tang, C., Dungey, M. M., Martin, M. V., González-Hermosillo, M. B., and Fry, M. R. (2010). Are financial crises alike? Working Paper 10-14, International Monetary Fund.

Teramoto, R., Saito, C., and Funahashi, S.-I. (2014). Estimating causal effects with a non-paranormal method for the design of efficient intervention experiments. *BMC Bioinformatics* **15**, 1–14 (http://doi.org/bhzp).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society* B **58**, 267–288.

Tsamardinos, I., Brown, L. E., and Aliferis, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning* **65**, 31–78 (http://doi.org/cvb5mq).

Verma, T. S., and Pearl, J. (1991). Equivalence and synthesis of causal models. In *Uncertainty in Artificial Intelligence*, Volume 6, pp. 220–227. Elsevier Science Publishers, Cambridge, MA.

Vogel, D., and Fried, R. (2011). Elliptical graphical modelling. *Biometrika* **98**, 935–951 (http://doi.org/d3423x).

Wainwright, M. J., and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning* **1**, 1–305 (http://doi.org/bpnwrm).

Wang, H., and Li, S. Z. (2012). Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electronic Journal of Statistics* **6**, 168–198 (http://doi.org/bhzq).

Wang, Y. J., and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* **82**, 8–19 (http://doi.org/bhzr).

Wright, L. (2007). Bayesian networks and probabilistic inference in forensic science. *Journal of the Royal Statistical Society* A **170**, 1187–1187 (http://doi.org/dhhnf6).

Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society* B **68**, 49–67 (http://doi.org/fvntrn).

Zha, H., He, X., Ding, C., Simon, H., and Gu, M. (2001). Bipartite graph partitioning and data clustering. In *Proceedings of the 10th International Conference on Information and Knowledge Management*, pp. 25–32. Association for Computing Machinery (http://doi.org/b2mqg5).

Zhao, T., Roeder, K., and Liu, H. (2012). Smooth-projected neighborhood pursuit for high-dimensional nonparanormal graph estimation. In *Advances in Neural Information Processing Systems*, pp. 162–170.

Zou, C., and Feng, J. (2009). Granger causality vs. dynamic Bayesian network inference: a comparative study. *BMC Bioinformatics* **10**(122), 1–17 (http://doi.org/cjfvm5).