Prior Specification in Bayesian Model Averaging: An application to Economic Growth

T.P Ogundunmade and A.A Adepoju

Department of Statistics, University of Ibadan, Oyo State, Nigeria.
ogundunmadetayo@yahoo.com and pojuday@yahoo.com

Abstract

Some recent cross-country cross-sectional analyses have employed Bayesian Model Averaging to tackle the issue of model uncertainty. Bayesian model averaging has become an important tool in empirical settings with large numbers of potential regressors and relatively limited number of observations. We examine the effect of a variety of prior assumptions on the inference, posterior inclusion probabilities of regressors and on predictive performance. Bayesian model averaging (BMA) has become a widely accepted way of accounting for model uncertainty in regression models. However, to implement BMA, a prior is usually specified in two parts: prior for the regression parameters and prior over the model space. Hence, the choice of prior specification becomes paramount in Bayesian inference, unfortunately, in practice, most Bayesian analyses are performed with the so-called non-informative priors (i.e. priors constructed by some formal rule). The arbitrariness in the choice of prior or choosing inappropriate priors often lead to badly behaved posteriors. It is therefore imperative to study the effect of choice of priors in Bayesian model averaging. Six candidate parameter priors namely, Unit information prior (UIP), Risk inflation criterion (RIC), Bayesian Risk Inflation criterion (BRIC), Hannan-Quinn criterion (HQ), Empirical Bayes (EBL) and hyper-g and three model priors: uniform, beta-binomial and binomial were examined in this study. The performances of the resulting eighteen cases were judged using posterior inference, posterior inclusion probabilities of regressors and predictive performance. Analyses were carried out using datasets with 8-potential drivers of growth for 126 countries from 2010 to 2014. Finally, our analysis shows that the EBL parameter prior with random model prior robustly identifies far more growth determinants than other priors.

Keywords: Prior specification, Bayesian Model Averaging, Economic growth, Predictive performance

1 Introduction

Variable selection process has received considerable attention in the econometrics and statistics literature over the years. The special issues of the Journal of

Econometrics (Vol. 16, No. 1, 1981) and Statistica Sinica (Vol. 7, No. 2, 1997) were devoted to the subject of model selection process in view of its relevance. In modeling, functional forms, whether linear or non linear; error structures, whether addictive or multiplicative; variables inclusion, whether redundant or relevant; and model choice are uncertain. When a single model is selected by considering its highest posterior probability value and later use to make inference, it ignores and underestimates the uncertainties involved in the models. The overall uncertainty about quantities of interest usually lead to unreliable standard error and confidence interval of such model.

A Bayesian tool for dealing with such uncertainties inherent in the model selection process is known as Bayesian Model Averaging (BMA). BMA is used in empirical settings with large amount of possible regressors and relatively limited number of observations to account for uncertainties. It is a weighted averaging method based on posterior distribution which the literature has shown that it provides an improved out - of sample predictive performance. Bayesian Model Averaging has been applied successfully to many econometric model classes especially the normal linear regression model [Raftery etal 1997; Raftery etal 2005; Fernandez etal (2001a, 2001b)].

BMA is a method that allows selecting models consistently from a model space, without having to analyse every model in order to determine which ones better fit the data or assist to predict more accurately a variable of interest. This can be achieved by drawing a sample of models from the distribution of the model space and order them according to the posterior probability, which relies on the likelihood of the model and a prior belief on each particular model. Thus, the weight attributed to each prediction to be combined is given by each posterior model probability. In Bayesian paradigm, posterior model probabilities (PMP) are very sensitive to Prior specification (Zellner (1986)), (Fernandez etal (FLS) (2001a)), (Olubusoye and olawale (2009)) and (Ley et al (2011)) under the uncertainty inherent in model selection. Also, the best priors to be elicited for both the quantities of interest and the choice of models in the BMA approach are major problems encountered by the Bayesian econometricians. Using the BMA, priors are elicited in two forms namely, model and parameter priors. Model priors can be fixed, random, uniform or even custom prior inclusion probability while the parameter priors can also be fixed, empirical Bayes (local) or hyper g prior.

From the literature, the Zellner g- structure in the parameter prior is expected to be as small as possible such that consistency of the true posterior model probability holds. And improving on work of the Fernandez etal (2001a) priors, therefore this study focuses on the g-parameter prior elicitation in the BMA approach to normal linear regression model. This modified prior structure allows the marginal likelihood be computed analytically and does not violate the probability rule (not rely on the sample model y). It is independent of input from the researcher or information in the data but depends only on the sample size. Three different model priors and six parameter priors were elicited for this study.

The rest of the paper is divided into four sections. Section II describes the BMA theoretical approach while section III elicits the determinants of economic growth. Section IV shows the results of the data and the analysis and finally, the discussion of the results and concluding remarks are given in section V.

2 Bayesian Model averaging Framework

Consider n independent random samples from a normal linear regression model with constant term, β_0, K potential explanatory variables in a matrix X of dimension n x K and a normal IID error term ϵ with variance $\sigma^2 \equiv (h^{-1})$. If X_j are the explanatory variables included in Model j, $j = 1, ..., M$ ($M = 2^K$ plausible models) which contains the choice of $0 < k_j < K$ (k_j is the number of regressors in model j) and then leads to the linear regression model, 'y' of the form:

$$y = \beta_{0i_n} + X_j\beta_j + \varepsilon \qquad \epsilon \sim N\left(0, h^{-1}I\right) \tag{1}$$

In Bayesian paradigm, we may be interested in a quantity or variable (β_j) across the entire model space. Hence, the posterior distribution of the quantity of interest, (β_j) given the data is as:

$$P\left(\beta_j/D\right) = \sum_{j=1}^{2^K} P\left(\beta_j/D, M_j\right) P\left(M_j/D\right) \tag{2}$$

where D is the sample data. Equation (2) is the mixture of the posterior distributions of that quantity under each of the models with the weighted probability model - $P\left(M_j/D\right)$

Thus,

$$P\left(M_j/D\right) = \frac{P\left(D/M_j\right)P\left(M_j\right)}{\sum\limits_{i=1}^{2^K} P\left(D/M_i\right)P\left(M_i\right)} = \left[\sum_{i=1}^{2^K} \frac{P\left(D/M_i\right)P\left(M_i\right)}{P\left(D/M_j\right)P\left(M_j\right)}\right]^{-1} \tag{3}$$

Where, $P\left(M_j\right)$ is prior probability that M_j is the true model and the Marginal Likelihood is given by

$$P\left(D/M_j\right) = \int_0^\infty P\left(D/\beta_0, \beta_j, h^{-1}, M_j\right) P\left(\beta_0, \beta_j, h^{-1}/M_j\right) d\beta_0, d\beta_j, dh^{-1} \tag{4}$$

where h is the model precision, $P\left(D/M_j\right)$ -Marginal likelihood, $P(D/\beta_0,\ \beta_j,\ h^{-1},$ $M_j)$ likelihood of the data and $P\left(\beta_0, \beta_j, h^{-1}/M_j\right)$ - the prior distribution for the parameters given the model. BMA gets the Posterior Inclusion Probability (PIP) of an explanatory variable by summing the Posterior Model Probabilities across those models that contain the explanatory variable.

Introducing the g parameter prior to (4) and taking its ratio gives us another very important tool in BMA called the Bayes factor (B_{jr}) for comparing two models M_j and M_r and can be computed analytically by

$$B_{jr} = \left(\frac{g_j}{g_j+1}\right)^{k_j/2} \left(\frac{g_r+1}{g_r}\right)^{k_r/2} \left[\frac{\frac{1}{g_r+1}y^1 Q_{X_r} y + \frac{g_r}{g_r+1}\left(y-\overline{y}_{i_n}\right)^1\left(y-\overline{y}_{i_n}\right)}{\frac{1}{g_j+1}y^1 Q_{X_j} y + \frac{g_j}{g_j+1}\left(y-\overline{y}_{i_n}\right)^1\left(y-\overline{y}_{i_n}\right)}\right]^{\frac{n-1}{2}} if\mathrm{k_j},\ \mathrm{k_r} \geq 1 \tag{5}$$

Source: Fernadez et al. (2001a)

2.1 Priors in BMA

Basically, in BMA two priors are specified namely, model priors and parameter priors. The model prior $P(M_j)$ is specified by the researcher which should reflect the prior belief about the model. A common choice is to elicit a uniform prior probability for each model to explain the lack of prior information and this follows from the rule of thumb. There are other model priors like binomial, beta-binomial and custom prior inclusion probabilities but for this study the binomial, beta-binomial and the uniform model prior are used and are stated as follow:

$$P\left(M_j\right) = 1/2^K, \qquad P\left(M_j\right) > 0 \qquad \text{and} \qquad \sum_{j=1}^{\mathrm{M}} \mathrm{P}\left(\mathrm{M_j}\right) = 1 \tag{6}$$

The g- parameter priors considered in this study are Unit Information Prior(UIP), Local empirical Bayes, Hyper-g or fixed (Zellner-g) prior,Bayesian Risk Inflation criterion (BRIC), Hannan-Quinn criterion (HQ), Empirical Bayes (EBL). The g-prior was first introduced under the BMA in Zellner (1986). Zellner assumed that covariance of the prior should be proportional to covariance expression $(X_j^{*'}X_j^*)^{-1}$ of the posterior gotten from the data with the scalar g (to be elicited by the researcher) to determine the degree of importance attributed to the prior precision. And from the literature, this g prior structure has shown that it leads to simple closed form expressions of posterior statistics.

Following from the rule of Thumb, the prior probability for

The intercept is

$$P(\beta_0) = 1 \tag{7}$$

The prior probability for precision is

$$P(h) = (1/h) \tag{8}$$

Then,
The parameter prior is

$$P\left(\beta_j/h\right) \sim N\left(0_{k_j}, h^{-1}\left[g_j X_j^{*'} X_j^*\right]^{-1}\right) \tag{9}$$

Source: Zellner, 1986

Where, X_j^* is the mean deviation of X_j, and the g-prior is proportional to the comparable data based quantity; the smaller the g, the fewer the prior parameter variance.

Table 1 shows the g-parameter priors elicited in the literature. These g-priors in the BMA are related to a natural conjugate prior with the scalar g to be elicited by the researcher (Zellner (1986)). The g-prior with unit information prior (UIP), $g = (1/n)$ and $g = (1/\sqrt{n})$ explaining that the priors contain information approximately equal to that contained in a single typical observation. Also, their resulting posterior model probabilities are closely approximated by Schwarz (SIC) or Bayesian Information Criterion (BIC). They have the same mean and precision except for UIP with maximum likelihood as its mean. Prior 2, it is called Bayesian Risk Inflation Criterion prior. Prior 3 explains the decrease in the prior information even slower with sample size and there is asymptotic convergence to the Hannan Quinn Criterion with $CHQ = 1$. But if the g-prior elicitation depends only on the regressors like $g = 1/K^2$, it is approximated by Risk Inflation Criterion (RIC), the larger the value of K the higher the prior information. Prior can also rely on information from the data with R^2 known, the closer the R^2 to 1 the smaller the precision and the higher the prior information. This type of g-class (prior 5 of Table 1) is called Hyper-g prior, the data dependent prior as elicited in the work of Raftery et al (1997). Another class of g with a natural conjugate prior structure which is subjectively elicited through predictive implications is prior six specification.

Table 1: The g-Parameter Prior Structures

Prior	Specification of g-prior	Comment	Source
1	Unit Information Prior	The prior contains information approximately equal to that contained in a single typical observation. The resulting posterior model probabilities are closely approximated by the Schwarz Criterion, BIC.	Kass and Wasserman (1995), Raftery (1995)
2	*BRIC*	A mechanism that asymptotically converges to the unit information prior (g =N) or the risk in ation criterion (g = K2). That is, the g prior is set to g = max(N;K2).	FLS (2001b)
3	*HQ*	The Hannan-Quinn criterion. $CHQ = 3$ as n becomes large.	Hannan-Quinn (1979)
4	*EBL*	Prior information decreases even slower with sample size and there is asymptotic convergence to the Hannan-Quinn criterion with CHQ=1.	Hannan-Quinn (1979)
5	RIC	Sets $g = K^2$ and conforms to the risk in ation criterion	Foster and George (1994)
6	$Hyper - g$	This option uses a family of priors on g that provides improved mean square risk over ordinary maximum likelihood estimates in the normal means problem. An advantage of the hyper-g prior is that the posterior distribution of g given a model is available in closed form.	Strawder (1971)

source: Eicher et al 2009

3 Determining Growth Determinants

Since economic growth is the fundamental driver of living standards, it is of great interest to economists and policymakers alike to identify which of the numerous theories proposed receive support from the data and which determinants have a significant effect on growth. Attempts to identify robust growth determinants date back to Levine and Renelt (1992), who used extreme bounds analysis. Formal BMA analysis was conducted by Brock and Durlauf (2001), FLS (2001a) and SDM (2004). The dataset used across studies always contains a core of at least 41 candidate regressors, motivated by Sala-i-Martin (1997) and FLS (2001a).

In this section a time series cross-sectional (panel) data of 126 countries has been used in the analysis. The annual time period ranges from 2010 to 2014. The variables considered for the countries are the GDP, Government Consumption rate, Inflation rate, Fiscal Policy Rate,Unemployment Rate, Industrial Production, Trade Opennes, Exchange Rate and Public Debt.

Data for all these variables, was obtained from World Bank World Development Indicators (WDI). Trade openness data was obtained using the simple measure (exports plus imports divided by GDP). For exports, imports and GDP data, these were obtained also from World Bank WDI. For datasets with small numbers of observations, priors play important role.

4 Analysis of Results

In this section we will present Posterior Inference, Posterior Inclusion Probability (PIP) and the Predictive Inference results for three model priors discussed in the methodology above. We focus, in particular, on the effect of the prior choices on posterior model distributions, the spread of the posterior mass over model space, posterior model probabilities and the inclusion of individual regressors.

4.1 Assessment of Prior Distributions Using Posterior Inference

Tables 2,3 and 4 present the posterior inference using the three stated model priors against each of the parameter priors discussd above. The response variable is only associated with the stochastic error term. The average posterior probability of the model used here indicates higher probability value for small samples but lower (0.52) compared to Table 2 as n tends to infinity (0.9996). It is also noticed that the average probability for the two models is equally likely with a value of 0.9678.

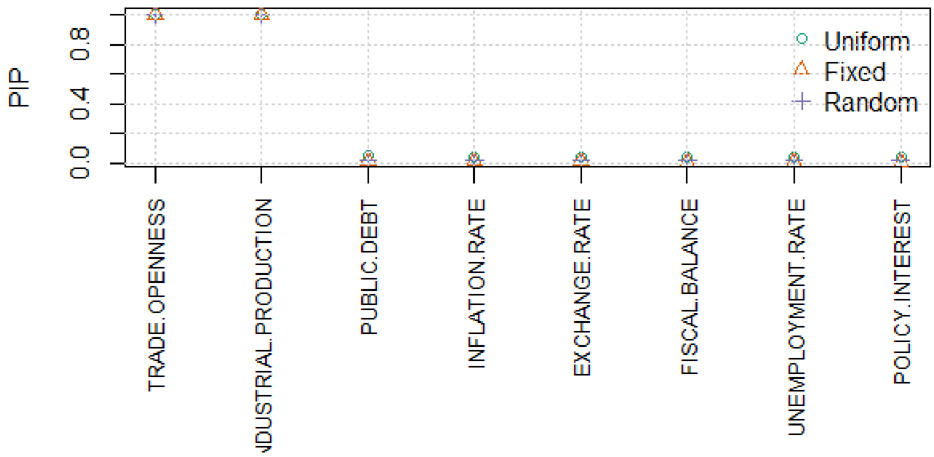

Figure 1: **Model Probabilities Composition**

Table 2: Posterior Inference using Uniform Model Prior(Mean and Standard Deviation)

	PARAMETER PRIORS					
VARIABLES	UIP	RIC	BRIC	HQ	EBL	HYPER
FISCAL BALANCE	1054.96	1038.78	1054.96	1050.17	995.17	996.29
	(168.19)	(167.06)	(168.19)	(167.83)	(164.45)	(164.53)
GOVERNMENT CONSUMPTION	-1634.04	-1533.83	-1634.04	-1586.27	-1450.64	-1456.21
	(417.83)	(386.94)	(417.83)	(403.86)	(371.42)	(373.83)
INDUSTRIAL PRODUCTION	-533.77	-623.26	-533.77	-583.41	-621.90	-617.37
	(343.94)	(285.33)	(343.94)	(318.01)	(257.54)	(262.12)
TRADE OPENNESS	-1159.97	-1380.19	-1159.97	-1280.33	-1384.02	-1373.64
	(795.83)	(665.49)	(795.83)	(740.27)	(598.76)	(609.71)
POLICY INTEREST	-32.64	-75.64	-32.64	-47.88	-111.24	-104.51
	(105.29)	(149.48)	(105.29)	(124.69)	(166.25)	(163.33)
INFLATION RATE	17.26	36.58	17.26	24.10	53.02	49.93
	(54.44)	(73.92)	(54.45)	(62.83)	(81.94)	(80.53)
PUBLIC DEBT	0.8025	2.0908	0.8025	1.2051	3.18	3.50
	(6.49)	(10.2306)	(6.46)	(7.87)	(13.42)	(12.91)
EXCHANGE RATE	-0.6794	-0.0176	-0.0068	0.0103	-0.031	-0.0285
	(0.0648)	(0.1026)	(0.0648)	(0.0792)	(0.13)	(0.129)
UNEMPLOYMENT RATE	3.0622	7.5331	3.0622	4.5115	13.04	12.05
	(30.40)	(47.57)	(30.40)	(36.88)	(62.26)	(59.898)

Table 3: Posterior Inference using Fixed Model Prior (Mean and Standard Deviation)

	PARAMETER PRIORS					
VARIABLES	UIP	RIC	BRIC	HQ	EBL	HYPER
FISCAL BALANCE	1054.96	1038.78	1054.96	1050.17	995.17	996.29
	(168.19)	(167.06)	(168.19)	(167.83)	(164.45)	(164.53)
GOVERNMENT CONSUMPTION	-1634.03	-1533.83	-1634.04	-1586.27	-1450.64	-1456.21
	(417.83)	(386.94)	(417.83)	(403.86)	(371.42)	(373.83)
INDUSTRIAL PRODUCTION	-533.77	-623.26	-533.77	-583.41	-621.90	-617.37
	(343.94)	(285.33)	(343.94)	(318.01)	(257.54)	(262.12)
TRADE OPENNESS	-1159.97	-1380.19	-1159.97	-1280.33	-1384.02	-1373.64
	(795.83)	(665.49)	(795.83)	(740.27)	(598.76)	(609.71)
POLICY INTEREST	-32.64	-75.64	-32.64	-47.88	-111.24	-104.51
	(105.29)	(149.48)	(105.29)	(124.69)	(166.25)	(163.33)
INFLATION RATE	17.26	36.58	17.26	24.10	53.02	49.93
	(54.44)	(73.92)	(54.45)	(62.83)	(81.94)	(80.53)
PUBLIC DEBT	0.8025	2.0908	0.8025	1.2051	3.18	3.50
	(6.49)	(10.2306)	(6.46)	(7.87)	(13.42)	(12.91)
EXCHANGE RATE	-0.6794	-0.0176	-0.0068	0.0103	-0.031	-0.0285
	(0.0648)	(0.1026)	(0.0648)	(0.0792)	(0.13)	(0.129)
UNEMPLOYMENT RATE	3.0622	7.5331	3.0622	4.5115	13.04	12.05
	(30.40)	(47.57)	(30.40)	(36.88)	(62.26)	(59.898)

Table 4: Posterior Inference using Random Model Prior(Mean and Standard Deviation)

	PARAMETER PRIORS					
VARIABLES	UIP	RIC	BRIC	HQ	EBL	HYPER
FISCAL BALANCE	1058.18	1039.04	1058.18	1051.89	984.51	986.43
	(168.52)	(167.12)	(168.52)	(168.04)	(163.94)	(164.14)
GOVERNMENT CONSUMPTION	-1697.67	-1538.33	-1697.67	-1620.22	-1421.73	-1429.57
	(433.59)	(386.88)	(433.59)	(416.77)	(364.69)	(368.17)
INDUSTRIAL PRODUCTION	-460.28	-618.27	-460.28	-544.398	-630.17	-625.17
	(367.38)	(290.14)	(367.38)	(339.34)	(242.69)	(248.75)
TRADE OPENNESS	-992.85	-1369.16	-992.85	-1190.22	-1404.63	-1392.40
	(839.21)	(674.78)	(839.21)	(784.01)	(561.42)	(575.96)
POLICY INTEREST	-29.29	-94.07	-29.29	-49.84	-164.32	-154.22
	(100.22)	(161.40)	(100.22)	(126.83)	(177.88)	(176.87)
INFLATION RATE	15.54	45.22	17.26	24.95	78.42	73.64
	(51.89)	(79.61)	(54.45)	(63.699)	(18.50)	(87.54)
PUBLIC DEBT	0.7644	32.30	0.7644	1.3956	8.1984	7.46
	(6.2923)	(12.49)	(6.29)	(8.42)	(18.50)	(17.81)
EXCHANGE RATE	-0.0063	-0.02595	-0.0063	-0.0116	-0.0632	-0.0577
	(0.0627)	(0.1238)	(0.0626)	(0.0839)	(0.1870)	(0.1794)
UNEMPLOYMENT RATE	2.853	10.898	2.85	5.0263	25.84	23.63
	(29.40)	(57.37)	(29.40)	(39.03)	(87.39)	(83.75)

Note: The values above shows the posterior mean of the economic variables under different parameter priors and the Posterior standard deviations in the brackets.

4.2 Assessment of Prior Distributions Using Posterior Inclusion Probability(PIP)

Tables 5,6 and 7 report the BMA posterior inclusion probabilities for all 6 prior distributions applied to the growth dataset. Table 5 shows the result of the uniform model priors against the 6 parameter priors. Table 6 shows the result for fixed model prior and table 7 for random model prior. Posterior inclusion probabilities and the number of regressors that exhibit evidence of an effect on growth vary substantially across priors. The number of regressors whose inclusion probability exceeds 50% ranges from a low of four regressors (Priors UIP, Hyper and RIC) to a high of 4 regressors (EBL) considering random model prior.

Table 5: Posterior Inclusion Probability(PIP) for Uniform Model Prior

	PARAMETER PRIORS					
VARIABLES	UIP	RIC	BRIC	HQ	EBL	HYPER
FISCAL BALANCE	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
GOVERNMENT CONSUMPTION	0.9966	0.9982	0.9966	0.9974	0.9986	0.9985
INDUSTRIAL PRODUCTION	0.7848	0.9164	0.7848	0.8551	0.9496	0.9427
TRADE OPENNESS	0.7578	0.9027	0.7578	0.8342	0.9409	0.9330
POLICY INTEREST	0.1227	0.2838	0.1227	0.1791	0.4368	0.4103
INFLATION RATE	0.1267	0.2762	0.1267	0.1788	0.4430	0.3977
PUBLIC DEBT	0.0479	0.1234	0.0479	0.0717	0.2297	0.2115
EXCHANGE RATE	0.0448	0.1154	0.0448	0.0673	0.2147	0.1977
UNEMPLOYMENT RATE	0.0441	0.1127	0.0441	0.0659	0.2098	0.1932

Table 6: Posterior Inclusion Probability(PIP) for Fixed Model Prior

	PARAMETER PRIORS					
VARIABLES	UIP	RIC	BRIC	HQ	EBL	HYPER
FISCAL BALANCE	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
GOVERNMENT CONSUMPTION	0.9966	0.9982	0.9966	0.9974	0.9986	0.9985
INDUSTRIAL PRODUCTION	0.7848	0.9164	0.7848	0.8551	0.9496	0.9427
TRADE OPENNESS	0.7578	0.9027	0.7578	0.8342	0.9409	0.9330
POLICY INTEREST	0.1227	0.2838	0.1227	0.1791	0.4368	0.4103
INFLATION RATE	0.1267	0.2762	0.1267	0.1788	0.423	0.3977
PUBLIC DEBT	0.0479	0.1234	0.0479	0.0717	0.2297	0.2115
EXCHANGE RATE	0.0448	0.1154	0.0448	0.0673	0.2147	0.1977
UNEMPLOYMENT RATE	0.0441	0.1127	0.0441	0.0659	0.2098	0.1932

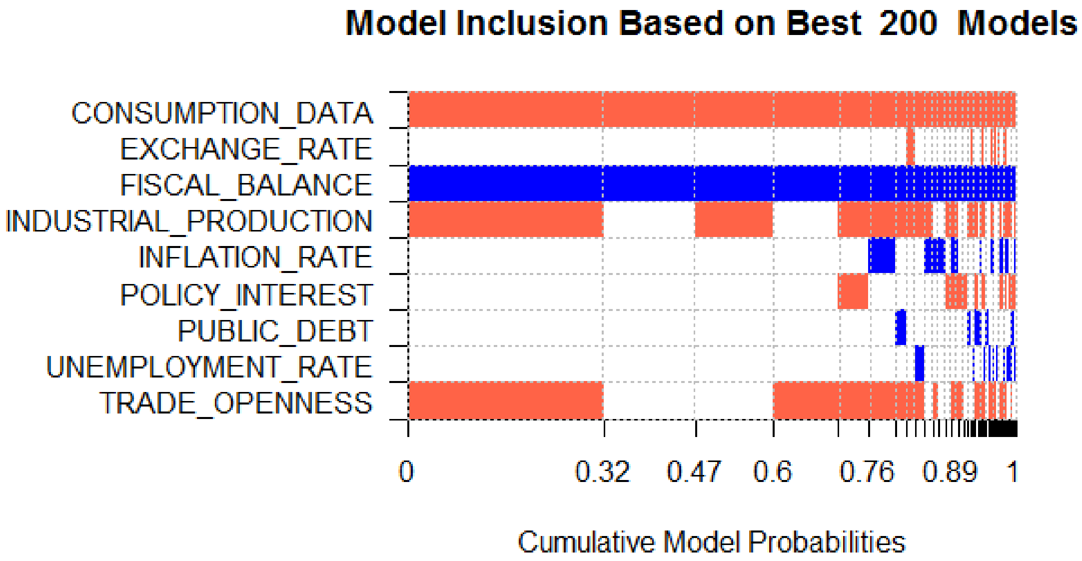

Figure 2: **Cummulative model inclusion probabilities Probabilties**

Table 7: Posterior Inclusion Probability(PIP) for Random Model Prior

	PARAMETER PRIORS					
VARIABLES	UIP	RIC	BRIC	HQ	EBL	HYPER
FISCAL BALANCE	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
GOVERNMENT CONSUMPTION	0.9964	0.9983	0.9964	0.9971	0.9992	0.999
INDUSTRIAL PRODUCTION	0.6791	0.9082	0.6791	0.7992	0.9689	0.9606
TRADE OPENNESS	0.6508	0.8950	0.6508	0.7767	0.9637	0.9544
POLICY INTEREST	0.1102	0.3438	0.1101	0.1858	0.6396	0.6112
INFLATION RATE	0.1142	0.2762	0.1267	0.1788	0.4430	0.3977
PUBLIC DEBT	0.0451	0.1837	0.0451	0.0814	0.4740	0.4324
EXCHANGE RATE	0.0420	0.1718	0.0420	0.0760	0.4532	0.4126
UNEMPLOYMENT RATE	0.0414	0.1680	0.0414	0.00744	0.4468	0.4065

Figure 2 below shows the cumulative model inclusion probabilities based on best 200 models. It also depicts the inclusion of a regressor with its sign in the model selection process. This image plot is based on the UIP parameter [rior against uniform model priro. Blue color corresponds to a positive coefficient, red to a negative coefficient, and white to non-inclusion(a zero coefficient) of the respective variable. The horizontal axis is scaled by the models' posterior model probabilities. It is confirmed that the selected best model with PMP of 29% includes only the real interest rate of red colour (a negative sign of the posterior mean).

4.3 Assessment of Prior Distributions Using Predictive Performance

We now compare the competing default priors on the basis of predictive performance on hold-out samples, a neutral criterion that allows the comparison of different methods on the same footing. We compare the performance of the full predictive distributions produced by the methods, as well as that of point predictions. We use a proposed method by Theo Eicher (”bma.compare” proposed by Theo Eicher(2010) , programmed in R) simultaneously evaluates all 6 different parameter priors and any specific prior expected model size, as well as their predictive performance. We divide the dataset randomly into a training set,D^T , which is used to estimate the BMA predictive distribution, and a hold-out set,D^H , which is used to assess the quality of the resulting predictive distributions. We use three different criteria, or scoring rules: the mean squared error (MSE) of prediction, the log predictive score (LPS; Good, 1952), and the continuous ranked probability score (CRPS; Matheson and Winkler , 1976). All our scoring rules are negatively oriented, that is, lower is better.

The MSE is the most popular measure to assess predictive performance in economics. It focuses on point estimation, while the LPS and the CRPS assess the entire predictive distribution. The CRPS and the LPS assess both the sharpness of a predictive distribution and its calibration, namely the consistency between the distributional forecasts and the observations. However, the LPS assigns harsh penalties to particularly poor probabilistic forecasts, and can be very sensitive to outliers and extreme events (Weigend and Shi, 2000; Gneiting and Raftery, 2007). This may be a factor when we split our small sample to examine predictive performance. The CRPS is more robust to outliers (Carney, Cunningham and Byrne, 2006; Gneiting and Raftery, 2007), and hence it is our preferred measure of the performance of the predictive distribution as a whole.
The MSE of prediction is conventionally used to assess the quality of point predictions. The BMA point prediction for an observation in the hold-out dataset ,y_{new}, with predictors x_{new}, is

$$y_{new}, BMA = \sum_{k=1}^{k} E[y_{new}|x_{new}, D^T, M_k]pr(M_k)|D^T$$

The MSE of prediction is then

$$1/n_H \sum_{y_{new} \in D^H} (y_{new} - \hat{y}_{new}, BMA)^2$$

where n_H is the number of observations in D^T.
The other two scoring rules measure the quality of the predictive distribution as a whole. The BMA predictive distribution is

$$pr_{BMA}, (y_{new}) = \sum_{k=1}^{k} pr[y_{new}|x_{new}, D^T, M_k]pr(M_k)|D^T$$

The LPS is then defined as

$$LPS = -\sum_{y_{new} \in D^H} logpr_{BMA}(y_{new})$$

Let $F_{BMA}(y_{new})$ be the cumulative distribution function corresponding to the BMA predictive density$F_{BMA}(y_{new})$. Then the CRPS for the single observation y_{new} is

$$CPRS(y_{new}) = \int_{-\inf}^{+\inf} (F_{BMA}(y) - 1\{y_{new} < y\})^2 dy$$

where $1\{y_{new} < y\} = 1$ if $y_{new} < y$ and 0 otherwise. The CRPS for the hold-out dataset as a whole is then

$$CPRS = 1/n_H \sum_{y_{new} \in D^H} CPRS(y_{new})$$

The CRPS measures the area between a step function at the observed value and the predictive cumulative distribution function. Unlike the LPS, it is defined when the prediction is deterministic; in that case it reduces to the mean absolute error (Hersbach, 2002). The LPS and the CRPS assess both the sharpness of a predictive distribution and its calibration, namely the consistency between the distributional forecasts and the observations. However, the LPS assigns particularly harsh penalties to poor probabilistic forecasts, and so can be very sensitive to outliers and extreme events (Weigend and Shi, 2000; Gneiting and Raftery, 2007). The CRPS is more robust to outliers (Carney et al., 2009; Gneiting and Raftery, 2007), and hence it is our preferred measure of the performance of the predictive distribution as a whole. We also report the LPS for comparability with previous work, notably that of FLS (2001b) and LS. We divided the dataset randomly into a training set that contains 80% of the data and thus leaves 20% of the data to be predicted, and we repeated the analysis for 400 different random splits, reporting the average over all splits.

Table 8 shows the predictive performance of the 6 parameter priors in conjunction with uniform, fixed and random model priors as evaluated by the MSE, LPS and CRPS. The MSE and the CRPS agree that our baseline UIP decisively outperformed all the other priors. The LPS suggests, however, that EBL and BRIC outperform UIP. Since this result runs counter to the results from the two other scoring rules, it seems possible that the difference is due to influential observations in the dataset or outliers in a particular subsample. Several of the regressors have extreme outlying values. When such cases are in the test set, they can have a large effect on the LPS, while the CRPS is more robust to individual cases. Given the known outlier sensitivity of the LPS, we discount the results it gives for this dataset, and conclude that EBL performs best in this case.

Table 8: Parameter priors and predictive performance: performance scores relative to parameter UIP

Prior	Uniform model	Fixed model	Random model
MSE			
BRIC	0.715	1.000	1.03
HQ	1.08	1.06	0.907
EBL	0.621	0.628	0.969
RIC	0.983	0.767	0.926
HYPER	0.808	0.903	0.786
CRPS			
BRIC	0.498	0.567	0.525
HQ	0.583	0.556	0.533
EBL	0.454	0.489	0.558
RIC	0.531	0.505	0.543
HYPER	0.539	0.551	0.524
LPS			
BRIC	160	179	177
HQ	183	183	172
EBL	155	154	176
RIC	170	163	174
HYPER	166	173	164

Table 9 represents the MCMC and the exact posterior probabilities for the first best 10 models. The numbers in the left-hand column represent analytical PMPs (PMP (Exact)) while the right-hand side displays MCMC-based PMPs (PMP (MCMC)). Both decline in roughly the same fashion, however sometimes the values for analytical PMPs differ considerably from the MCMC-based ones. This comes from the fact that MCMC-based PMPs derive from the number of iteration counts, while the "exact" PMPs are calculated from comparing the analytical likelihoods of the best models. Both columns sum up to the same number and show that in total, the top 2,000 models of posterior model mass.

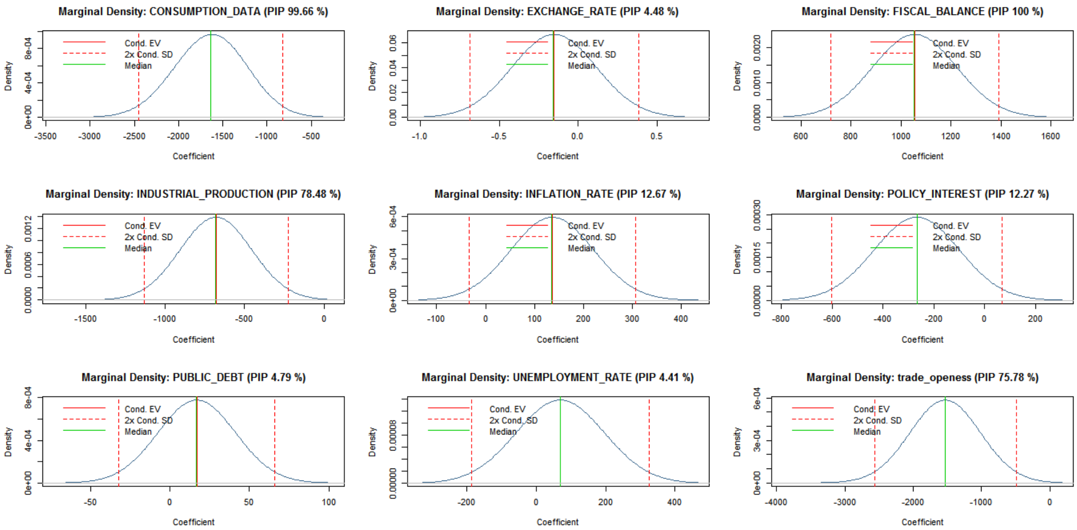

Figure 3: **Posterior Marginal Density**

Table 9: The MCMC and the Exact Posterior probabilities for the First Best 10 Models

Models	PMP (Exact)	PMP (MCMC)
021	0.8670	0.8670
025	0.0255	0.0260
031	0.0212	0.02124
0a1	0.0199	0.020
061	0.01879	0.0188
023	0.01766	0.0177
029	0.0177	0.0176
035	0.00097	0.00097
0a5	0.00094	0.00094
065	0.00092	0.00092

Figure 3 shows the computed marginal posterior densities are a Bayesian model averaging mixture of the marginal posterior densities of the individual models. The accuracy of the result therefore depends on the number of "best" models. Note that the marginal posterior density can be interpreted as "conditional on inclusion": If the posterior inclusion probability of a variable is smaller than one, then some of its posterior density is Dirac at zero. Therefore the integral of the returned density vector adds up to the posterior inclusion probability, i.e., the probability that the coefficient is not zero.

Figure 4,5 and 6 below show the posterior inclusion probabilities plot of the 6 parameter priors under different model priors. The trend shows the posterior inclusion probabilities of the economic variables captures by the different parameter priors. Figure 4 shows the trend of the priors under uniform model prior. Figure 5 shows the trend under fixed model prior and figure 6 for the random model prior.

5 Conclusion

To identify the best prior for our growth dataset, we examine the predictive performance of 6 candidate default parameter priors that have been proposed in the economics and statistics literature, as well as three candidate model priors. We argue that predictive performance is a neutral criterion for comparing different priors, and we introduce an improved scoring rule. In addition, we examine these priors success in identifying the right determinants in the datasets. The Empirical Bayes Local(EBL) for the parameters performed consistently better than the other 5 priors in the growth data, and in the data, and as measured by all three scoring rules. We view the random model prior together with the Empirical Bayes Local(EBL) as a reasonable default prior and starting place, but our results also highlight that researchers should also assess other possibilities that may be more appropriate for their data.

In spite of widespread doubts about the ability of the small cross-country growth dataset to provide a rich set of growth determinants, our analysis shows that the random model prior together with the Empirical Bayes Local(EBL) robustly identifies far more growth determinants than other priors. The random model prior discovers substantial evidence for 5 growth determinants as compared to others considered. Hence we show that the appropriate prior in the growth context delivers a rich set of robust growth determinants that also generate good predictive performance. The new regressors prominently feature fiscal balance,trade openness, government consumption and industrial production. Thus our results provide support for several new growth theories.

References

[1] Albert.J. 2007. Bayesian Computation with R. Springer. ISBN 9780-387-71384-7.

[2] Barbieri, M.M. and Berger, J.O. (2004). Optimal predictive model selection. Annals of Statistics 32, 870-897

[3] Brock W, Durlauf SN. 2001. Growth empirics and reality. World Bank Economic Review 15: 229272.

[4] Clyde, M. and E.I. George. (2004). "Model Uncertainty," Statistical Science 19, 81-94.

[5] Clyde, M., H. DeSimone and G. Parmigiani. (1996). "Prediction Via Orthogonalized Model Mixing," Journal of the American Statistical Association 91, 1197-1208.

[6] Eicher, T.S., C. Papageorgiou and A.E. Raftery (2009) "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, Seattle.

[7] Fernndez C., E. Ley and M.F.J. Steel.(FLS) (2001a). "Benchmark Priors for Bayesian Model Averaging", Journal of Econometrics 100, 381-427.

[8] Fernndez C., E. Ley and M.F.J. Steel. (2001b). Model Uncertainty in Cross-Country Growth Regressions", Journal of Applied Econometrics 16, 563-576.

[9] Furnival G.M. and R.W. Wilson. (1974). "Regressions by Leaps and Bounds",Technometrics 16, 499-511.

[10] George, E.I. and R.E. McCulloch. (1993). "Variable Selection via Gibbs Sampling", Journal of the American Statistical Association 88, 881-889.

[11] Hoeting, J.A., D. Madigan, A.E. Raftery and C.T. Volinsky. (1999). "Bayesian Model Averaging: A Tutorial", Statistical Science 14, 382-417.

[12] Levine, R.and D.Renelt(1992), "A sensitivity analysis of cross-country growth regressions", American Economic Review, 82: 942-63.

[13] Madigan, D. and A.E. Raftery. (1994). "Model Selection and Accounting for Model Uncertainty in Graphical Models using Occams Window", Journal of the American Statistical Association 89, 1535-1546.

[14] Madigan, D., Gavrin, J. and Raftery, A.E. (1995), "Eliciting Prior Information to Enhance the Predictive Performance of Bayesian Graphical Models", Communications in Statistics, Theory and Methods, 24, 2271-2292.

[15] McMahon T,(2008), 'Inflation: 'Causes and effects". In inflatioData.com

[16] Mishkin F.S and K. Schmidt-Hebbel, (2001), "One Decade of inflation Targeting in the world. What do we know and what do we to know?" Cental Bank of Chile working paper No 101, July.

[17] Olubusoye, O.E and Okewole, D.M (2009). Prior Sensitivity in Bayesian Linear regression Model. Int. Journal (Sciences) Vol. 3, No. 1, pp21-29

[18] Olubusoye and Oyaromade (2008). "Inflation modelling process in Nigeria Afrca Economic Reseach Consortium, Nairobi, AERC Reseach Paper 182.

[19] Raftery, A.E. (1995). "Bayesian Model Selection for Social Research", Sociological Methodology 25, 111-163.

[20] Raftery, A.E. (1996). "Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models". Biometrika 83, 251-266.

[21] Raftery, A.E. (1999). "Bayes Factors and BIC: Comment on Weakliem", Sociological Methods and Research 27, 411-427.

[22] Raftery, A.E., D. Madigan and J.A. Hoeting. (1997). "Bayesian Model Averaging for Linear Regression Models", Journal of the American Statistical Association 92, 179-191.

[23] Raftery, A.E., Painter, I. and Volinsky, C.T. (2005). "BMA: An R package for Bayesian Model Averaging." R News 5, no. 2, 2-8.

[24] Raftery, A.E., J.A. Hoeting, C.T. Volinsky, I. Painter and K.Y. Yeung (2009). "BMA: An R package for Bayesian Model Averaging.", http://cran.r-project.org/web/packages/BMA/

[25] Sala-i-Martin, X. (1997). "I Just Ran Two Million Regressions", AEA Papers and Proceedings 87, 178-183.

[26] Sala-i-Martin, X. (1997b), "I have just run four million regressions", unpublished typescript, Economic Department , Columbia University.

[27] Sala-i-Martin, X., G. Doppelhofer and R.I. Miller.(SDM) (2004). "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," American Economic Review 94, 813-835.

[28] Selialia, F.L. 1995. The Dynamics of Inflation in Lesotho. Unpublished M.A. Thesis. University College, Dublin.

[29] Zellner A. 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, Goel PK, Zellner A (eds). North-Holland: Amsterdam; 233-243.

6 Appendix

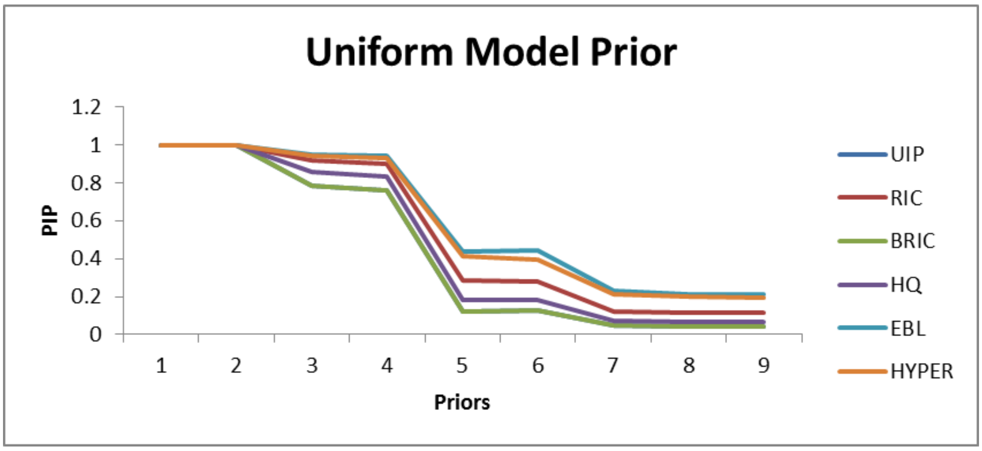

Figure 4: **Posterior Inclusion Probability for Uniform Model Prior**

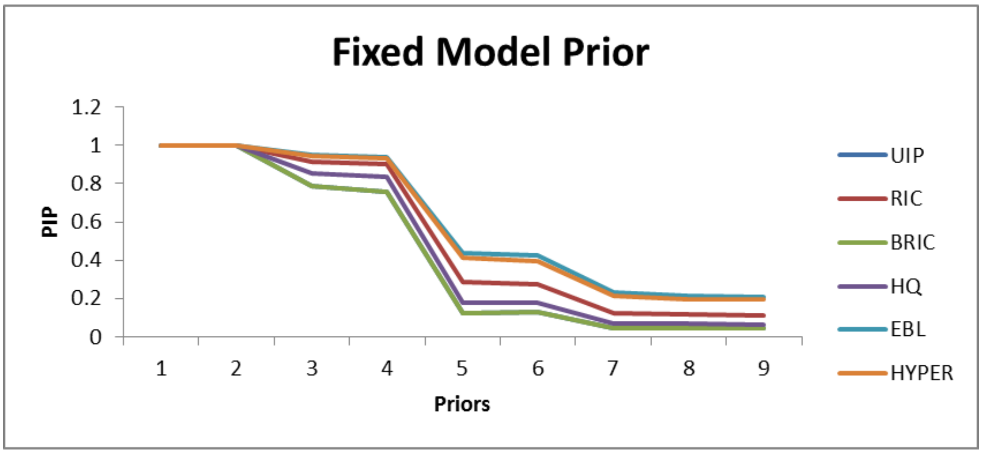

Figure 5: **Posterior Inclusion Probability for Fixed Model Prior**

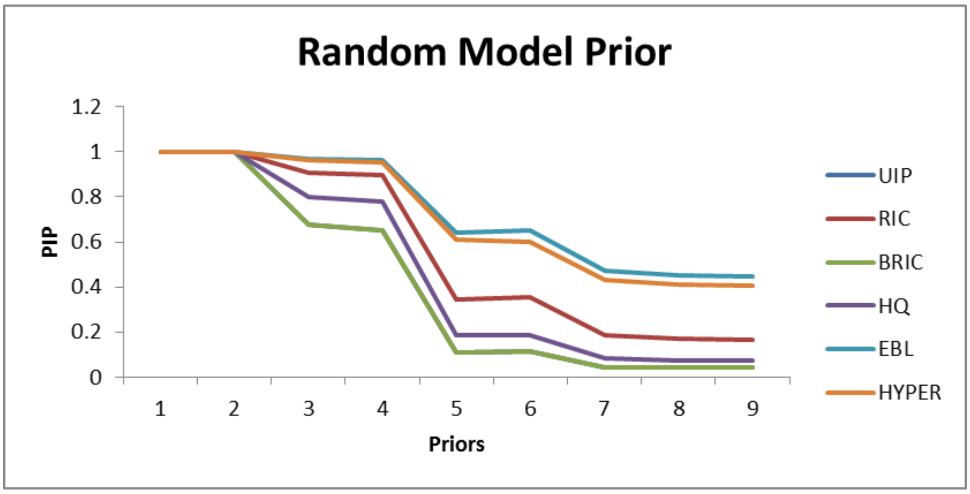

Figure 6: **Posterior Inclusion Probability for Random Model Pror**

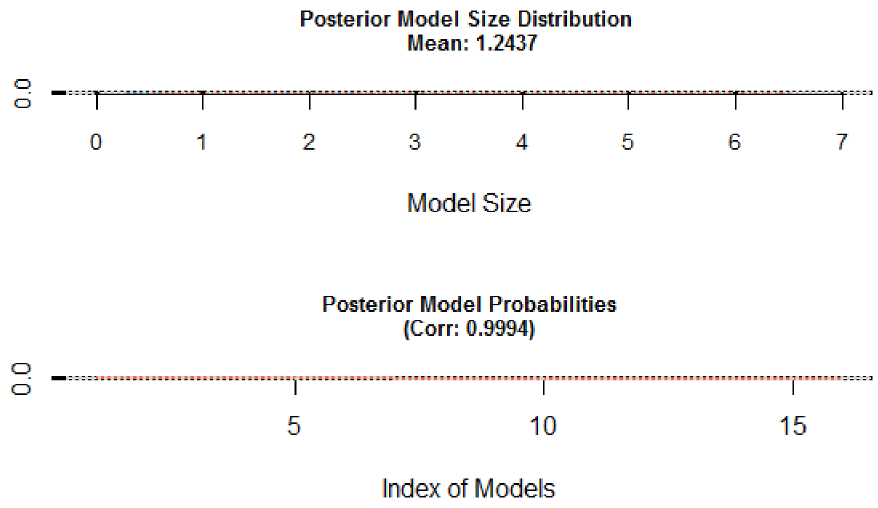

Figure 7: **Posterior Model size Distribution for uniform model prior**

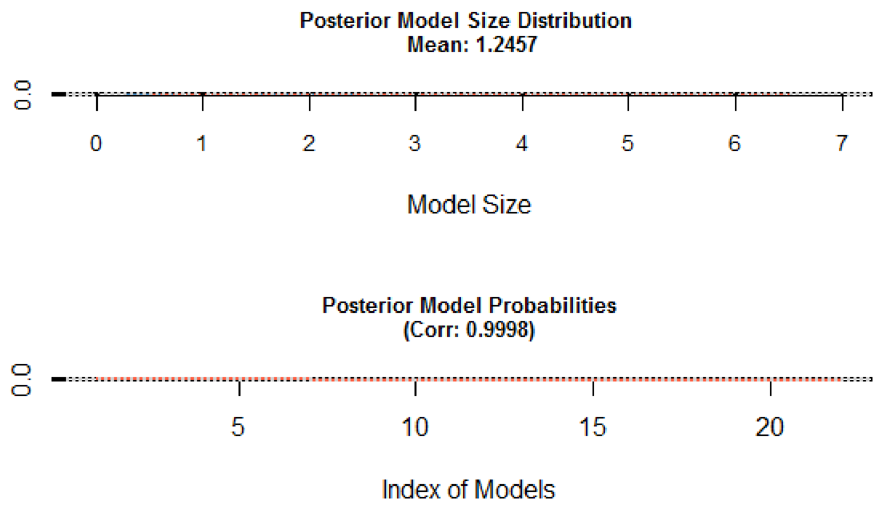

Figure 8: **Posterior Model size Distribution for fixed model prior**

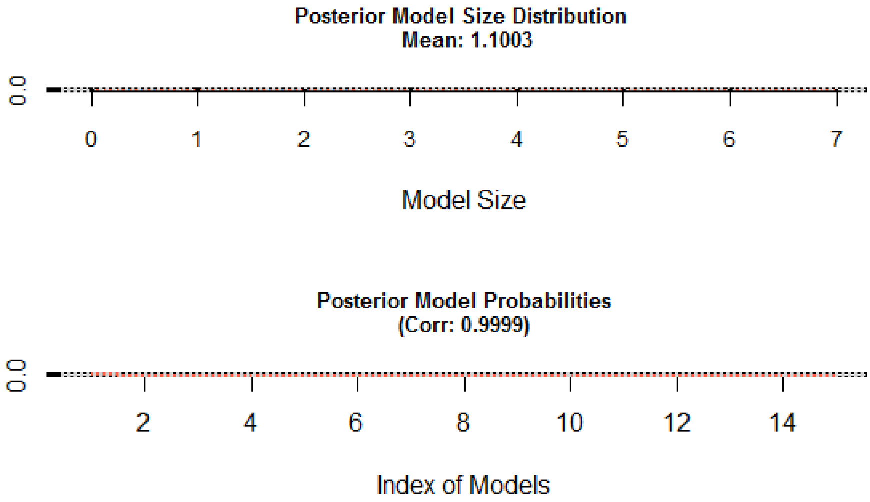

Figure 9: **Posterior Model size Distribution for random model prior**