WPS4063

Jointness in Bayesian Variable Selection with Applications to Growth Regression

Eduardo Ley
The World Bank, Washington DC, U.S.A.

Mark F.J. Steel
Department of Statistics, University of Warwick, U.K.

Abstract. We present a measure of jointness to explore dependence among regressors, in the context of Bayesian model selection. The jointness measure proposed here equals the posterior odds ratio between those models that include a set of variables and the models that only include proper subsets. We illustrate its application in cross-country growth regressions using two datasets from the model-averaging growth literature.

Keywords. Bayesian model averaging; Complements; Model uncertainty; Posterior odds; Substitutes

JEL Classification System. C11, C7

World Bank Policy Research Working Paper 4063, November 2006

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at http://econ.worldbank.org.

Address. E. Ley, World Bank, 1818 H St NW, Washington DC 20433. Email: eduley@gmail.com. M.F.J. Steel, Department of Statistics, University of Warwick, Coventry CV4 7AL, Vox: +44(0)24 7652 3369. U.K. Email: M.F.Steel@stats.warwick.ac.uk

1. Introduction

Performing inference on the determinants of GDP growth is challenging because, in addition to the complexity and heterogeneity of the objects of study, a key characteristic of the empirics of growth lies in its open-endedness (Brock and Durlauf, 2001). Open-endedness entails that, at a conceptual level, alternative theories may suggest additional determinants of growth without necessarily excluding determinants proposed by other theories. The absence, at the theoretical level, of such tradeoff leads to substantial model uncertainty, at the empirical level, about which variables should be included in a growth regression. In practice, a substantial number of growth determinants may be included as explanatory variables. If two such variables are capturing different sources of relevant information and should both be included, we will talk of *jointness* (as defined later), whereas if they perform very similar roles they should not appear jointly, which we will denote by *disjointness*. We could think of these situations as characterized by the covariates being complements or substitutes, respectively.

Various approaches to deal with this model uncertainty have appeared in the literature: early contributions are the extreme-bounds analysis in Levine and Renelt (1992) and the confidence-based analysis in Sala-i-Martin (1997). Fernández *et al.* (2001b, FLS henceforth) use Bayesian model averaging (BMA, see Hoeting *et al.*, 1999) to handle the model uncertainty that is inherent in growth regressions, as discussed above. BMA naturally deals with model uncertainty by averaging posterior inference on quantities of interest over models, with the posterior model probabilities as weights. Other papers using BMA in this context are León-González and Montolio (2004) and Papageorgiou and Masanjala (2005). Alternative ways of dealing with model uncertainty are proposed in Sala-i-Martin *et al.* (2004, SDM henceforth),[1] and Tsangarides (2005). Insightful discussions of model uncertainty in growth regressions can be found in Brock and Durlauf (2001) and Brock, Durlauf and West (2003). All of these studies adopt a Normal linear regression model and consider modelling n growth observations in y using an intercept and explanatory variables from a set of k variables in Z, allowing for any subset of the variables in Z to appear in the model. This results in 2^k possible models, which will thus be characterized by the selection of regressors. We call model M_j the model with the $0 \leq k_j \leq k$ regressors grouped in Z_j, leading to

$$y|\alpha, \beta_j, \sigma \sim \mathrm{N}(\alpha\iota_n + Z_j\beta_j, \sigma^2 I), \tag{1}$$

where ι_n is a vector of n ones, $\beta_j \in \Re^{k_j}$ groups the relevant regression coefficients and $\sigma \in \Re_+$ is a scale parameter.

Based on theoretical considerations and simulation results in Fernández *et al.* (2001a), FLS adopt the following prior distribution for the parameters in M_j:

$$p(\alpha, \beta_j, \sigma|M_j) \propto \sigma^{-1} f_N^{k_j}(\beta|0, \sigma^2(gZ_j'Z_j)^{-1}), \tag{2}$$

We thank Gernot Doppelhofer for stimulating our interest on these issues, and for kindly making a preliminary draft of Doppelhofer and Weeks (2006) available to us.

[1] SDM's procedure BACE in fact uses approximate Bayesian posterior probabilities of regression models based on the Schwarz criterion, as proposed by Raftery (1995).

where $f_N^q(w|m, V)$ denotes the density function of a q-dimensional Normal distribution on w with mean m and covariance matrix V and they choose $g = 1/\max\{n, k^2\}$. Finally, the components of β not appearing in M_j are exactly zero, represented by a prior point mass at zero.

The prior model probabilities are specified by $P(M_j) = \theta^{k_j}(1-\theta)^{k-k_j}$, which implies that each regressor enters a model independently of the others with prior probability θ. Thus, the prior expected model size is θk. We follow Fernández *et al.* (2001a) and FLS in choosing $\theta = 0.5$, which is a benchmark choice—implying that $P(M_j) = 2^{-k}$ and that expected model size is $k/2$. Throughout this paper, we shall use the same prior as in FLS. An explicit analysis of alternative priors in this context is the subject of current research.

FLS use a Markov chain Monte Carlo (MCMC) sampler to deal with the very large model space ($k = 41$ for their data set, leading to 2.2×10^{12} possible models to consider), which can easily be implemented through a Metropolis algorithm over model space alone—since the marginal likelihoods and thus the posterior model odds between any two models are analytically available; see FLS for details. Thus, the MCMC sampler is only implemented to deal with the practical impossibility of exhaustive analysis of the model space: the chain will visit only the most promising models, which are the ones with non-negligible posterior probability.

The results in FLS indicate that the posterior probability is spread widely among many models (Table 1), which implies the superiority of BMA over choosing a single model but, precisely because of its richness, also makes it harder to summarise posterior results.

Table 1. FLS Data: distribution of posterior model probability and model size.
(MCMC sampler with 2 million recorded draws after a burn-in of 1 million draws.)

Number	Posterior	Number of Regressors	
of Models	Prob.	Mean	St.Dev.
1	0.01	10	-
5	0.04	8.7	1.1
25	0.10	8.9	1.4
50	0.14	9.0	1.4
100	0.19	9.1	1.4
190	0.25	9.1	1.4
1,606	0.50	9.4	1.5
8,688	0.75	9.6	1.5
25,269	0.90	9.8	1.6
39,839	0.95	9.8	1.6
71,493	0.99	9.9	1.6
148,342	1.00	9.9	1.6

In related work, Doppelhofer and Weeks (2005, DW henceforth) analyse the same linear regression model (using the BACE method of SDM) and introduce a measure of jointness "to address dependence among explanatory variables." DW propose a jointness statistic defined as the log ratio of the joint posterior inclusion probability of a set of variables over the product of the individual posterior inclusion probabilities. Below we discuss this measure as well as another measure

proposed by the same authors in a later paper.

In this paper we propose alternative measures of jointness for Bayesian variable selection, based on probabilistic arguments and illustrate their application in the context of BMA. The next section discusses measures of jointness, which will be applied to two different growth data sets: the FLS data and the DW data in Section 3. The final section concludes.

2. Jointness in Bayesian Variable Selection

Consider two variables, i and j, and let $P(i)$ denote the posterior probability of inclusion for regressor i.[2] $P(i \cap j)$ is the posterior probability of including both variables i and j. These probabilities are simply defined as the sum of the posterior probabilities of all models that contain regressor i (for $P(i)$) or regressors i and j (for $P(i \cap j)$). Clearly $P(i) \geq P(i \cap j)$.

DW define jointness as (the logarithm of):

$$\mathcal{J}_{ij}^{DW} = \frac{P(i \cap j)}{P(i)P(j)} \tag{3}$$

There are several problems associated with this measure. First, since $\mathcal{J}_{ij}^{DW} = P(i|j) \div P(i) = P(j|i) \div P(j)$, it can be verified that as either $P(i)$ or $P(j)$ approaches unity, then $\mathcal{J}_{ij}^{DW}$ also approaches unity—regardless of the behaviour of the other variable. In fact, it is bounded from above as $\mathcal{J}_{ij}^{DW} \leq 1/\max\{P(i), P(j)\}$. This means that when the inclusion probability of one of the variables exceeds $e^{-1} = 0.37$ this variable can never be a "significant complement" (corresponding to $\log(\mathcal{J}_{ij}^{DW}) > 1$) of any other variable in the data set according to the DW jointness measure.

Secondly, consider a case of extreme jointness where variables i and j always appear together; then $\mathcal{J}_{ij}^{DW} = P(i)^{-1} = P(j)^{-1} = P(i \cap j)^{-1}$ so the DW measure equals the reciprocal of $P(i \cap j)$. This makes comparisons of DW-jointness across different pairs of variables quite difficult, and is not consistent with the fixed critical level used by DW. It also raises questions about interpretation, since while $P(i \cap j)$ could itself be considered a natural measure of jointness, that does not apply to its reciprocal.

Finally, contrary to the claim in DW, $\mathcal{J}_{ij}^{DW}$ does not correspond to the posterior odds of models including i and j *vs* models that include them individually. This ratio is properly defined in equation (6) below.

In subsequent work, Doppelhofer and Weeks (2006, DWa henceforth) propose an alternative jointness measure, based on the cross-product ratio of the binary indicators of variable inclusion. For two regressors i and j this measure corresponds to (the logarithm of):

$$\mathcal{J}_{ij}^{DWa} = \frac{P(i \cap j)P(\tilde{i} \cap \tilde{j})}{P(i \cap \tilde{j})P(\tilde{i} \cap j)}, \tag{4}$$

[2] To economize on notation, we shall not make the dependence on the sample explicit, but it should be understood that all posterior quantities are conditional on the observed sample.

where $\tilde{\imath}$ and $\tilde{\jmath}$ stand for the *exclusion* of i and j, respectively. This measure can be written as

$$\mathcal{J}_{ij}^{DWa} = \frac{P(i|j)}{P(\tilde{\imath}|j)} \div \frac{P(i|\tilde{\jmath})}{P(\tilde{\imath}|\tilde{\jmath})},$$

which clearly shows that $\mathcal{J}_{ij}^{DWa}$ can be interpreted as the posterior odds of including i given that j is included divided by the posterior odds of including i given that j is not included—rather than the posterior odds interpretation given in DWa.

The measure $\mathcal{J}_{ij}^{DWa}$ is undefined both (i) when either of the regressors is always included, and (ii) when one of the regressors is never included. Furthermore, in practical situations where, say, $P(i) \to 1$, the measure depends crucially on the limit of the ratio $P(\tilde{\imath} \cap \tilde{\jmath})/P(\tilde{\imath} \cap j)$ which means that the (few) low-probability models without i can make the measure range all the way from 0 (if they all include j) to ∞ (if they all exclude j). This critical dependence of the jointness involving important variables on models with very low posterior probability seems an undesirable characteristic.

We now present our own alternative jointness measures. Consider Fig. 1, where the posterior probabilities of two variables are represented in a Venn diagram. As mentioned, a raw, yet natural, measure of jointness is simply probability of joint inclusion, $P(i \cap j)$—*i.e.*, the intersection shown in the diagram. The diagram also suggests two better measures of jointness: (i) the joint probability (*i.e.*, intersection) relative to the probability of including either one (*i.e.*, the union), and (ii) the joint probability relative to the probability of including either one, but not both—*i.e.*, excluding the intersection itself. We shall denote these measures of jointness by $\mathcal{J}_{ij}^{\star}$ and $\mathcal{J}_{ij}$.

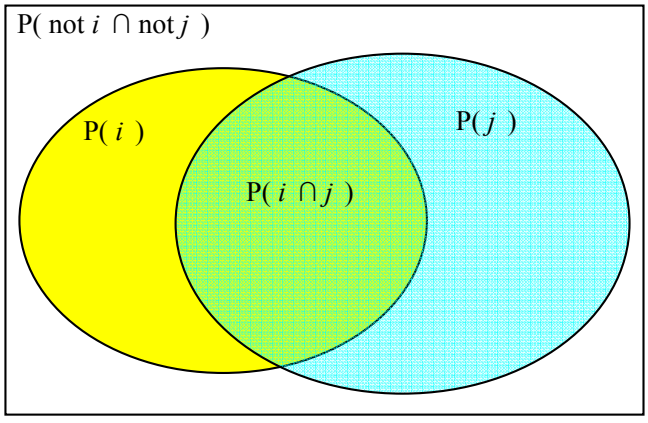


Fig. 1. Jointness.

Thus, the measures of jointness proposed in this paper are:

$$\mathcal{J}_{ij}^{\star} = \frac{P(i \cap j)}{P(i \cup j)} = \frac{P(i \cap j)}{P(i) + P(j) - P(i \cap j)} \in [0, 1] \tag{5}$$

$$\mathcal{J}_{ij} = \frac{P(i \cap j)}{P(i \cap \tilde{\jmath}) + P(j \cap \tilde{\imath})} = \frac{P(i \cap j)}{P(i) + P(j) - 2P(i \cap j)} \in [0, \infty) \tag{6}$$

In the extreme jointness case discussed above (when i and j appear always together), while $\mathcal{J}_{ij}^{DW} = P(i \cap j)^{-1}$, we have that $\mathcal{J}_{ij}^{DWa} = \infty$, $\mathcal{J}_{ij}^{\star} = 1$ and $\mathcal{J}_{ij} = \infty$, which are more intuitive and practical as extreme values. In case that $P(i) \to 1$, we obtain $\mathcal{J}_{ij}^{\star} \to P(j)$ and $\mathcal{J}_{ij} \to P(j)/(1 - P(j))$, which are well-behaved quantities in line with intuition.

Furthermore, note that $\mathcal{J}_{ij}$ does indeed correspond to the posterior odds ratio of the models including both i and j *vs* the models that include them only individually. Since this fact leads to a straightforward interpretation and to a statistically meaningful metric, $\mathcal{J}_{ij}$ is our preferred measure of jointness. In particular, there is no need for calibration, as this measure has a clear and direct interpretation in terms of model probabilities.

Another advantage of the measures proposed here is that they are easily extended to the case of more than two regressors. In particular, we define *multivariate jointness* for general sets of regressors $\mathcal{S}$ through two quantities: $P(\mathcal{S})$, which is the total posterior probability assigned to those models having all regressors in $\mathcal{S}$, and $P(\subset \mathcal{S})$, defined as the posterior mass assigned to all models including only proper subsets of $\mathcal{S}$. Then we can generalise the measures above as follows:

$$\mathcal{J}_{\mathcal{S}}^{\star} = \frac{P(\mathcal{S})}{P(\subset \mathcal{S}) + P(\mathcal{S})} \in [0, 1] \tag{7}$$

$$\mathcal{J}_{\mathcal{S}} = \frac{P(\mathcal{S})}{P(\subset \mathcal{S})} \in [0, \infty) \tag{8}$$

Thus, the measure $\mathcal{J}_{\mathcal{S}}^{\star}$ is the posterior mass of the models containing all of $\mathcal{S}$ as a fraction of the posterior mass assigned to all models having any (or all) of the regressors in $\mathcal{S}$. As before, the measure $\mathcal{J}_{\mathcal{S}}$ is the posterior odds ratio between those models having all the variables in the set $\mathcal{S}$ and the models including only proper subsets of these variables.

3. Jointness of Growth Determinants

3.1. The FLS Data

We first illustrate the behaviour of our proposed measures of jointness using the growth data of FLS. The latter data set contains $k = 41$ potential regressors to model the average per capita GDP growth over 1960-1992 for a sample of $n = 72$ countries. [3]

The results are based on an MCMC chain on model space with 2 million recorded draws, after a burn-in of 1 million. The correlation between model visit frequencies and probabilities computed on the basis of the exact posterior odds of the visited models is 0.992, indicating excellent convergence. An estimate of the total model probability captured by the chain, computed as in George and McCulloch (1997), is 70%.[4] Results are virtually identical to those obtained in

[3] The dataset and the code used in FLS are available on the *Journal of Applied Econometrics* code and data archive at http://qed.econ.queensu.ca/jae/. The updated code used here has been uplodaded to that website, and is also available at http://www.warwick.ac.uk/go/msteel/steel_homepage/software/.

[4] This is quite high, given that we only visit one in every 15 million models. Longer chains will capture marginally more of this posterior probability, but they will only add models with very small posterior probabilities and this will not affect any of the conclusions.

FLS on the basis of a chain of the same length.[5]

Table 2 displays the marginal posterior probabilities of inclusion of each regressor—*i.e.*, $P(i)$. Fig. 2 displays scatter plots of the logarithms of $P(i \cap j)$, $\mathcal{J}_{ij}^{\star}$, and $\mathcal{J}_{ij}$ for all possible combinations of two regressors. Fig. 2 shows that $\log(\mathcal{J}_{ij}^{\star})$ and $\log(\mathcal{J}_{ij})$ are linearly related for a range of values, but the relation turns non-linear as $P(i \cap j) \to 1$. The ranking in terms of both jointness measures is, however, identical, so the only issue for identifying jointness is to decide on suitable "critical" values. This will be easier for $\mathcal{J}_{ij}$, given its interpretation as a posterior odds ratio. Another notable feature of Fig. 2 is that both jointness measures tend to be increasing in $P(i \cap j)$. This relationship is less exact for average values of $P(i \cap j)$, but it does seem that extremely high and low values of $P(i \cap j)$ correspond to similar extremes for both jointness measures. The main features of Fig. 2 are closely mirrored by a similar graph for the SDM data set used in the next subsection, thus lending somewhat more generality to our findings.

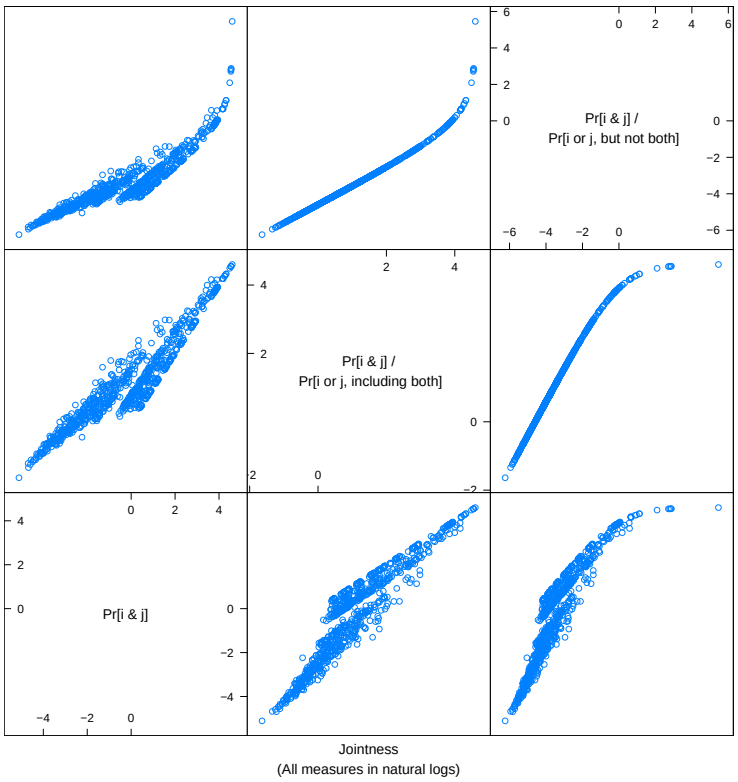


Fig 2. FLS data: Joint inclusion probabilities, bivariate jointness $\mathcal{J}_{ij}^{\star}$ and $\mathcal{J}_{ij}$.

[5] Note that in the applications in this paper, the regressors have been standardized to have (zero mean and) unit variance to ease the comparison between regression coefficients. The dependent variable is expressed in percentage points of annual per capita GDP growth and displays similar behaviour in both datasets—means of 2.1 and 1.8, and standard deviations of 1.9 and 1.8.

Table 2. FLS data: Marginal posterior summary of the β's.

(Regressors are standardized.)

						Conditional on Inclusion			
	Regressors	Prob.	$E(\beta_i\|y)$	$\sqrt{\text{Var}}(\beta_i\|y)$	$\frac{\|E(\beta_i\|y)\|}{\sqrt{\text{Var}}(\beta_i\|y)}$	$E(\beta_i\|y)$	$\sqrt{\text{Var}}(\beta_i\|y)$	$\frac{\|E(\beta_i\|y)\|}{\sqrt{\text{Var}}(\beta_i\|y)}$	Sign
1	log GDP in 1960	1.00	-1.4180	0.269	5.3	-1.4180	0.267	5.3	1.0
2	Fraction Confucian	1.00	0.4900	0.117	4.2	0.4922	0.113	4.4	1.0
3	Life expectancy	0.95	0.9574	0.371	2.6	1.0120	0.300	3.4	1.0
4	Equipment investment	0.94	0.5575	0.222	2.5	0.5920	0.179	3.3	1.0
5	Sub-Saharan dummy	0.76	-0.4772	0.334	1.4	-0.6303	0.225	2.8	1.0
6	Fraction Muslim	0.66	0.2565	0.219	1.2	0.3909	0.143	2.7	1.0
7	Rule of law	0.52	0.2594	0.280	0.9	0.5027	0.172	2.9	1.0
8	Number of years open economy	0.50	0.2556	0.283	0.9	0.5088	0.174	2.9	1.0
9	Degree of capitalism	0.47	0.1577	0.184	0.9	0.3347	0.114	2.9	1.0
10	Fraction Protestant	0.46	-0.1441	0.176	0.8	-0.3129	0.120	2.6	1.0
11	Fraction GDP in mining	0.44	0.1384	0.176	0.8	0.3136	0.122	2.6	1.0
12	Non-Equipment investment	0.43	0.1346	0.173	0.8	0.3130	0.116	2.7	1.0
13	Latin American dummy	0.19	-0.0729	0.175	0.4	-0.3829	0.205	1.9	1.0
14	Primary school enrollment, 1960	0.18	0.0941	0.224	0.4	0.5126	0.241	2.1	1.0
15	Fraction Buddhist	0.17	0.0394	0.100	0.4	0.2356	0.114	2.1	1.0
16	Black market premium	0.16	-0.0355	0.092	0.4	-0.2250	0.106	2.1	1.0
17	Fraction Catholic	0.11	-0.0123	0.113	0.1	-0.1118	0.326	0.3	0.7
18	Civil liberties	0.10	-0.0388	0.134	0.3	-0.3879	0.212	1.8	1.0
19	Fraction Hindu	0.10	-0.0247	0.094	0.3	-0.2556	0.182	1.4	1.0
20	Primary exports, 1970	0.07	-0.0209	0.089	0.2	-0.2916	0.178	1.6	1.0
21	Political rights	0.07	-0.0205	0.090	0.2	-0.2994	0.189	1.6	1.0
22	Exchange rate distortions	0.06	-0.0134	0.063	0.2	-0.2202	0.142	1.6	1.0
23	Age	0.06	-0.0098	0.048	0.2	-0.1714	0.110	1.6	1.0
24	War dummy	0.05	-0.0097	0.051	0.2	-0.1886	0.127	1.5	1.0
25	Fraction of Pop. Speaking English	0.05	-0.0071	0.039	0.2	-0.1512	0.104	1.5	1.0
26	Fraction speaking foreign language	0.05	0.0089	0.051	0.2	0.1893	0.146	1.3	0.9
27	Size labor force	0.05	0.0099	0.069	0.1	0.2107	0.240	0.9	0.8
28	Ethnolinguistic fractionalization	0.04	0.0059	0.042	0.1	0.1718	0.154	1.1	1.0
29	Spanish Colony dummy	0.03	0.0058	0.050	0.1	0.1693	0.211	0.8	0.9
30	S.D. of black-market premium	0.03	-0.0041	0.031	0.1	-0.1337	0.117	1.1	1.0
31	French Colony dummy	0.03	0.0042	0.031	0.1	0.1358	0.117	1.2	1.0
32	Absolute latitude	0.02	0.0005	0.040	0.0	0.0212	0.257	0.1	0.5
33	Ratio workers to population	0.02	-0.0030	0.031	0.1	-0.1220	0.158	0.8	0.9
34	Higher education enrollment	0.02	-0.0041	0.039	0.1	-0.1720	0.187	0.9	1.0
35	Population growth	0.02	0.0032	0.035	0.1	0.1513	0.188	0.8	0.9
36	British colony dummy	0.02	-0.0019	0.022	0.1	-0.0913	0.121	0.8	0.9
37	Outward orientation	0.02	-0.0018	0.021	0.1	-0.0865	0.117	0.7	0.9
38	Fraction Jewish	0.02	-0.0014	0.020	0.1	-0.0701	0.127	0.6	0.8
39	Revolutions and coups	0.02	0.0000	0.017	0.0	0.0009	0.136	0.0	0.5
40	Public education share	0.02	0.0004	0.017	0.0	0.0230	0.130	0.2	0.6
41	Area (scale effect)	0.02	-0.0006	0.014	0.0	-0.0427	0.108	0.4	0.8

Fig. 3 displays the log posterior odds, $\log(\mathcal{J}_{ij})$, for all the $820 = 41(41-1)/2$ pairs of variables. In the plot, the log Bayes factors, $\log(\mathcal{J}_{ij})$, are sorted lexicographically along the horizontal axis, first by $P(i)$ and then by $P(j)$. Thus, since variable 1 has the highest inclusion probability, all its pairings appear first. Among those, variable 2 appears first, then 3, etc.

Posterior odds in Fig. 3 are classified as conveying positive, strong, very strong or decisive evidence of jointness when they exceed 3, 10, 30 or 100, and horizontal lines separate these

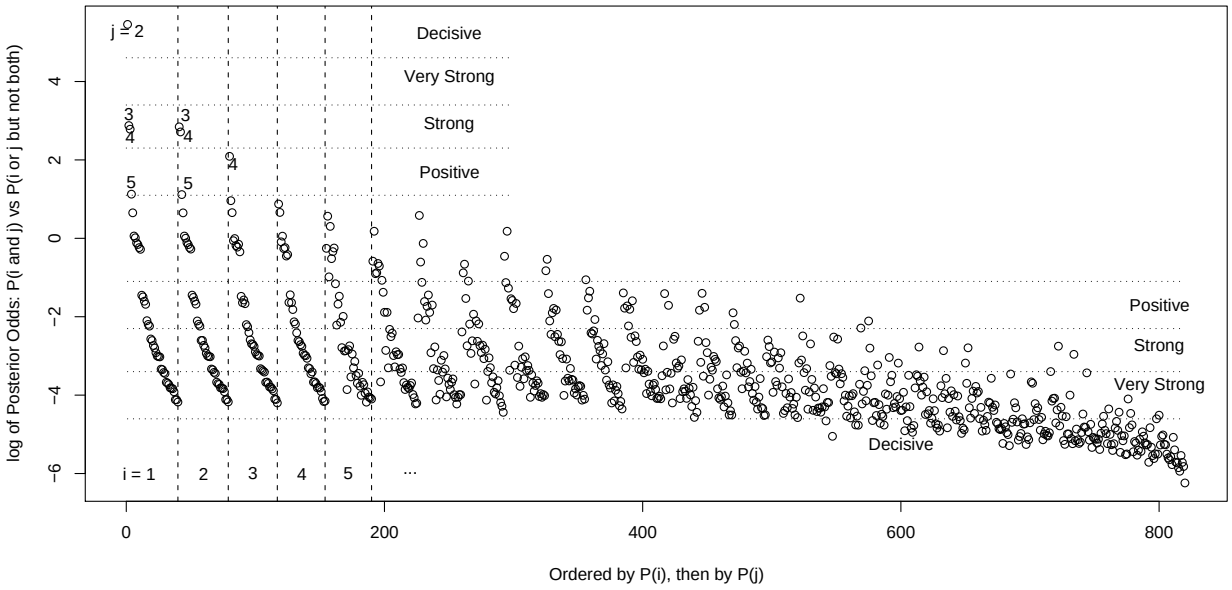


Fig 3. FLS data: Log posterior odds in favor of jointness, $\log(\mathcal{J}_{ij})$.

regions.[6] Only 8 pairs (1% of the total) display some degree of evidence for jointness. These are variables 1 and 2 (GDP in 1960 and Fraction Confucian), which show decisive jointness, the pairs $(1, 3)$, $(1, 4)$, $(2, 3)$, $(2, 4)$ display strong evidence of jointness, while there is positive evidence for $(1, 5)$, $(2, 5)$, and $(3, 4)$. As was already suggested by Fig. 2, only the regressors with high values for $P(i \cap j)$ and thus, high marginal inclusion probabilities tend to appear jointly in the growth regressions. Both of these probabilities are over 0.75 for the pairs mentioned above, and over 0.93 for those pairs with strong or decisive jointness. This means that we do not have, for this data set, any regressor pairs that virtually always appear together, while appearing in less than 75% of the visited models. Fig. 4 summarizes the jointness results graphically. The nodes are proportional to posterior probabilities of inclusion and the thickness of the joining lines is proportional to $\log(\mathcal{J}_{ij})$ for any pair.

Examining sets of more than two regressors, we note that jointness of all four triplets of the first four variables is supported by posterior odds of 7 to one and higher, whereas the jointness of all four of these variables is favorably supported by posterior odds of 7.8. Posterior odds pertaining to jointness for the first five variables, however, are only 2 to one.

Table 1 shows that the visited models typically have 9 or 10 regressors. This means that models tend to have five or six regressors in addition to the four or five that are usually present. Since we

[6] These cutoff points are easily interpretable—*e.g.*, positive evidence implies the models with both regressors get more than three times as much posterior probability as those with only one of the regressors. The actual values of the cutoff points are inspired by Jeffreys (1961, p. 432).

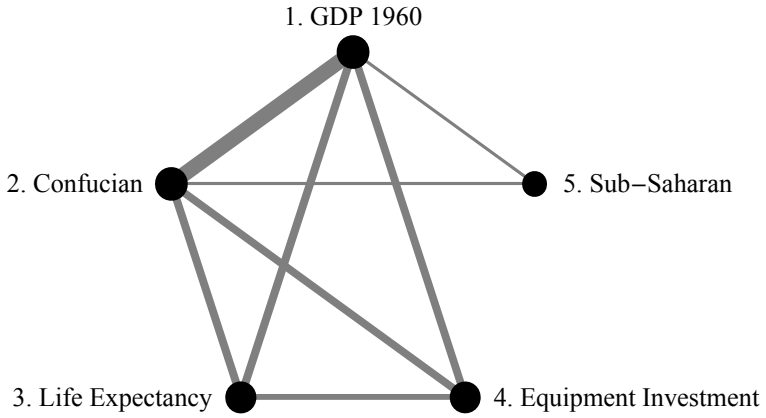


Fig 4. Pairwise joint growth determinants.
(Nodes and graphs are proportional to posterior probabilities.)

see no jointness beyond the first five variables, these additional regressors have to be alternating. Evidence suggesting that variables do appear on their own, but not jointly will be denoted by *disjointness*. This can occur, *e.g.*, when variables are highly collinear and are proxies for each-other. We can, again, interpret posterior odds directly, and similar thresholds (now favoring models with separate inclusion) can be used for disjointness. See Fig. 3 for a graphical display. Table 3 summarizes the posterior odds for assessing disjointness. Now 757 pairs of variables (92.3% of total) display some degree of disjointness—leaving only 63 pairs with posterior odds, *i.e.*, $\mathcal{J}_{ij}$, larger than $1/3$. In line with Fig. 2, decisive disjointness occurs only for determinants with small posterior inclusion probabilities (usually under 5%). The only exception is decisive disjointness between Civil liberties and Political rights, which are indeed likely to be good substitutes for each other. For triplets of variables, we get even more evidence of disjointness: 8,752 of the 10,660 possible triplets (*i.e.*, 82.1%) indicate decisive disjointness.

Table 3. Pairwise Disjointness

Evidence	posterior odds	Number	Percentage
Favorable	$\frac{1}{10} < PO \leq \frac{1}{3}$	79	10.4
Strong	$\frac{1}{30} < PO \leq \frac{1}{10}$	182	24.0
Very Strong	$\frac{1}{100} < PO \leq \frac{1}{30}$	353	46.6
Decisive	$PO \leq \frac{1}{100}$	143	18.9
Total		757	92.3

In conclusion, the five most important variables (at the top of Table 2) are not mutually exclusive determinants and tend to appear jointly. This is perhaps not surprising as they seem to capture rather different explanations for growth. The variables in the data set with moderate marginal inclusion probabilities do not have strong jointness or disjointness relations (with one exception), while the variables with least marginal importance tend to avoid occurring jointly and alternate their presence in the visited models.

Table 2 also presents the first two moments of the regression coefficients. The first set of moments is marginalized over all models visited, including the zeros for those models that exclude the regressor in question. These numbers reveal a strong positive correlation (equal to 0.89) between the absolute values of the normalized posterior mean and the inclusion probabilities. This ensures us that there are no variables with high inclusion probability but counteracting effects in different substantial sets of models. Any serious doubt as to the sign of the effect would be reflected in a posterior mean close to zero and a large posterior variance. This is reassuring as it implies that the direction of the influence of variables with high inclusion probabilities tends to be clear. Plots of the posterior densities of the β_i's can be found in FLS, which shows that the variable "Fraction Catholic" has modes on different sides of the origin. Indeed, this is the one variable where the value of the normalized posterior mean is a lot lower than could be expected on the basis of the (moderate) inclusion probability.

The second set of moments presented in Table 2 are computed conditionally upon inclusion—*i.e.*, by averaging over only those models which actually include the coefficient in question. These moments are important for assessing the effect of each regressor, given that it is included in the model. Clearly, differences with the unconditional moments can be very large if the regressor has a low or even moderate posterior inclusion probability.

3.2. Application to the Data Set of SDM and DW

SDM, DW and DWa use a larger data set, and model annual GDP growth per capita between 1960 and 1996 for n = 88 countries as a function of k = 67 potential drivers. This leads to an even larger model space with $2^{67} = 1.5 \times 10^{20}$ models, and requires a slight update to the code used for FLS. In the Fortran code used in FLS, an unsigned double-precision variable was used to index models for all the MCMC accounting. This results in an upper limit of 2^{52} models.[7] We have updated the code by using instead two model indices, making it suitable for any $k \leq 104$. The revised code is available at http://www.warwick.ac.uk/go/msteel/steel_homepage/software/.

This very large model space is remarkably well explored by an MCMC chain with 2 million retained model visits, after discarding the first million as a burn-in, as evidenced by a correlation between model visit frequencies and actual posterior probabilities of the set of visited models equal to 0.995. The chain visits 126,844 models and is estimated to cover 47% of the total posterior mass. This is a very high coverage, given that we visit less than one model for every 10^{15} models in the space. Any additional models (visited by running longer chains) will have virtually zero posterior probabilities, and will not affect the conclusions in any way.

Despite the much larger model space, Table 4 shows that posterior model probabilities are more concentrated than in the previous case and that the models with high posterior probability tend to be smaller, containing around 6–7 regressors.

[7] Double-precision numbers cannot distinguish between 2^{52} and $1+2^{52}$ because they have the same internal representation. The 64-bit double format uses 1 bit for the sign and 11 for the exponent leaving 52 bits available for the fraction—thus the upper limit on 'counting' of 2^{52}. An additional index will double the upper limit to 2^{104}, which should be enough in most practical applications.

Table 4. SDM Data: Distribution of posterior model probability and model size.
(MCMC sampler with 2 million recorded draws after a burn-in of 1 million draws.)

Number	Posterior	Number of Regressors	
of Models	Prob.	Mean	St.Dev.
1	0.07	6	-
5	0.10	6.3	0.5
25	0.18	6.4	0.7
50	0.23	6.3	0.9
100	0.29	6.4	1.0
752	0.50	6.6	1.2
4,627	0.75	6.8	1.3
13,527	0.90	6.9	1.4
21,948	0.90	6.9	1.4
42,363	0.99	7.0	1.4
126,844	1.00	7.0	1.4

Table 5 presents the marginal inclusion probabilities for those regressors that are included more than 1% of the time[8] as well as the moments of the corresponding β's in the same format as Table 2. Again, note the high correlation between the absolute values of the unconditional standardized posterior means of the β's and the posterior inclusion probabilities. There are, however, some interesting differences with the results based on the FLS data set. In spite of the fact that the posterior model probability is more concentrated with the SDM data, there are no regressors that are virtually always included. Nevertheless, fewer regressors have marginal inclusion probabilities exceeding 10% and there is a longer tail of marginally less important variables (53 out of the possible 67). In addition, the important determinants of growth are often not the same ones: only Past GDP, Primary school enrollment, Fraction Confucian, Life expectancy, Fraction GDP in mining and the Sub-Saharan and Latin American dummies receive inclusion probabilities above 10% for both data sets, and appear in a very different ordering in both. By comparing the moments of β conditionally upon inclusion, we can assess whether the effects of these regressors are similar in both datasets.[9] Comparing Tables 2 and 5 reveals that they all have the same signs, but some have rather different magnitudes: the main difference is in the effect of Past GDP, which is almost twice as large with the FLS data set (the difference is more than 2.5 times the posterior standard deviation).

The results on posterior inclusion probabilities and posterior moments of β are,[10] however,

[8] This is merely for ease of presentation, but the sampler was run over the entire model space with 67 potential regressors.

[9] These particular regressors were all measured in the same way in both data sets, and the raw moments were very similar, so that the standardisation does not affect this comparison.

[10] For the moments of β this can not be assessed directly from Table 5, but is clear from running the sampler without standardizing the data.

Table 5. DW data: Marginal posterior summary of the β's.

(Regressors are standardized.)

						Conditional on Inclusion			
	Regressors	Prob.	$E(\beta_i\|y)$	$\sqrt{Var(\beta_i\|y)}$	$\frac{\|E(\beta_i\|y)\|}{\sqrt{Var(\beta_i\|y)}}$	$E(\beta_i\|y)$	$\sqrt{Var(\beta_i\|y)}$	$\frac{\|E(\beta_i\|y)\|}{\sqrt{Var(\beta_i\|y)}}$	Sign
1	Primary school enrollment, 1960	0.87	0.7080	0.338	2.1	0.8101	0.218	3.7	1.0
2	Investment price	0.86	-0.3900	0.195	2.0	-0.4530	0.125	3.6	1.0
3	East Asian dummy	0.84	0.5647	0.293	1.9	0.6693	0.179	3.7	1.0
4	log GDP in 1960	0.78	-0.5828	0.371	1.6	-0.7444	0.235	3.2	1.0
5	Fraction in tropical area	0.66	-0.4703	0.370	1.3	-0.7180	0.175	4.1	1.0
6	Population coastal density	0.56	0.2516	0.246	1.0	0.4527	0.135	3.4	1.0
7	Malaria prevalence	0.20	-0.1411	0.302	0.5	-0.6950	0.256	2.7	1.0
8	Fraction Confucian	0.17	0.0725	0.176	0.4	0.4326	0.168	2.6	1.0
9	Life expectancy	0.16	0.1556	0.390	0.4	0.9554	0.415	2.3	1.0
10	Sub-Saharan dummy	0.15	-0.1068	0.282	0.4	-0.7277	0.300	2.4	1.0
11	Latin American dummy	0.14	-0.0832	0.219	0.4	-0.5825	0.214	2.7	1.0
12	Spanish colony dummy	0.11	-0.0454	0.143	0.3	-0.4216	0.173	2.4	1.0
13	Fraction GDP in mining	0.10	0.0305	0.102	0.3	0.3023	0.147	2.1	1.0
14	Ethnolinguistic fractionalization	0.10	-0.0331	0.111	0.3	-0.3316	0.155	2.1	1.0
15	Fraction Buddhist	0.09	0.0315	0.112	0.3	0.3505	0.167	2.1	1.0
16	Fraction Muslim	0.09	0.0319	0.116	0.3	0.3705	0.178	2.1	1.0
17	Population density	0.08	0.0201	0.077	0.3	0.2535	0.122	2.1	1.0
18	Government consumption share	0.08	-0.0231	0.092	0.3	-0.3078	0.157	2.0	1.0
19	Number of years open economy	0.07	0.0286	0.115	0.2	0.3851	0.198	1.9	1.0
20	Political rights	0.07	-0.0278	0.115	0.2	-0.3966	0.204	1.9	1.0
21	Fraction speaking foreign language	0.06	0.0156	0.073	0.2	0.2762	0.150	1.8	1.0
22	Openness measure 1965–74	0.05	0.0148	0.075	0.2	0.2989	0.165	1.8	1.0
23	Real exchange rate distortions	0.05	-0.0142	0.075	0.2	-0.3038	0.176	1.7	1.0
24	Higher education enrollment	0.05	-0.0158	0.084	0.2	-0.3545	0.191	1.9	1.0
25	Government share of GDP	0.04	-0.0115	0.063	0.2	-0.2578	0.159	1.6	1.0
26	Public investment share	0.04	-0.0115	0.063	0.2	-0.2709	0.155	1.7	1.0
27	Air distance to big cities	0.04	-0.0096	0.057	0.2	-0.2583	0.150	1.7	1.0
28	Primary exports	0.03	-0.0090	0.063	0.1	-0.3144	0.208	1.5	1.0
29	Fraction population under 15	0.03	0.0091	0.069	0.1	0.3383	0.259	1.3	1.0
30	Fraction population in tropics	0.03	-0.0101	0.072	0.1	-0.3673	0.242	1.5	1.0
31	Fraction Protestant	0.02	-0.0072	0.060	0.1	-0.3064	0.242	1.3	1.0
32	Fraction Hindu	0.02	0.0048	0.038	0.1	0.2047	0.146	1.4	1.0
33	Nominal Government share	0.02	-0.0032	0.030	0.1	-0.1760	0.141	1.2	1.0
34	Outward orientation	0.02	-0.0030	0.027	0.1	-0.1611	0.122	1.3	1.0
35	Revolutions and coups	0.02	-0.0026	0.026	0.1	-0.1570	0.128	1.2	1.0
36	Civil liberties	0.02	-0.0039	0.042	0.1	-0.2438	0.231	1.1	1.0
37	Fertility	0.02	-0.0049	0.061	0.1	-0.3021	0.379	0.8	0.9
38	Colony dummy	0.02	-0.0029	0.032	0.1	-0.1924	0.177	1.1	1.0
39	European dummy	0.01	-0.0004	0.047	0.0	-0.0315	0.397	0.1	0.5
40	Absolute latitude	0.01	0.0021	0.050	0.0	0.1583	0.402	0.4	0.6
41	Hydrocarbon deposits	0.01	0.0018	0.023	0.1	0.1378	0.152	0.9	0.9
42	Fraction Catholic	0.01	-0.0035	0.048	0.1	-0.2790	0.329	0.8	0.9
43	British colony dummy	0.01	0.0017	0.022	0.1	0.1406	0.152	0.9	1.0
44	Religion measure	0.01	-0.0013	0.018	0.1	-0.1137	0.123	0.9	1.0
45	Average inflation 1960–90	0.01	-0.0014	0.018	0.1	-0.1209	0.122	1.0	1.0
46	Landlocked country dummy	0.01	-0.0011	0.018	0.1	-0.1072	0.138	0.8	0.9
47	Terms of trade growth in 1960s	0.01	0.0012	0.019	0.1	0.1174	0.150	0.8	0.9
48	Defense spending share	0.01	0.0013	0.020	0.1	0.1238	0.150	0.8	1.0
49	Square of inflation 1960—90	0.01	-0.0010	0.016	0.1	-0.1037	0.121	0.9	1.0
50	Fraction of population over 65	0.01	0.0005	0.032	0.0	0.0493	0.313	0.2	0.6
51	Public education spending share	0.01	0.0011	0.018	0.1	0.1156	0.146	0.8	1.0

quite close to those in SDM, which are replicated in DW and DWa. It is, therefore, interesting to compare our findings on jointness with those in DW, based on their criterion $\mathcal{J}_{ij}^{DW}$ in (3) and those in DWa, based on $\mathcal{J}_{ij}^{DWa}$ in (4). We do not find much evidence of jointness (see Fig. 5), with only pairs $(1,2)$ and $(2,4)$ displaying posterior odds in favor of jointness greater than 3. Weak jointness (with posterior odds over 2.5) exists for pairs $(1,3)$, $(1,4)$, $(2,3)$, $(3,5)$ and $(5,6)$. No jointness is found for triplets, with 2.2 being the highest posterior odds in favor of trivariate jointness, corresponding to the triplet (1,2,4). In contrast, evidence for disjointness abounds, with 99.1% of possible pairs indicating some disjointness and 70% displaying decisive disjointness. Such disjointness is found almost exclusively between relatively unimportant regressors. There is only one occasion of decisive disjointness between two variables each having over 10% marginal inclusion probability: the Latin American and Spanish colony dummies. Clearly, these dummies are likely to be substitutes, given the large amount of former Spanish colonies in Latin America. If we consider triplets of variables, we even find decisive disjointness in 98.2% of all combinations. Again, this tends to affect mostly combinations with relatively unimportant variables. Exceptions are the triplets (Fraction in tropical area, Malaria prevalence, Life expectancy) and (Fraction in tropical area, Malaria prevalence, Sub-Saharan dummy), where it is clear not all three variables involved are required to provide the information. We do, however, need two of them, since the variables are not found to be decisively pairwise disjoint.

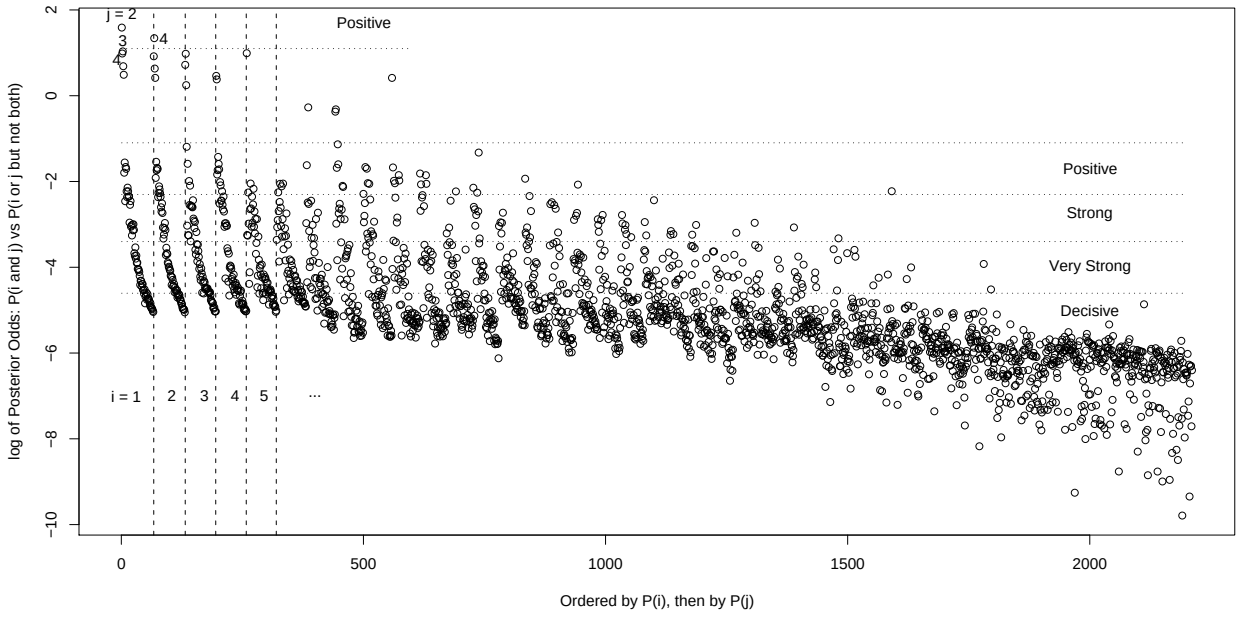


Fig 5. DW Data: Log posterior odds in favor of jointness, $\log(\mathcal{J}_{ij})$.

In sharp contrast, DW and DWa find disjointness only between variables with more than 10% inclusion probabilities. However, they both identify the disjointness between the Latin American

and Spanish colony dummies. They find pairwise disjointness between the Fraction tropical area and Malaria prevalence and between the Sub-Saharan dummy and Malaria prevalence (whereas we found this disjointness to exist only through the triplet). In addition, none of the jointness relations we find are identified by the DW criterion. Of course, finding conclusive evidence of jointness in the sense of DW in cases where the inclusion probability of one of the variables exceeds 0.37 is precluded by the very definition of $\mathcal{J}_{ij}^{DW}$, as explained in Section 2. The criterion in DWa identified jointness in four of the seven pairs in which we encountered at least weak jointness. However, two of the three remaining pairs (East Asian dummy with the investment price and with primary school enrollment) are labelled significantly disjoint according to $\mathcal{J}_{ij}^{DWa}$. As the three variables involved are the ones with highest posterior inclusion probabilities, this may be related to the potentially erratic behaviour of $\mathcal{J}_{ij}^{DWa}$ for important regressors, as discussed in Section 2. Both the measures of DW and especially that of DWa indicate many bivariate jointness links beyond those found through $\mathcal{J}_{ij}$. All this clearly shows that the DW and DWa jointness measures provide very different summaries of the posterior distribution than the measure introduced here.

4. Concluding Remarks

Growth regressions are affected by uncertainty regarding which regressors to include from a set of k potential covariates, where k could be quite large (k = 41 and k = 67 in our applications). Bayesian Model Averaging is an effective and intuitive way of dealing with this problem and leads to a k-dimensional posterior distribution for the vector of regression coefficients β, which consists of a mixture of continuous distributions and point masses in each dimension. Summarizing such a posterior distribution beyond the information contained in these marginals is quite challenging and requires well-honed tools for extracting relevant pieces of information. It is important that these tools provide us additional insight into properties of the posterior that are particularly interesting for both researchers and policymakers, and that they are easy to interpret. We argue that our jointness measure proposed in (8) satisfies both criteria: it addresses relevant questions and is directly interpretable as the posterior odds ratio between models that include a set of variables and models that include only proper subsets. In addition, it is naturally defined for any set of regressors and is not restricted to bivariate jointness evaluations (between variables or sets of variables).

We applied this jointness measure to two data sets used for growth regressions in the recent literature. In both data sets we encounter jointness only between important determinants of growth. The regressors involved are complements in that each of them has a separate role to play in explaining growth. Much more frequently, we encounter situations of disjointness, where regressors are substitutes and really should not appear together in interesting growth models. However, the regressors displaying disjointness relationships tend to be fairly unimportant drivers of growth. When we consider triplets rather than pairs of variables, these conclusion are strengthened in that we find even less jointness and more disjointness. However, even then decisive disjointness is mostly confined to variables with relatively small posterior inclusion probabilities.

Thus, the data sets analysed here seem to contain a few key growth determinants that have a clearly defined and separated role to play, and should, thus, occur jointly in growth regressions, while a substantial fraction of the regressors is of relatively small importance and captures effects that can also be accounted for by other regressors. In between we have a number of variables with typically moderate explanatory power and no clear jointness or disjointness relationships. It

is perhaps mostly due to this latter group that model uncertainty is such an important feature of growth regression.

We hope the simple jointness measure introduced here can provide a useful tool for further exploration of the posterior distribution in problems involving Bayesian variable selection. In particular, we hope it may contribute to our understanding of the growth of countries or regions.

5. References

Brock, William, and Steven Durlauf (2001), "Growth Empirics and Reality," *World Bank Economic Review*, 15: 229–72.

Brock, William, Steven Durlauf and Kenneth West (2003), "Policy Evaluation in Uncertain Economic Environments," *Brookings Papers of Economic Activity*, 235–322.

Doppelhofer, Gernot, and Melvyn Weeks (2005), "Jointness of Growth Determinants," unpublished (University of Cambridge: CWPE #0542).

Doppelhofer, Gernot, and Melvyn Weeks (2006), "Jointness of Growth Determinants," unpublished (University of Cambridge: mimeo).

Fernández, Carmen, Eduardo Ley and Mark F.J. Steel (2001a), "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, 100: 381–427.

Fernández, Carmen, Eduardo Ley and Mark F.J. Steel (2001b), "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, 16: 563–76.

George, Edward I., and Robert E. McCulloch (1997), "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7: 339–373.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery and Chris T. Volinsky (1999), "Bayesian model averaging: A tutorial," *Statistical Science* 14: 382–401.

Jeffreys, Harold (1961), *Theory of Probability*, 3rd ed., Clarendon Press: Oxford.

León-González, Roberto and Daniel Montolio (2004), "Growth, Convergence and Public Investment: A BMA Approach," *Applied Economics*, 36: 1925–36.

Levine, Ross, and David Renelt (1992), "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, 82: 942–963.

Papageorgiou, Chris, and Winford Masanjala (2005), "Initial Conditions, European Colonialism and Africa's Growth," unpublished (Baton Rouge: Department of Economics, Louisiana State University)

Raftery, Adrian E. (1995), "Bayesian Model Selection in Social Research," *Sociological Methodology*, 25: 111–63.

Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting (1997), "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92: 179–91.

Sala-i-Martin, Xavier X. (1997), "I Just Ran Two Million Regressions," *American Economic Review*, 87: 178–183.

Sala-i-Martin, Xavier X., Gernot Doppelhofer and Ronald I. Miller (2004), "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach." *American Economic Review*, 94: 813–835.

Tsangarides, Charalambos G. (2005) "Growth Empirics under Model Uncertainty: Is Africa Different?," unpublished (Washington DC: IMF Working Paper: 05/18).