Modeling uncertainty in macroeconomic growth determinants using Gaussian graphical models [1]

Adrian Dobra
University of Washington, Seattle

Theo S. Eicher
University of Washington, Seattle

Alex Lenkoski
University of Washington, Seattle

[1]Adrian Dobra is Assistant Professor of Statistics and Nursing, Department of Statistics and Biobehavioral Nursing and Health Systems, University of Washington. E-mail: adobra@stat.washington.edu; Web: www.stat.washington.edu/adobra. Theo S. Eicher is Professor and Robert R. Richards Distinguished Scholar, Department of Economics, University of Washington. E-mail: te@u.washington.edu; Web: http://faculty.washington.edu/~te. Alex Lenkoski is Graduate Student, Department of Statistics, University of Washington. E-mail: lenkoski@stat.washington.edu

MODELING UNCERTAINTY IN MACROECONOMIC GROWTH DETERMINANTS USING GAUSSIAN GRAPHICAL MODELS

By Adrian Dobra, Theo S. Eicher and Alex Lenkoski

University of Washington

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 the utilized in numerous previous analyses to account for model uncertainty as well as jointness.

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).

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 (25), subsequent approaches have focused on Bayesian Model Averaging (BMA) to resolve model uncertainty – see Eicher et al

1

(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 dillusion 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 (24). 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 (23). 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 convariates. 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 therefore 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 (19; 21).

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; 26), that is consistent between the two approaches. The method for scoring models – either regression or graphical – therefore relies on a single common prior parameterization. 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) (here-after 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 method-ological advances that investigate model uncertainty in growth regressions – extreme bound analysis in Levine and Renelt (28), Bayesian model av-eraging, 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 (20), to explore panel regressions (1), or to investigate parameter hetero-geneity (33; 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 con-sistent with the current literature. We then consider a GGM search and show that by enriching the model space to allow for more complicated inter-action 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 develop our Bayesian statistical framework for performing variable selection using GGMs, while in Section 3 we discuss a stochastic search algorithm for GGMs called the mode oriented stochastic search (MOSS). 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 per-forms in situations of highly correlated candidate predictors and parsimo-nious interaction structures. In Section 6 we illustrate our methodology in the analysis of the FLS data. In Section 7 we conclude.

2. Variable selection with Gaussian graphical models. Formally, we assume that the observed data $x = (x^{(1)}, \ldots, x^{(n)})^T$ are independent ran-

dom samples from a p-dimensional multivariate normal distribution $\mathrm{N}_p(0, \Sigma)$ with $\Sigma = (\Sigma_{ij})_{1 \leq i,j \leq p}$. The response variable Y is associated with the first component of the random vector $X = (X_1, \ldots, X_p)$, while the remaining components are the candidate explanatory covariates. Let $V = \{1, 2, \ldots, p\}$. The likelihood function is proportional to

$$(2.1) \qquad L(x|\Sigma) \propto (\det \Sigma)^{-n/2} \exp\left\{ -\frac{1}{2}\langle \Sigma^{-1}, U \rangle \right\},$$

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 (2.1). Its density is

$$(2.2) \qquad p(K) \;=\; \frac{1}{I_p(\delta, D)}(\det K)^{(\delta-2)/2} \exp\left\{ -\frac{1}{2}\langle K, D \rangle \right\},$$

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

$$(2.3) \quad I_p(\delta, D) \;=\; 2^{(\delta+p-1)p/2}\Gamma_p\left\{(\delta + p - 1)/2\right\}(\det D)^{-(\delta+p-1)/2},$$

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 $a > (p-1)/2$ (34). By comparing (2.1) and (2.2), 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

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

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 2.1. In Section 2.2 we show that linear regressions as they are currently defined in the literature are a special type of GGMs.

2.1. *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 $\mathrm{nbd}_G(Y) = \{i \geq 2|(1,i) \in E\}$ 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 $X_{\mathrm{nbd}_G(Y)}$, which implies that the conditional of Y given $X_{\{2,\ldots,p\}}$ coincides with the conditional of Y given $X_{\mathrm{nbd}_G(Y)}$ (24). 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 $\mathrm{nbd}_G(Y)$ 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 $X_{\mathrm{nbd}_G(Y)}$. 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 $W_p(\delta, D)$ for K becomes a G-Wishart prior $W_G(\delta, D)$ 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 $W_G(\delta + n, D^*)$. 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:

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

If G is decomposable (24) with cliques $\{C_1, \ldots, C_k\}$ and separators $\{S_2, \ldots, S_k\}$ the marginal likelihood can be explicitly calculated (38):

$$(2.4) \qquad p(x|G) = p(x_{C_1}) \prod_{j=2}^{k} [p(x_{C_j})/p(x_{S_j})].$$

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 (26) 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 this paper 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 specification on GGMs if such choices seem to be more suitable for some reason. In Section 3 we describe a stochastic search algorithm called MOSS that identifies a subset of GGMs $\mathcal{S}$ having the highest posterior probabilities.

The joint posterior distribution of X given the 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 (22) as a mixture of G-Wishart posterior distributions

$$(2.5) \qquad p(K_{x,\mathcal{S}}) = \sum_{G \in \mathcal{S}} \mathrm{W}_G(\delta + n, D^*)p(G|x, \mathcal{S}),$$

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

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

Sampling from the mixture (2.5) as well as estimation of $K_{x,\mathcal{S}}$ is discussed in Lenkoski and Dobra (26). Given an estimator $\widehat{K}_{x,\mathcal{S}} = (\widehat{K}_{ij})_{1 \leq i,j \leq p}$ of $K_{x,\mathcal{S}}$, the conditional posterior distribution of Y given $X_{(2:p)} = x_{(2:p)}$ is:

$$(2.6) \qquad p(Y|X_{(2:p)} = x_{(2:p)}) = \mathrm{N}\left(-\sum_{i=2}^{p} \frac{\widehat{K}_{1i}}{\widehat{K}_{11}}x_i, \frac{1}{\widehat{K}_{11}}\right).$$

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 ($2.5$) we see that the variables X_i, $i > 2$, with a zero posterior inclusion probability of the edge $(1, i)$ have $\widehat{K}_{1i} = 0$ and consequently do not appear in the regression (2.6).

2.2. *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 $X_{(2:p)\backslash A}$ given X_A, which implies:

$$(2.7) \qquad\qquad p(Y|X_{(2:p)} = x_{(2:p)}) = p(Y|X_A = x_A).$$

There is a unique GGM $G^{(A)} = (V, E^{(A)})$, where

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

that implies the same conditional independence relationship and does not imply any other conditional independence relationships that are not a direct consequence of (2.7). The graph $G^{(A)}$ is decomposable with two cliques $\{1\}\cup A$ and $(2:p)$ and one separator A. From (2.4) it follows that the marginal likelihood of the regression γ_A is given by

$$(2.8) \qquad p(x|\gamma_A) = p(x|G^{(A)}) = p(x_{\{1\}\cup A})p(x_{(2:p)})/p(x_A).$$

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 (2.8) becomes

$$(2.9) \qquad p(x|\gamma_A) \;\; \propto \;\; \frac{I_{1+|A|}(\delta + n, (D+U)_{\{1\}\cup A})I_{|A|}(\delta, D_A)}{I_{1+|A|}(\delta, D_{\{1\}\cup A})I_{|A|}(\delta + n, (D+U)_A)}.$$

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

$$(2.10) \qquad p(x|\gamma_\emptyset) \;\; \propto \;\; 2^{n/2}\frac{\Gamma\{(\delta + n)/2\}(D_{11} + U_{11})^{-(\delta+n)/2}}{\Gamma(\delta/2)(D_{11})^{-\delta/2}}.$$

Geiger and Heckerman (15) and Dobra et al. (7) show that the Wishart prior (2.2) 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 (2.9) or (2.10).

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 graphs. We let $\mathcal{S}$ be the set of regressions identified by MOSS. The revelance 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 $p(\gamma_A|x, \mathcal{S}) \propto p(G^{(A)}|x, \mathcal{S})$ in which X_i appears, i.e. $i \in A$.

2.3. *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, 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 $\{(1, 2), (2, 3), (3, 4), (4, 1)\}$ 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.

It is important to remark that the graphs 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 relavance of each regression model.

3. Stochastic search for GGMs. The Bayesian approach to model selection involves determining GGMs with high posterior probability. In this paper we make use of the mode oriented stochastic search (MOSS) algorithm described by Lenkoski and Dobra (26). MOSS is proven to find high posterior probability graphs faster than the shotgun stochastic search algorithm (SSS) of Jones et al. (21) or any Markov chain Monte Carlo stochastic search algorithm (e.g., the Markov chain Monte Carlo model composition algorithm of Madigan and York (32)) that visits models by sampling from the posterior distribution over the candidate models space. Since the size of the search space significantly increases from $|S_1| = 2^{p-1}$ for regression graphs to $|S_3| = 2^{p(p-1)/2}$ for unrestricted GGMs, it is critical to employ an algorithm that rapidly moves towards regions of high posterior probability graphs by visiting a reduced number of graphs along the way.

Let $\mathcal{M}$ be a set of candidate graphs. The objective of MOSS is to identify those graphs having the ratio between their posterior probability and the largest posterior probability of the graphs in $\mathcal{M}$ greater than a threshold $c \in (0, 1)$. We denote this subset of models with $\mathcal{M}(c)$. Lower posterior probability graphs in $\mathcal{M} \setminus \mathcal{M}(c)$ are discarded as suggested by the "Occam's

window" principle of Madigan and Raftery (31). The other parameters that appear in the description of MOSS are used to balance the running time of the procedure with thoroughness of the exploration of $\mathcal{M}$ (26). Inference can be subsequently performed with respect to the models in $\mathcal{M}(c)$ by Bayesian model averaging (22) as explained in Section 2.1.

There are two critical requirements for a successful application of MOSS. The first requirement is the existence of a method to rapidly calculate the marginal likelihood of each graph in $\mathcal{M}$ – see Section 2.1. The second requirement is the specification of a neighborhood set $\mathrm{nbd}(M) \subset \mathcal{M}$ for every $M \in \mathcal{M}$. These neighborhood sets must be symmetric (i.e., $M \in \mathrm{nbd}(M')$ if and only if $M' \in \mathrm{nbd}(M)$) and must link any two graphs through a path of graphs such that two consecutive graphs on this path are neighbors of each other. We define the neighborhood of a graph with respect to the candidate set of graphs as follows:

Regression GGMs. Let γ_A, $A \subset (2 : p)$, be regression and let $B = (2 : p) \backslash A$ be the indices of the explanatory variables that are not in the regression γ_A. As suggested in Hans et al. (18), the neighborhood of γ_A comprises the regressions obtained by (i) deleting one variable from γ_A, (ii) adding one variable X_j, $j \in B$, to γ_A, and (iii) replacing any one variable X_i, $i \in A$ with any one variable X_j, $j \in B$. The replacement of a variable currently in the model with another variable that is not in the model is especially important if we focus on small subsets regressions having a maximum number of predictors that is much smaller than $p - 1$.

Decomposable GGMs. The neighborhood of a decomposable graph G is comprised of all the decomposable graphs obtained from G by adding or deleting one edge.

GGMs. The neighborhood of a graph G is comprised of all the graphs obtained from G by adding or deleting one edge.

4. Jointness measures. The relevance of each predictor X_i, $i > 2$ 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 or GDAGs, 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:

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

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. Remark 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 ([24]).

5. Simulated data. When conditional independencies exists amongst covariates, a search in the space of GGMs will perform better than regression variable selection at determining which covariates interact directly with growth. This section considers two simulation studies that bear out the utility of GGMs when compared to regression models. When considering various techniques for evaluation of posterior model probabilities, a number of factors will clearly affect results. For instance, different settings of prior parameters will yield richer or sparser models. Therefore, if the methodology for scoring graphical models were fundamentally different than that used for regression models, it would be unclear whether the conclusions of a simulation study were the result of true methodological differences or simply an artifact of model settings. The benefit of the model scoring framework outlined in Section 2 is that both regression models and GGMs are scored using

the same underlying prior parameters. This allows for a direct comparison of the regression models and GGMs in placing high posterior probability on the factors that interact with growth. As such, the interest in the following will be to make comparisons of posterior variable inclusion probabilities under the same prior settings of regressions models and GGMs when it is known whether a given variable is independent of a response. The first simulation study examines a case when the covariates exhibit a strong degree of multi-collinearity, and focuses on the ability of the model selection techniques to place zero inclusion probability on variables that have no relationship to the response. The second study examines the relative performance of regression variable selection to graphical model selection when a graphical model underlies the data generation process.

5.1. *First simulated example.* This is a variation of the example suggested by Nott and Green ([35]). As in George and McCulloch ([16]), we generate $Z_1, \ldots, Z_{15}, Z \sim N_{300}(0, I_{300})$. 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 a regression and a decomposable graphical models search. The interest in both cases is to study whether predictors that belong to $X^{(2)}$ are not selected by MOSS, and to compare which technique has a larger tendency to exclude these variables. Due to the complex correlation structure amongst the predictors in $X^{(1)}$, variables may be selected even if their regression coefficients are zero. We employed MOSS with $c = 0.3125$, $c' = 0.01$, a pruning probability $q = 0.1$ and five different starting points. We allowed MOSS to explore regressions containing at most 30 predictors. In order to reduce the sampling variability, we report the results we obtained by averaging across 100 replicates of this experiment.

Figure 5.1 shows the mean posterior inclusion probabilities for each variable across the 100 simulated datasets after running a regression model search and a decomposable GGM search. The figure shows that in both
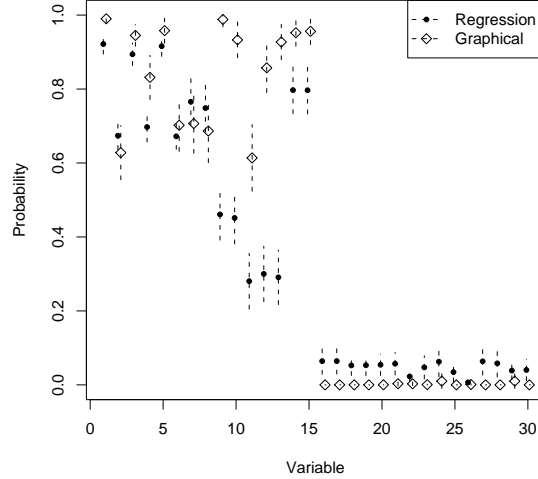
FIG 1. *Average posterior inclusion probabilities and (95%) confidence bands by variable for the simulation study described in Section 5.1 after employing MOSS for decomposable GGMs and regression models.*

cases 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. While this fact alone suggests that both the regression and GGM frameworks are well suited for discovering growth determinants, comparing the two methods reveals the usefulness of considering GGMs. We see that instead of receiving low inclusion probabilities, 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 GGM case than the regression case, particularly for variables 11 through 13, which is crucial since these covariates interact directly with the response. We again remark that the comparison of these two searches is the key component of the simulation study. Alternate prior specifications would have perhaps included or excluded variables to different degrees, however the result that the GGM search will be more likely to eliminate excess regressors and place higher probability on variables that interact with the response remains robust to such alterations.

5.2. *Second simulated example.* The second simulation study considers the case when the data are truly generated from an undirected graphical model. This situation was considered in-depth in Lenkoski and Dobra (26) and originally stems from a study in Yuan and Lin (44). In the following we consider four different, ten-dimensional graphical models:

- Independence: the graph G consists of individual nodes with no edges between nodes.
- AR(1): the graph G is such that there exist edges from variable i to variables $i - 1$ and $i + 1$.
- AR(2): similar to the graph above, except that i is also connected to $i - 2$ and $i + 2$.
- Circle: In this graph, variable i is connected to variable $i + 1$, and, in addition, there exists an edge between variables 1 and 10.

These represent four of the eight models considered in Yuan and Lin (44) and Lenkoski and Dobra (26) and were chosen because they have a significant degree of sparsity in their structure. To frame the simulation study in a regression context, we considered variable 1 to be the "response" and variables 2 to 10 to be potential covariates. 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 model, we generated 100 datasets of 100 observations each, in the manner described in Yuan and Lin (44) and ran MOSS for undirected GGMs and regression models. In both cases the MOSS settings were $c = 0.1, c' = 0.01, q = 0.1, m = \infty$.

Figure 5.2 shows the results of the simulation study for these four models. The results are consistent across the four model types: when considering either GGM or regression models, variables linked to the "response" are correctly given high inclusion probabilities. However, as was the case in Section 5.1, the 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 model, where the regression search puts an average of approximately 0.5 probability on variables 3 through 9, while the GGM search gives these variables essentially zero probability. This result is important, as the circle model is the only nondecomposable graph considered in Yuan and Lin (44), and it shows the importance of performing a model search in the correct space, even when the interest lies in a regression-type interpretation of interactions by ultimately focusing on a single variable.

6. Results: FLS data. Our analysis of the FLS growth data proceeds as follows. The parameters for MOSS were chosen to be $c = 0.3125, c' = 0.01,$
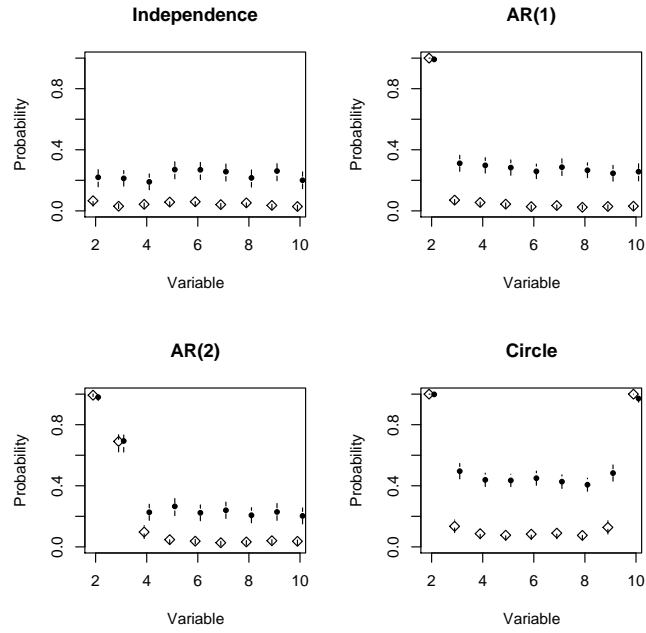
FIG 2. *Average posterior inclusion probabilities and (95%) confidence bands by variable for the four models described in Section 5.2. As in Figure 5.1, the dots represent the mean inclusion probability for each variable across 100 datasets for the regression search, while the diamonds represent the mean for the GGM search*

$q = 0.1$ – see Lenkoski and Dobra (26) for detailed explanations. We performed an initial linear regression search involving the entire set of potential growth determinants. Five separate instances of MOSS were run from random starting regressions. MOSS identified 595 regressions having a Bayes factor greater than $c = 0.3125$ with respect to the highest posterior probability regression. 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).

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 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 MOSS for GGMs, using the settings above. In order to cope with the large number of models that will be returned, we set the parameter $m = 1000$, see Lenkoski and Dobra (26) for a discussion of this parameter.

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 and SubSahara 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	0.0011	LatAmerica	0.74	-0.0062	NA	NA
		(0.0045)		(0.005)			(0.024)		(NA)
GDPsh560	1	-0.0134	1	-0.0116	EthnoLFrac	0.722	0.0068	0	0.0017
		(0.0153)		(0.0127)			(0.0221)		(0.0127)
EquipInv	1	0.1162	1	0.1318	PublEdu	0.651	0.1405	0	0.1477
		(0.0477)		(0.0427)			(0.0909)		(0.0838)
SubSahara	1	-0.0186	NA	NA	PrSc	0.42	0.0063	0.035	0
		(0.0201)		(NA)			(0.0334)		(0)
Confucious	1	0.0646	1	0.0814	Age	0.397	0	0	0
		(0.0239)		(0.023)			(0.0013)		(0.0011)
RuleofLaw	1	0.0113	0.274	0.0055	OutwarOr	0.34	-0.0012	NA	NA
		(0.0158)		(0.0137)			(0.0102)		(NA)
Mining	0.981	0.0352	0	0.0274	CivlLib	0.33	-5e-04	1	1e-04
		(0.0252)		(0.0239)			(0.0084)		(0.0146)
Protestants	0.968	-0.0129	1	-0.0104	Buddha	0.328	0.0032	0.086	0.0177
		(0.0173)		(0.0171)			(0.0165)		(0.0144)
Hindu	0.943	-0.0567	0	-0.0385	Muslim	0.32	0.0035	0	0.018
		(0.0508)		(0.0382)			(0.0233)		(0.0146)
EcoOrg	0.936	0.002	0	0.0021	PolRights	0.282	-3e-04	0	-0.0016
		(0.0071)		(0.0052)			(0.0068)		(0.0143)
NEquipInv	0.91	0.0415	0	0.0542	Catholic	0.236	-0.0015	1	0.0072
		(0.0353)		(0.0299)			(0.0146)		(0.0151)
LabForce	0.898	0.1818	0	0.1987	SpanishCol	0.234	0.0013	NA	NA
		(0.1024)		(0.0826)			(0.0135)		(NA)
BlMktPm	0.894	-0.0071	0	-0.0073	Popg	0.059	0.008	0	-0.0102
		(0.0144)		(0.0116)			(0.0331)		(0.1128)
HighEnroll	0.873	-0.0736	0	-0.0637	PrExports	0.057	-3e-04	0	-5e-04
		(0.0654)		(0.0495)			(0.0064)		(0.0178)

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 also show two instances, similar to the simulation study in Secton 5.1 where variables, in this case 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. By conducting this analysis, 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.

The resulting 10 variables represent an astounding support for the Neo-classical 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 covary 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 covary 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 returned. 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 .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	29.87959	Strong	∞	Decisive
GDPsh560	Protestants	29.87959	Strong	∞	Decisive
EquipInv	Protestants	29.87959	Strong	∞	Decisive
Confucious	Protestants	29.87959	Strong	∞	Decisive
Life	CivlLib	0.4930215	Little	∞	Decisive
GDPsh560	CivlLib	0.4930215	Little	∞	Decisive
EquipInv	CivlLib	0.4930215	Little	∞	Decisive
CivlLib	Confucious	0.4930215	Little	∞	Decisive
CivlLib	Protestants	0.427147	Little	∞	Decisive
Catholic	Protestants	0.3225106	Little	∞	Decisive
Life	Catholic	0.3088408	Little	∞	Decisive
GDPsh560	Catholic	0.3088408	Little	∞	Decisive
EquipInv	Catholic	0.3088408	Little	∞	Decisive
Catholic	Confucious	0.3088408	Little	∞	Decisive
CivlLib	Catholic	0.1768784	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 as well as the stochastic search techniques we used seem to work well and allow the development of a coherent framework that could be adapted to other prior specifications or stochastic search methods.

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 cares to employ little or no subjective prior information. For this case Fernández et al. 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. 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. In essence the authors proposed an automatic data dependent prior that is more transparent than the data dependent prior suggested by Raftery et al. (36) that was a function of the number of observations and required the researchers to set a hyper parameter (phi) with very little information.

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 2. The Gaussian graphical models induce parsimony in the structure of the covariance matrix by identifying the most relevant conditional independence relationships. This is the reason why we are able to identify fewer growth determinants while employing the same prior specification for all the classes of models we consider.

Software implementing the methods described in this paper is available upon request from the authors.

References.

[1] ACOSTA-GONZÁLEZ, E. AND FERNÁNDEZ-RODRÍGUEZ, F. (2007). Model selection via genetic algorithms illustrated with cross-country growth data. *Empirical economics 33*, 313–337.

[2] ATAY-KAYIS, A. AND 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. AND DURLAUF, S. N. (2001). Growth empirics and reality. *World Bank Economic Review 15*, 229–272.

[4] BROECK, S. AND 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. http://ideas.repec.org/p/sce/scecf5/343.html.

[5] DELLAPORTAS, P., GIUDICI, P., AND 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., AND WEST, M. (2004). Sparse graphical models for exploring gene expression data. *J. Multivariate Anal. 90*, 196–212.

[8] DOPPELHOFER, G. AND WEEKS, M. (2007). Jointness of growth determinants. CESifo Working Paper Series No. 1978. Available at http://ssrn.com/abstract=986545.

[9] DURLAUF, S. N., JOHNSON, P., AND TEMPLE, J. (2005). Growth econometrics. In *Handbook of Economic Growth*. P. Aghion and N. Durlauf, eds., North Holland, Amsterdam.

[10] DURLAUF, S. N., KOURTELLOS, A., AND TAN, C.-M. (2007). Are any growth theoris robust? Discussion Papers Series, Department of Economics, Tufts University 0703.

[11] EICHER, T. S., PAPAGEORGIOU, C., AND RAFTERY, A. E. (2007). Determining growth determinants: default priors and predictive performance in Bayesian Model Averaging. Working Paper no. 76, Center for Statistics and the Social Sciences, University of Washington.

[12] EICHER, T. S., PAPAGEORGIOU, C., AND 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., AND STEEL, M. F. J. (2001a). Benchmark priors for Bayesian model averaging. *Journal of Econometrics 100*, 381–427.

[14] FERNÁNDEZ, C., LEY, E., AND STEEL, M. F. J. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics 16*, 563–576.

[15] GEIGER, D. AND 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. AND 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 *Bayesian Statistics*, A. P. D. J. M. Bernardo, J. O. Berger and A. F. M. Smith, Eds. Vol. **6**. Oxford University Press, 175–177. Invited discussion of "Bayesian model averaging and model search strategies" by M. A. Clyde.

[18] HANS, C., DOBRA, A., AND WEST, M. (2007). Shotgun stochastic search for "large p" regression. *J. Am. Statist. Assoc. 102*, 507–16.

[19] HECKERMAN, D., GEIGER, D., AND CHICKERING, D. (1995). Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning 20*, 197–243.

[20] HENDRY, D. F. AND KROLZIG, H.-M. (2004). We ran one regression. *Oxford Bulletin of Economics and Statistics 66*, 799–810.

[21] JONES, B., CARVALHO, C., DOBRA, A., HANS, C., CARTER, C., AND WEST, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statist. Sci. 20*, 388–400.

[22] KASS, R. AND RAFTERY, A. E. (1995). Bayes factors. *J. Am. Statist. Assoc. 90*, 773–95.

[23] KENNEDY, P. (1998). *A Guide to Econometrics*. The MIT Press.

[24] LAURITZEN, S. L. (1996). *Graphical Models*. Oxford University Press.

[25] LEAMER, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. Wiley, New York.

[26] LENKOSKI, A. AND DOBRA, A. (2008). Bayesian structural learning and estimation

in Gaussian graphical models. Technical Report no. 545, Department of Statistics, University of Washington.

[27] LETAC, G. AND MASSAM, H. (2007). Wishart distributions for decomposable graphs. *Ann. Statist. 35*, 1278–323.

[28] LEVINE, R. AND RENELT, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review 82*, 942–963.

[29] LEY, E. AND STEEL, M. F. (2007). Jointness in Bayesian variable selection with applications to growth regression. *Journal of Macroeconomics 29*, 476–493.

[30] LEY, E. AND STEEL, M. F. (2008). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*. forthcoming.

[31] MADIGAN, D. AND RAFTERY, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Am. Statist. Assoc. 89*, 1535–45.

[32] MADIGAN, D. AND YORK, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review 63*, 215–232.

[33] MASANJALA, W. H. AND PAPAGEORGIOU, C. (2006). A rough and lonely road to prosperity: a reexamination of sources of growth in Africa using Bayesian model averaging. Working Paper, Louisiana State University.

[34] MUIRHEAD, R. J. (2005). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons.

[35] NOTT, D. AND 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., AND 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., AND 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] YUAN, M. AND LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika 94*, 19–35.

[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 *Bayesian Inference and Decision Techniques: essays in honour of Bruno de Finetti*, P. K. Goel and A. Zellner, Eds. Amsterdam, North Holland, 233–243.

University of Washington
Department of Statistics
Seattle, WA 98195
USA
E-mail: adobra@u.washington.edu
URL: http://www.stat.washington.edu/adobra

University of Washington
Department of Economics
1100 Campus Parkway
Seattle, WA 98195
USA
E-mail: te@u.washington.edu
URL: http://faculty.washington.edu/~te

University of Washington
Department of Statistics
Seattle, WA 98195
USA
E-mail: lenkoski@stat.washington.edu
URL: http://www.stat.washington.edu/lenkoski