Modeling uncertainty in macroeconomic growth determinants using Gaussian graphical models

Adrian Dobra[*,a,b], Theo S. Eicher[c], Alex Lenkoski[a]

[a]*Department of Statistics, University of Washington, Seattle, WA 98155, USA*
[b]*Department of Biobehavioral Nursing and Health Systems, School of Nursing, University of Washington, Seattle, WA 98155, USA*
[c]*Department of Economics, University of Washington, Seattle, WA 98155, USA*

Abstract

Model uncertainty has become a central focus of policy discussion surrounding the determinants of economic growth. Over 140 regressors have been employed in growth empirics due to the proliferation of several new growth theories in the past two decades. Recently Bayesian model averaging (BMA) has been employed to address model uncertainty and to provide clear policy implications by identifying robust growth determinants. The BMA approaches were, however, limited to linear regression models that abstract from possible dependencies embedded in the covariance structures of growth determinants. The recent empirical growth literature has developed jointness measures to highlight such dependencies. We address model uncertainty and covariate dependencies in a comprehensive Bayesian framework that allows for structural learning in linear regressions and Gaussian graphical models. A common prior specification across the entire comprehensive framework provides consistency. Gaussian graphical models allow for a principled analysis of dependency structures, which allows us to generate a much more parsimonious set of fundamental growth determinants. Our empirics are based on a prominent growth dataset with 41 potential economic factors that has been utilized in numerous previous analyses to account for model uncertainty as well as jointness.

Key words: Covariance estimation, Gaussian graphical models, Variable selection, Stochastic search

1. Introduction

The advent of the New Growth Theory [37] produced a dramatic increase in potential growth determinants that have been motivated by economic theory. After focusing heavily on about five growth determinants from the 1950s to the 1980s [41], the candidate regressors considered in seminal growth empirics has risen rapidly to 42 [28], 56 [39], 67 [40], and most recently to 140 [9].

[*]Corresponding author at: Department of Statistics, Box 354322, University of Washington, Seattle, WA 98155, USA

Email addresses: adobra@u.washington.edu (Adrian Dobra), te@u.washington.edu (Theo S. Eicher), lenkoski@stat.washington.edu (Alex Lenkoski)

To no surprise, growth empirics have since become a case study of model uncertainty. After initial attempts to apply the extreme bound analysis proposed by Leamer [24], subsequent approaches have focused on Bayesian Model Averaging (BMA) to resolve model uncertainty – see Eicher et al [11] for a review. The recent literature has pointed out, however, that early assumptions regarding linear models and strict independence of candidate regressors are inconsistent with growth theory. Durlauf et al. [9] survey the list of nonlinearities and interactions suggested by growth theories, and Brock and Durlauf [3] argue forcefully that the resolution of model uncertainty must take into account the probability that the effect of one growth determinant may depend on the inclusion of another.

Beyond the obvious collinearity, Brock and Durlauf [3] outline several additional examples of interactions, including parameter heterogeneity and exchangeability, where regressor interdependence determines explanatory power. For example, the effect of a particular growth determinant might (dis)appear only with the inclusion of specific covariates. Durlauf et al. [10] introduce dilution priors designed to address interdependencies among redundant, collinear, or exchangeable regressors – see George [17]. Doppelhofer and Weeks [8] and Ley and Steel [29] define measures of dependence (which they call "jointness") among explanatory variables that appear in linear regression models. Aside from the related statistical issues, accounting for interdependencies may also deliver more parsimonious models with equal or only minimally reduced explanatory power.

In this paper we propose a novel approach for selecting growth determinants by considering regressions induced by Gaussian graphical models (GGMs) that take advantage of covariance structures [6]. The graphical models approach has the advantage that it relaxes the strict conditional independence constraints implied by Normal linear regression models [14, 40, 29, 11, 12] and explicitly accounts for the complex dependency patterns that exist among the growth determinants. We are able to differentiate between factors that affect growth directly and other factors that affect growth indirectly by influencing other covariates. The overall consistency of our methodology comes from a common prior specification for the model parameters of the various families of GGMs considered.

While the use of GGMs in econometric modeling has been relatively unexplored, the concepts of covariation and conditional independence amongst macroeconomic variables have long been considered in the literature. It has been shown that graphical models can be seen as structural or simultaneous equations models (SEMs) and can involve any combination of discrete, continuous or categorical variables [23]. SEMs are often used to model the evolution of complicated systems in a variety of scientific fields where each factor is believed to depend in some way on the state of other factors in the system. Macroeconomic modeling frequently employs such techniques [22]. Often in such models, terms are included in selected equations and excluded from others, evidence that macroeconomists have long believed conditional independencies exist amongst their covariates. Despite this assumption, uncertainty in growth determinants has typically been considered in the context of regression modeling. Through simulation studies, we show that such an approach will have a tendency to link too many terms to the response, especially as the interaction amongst the covariates grows sparser and more complicated. This over-inclusion provides the motivation for enriching the model space by considering graphical models.

Modeling the covariance structure among candidate regressors using graphical models is there-

fore an alternative to standard SEM modeling techniques – see, for example, Broeck and Binder [4]. A key advantage of the GGM approach is that model uncertainty can be easily incorporated into the framework, especially when suitable conjugate priors are chosen and can therefore help in identifying hypotheses regarding growth interactions that are supported by the observed data. Of course, the full set of candidate hypotheses allowed by GGMs cannot be exhaustively enumerated and as a consequence, the space of graphical models must be explored using suitable stochastic search techniques [18, 20].

The focus of this paper is thus in developing a coherent methodology for modeling uncertainty in macroeconomic growth determinants using a framework that accurately represents the interaction structure believed to underlie macroeconomic variables. As discussed above, we believe that the GGM framework is better suited to this task than the standard regression methodology currently used. As such, results in this paper will typically be comparative in nature: in both simulation studies and the analysis of real data we focus on comparing the results of GGM searches to regression variable selection techniques. By characterizing regression models as a specific type of undirected graphical model, we are able to specify a conjugate prior framework, the G-Wishart distribution [38, 2, 25], that is consistent between the two approaches. The method for scoring models – either regression or graphical – therefore relies on a single common prior parametrization. This fact is crucial as it allows for direct comparisons between the methodologies that is not confused by arbitrary factors such as the use of different priors. As an indirect, but desirable, consequence of this development we propose a new methodology for regression variable selection that performs well for high-dimensional problems.

After using simulation studies to show the propensity of regression variable selection techniques to over-include variables when the covariates exhibit complicated independence structures, we turn to the modeling of growth determinants. Our results are based on a well-known growth dataset with 41 potential determinants originally compiled by Fernández et al. [14] (hereafter FLS data). This dataset has become a consensus dataset in growth empirics to examine growth determinants. It contains a consensus set of growth regressors that have been utilized in the most important methodological advances that investigate model uncertainty in growth regressions – extreme bound analysis in Levine and Renelt [28], Bayesian model averaging, and Bayesian averaging of classical estimators [14, 40, 29, 11, 12] – as well as sensitivity to priors and jointness measures [11, 29, 30, 42]. In addition, the FLS dataset has also been widely used in growth empirics to resolve model uncertainty using the PcGets general-to-specific approach [19], to explore panel regressions [1], or to investigate parameter heterogeneity [32, 12].

As a first step, we consider the FLS data using our new technique for selecting regression terms. We show that the results of this search perform well when compared to existing techniques and receive results broadly consistent with the current literature. We then consider a GGM search and show that by enriching the model space to allow for more complicated interaction structures, the set of potential growth determinants is significantly reduced. A consequence of this refinement is that the core set of variables interacting with growth are those associated with neoclassical theories or purely exogenous characteristics of the countries. This is an encouraging result: it suggests that the use of GGMs recovers those factors that have the longest tradition of being associated with growth, and separates newer covariates that may have been included in other searches types due

to their indirect association with such causes. Based on these results, we feel that as datasets of increasing size and quality are developed, the GGM framework will be better suited for modeling uncertainty in growth determinants than traditional regression variable selection techniques.

The structure of this paper is as follows. In Section 2 we explain how to perform variable selection using conditional independence graphs and contrast this approach with linear regressions. In Section 3 we develop our Bayesian statistical framework for performing variable selection using GGMs. In Section 4 we study jointness measures associated with sets of regressions induced by GGMs. Section 5 contains simulation studies that show how our proposed methodology performs in situations of highly correlated candidate predictors and parsimonious interaction structures. In Section 6 we illustrate our methodology in the analysis of the FLS data. In Section 7 we conclude.

2. Variable selection using conditional independence graphs

We let $X = X_V$, $V = \{1, 2, \ldots, p\}$, be a random vector. The conditional independence relationships among $\{X_v : v \in V\}$ can be summarized in a graph $G = (V, E)$, where each vertex $v \in V$ corresponds with a random variable X_v and $E \subset V \times V$ are undirected edges [44]. A brief summary of relevant definitions from graph theory are given in the Appendix.

The absence of an edge between X_{v_1} and X_{v_2} corresponds with the conditional independence of these two random variables given the rest and is denoted by

$$X_{v_1} \perp\!\!\!\perp X_{v_2} \mid X_{V\setminus\{v_1,v_2\}}.$$

This is called the parwise Markov property relative to G, which in turn implies the local as well as the global Markov properties relative to G [23]. The local Markov property plays a key role in this framework since it gives the regression model induced by G on each variable X_v. More explicitly, consider the neighbors of v in G, that is, the set of vertices $v' \in V$ such that $(v, v') \in E$. We denote this set by $\mathrm{bd}_G(v)$. The local Markov property relative to G says that

$$X_v \perp\!\!\!\perp X_{V\setminus\{\{v\}\cup \mathrm{bd}_G(v)\}} \mid X_{\mathrm{bd}_G(v)}.$$

This statement is precisely the statement we make when we drop the random variables $\{X_{v'} : v' \in V \setminus \mathrm{bd}_G(v)\}$ from the regression of X_v on $\{X_{v'} : v' \in V \setminus \{v\}\}$. However, the local Markov properties associated with each variable in V are only a subset of the conditional independence relationships implied by G.

In general, the variable selection problem for X_v, $v \in V$, determines which are the best subsets of variables $A \subset V \setminus \{v\}$ such that

$$p(X_v|X_{V\setminus\{v\}} = x_{V\setminus\{v\}}) = p(X_v|X_A = x_A), \tag{1}$$

which is equivalent with the conditional independence relation:

$$X_v \perp\!\!\!\perp X_{V\setminus(\{v\}\cup A)} \mid X_A. \tag{2}$$

In the context of regression, the best subsets of variables A are determined exclusively by modeling the conditional $p(X_v|X_{V\setminus\{v\}} = x_{V\setminus\{v\}})$. The marginal distribution $p(X_{V\setminus\{v\}})$ is overlooked in this

decision process since it is not explicitly modeled. The only conditional independence relationship implied by regression models is (2). However, (2) is implied by many other graphs. As such, there is no apparent reason to solve the variable selection problem for X_v by looking only in the set of regressions.

Example. The four graphs in Figure 1 involve the same four random variables Y, Z, T and U. In graph (a) the edge between Y and U is missing, which means that Y and U are conditionally independent given Z and T. This is written as:

$$Y \perp\!\!\!\perp U \mid \{Z, T\}.$$

Since all paths linking Y and U in graph (a) go through Z but do not necessarily include T, we also have $Y \perp\!\!\!\perp U \mid Z$. Similarly, we have $Y \perp\!\!\!\perp U \mid T$ in graph (b) and $T \perp\!\!\!\perp U \mid Z$ in graph (c).

Consider a linear regression model on Y given Z, T and U:

$$Y = \beta_Z \cdot Z + \beta_T \cdot T + \beta_U \cdot U + \epsilon, \tag{3}$$

where ϵ is an error term. Assume that we made the determination that T and U should be dropped from equation (3). Setting $\beta_T = \beta_U = 0$ implies that

$$p(Y|Z, T, U) = p(Y|Z),$$

hence $Y \perp\!\!\!\perp \{T, U\} \mid Z$. This is the conditional independence statement we make by eliminating T and U from regression (3). Figure 1 shows that the conditional independence statement $Y \perp\!\!\!\perp \{T, U\} \mid Z$ is actually implied by all four graphs. However, the graphs (b), (c) and (d) embed other conditional independence statements that are not implied by $Y \perp\!\!\!\perp \{T, U\} \mid Z$. Some of these statements are: $Z \perp\!\!\!\perp U \mid T$ in graph (b), $T \perp\!\!\!\perp U \mid Z$ in graph (c) and $Z \perp\!\!\!\perp T|U$ in graph (d).

3. Variable selection with Gaussian graphical models

Formally, we assume that the observed data $x = (x^{(1)}, \ldots, x^{(n)})^T$ are independent random samples from a p-dimensional multivariate normal distribution $\mathrm{N}_p(0, \Sigma)$ with $\Sigma = (\Sigma_{ij})_{1 \le i,j \le p}$. The response variable $Y = X_1$ is associated with the first component of the random vector $X = X_V$, while the remaining components are the candidate explanatory covariates. The likelihood function is proportional to

$$L(x|\Sigma) \propto (\det \Sigma)^{-n/2} \exp\left\{-\frac{1}{2}\langle \Sigma^{-1}, U\rangle\right\}, \tag{4}$$

where $U = \sum_{i=1}^{n} x^{(i)} x^{(i)T}$, and $\langle A, B\rangle = \mathrm{tr}(A^T B)$ denotes the trace inner product. We take the prior for the precision matrix $K = \Sigma^{-1}$ to be a Wishart distribution $\mathrm{W}_p(\delta, D)$ that is conjugate to the likelihood (4). Its density is

$$p(K) \;=\; \frac{1}{I_p(\delta, D)} (\det K)^{(\delta-2)/2} \exp\left\{-\frac{1}{2}\langle K, D\rangle\right\}, \tag{5}$$

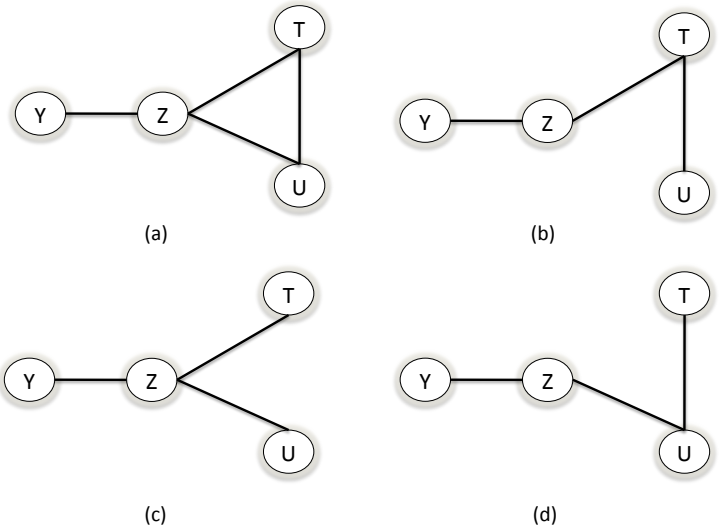

Figure 1: Four conditional independence graphs.

with respect to the Lebesgue measure on the cone P_p of p-dimensional symmetric positive definite matrices. The normalizing constant

$$I_p(\delta, D) \quad = \quad 2^{(\delta+p-1)p/2}\Gamma_p\left\{(\delta+p-1)/2\right\}(\det D)^{-(\delta+p-1)/2}, \tag{6}$$

is finite if $\delta > 2$ and $D^{-1} \in P_p$. Here $\Gamma_p(a) = \pi^{p(p-1)/4}\prod_{i=0}^{p-1}\Gamma\left(a-\frac{i}{2}\right)$ for $a > (p-1)/2$ [33]. By comparing (4) and (5), we remark that the prior parameters can be interpreted as being associated with a fictive dataset with a sample size of $\delta - 2$ and a sample covariance matrix D. Below we discuss the impact of the prior specification on the set of chosen regressors by comparing it to well known results that examine the sensitivity of the resulting growth determinants to various choices of diffuse priors. We assume that the data x has been scaled to have unit variance and subsequently set $\delta = 3$ and $D = I_p$, the p-dimensional identity matrix. The interpretation of this prior is that the components of X are independent apriori and that the "weight" of the prior is equivalent with one observed sample.

The induced prior for the covariance matrix Σ is inverse Wishart $\mathrm{IW}_p(\delta, D)$ with density

$$p(\Sigma) \quad = \quad \frac{1}{I_p(\delta, D)}(\det\Sigma)^{-(\delta+2p)/2}\exp\left\{-\frac{1}{2}\langle\Sigma^{-1}, D\rangle\right\},$$

with respect to the Lebesgue measure on P_p [2]. It follows that the posterior distribution of K is $\mathrm{W}_p(\delta+n, D+U)$, while the posterior distribution of Σ is $\mathrm{IW}_p(\delta+n, D+U)$.

Our goal is to reduce the large set of candidate regressors to a smaller subset of explanatory variables that are robustly related to the variable of interest Y by explicitly modeling the multivariate dependency patterns among the observed variables X. These patterns are identified by constraining to zero some of the off-diagonal elements of K. The remaining elements of K are associated with edges in an undirected graph G which is called a Gaussian graphical model (GGM) – see Dempster [6] and Wermuth [43]. We discuss GGMs in our Bayesian framework in Section 3.1. In Section 3.2 we show that linear regressions as they are currently defined in the literature are a special type of GGMs.

3.1. Gaussian graphical models

Let $G = (V, E)$ be a GGM with $E = \{(i, j)|K_{ij} \neq 0\}$. The precision matrix K is now constrained to the cone P_G of symmetric positive definite matrices with entries K_{ij} equal to zero for all $(i, j) \notin E$. Let $\text{nbd}_G(Y) = \{i \geq 2|(1, i) \in E\}$ be the neighbors of Y in G. The local Markov property associated with G shows that Y is independent of the remaining explanatory variables given its neighbors $X_{\text{nbd}_G(Y)}$, which implies that the conditional of Y given $X_{\{2,\ldots,p\}}$ coincides with the conditional of Y given $X_{\text{nbd}_G(Y)}$ [23]. Therefore G reduces the set of explanatory variables that seem to be related with Y by distinguishing between direct and indirect associations. The variables in $\text{nbd}_G(Y)$ are directly related with Y. The variables that can be reached from Y by following paths in G of length at most two influence Y indirectly through some subset of $X_{\text{nbd}_G(Y)}$. The variables that cannot be reached from Y (i.e., that belong to a different connected component of G) are independent of Y.

Conditional on G, the Wishart prior $\text{W}_p(\delta, D)$ for K becomes a G-Wishart prior $\text{W}_G(\delta, D)$ defined on P_G [38, 2, 27]. We need to have $D^{-1} \in P_G$ that is satisfied if we choose $D = I_p$. The posterior distribution of K given G is $\text{W}_G(\delta + n, D^*)$. Here D^* coincides with $D + U$ on the diagonal and on elements associated with the edges of G, while the elements of $(D^*)^{-1}$ are set to zero for all the other entries. The marginal likelihood of G is the ratio of the normalizing constants of the G-Wishart posterior and prior:

$$p(x|G) \quad = \quad I_G(\delta + n, D^*)/I_G(\delta, D).$$

If G is decomposable [23] with cliques $\{C_1, \ldots, C_k\}$ and separators $\{S_2, \ldots, S_k\}$ the marginal likelihood can be explicitly calculated [38]:

$$p(x|G) = p(x_{C_1}) \prod_{j=2}^{k} [p(x_{C_j})/p(x_{S_j})]. \tag{7}$$

where x_A are the rows of the $p \times n$ observed data matrix x specified by the indices $A \subset V$, $|A|$ is the size of A and $p(x_A) = I_{|A|}(\delta + n, D_A^*)/I_{|A|}(\delta, D_A)$.

If G is not decomposable, numerical approximation methods for $p(x|G)$ have to be employed [38, 5, 2]. We use the Laplace approximation developed in Lenkoski and Dobra [25] for $I_G(\delta + n, D^*)$ because it is fast and accurate, and the Monte Carlo method of Atay-Kayis and Massam [2] for $I_G(\delta, D)$.

We assume throughout that the GGMs are apriori equally likely. Therefore the GGMs with

the highest posterior probability are those GGMs with the largest marginal likelihoods. Our framework can be easily used to accommodate other prior specifications on GGMs if such choices seem to be more suitable.

The joint posterior distribution of X given a set of graphs $\mathcal{S}$ is $\mathrm{N}_p(0, (K_{x,\mathcal{S}})^{-1})$ where the distribution of $K_{x,\mathcal{S}}$ is obtained by Bayesian model averaging [21] as a mixture of G-Wishart posterior distributions

$$p(K_{x,\mathcal{S}}) = \sum_{G \in \mathcal{S}} \mathrm{W}_G(\delta + n, D^*) p(G|x, \mathcal{S}), \tag{8}$$

with weights equal to the marginal likelihoods of the graphs normalized within $\mathcal{S}$:

$$p(G|x, \mathcal{S}) = p(x|G) / \left[\sum_{G' \in \mathcal{S}} p(x|G') \right].$$

Given an estimator $\widehat{K}_{x,\mathcal{S}} = (\widehat{K}_{ij})_{1 \le i,j \le p}$ of $K_{x,\mathcal{S}}$, the conditional posterior distribution of Y given $X_{(2:p)} = x_{(2:p)}$ is:

$$p(Y|X_{(2:p)} = x_{(2:p)}) = \mathrm{N}\left(-\sum_{i=2}^{p} \frac{\widehat{K}_{1i}}{\widehat{K}_{11}} x_i, \frac{1}{\widehat{K}_{11}} \right). \tag{9}$$

Here $(2 : p) = \{2, \ldots, p\}$. The relevance of the direct interaction between two variables X_i and X_j is given by the posterior inclusion probability of the edge (i, j) in the graphs $\mathcal{S}$ defined as the sum of $p(G|x, \mathcal{S})$ such as $G \in \mathcal{S}$ and G contains the edge (i, j). This posterior inclusion probability represents the posterior probability that X_i and X_j are conditionally independent given the rest of the variables $X_{V \setminus \{i,j\}} = x_{V \setminus \{i,j\}}$. From (8) we see that the variables X_i, $i \ge 2$, with a zero posterior inclusion probability of the edge $(1, i)$ have $\widehat{K}_{1i} = 0$ and consequently do not appear in the regression (9).

3.2. Linear regressions

We consider the regression model specified by the p-dimensional indicator vector γ_A with $A = \{i_1, \ldots, i_{|A|}\} \subseteq (2 : p)$. We have $\gamma_i = 1$ if X_i is in the regression model and $\gamma_i = 0$ otherwise. The only conditional independence assumption implied by the regression γ_A is that Y is conditionally independent of $X_{(2:p) \setminus A}$ given X_A, which implies:

$$p(Y|X_{(2:p)} = x_{(2:p)}) = p(Y|X_A = x_A). \tag{10}$$

The GGM $G^{(A)} = (V, E^{(A)})$, where

$$E^{(A)} = \{(i, j) : i > 1, j > 1\} \cup \{(1, i) : i \in A\},$$

implies (10) and does not imply any other conditional independence relationships that are not a direct consequence of (10). The graph $G^{(A)}$ is decomposable with two cliques $\{1\} \cup A$ and $(2 : p)$

and one separator A. From (7) it follows that the marginal likelihood of the regression γ_A is given by

$$p(x|\gamma_A) = p(x|G^{(A)}) = p(x_{\{1\}\cup A})p(x_{(2:p)})/p(x_A). \tag{11}$$

The term $p(x_{(2:p)})$ appears in the marginal likelihood of any regression and consequently it is not needed when comparing the values of the marginal likelihoods of two regression models. It follows that (11) becomes

$$p(x|\gamma_A) \quad \propto \quad \frac{I_{1+|A|}(\delta+n,(D+U)_{\{1\}\cup A})I_{|A|}(\delta,D_A)}{I_{1+|A|}(\delta,D_{\{1\}\cup A})I_{|A|}(\delta+n,(D+U)_A)}. \tag{12}$$

In particular, the marginal likelihood of the null regression that does not contain any predictors is given by

$$p(x|\gamma_\emptyset) \quad \propto \quad 2^{n/2}\frac{\Gamma\{(\delta+n)/2\}(D_{11}+U_{11})^{-(\delta+n)/2}}{\Gamma(\delta/2)(D_{11})^{-\delta/2}}. \tag{13}$$

Geiger and Heckerman [15] and Dobra et al. [7] show that the Wishart prior (5) induces consistent normal/inverse Gamma priors for the regression parameters, that lead to conjugate normal/inverse Gamma posterior distributions. See also Zellner [45] for related results. The corresponding marginal likelihood of regression models is again given by (12) or (13).

This implies that linear regression models are a particular case of decomposable graphical models since there is a one-to-one correspondence between the set of regressions $\{\gamma_A : A \subset (2:p)\}$ and the set of decomposable graphs $\{G^{(A)} : A \subset (2:p)\}$. We call these regression GGMs. The relevance of each X_i, $i \in (2:p)$, with respect to Y is given by its posterior inclusion probability defined as the sum of the model probabilities in which X_i appears, i.e. $i \in A$.

3.3. Regressions induced by GGMs

We have described three families of GGMs of increasing size and complexity: S_1 comprises the regression GGMs, S_2 comprises the decomposable GGMs (i.e., the GGMs whose conditional independence graphs are decomposable), while S_3 comprises all GGMs. The following inclusion relationships hold:

$$S_1 \subset S_2 \subset S_3.$$

These inclusions are strict if $p \geq 4$. To be more precise, there exist decomposable graphs with more than two cliques and hence $S_2 \setminus S_1 \neq \emptyset$. The graph

$$(\{1,2,3,4\},\{(1,2),(2,3),(3,4),(4,1)\})$$

is not decomposable, hence $S_3 \setminus S_2 \neq \emptyset$.

There is a unique regression graph in S_1 associated with a regression γ_A, $A \subset (2:p)$. On the other hand, it is likely that one, two, or possibly more GGMs in the other sets S_2 and S_3 lead to the same regression γ_A. These are graphs G such that $\mathrm{nbd}_G(Y) = A$. The richer the set of admissible

graphs, the more likely it is that more GGMs are associated with γ_A. The posterior probability of γ_A is therefore the sum of the posterior probabilities of all the graphs that induce it.

The graphs in S_1 embed the implicit assumption that any pair of explanatory variables X_i and X_j are not conditionally independent given the rest. By relaxing this assumption we consider various possible patterns of dependencies among $X_{(2:p)}$ and consequently obtain a more accurate measure of the relative relevance of each regression model.

4. Jointness measures

The relevance of each predictor X_i, $i \geq 2$ with respect to the outcome Y is measured through the posterior inclusion probability of X_i in a regression for Y. In the context of GGMs, this is the posterior probability of the undirected edge $(1, i)$ since it represents the probability that X_i is in the set of neighbors of Y which, in turn, are the regressors present in the implied regression for Y. As an aside, the posterior probability that X_i is a regressor for Y equals the posterior probability that Y is a regressor for X_i.

Doppelhofer and Weeks [8] as well as Ley and Steel [29] raise the question of how to measure the co-occurence (or jointness) of two explanatory variables X_i and X_j in the context of linear regressions. In particular, Ley and Steel [29] argue that any useful jointness measure should satisfy four criteria: (i) *interpretability*: any jointness measure should have either a formal statistical or a clear intuitive meaning in terms of jointness; (ii) *calibration*: values of the jointness measure should be calibrated against some clearly defined scale, derived from either formal statistical or intuitive arguments; (iii) *extreme jointness*: the situation where two variables always appear together should lead to the jointness measure reaching its value reflecting maximum jointness; and (iv) *definition*: the jointness measure should always be defined whenever at least one of the variables considered is included with positive probability. Ley and Steel [29] propose two jointness measures as follows:

$$\begin{aligned} \mathcal{J}^*_{ij} &= \frac{p(i \cap j)}{p(i) + p(j) - p(i \cap j)} \in [0, 1], \\ \mathcal{J}_{ij} &= \frac{p(i \cap j)}{p(i) + p(j) - 2p(i \cap j)} \in [0, \infty). \end{aligned}$$

Here $p(i \cap j)$ is the sum of the posterior probabilities of the regression models that contain both X_i and X_j, while $p(i)$ and $p(j)$ are the posterior inclusion probabilities of X_i and X_j, respectively. The disjointness of X_i and X_j is the reciprocal of $\mathcal{J}^*_{ij}$ or $\mathcal{J}_{ij}$ [29].

One can evaluate $\mathcal{J}^*_{ij}$ and $\mathcal{J}_{ij}$ for GGMs by considering the set of neighbors of Y in each corresponding undirected graph. In this case $p(i)$ is the posterior probability of the edge $(1, i)$, while $p(i \cap j)$ represents the sum of the posterior probabilities of the graphs in which both the undirected edges $(1, i)$ and $(1, j)$ appear. Note that $p(i \cap j)$ should not be confused with the posterior inclusion probability of the edge (i, j) which is equal with the posterior probability that X_i and X_j are independent given the remaining variables [23].

5. Simulated data

This section considers two simulation studies that bear out the utility of considering GGMs as opposed to regression models when performing variable selection. We show that, given the same prior specification on model parameters, we obtain better results by considering graphs whose structure is less restricted than the structure of regression graphs. In the first simulation study, variable selection is compounded by a very complex pattern of correlation among the possible predictors. In the second study, the data are generated by various GGMs that are not regression graphs.

5.1. First simulated example

This is a variation of the example suggested by Nott and Green [35]. We generate $Z_i \sim N_{300}(0, I_{300})$, $i = 1, 2, \ldots, 15$. Let $X_i = Z_i + 2Z$, $i = 1, 3, 5, 8, 9, 10, 12, 13,14, 15$, $X_2 = X_1 + 0.15Z_2$, $X_4 = X_3 + 0.15Z_4$, $X_6 = X_5 + 0.15Z_6$, $X_7 = X_8 + X_9 - X_{10} + 0.15Z_7$ and $X_{11} = X_{14} + X_{15} - X_{12} - X_{13} + 0.15Z_{11}$. George and McCulloch [16] point out that this design matrix leads to correlations of about 0.998 between X_i and X_{i+1} for $i = 1, 3, 5$. There are also strong linear associations between (X_7, X_8, X_9, X_{10}) and $(X_{11}, X_{12}, X_{13}, X_{14}, X_{15})$. We let $\tilde{X} = [X^{(1)}X^{(2)}]$ be a 300×30 design matrix obtained by independently simulating two instances $X^{(1)}$ and $X^{(2)}$ of the 300×15 design matrix X. Consider the 30-dimensional vector of regression coefficients β defined by $\beta_j = 1.5$, if $j = 1, 3, 5, 7, 11, 12, 13$, $\beta_8 = -1.5$ and $\beta_j = 0$ otherwise. We generate the response vector as $Y = \tilde{X}\beta + \epsilon$ where $\epsilon \sim N_{300}(0, 2.5 \cdot I_{300})$.

We perform two separate searches over the space of regression graphs and over the space of decomposable GGMs. The interest in both cases is to study whether predictors that belong to $X^{(2)}$ are not selected. Due to the complex correlation structure amongst the predictors in $X^{(1)}$, variables may be selected even if their true regression coefficients are zero. We explored the space of candidate graphs using the Markov chain Monte Carlo model composition (MC^3) algorithm of Madigan and York [31]). We performed 25000 iterations with a burn-in time of 5000 iterations. In order to reduce the sampling variability, we report the results we obtained by averaging across 100 replicates of this experiment.

Figure 2 shows the mean posterior inclusion probabilities for each variable across the 100 simulated datasets. For both searches all fifteen covariates in $X^{(1)}$ have non-negligible average posterior inclusion probabilities, while the posterior inclusion probabilities of the variables in $X^{(2)}$ are consistently low. However, in the decomposable GGMs search, the variables in $X^{(2)}$ receive essentially zero inclusion probability, therefore completely eliminating variables from this block. Furthermore, the inclusion probabilities of the first fifteen variables are larger in the decomposable GGM search than in the regression search.

5.2. Second simulated example

The second simulation study considers the case when the data are generated from GGMs whose underlying conditional independence graphs have ten vertices and edges defined as follows:

- Independence: G_1 does not have any edges.

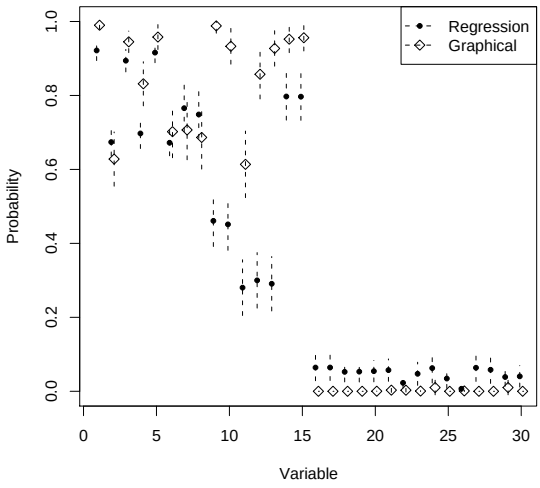

Figure 2: Average posterior inclusion probabilities and (95%) confidence bands by variable for the simulation study described in Section 5.1.

- AR(1): the edges of G_2 are $\{(i-1, i) : i \geq 2\} \cup \{(i, i+1) : i \leq 9\}$.
- AR(2): G_3 contains the edges of G_2 and also $\{(i-2, i) : i \geq 3\} \cup \{(i, i+2) : i \leq 8\}$.
- Circle: the edges of G_4 are $\{(1, 10)\} \cup \{(i, i+1) : i \leq 9\}$.

These four models were chosen because they have a significant degree of sparsity in their structure. To frame the simulation study in a regression context, we take variable 1 to be the "response" and variables 2 to 10 to be candidate explanatory variables. The true regression implied by G_1 for variable 1 is the regression with no predictors. The true regression implied by G_2 involves variable 2, the true regression implied by G_3 involves variables 2 and 3, while the true regression implied by G_4 involves variables 2 and 10.

Interest again lies in comparing the results of a graphical model search to a regression search, in particular the inclusion probabilities linking variable 1 to the remaining variables. For each graph we generated 100 datasets of 100 observations each, and employed MC^3 to perform two separate searches in the space of unrestricted GGMs and the space of regression models. In both cases MC^3 was run for 25000 iterations with a burn-in of 5000 iterations.

The results of the simulation study are consistent across the four graphs G_1, G_2, G_3, G_4: when considering either unrestricted GGMs or regression models, the variables linked to the "response" in the true graph are correctly given high inclusion probabilities – see Figure 3. However, the unrestricted GGM search consistently places lower inclusion probability on variables that are not connected to variable 1. This phenomenon is perhaps most striking in the circle graph G_4.

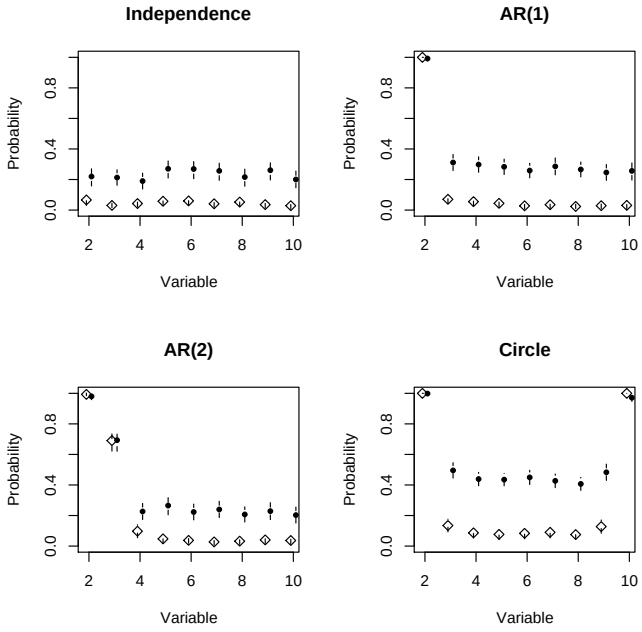

Figure 3: Average posterior inclusion probabilities and (95%) confidence bands by variable for the four models described in Section 5.2. As in Figure 2, the dots represent the mean inclusion probability for each variable across 100 datasets for the regression graph search, while the diamonds give these probabilities for the unrestricted GGM search.

6. Results: FLS data

Our analysis of the FLS growth data proceeds as follows. We performed an initial linear regression search involving the entire set of potential growth determinants. We run five instances of MC^3 starting from random regressions. Each instance comprised 250000 iterations with a burn-in of 25000 iterations. There are 28 growth determinants with a posterior inclusion probability greater than 0.05 – see the column "Regression" of Table 1 and 17 variables with posterior inclusion probability greater than 0.5.

We compare our findings with those from Eicher et al. [11] who performed an extensive sensitivity analysis of Bayesian model averaging in linear regressions across 12 diffuse priors. Eicher et al. [11] identified between 7 and 22 regressors having posterior inclusion probabilities greater than 0.5. Sixteen growth determinants we identified having inclusion probability greater than 0.5 in this initial search have been discovered before in the literature – see Table 2 in Eicher et al. [11]. Percent expenditure on public education (PublEdu) is a new potential predictor. It is a fundamental tenant of the new growth theory and therefore a welcomed addition to the set of growth determinants. Interestingly, most of the other regressors that are excluded (French, Spanish, and British colonial history) are regressors that are often introduced to instrument for the endogeneity of institutions. They are certainly not thought of as direct causal growth determinants for the 1960-1990 period. Political Rights is an additional regressor that is suggested by a minority of priors. It is no surprise to see this excluded because of the high collinearity with "Rule of Law" and it is commonly expected that one of these political regressors is identified as relevant. More

importantly, the Wishart prior for model parameters we are employing is one of only four priors that discover key growth variables public education and the degree of capitalism (EcoOrg) that are considered fundamental to growth. The posterior means of the regressors are all within the same range uncovered in Eicher et al. [11].

Table 1: Posterior inclusion probabilities ($p(i)$) and regression coefficient estimates ($\hat{\beta}_i$) by search type. Standard deviations of the estimates are given in parentheses. The values of LabForce were divided by $100,000$ so that all the coefficients are on a similar scale. Since LatAmerica, SubSahara, SpanishCol and OutwarOr are binary variables, they are excluded from the GGM search and results are listed as "NA".

Variable	Regression		GGMs		Variable	Regression		GGMs	
	$p(i)$	$\hat{\beta}_i$	$p(i)$	$\hat{\beta}_i$		$p(i)$	$\hat{\beta}_i$	$p(i)$	$\hat{\beta}_i$
Life	1	7e-04	1	0.001	LatAmerica	0.74	-0.006	NA	NA
		(2e-04)		(2e-04)			(0.0047)		(NA)
GDPsh560	1	-0.0136	1	-0.0139	EthnoLFrac	0.722	0.0068	0	0
		(0.0021)		(0.0023)			(0.0055)		(0)
EquipInv	1	0.1162	1	0.2018	PublEdu	0.651	0.1431	0	0
		(0.0368)		(0.0382)			(0.1262)		(0)
SubSahara	1	-0.0188	NA	NA	PrSc	0.42	0.0059	0.035	3e-04
		(0.0044)		(NA)			(0.009)		(0.0019)
Confucious	1	0.0638	1	0.0527	Age	0.397	0	0	0
		(0.0109)		(0.0119)			(0)		(0)
RuleofLaw	1	0.0113	0.274	0.0018	OutwarOr	0.34	-0.0011	NA	NA
		(0.0036)		(0.0036)			(0.0019)		(NA)
Mining	0.981	0.0352	0	0	CivlLib	0.33	-5e-04	1	-0.0027
		(0.0119)		(0)			(9e-04)		(8e-04)
Protestants	0.968	-0.0125	1	-0.0153	Buddha	0.328	0.0032	0.086	9e-04
		(0.0048)		(0.004)			(0.0053)		(0.0032)
Hindu	0.943	-0.056	0	0	Muslim	0.32	0.0036	0	0
		(0.0225)		(0)			(0.0059)		(0)
EcoOrg	0.936	0.002	0	0	PolRights	0.282	-3e-04	0	0
		(8e-04)		(0)			(7e-04)		(0)
NEquipInv	0.91	0.0423	0	0	Catholic	0.236	-0.0014	1	-0.0073
		(0.021)		(0)			(0.003)		(0.0021)
LabForce	0.898	0.1819	0	0	SpanishCol	0.234	0.0011	NA	NA
		(0.0867)		(0)			(0.0026)		(NA)
BlMktPm	0.894	-0.0071	0	0	Popg	0.059	0.0084	0	0
		(0.0035)		(0)			(0.0451)		(0)
HighEnroll	0.873	-0.0735	0	0	PrExports	0.057	-4e-04	0	0
		(0.0386)		(0)			(0.002)		(0)

We conduct a more extensive analysis of the 28 regressors displayed in Table 1, which have posterior inclusion probabilities greater than 0.05. Among this reduced set of regressors, four are binary. The presence of discrete variables cannot be incorporated into the GGM framework, and therefore the variables must also be eliminated (see Section 7 for a discussion of future work related to this issue) from the reduced dataset, leaving 24 covariates. With this smaller dataset, we then run MC^3 for GGMs, using the same settings as before.

Table 1 is indicative of the parameter reduction properties discussed in Section 1. For the most part, as more refined searches are considered, variables with lower inclusion probabilities are recognized to be conditionally independent of the response and a smaller set of core variables is ultimately returned. Table 1 shows that CivlLib and Catholic begin with relatively low inclusion probabilities in the regression search and ultimately receive probabilities of one when the model

space is enriched. We reduced the set of potential covariates from 42, to 28 and ultimately to 10, not by artificially adjusting prior parameters to favor greater sparsity, but by expanding the richness of the model space to resemble conditional independencies that are more likely to exist in economic data.

An interesting feature of the results reported in Table 1 is that inclusion probabilities appear to be pushed to either near zero, or near one. This follows from the richness of the model space considered in the GGM search. In the course of the MC^3 search, over 15000 different GGMs were visited. All these graphs included the same set of core variables linked to growth but were otherwise different in terms of the edges linking two other variables.

The resulting 10 variables represent an astounding support for the Neoclassical growth model that dominated economic growth from the 1950s to the 1980s. Life expectancy, initial GDP and Equipment investment are the fundamental variables of that model. All other regressors identify characteristics of the economy and can be grouped into pure exogenous and broadly speaking "policy variables". Rule of Law and Civil Liberties are variables that constitutions and institutions can affect, while purely exogenous religious identifiers such as protestant, confucious and catholic, are the core fundamental variables that seem to co-vary with many of the variables suggested by the new growth theory. The insight here is that the large set of economic regressors speaking to education, mining, market premia, or trade, for example are not the fundamental causes of growth, but they co-vary with it. Once purely exogenous religious and policy variables are accounted for, the growth determinants have been established.

Table 2 gives estimates of the jointness measures $\mathcal{J}_{ij}$ of Ley and Steel [29] from the regression search, and the GGM search. In order to keep the size of the table manageable, we have only displayed the jointness calculations between the seven covariates that have inclusion probabilities above .5 in the undirected GGM search. Jointness measures attempt to determine the extent to which pairs of variables act together to "jointly" affect the response in a regression setting. The first column of Table 2 shows that the 21 pairs have varying degrees of jointness in the regression search. However, after the GGM search, jointness calculations show a significantly different interpretation. In the course of the GGM searches, a number of GGMs are visited. Each of these typically differs in links between the covariates, particularly the manner in which these seven covariates interact with the remaining variables not connected to growth. However, in the GGM search, the covariates considered in Table 2 are linked directly to growth, and therefore every pair shown receives a $\mathcal{J}_{ij}$ value of $+\infty$.

We contrast our findings with the results of Ley and Steel [29] who classify the posterior odds of jointness $\mathcal{J}_{ij}$ as conveying positive, strong, very strong or decisive evidence of jointness when $\mathcal{J}_{ij}$ exceeds 3, 10, 30 or 100, respectively. Ley and Steel [29] argue that "only 8 pairs (1% of the total) display some degree of evidence of jointness" and determine that only the pair Confucious and GDPsh560 (initial GDP) exhibit decisive evidence for jointness in a regression search. We find in Table 2 that in the GGM decisive evidence for jointness is much more prevalent especially among the fundamental growth regressors that we have identified. While we used the same data and the same measure for jointness, our results differ from Ley and Steel [29] because the dependency constraints among candidate growth factors are explicitly taken into account, which highlights the differences between the methods and the insights that can be derived using a GGM instead of a

regression search.

Table 2: Comparison of jointness measures by search type. We give estimates of $\mathcal{J}_{ij}$ and their interpretation for the 21 pairs involving the 7 growth determinants identified using GGMs with posterior inclusion probability greater than 0.5.

Variable 1	Variable 2	Regression		GGMs	
		$\mathcal{J}_{ij}$	Interpretation	$\mathcal{J}_{ij}$	Interpretation
Life	GDPsh560	∞	Decisive	∞	Decisive
Life	EquipInv	∞	Decisive	∞	Decisive
Life	Confucious	∞	Decisive	∞	Decisive
GDPsh560	EquipInv	∞	Decisive	∞	Decisive
GDPsh560	Confucious	∞	Decisive	∞	Decisive
EquipInv	Confucious	∞	Decisive	∞	Decisive
Life	Protestants	29.88	Strong	∞	Decisive
GDPsh560	Protestants	29.88	Strong	∞	Decisive
EquipInv	Protestants	29.88	Strong	∞	Decisive
Confucious	Protestants	29.88	Strong	∞	Decisive
Life	CivlLib	0.49	Little	∞	Decisive
GDPsh560	CivlLib	0.49	Little	∞	Decisive
EquipInv	CivlLib	0.49	Little	∞	Decisive
CivlLib	Confucious	0.49	Little	∞	Decisive
CivlLib	Protestants	0.43	Little	∞	Decisive
Catholic	Protestants	0.32	Little	∞	Decisive
Life	Catholic	0.31	Little	∞	Decisive
GDPsh560	Catholic	0.31	Little	∞	Decisive
EquipInv	Catholic	0.31	Little	∞	Decisive
Catholic	Confucious	0.31	Little	∞	Decisive
CivlLib	Catholic	0.18	Little	∞	Decisive

7. Conclusions

The methodology proposed in this paper allows the identification of a reduced set of growth determinants by modeling joint distributions of the observed variables. The linear regressions search is proven to provide good initial results that are further refined when considering Gaussian graphical models. We showed that relaxing the constraints on the dependency patterns of the candidate growth determinants leads to a significant decrease in the set of determinants that are ultimately selected. We emphasize that this overarching idea to variable selection should be viewed as the main contribution of our work. The choice of priors for model parameters we used seem to work well and allow the development of a coherent framework that could be adapted to other prior specifications.

The relevance of our work can be understood by carefully examining the structure of the priors for regression parameters proposed in first-rate papers focusing on growth regression. For example, Fernández et al. [14, 13] consider automatic priors based on the g-prior of Zellner [46]. Posterior model probabilities can be sensitive to prior specifications, which is particularly relevant in the growth context when the researcher must entertain a large number of sampling models with little or no subjective prior information. For this case Fernández et al. [14, 13] developed an automatic, partly non-informative g-prior structure where the amount of subjective information requested from the user is limited to the choice of a single scalar hyperparameter g. The choice is

automatic because Fernández et al. [14, 13] simplify matters even further, by facilitating the choice of the hyperparameter g as a function of only the model size and the number of observations. An alternative automatic data dependent prior was proposed in Raftery et al. [36].

The g-priors include in their specification the full sample covariance matrix. As such, they implicitly assume that the observed data can be summarized in a covariance matrix with no constraints imposed on its structure. This assumption is consistent with the multivariate normal assumption required by the Gaussian graphical models as it is explicitly stated in Section 3. The Gaussian graphical models induce parsimony in the structure of the covariance matrix by identifying the most relevant conditional independence relationships. This is why we are able to identify fewer growth determinants while employing the same prior specification.

The multivariate normal assumption for the joint distribution of the observed variables is quite restrictive and does not allow the presence of discrete variables. This serious limitation can be addressed with the Copula Gaussian graphical models (CGGMs) recently developed by Dobra and Lenkoski [26]. The CGGMs embed graphical model selection inside a semiparametric Gaussian copula. Copulas [34] provide the theoretical framework in which multivariate associations can be modeled separately from the univariate distributions of the observed variables. The CGGMs treat the univariate marginal distributions as nuisance parameters and impose a multivariate normal assumption only for a set of latent variables, thereby allowing the presence of binary and ordinal variables in addition to continuous variables. For each observed variable, the CGGMs include a latent variable, and vice versa. We plan to make use of the CGGMs in the context of growth empirics and to develop other extensions of our modeling framework.

Appendix

A graph G is a pair (V, E), where V is a finite set of vertices and $E \subset V \times V$ is a set of edges linking the vertices. In an undirected graph $(v_1, v_2) \in E$ implies $(v_2, v_1) \in E$. The subgraph of G induced by $A \subset V$ is $G(A) = (A, E(A))$, where $E(A) = \{(v_1, v_2) \in E | v_1, v_2 \in A\}$. Two vertices $v_1, v_2 \in V$ are adjacent if $(v_1, v_2) \in E$. A subgraph $G(A)$ is complete if the vertices in A are pairwise adjacent in G. We also say that A is complete in G. A clique is a complete vertex set that is maximal, that is, there does not exist another complete vertex set that includes it.

A path links two vertices by a sequence of edges. The path is a cycle if its end points coincide. Two vertices are connected if they are linked by a path, otherwise they are disconnected. The connected component of a vertex is the set of all vertices connected with it. A set $A \subset V$ is an $v_1 v_2$-separator if all the paths from v_1 to v_2 intersect A. A vertex set A separates A_1 from A_2 if it is a $v_1 v_2$-separator for every $v_1 \in A_1$, $v_2 \in A_2$. A vertex set A is a separator of G if two vertices in the same connected component of G are in two distinct connected components of $G(V \setminus A)$. Moreover, A is a minimal separator if no proper subset of A separates G.

The partition (A_1, S, A_2) of V is a decomposition of G if S is a minimal separator of A_1 and A_2. The graph G is decomposable if it is complete or if it has a proper decomposition (A_1, S, A_2) and the subgraphs $G(A_1 \cup S)$ and $G(A_2 \cup S)$ are also decomposable. Any decomposable graph has a unique set of cliques and a unique set of separators. Other relevant graph theory results can be found in Ch. 2 of Lauritzen [23].

References

[1] Acosta-González, E., Fernández-Rodríguez, F., 2007. Model selection via genetic algorithms illustrated with cross-country growth data. Empirical economics 33, 313–337.

[2] Atay-Kayis, A., Massam, H., 2005. A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. Biometrika 92, 317–35.

[3] Brock, W. A., Durlauf, S. N., 2001. Growth empirics and reality. World Bank Economic Review 15, 229–272.

[4] Broeck, S., Binder, M., 2005. A re-examination of the determinants of economic growth using simultaneous equation dynamic panel data models. Computing in Economics and Finance 2005, Technical Report no. 343, Society for Computational Economics.
URL `http://ideas.repec.org/p/sce/scecf5/343.html`

[5] Dellaportas, P., Giudici, P., Roberts, G., 2003. Bayesian inference for nondecomposable graphical Gaussian models. Sankhya 65, 43–55.

[6] Dempster, A. P., 1972. Covariance selection. Biometrics 28, 157–75.

[7] Dobra, A., Hans, C., Jones, B., Nevins, J. R., Yao, G., West, M., 2004. Sparse graphical models for exploring gene expression data. J. Multivariate Anal. 90, 196–212.

[8] Doppelhofer, G., Weeks, M., 2009. Jointness of growth determinants. Journal of Applied Econometrics 24, 209–244.

[9] Durlauf, S. N., Johnson, P., Temple, J., 2005. Growth econometrics. In: Handbook of Economic Growth. Vol. 1A. P. Aghion and S. N. Durlauf, eds., North Holland: Amsterdam, pp. 555–677.

[10] Durlauf, S. N., Kourtellos, A., Tan, C. M., 2008. Are any growth theoris robust? The Economic Journal 118, 329–346.

[11] Eicher, T. S., Papageorgiou, C., Raftery, A. E., 2009. Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. Journal of Applied EconometricsDOI: 10.1002/jae.1112.

[12] Eicher, T. S., Papageorgiou, C., Roehn, O., 2007. Unraveling the fortunes of the fortunate: an iterative Bayesian model averaging (IBMA) approach. Journal of Macroeconomics 29, 494–514.

[13] Fernández, C., Ley, E., Steel, M. F. J., 2001. Benchmark priors for Bayesian model averaging. Journal of Econometrics 100, 381–427.

[14] Fernández, C., Ley, E., Steel, M. F. J., 2001. Model uncertainty in cross-country growth regressions. Journal of Applied Econometrics 16, 563–576.

[15] Geiger, D., Heckerman, D., 2002. Parameter priors for directed acyclic graphical models and the characterization of several probability distributions. Ann. Statist. 30, 1412–1440.

[16] George, E., McCulloch, R., 1997. Approaches for Bayesian variable selection. Statistica Sinica 7, 339–373.

[17] George, E. I., 1999. Sampling considerations for model averaging and model search. In: J. M. Bernardo, J. O. Berger, A. P. D., Smith, A. F. M. (Eds.), Bayesian Statistics. Vol. 6. Oxford University Press, pp. 175–177, invited discussion of "Bayesian model averaging and model search strategies" by M. A. Clyde.

[18] Heckerman, D., Geiger, D., Chickering, D., 1995. Learning Bayesian networks: the combination of knowledge and statistical data. Machine Learning 20, 197–243.

[19] Hendry, D. F., Krolzig, H.-M., 2004. We ran one regression. Oxford Bulletin of Economics and Statistics 66, 799–810.

[20] Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., West, M., 2005. Experiments in stochastic computation for high-dimensional graphical models. Statist. Sci. 20, 388–400.

[21] Kass, R., Raftery, A. E., 1995. Bayes factors. J. Am. Statist. Assoc. 90, 773–95.

[22] Kennedy, P., 1998. A Guide to Econometrics. The MIT Press.

[23] Lauritzen, S. L., 1996. Graphical Models. Oxford University Press.

[24] Leamer, E. E., 1978. Specification Searches: Ad Hoc Inference with Nonexperimental Data. Wiley, New York.

[25] Lenkoski, A., Dobra, A., 2008. Bayesian structural learning and estimation in Gaussian graphical models, Technical Report no. 545, Department of Statistics, University of Washington.

[26] Lenkoski, A., Dobra, A., 2009. Copula Gaussian graphical models, Technical Report no. 555, Department of Statistics, University of Washington.

[27] Letac, G., Massam, H., 2007. Wishart distributions for decomposable graphs. Ann. Statist. 35, 1278–323.

[28] Levine, R., Renelt, D., 1992. A sensitivity analysis of cross-country growth regressions. American Economic Review 82, 942–963.

[29] Ley, E., Steel, M. F., 2007. Jointness in Bayesian variable selection with applications to growth regression. Journal of Macroeconomics 29, 476–493.

[30] Ley, E., Steel, M. F., 2009. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. Journal of Applied Econometrics 24, 651–674.

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

[32] Masanjala, W. H., Papageorgiou, C., 2008. A rough and lonely road to prosperity: a reexamination of sources of growth in Africa using Bayesian model averaging. Journal of Applied Econometrics 23, 671–682.

[33] Muirhead, R. J., 2005. Aspects of Multivariate Statistical Theory. John Wiley & Sons.

[34] Nelsen, R. B., 1999. An Introduction to Copulas. New York: Springer-Verlag.

[35] Nott, D., Green, P., 2004. Bayesian variable selection and the Swendsen-Wang algorithm. Journal of Computational and Graphical Statistics 13, 1–17.

[36] Raftery, A. E., Madigan, D., Hoeting, J. A., 1997. Bayesian model averaging for linear regression models. J. Am. Statist. Assoc. 92, 179–91.

[37] Romer, P., 1986. Increasing returns and long-run growth. Journal of Political Economy 94, 1002–1037.

[38] Roverato, A., 2002. Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. Scand. J. Statist. 29, 391–411.

[39] Sala-i Martin, X., 1997. I just ran two million regressions. AEA Papers and Proceedings 87, 178–183.

[40] Sala-i Martin, X., Doppelhofer, G., Miller, R. I., 2004. Determinants of long-term growth: a Bayesian averaging of classical estimates (BACE) approach. American Economic Review 94, 813–835.

[41] Solow, R. A., 1956. A contribution to the theory of economic growth. Quarterly Journal of Economics 70, 65–94.

[42] Strachan, R. W., 2007. Comment on "Jointness of growth determinants" by E. Ley and M. F. Steel. Journal of Macroeconomics 29.

[43] Wermuth, N., 1976. Analogies between multiplicative models in contigency tables and covariance selection. Biometrics 32, 95–108.

[44] Whittaker, J., 2009. Graphical Models in Applied Multivariate Statistics. Wiley.

[45] Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.

[46] Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distribution. In: Goel, P. K., Zellner, A. (Eds.), Bayesian Inference and Decision Techniques: essays in honour of Bruno de Finetti. Amsterdam, North Holland, pp. 233–243.